Thermoelectric Figure of Merit of Degenerate and Nondegenerate Semiconductors

A Dissertation Presented

by

MICHAEL CONSTANTINE NICOLAOU

to

The Department of Mechanical and Industrial Engineering

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In Mechanical Engineering

In the field of

Direct Thermoelectric Energy Conversion

Northeastern University
Boston, Massachusetts

December 2008
Abstract

When the electronic component of the thermal conductivity is much greater than the lattice or phonon component thereof, the dimensionless thermoelectric figure of merit, ZT, may be expressed as $ZT = \frac{S^2}{\sigma \kappa}$, where $S$ is the Seebeck coefficient or thermoelectric power, $\sigma$ is the electrical conductivity, $\kappa$ is the thermal conductivity, $L$ is the Lorenz number, $T$ is the absolute temperature, and $Z = \frac{S^2 \sigma}{\kappa}$. The starting point of this dissertation was to pick values of ZT and find corresponding values of $S$, using the foregoing equation. On this basis, a major investigation was carried out to find the upper limit that would be reasonably expected on ZT. We were able to show that such a limit does exist and amounts to 23.9 for degenerate semiconductors and 24.5 for nondegenerate semiconductors. Before reaching this conclusion, several intermediate paths were pursued. We investigated Mott’s equation for the Seebeck coefficient and found it to be inappropriate for nondegenerate semiconductors and partly correct for degenerate ones, although the $S$ values it yields for the latter still need to multiplied by a correction factor of 2. This should come as no surprise, since we already know that Mott’s equation was derived for metals. Combining the Wiedemann-Franz and Eucken laws, we were able to derive a number of theoretical correlations for the ratio of the electronic to the lattice thermal conductivity. These were used to help verify the validity of the equation $ZT = \frac{S^2}{\sigma \kappa}$ under particular or extreme conditions. We ended up deriving equations expressing ZT as a function of the free charge carrier concentration, $n$, and the absolute temperature, $T$, for both degenerate and nondegenerate semiconductors. Numerical results obtained on the basis of these equations are presented in tables. These were eventually used to specify the conditions under which the limit on ZT would be achieved for degenerate and nondegenerate semiconductors.
Acknowledgement

I wish hereby to express my gratitude to Professor Welville B. Nowak for his willingness to act as my academic advisor as well as for his numerous and invaluable suggestions during the execution of this dissertation.
Contents

Abstract i

Acknowledgement ii

Table of Contents iii

List of Figures vi

List of Tables vii

Special Note ix

1 Theoretical Fundamentals of Thermoelectric Energy Conversion 1

1.1 Introduction 1

1.2 The Kelvin Relations 1

1.3 Theory of Thermoelectric Devices 6

1.4 Coefficient of Performance of a Thermoelectric Refrigerator 6

1.5 The Figure of Merit Z 11

1.6 Efficiency of a Thermoelectric Generator 12

1.7 Alternative Analysis of a Thermoelectric Power Generator 14
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.8 Early Developments</td>
<td>21</td>
</tr>
<tr>
<td>1.9 Materials for Semiconductor Thermoelements</td>
<td>24</td>
</tr>
<tr>
<td>2 Minimizing Thermal Conductivity</td>
<td>29</td>
</tr>
<tr>
<td>2.1 Atomic Weight and Melting Point</td>
<td>29</td>
</tr>
<tr>
<td>2.2 Semiconductor Solid Solutions</td>
<td>32</td>
</tr>
<tr>
<td>3 Electron Transport in a Zero Magnetic Field</td>
<td>36</td>
</tr>
<tr>
<td>3.1 Single Bands</td>
<td>36</td>
</tr>
<tr>
<td>3.2 Degenerate and Nondegenerate Conductors</td>
<td>40</td>
</tr>
<tr>
<td>4 The Role of Solid-State Chemistry in the Discovery of New Thermoelectric Materials</td>
<td>45</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>45</td>
</tr>
<tr>
<td>4.2 Necessary Criteria for Thermoelectric Materials</td>
<td>49</td>
</tr>
<tr>
<td>4.3 Solid-State Chemistry as a Tool for Targeted Materials Discovery</td>
<td>51</td>
</tr>
<tr>
<td>4.4 Thermoelectric Materials Discovery</td>
<td>53</td>
</tr>
<tr>
<td>4.5 Conclusions and Outlook</td>
<td>57</td>
</tr>
<tr>
<td>5 General Review of the Thermoelectric Figure of Merit</td>
<td>61</td>
</tr>
<tr>
<td>6 Critical Analysis of the Thermoelectric Figure of Merit as Related to the Wiedemann-Franz and Eucken Laws</td>
<td>81</td>
</tr>
<tr>
<td>7 Correlations of the Reduced Fermi Energy and the Thermoelectric Figure of Merit as Functions of the Free Charge Carrier Concentration and the Absolute Temperature</td>
<td>93</td>
</tr>
<tr>
<td>8 Comparison of the Results and Findings of this Dissertation with Those of Other Investigators</td>
<td>125</td>
</tr>
</tbody>
</table>
9 Conclusions 143

Appendix A1 General Review of the Brillouin Zone Concept 148

Appendix A2 Transport of Heat and Electricity in Solids 163
A2.1 Crystalline Solids 163
A2.2 Heat Conduction by the Lattice 165
A2.3 Scattering of Phonons by Defects 171

Appendix A3 Band Theory of Solids 177
A3.1 Energy Bands in Metals, Insulators and Semiconductors 177
A3.2 Electron-Scattering Processes 182

References 184

Appendix References 186
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure (1)</td>
<td>The Thermoelectric Effects</td>
<td>2</td>
</tr>
<tr>
<td>Figure (2)</td>
<td>A simple Thermoelectric Circuit</td>
<td>3</td>
</tr>
<tr>
<td>Figure (3)</td>
<td>Thermocouple used as a refrigerator</td>
<td>7</td>
</tr>
<tr>
<td>Figure (4)</td>
<td>Thermocouple used as Power Generator</td>
<td>14</td>
</tr>
<tr>
<td>Figure (5)</td>
<td>Thermoelectric Power Generator</td>
<td>14</td>
</tr>
<tr>
<td>Figure (6)</td>
<td>One possibility of a generic yet simplistic band structure of a “nearly ideal” thermoelectric material. The high number of valleys in the conduction band or peaks in the valence band, coupled with a high band degeneracy parameter, $\gamma$ (deriving from high crystal symmetry), could produce enormous power factors.</td>
<td>59</td>
</tr>
<tr>
<td>Figure (A1.1)</td>
<td>Perturbed and Unperturbed Wave Frequency</td>
<td>151</td>
</tr>
<tr>
<td>Figure (A1.2)</td>
<td>Bragg’s Reflection</td>
<td>151</td>
</tr>
<tr>
<td>Figure (A1.3)</td>
<td>Ewald’s Construction</td>
<td>160</td>
</tr>
<tr>
<td>Figure (A1.4)</td>
<td>Waves in a Discontinuous Lattice</td>
<td>160</td>
</tr>
<tr>
<td>Figure (A2.2.1)</td>
<td>Two-dimensional representation of: (a) N-process and (b) U-process</td>
<td>170</td>
</tr>
</tbody>
</table>
List of Tables

Table (1) Values of the thermoelectric power, $S, \mu VK^{-1}$ 68

Table (2) Values of the reduced Fermi energy, $\eta$ (Goldsmid) 74

Table (3) Values of the reduced Fermi energy, $\eta$ (Mott) 74

Table (4) Modified Values of the reduced Fermi energy, $\eta$ (Mott) 80

Table (5) Values of the exponential term, $e^\eta$, in Eq. (6.8) for nondegenerate semiconductors 85

Table (6) Reduced Fermi energy, $\eta$, according to Eqs. (6.8)-(6.10), as well as Table (5), for nondegenerate semiconductors 85

Table (7) Values of the free charge carrier concentration, $n$, carriers cm$^{-3}$, as a function of the temperature, $T$, K, and the reduced Fermi energy, $\eta$, along with the corresponding values of the dimensionless thermoelectric figure of merit, $ZT$, and the Seebeck coefficient, $S, \mu VK^{-1}$, for nondegenerate semiconductors for which $r = -1/2$ 122

Table (8) Values of the free charge carrier concentration, $n$, carriers cm$^{-3}$, as a function of the temperature, $T$, K, and the reduced Fermi energy, $\eta$, along with the corresponding values of the dimensionless thermoelectric figure of merit, $ZT$, and the Seebeck coefficient, $S, \mu VK^{-1}$, for nondegenerate semiconductors for which $r = 1/2$ 122

Table (9) Values of the free charge carrier concentration, $n$, carriers cm$^{-3}$, as a function of the temperature, $T$, K, and the reduced Fermi energy, $\eta$, along with the corresponding values of the dimensionless thermoelectric figure of merit, $ZT$, and the Seebeck coefficient, $S, \mu VK^{-1}$, for nondegenerate semiconductors for which $r = 3/2$ 123

Table (10) Values of the free charge carrier concentration, $n$, carriers cm$^{-3}$, as a function of the dimensionless thermoelectric figure of merit, $ZT$, and the temperature, $T$, K, for degenerate semiconductors. 123

Table (11) Values of the minimum reduced Fermi energy, $\eta_{\text{min}}$, as a function of the temperature, $T$, K, along with the corresponding
values of the dimensionless thermoelectric figure of merit, $ZT$, the Seebeck coefficient, $S$, $\mu VK^{-1}$, and the concentration of the free charge carriers, $n$, carriers cm$^{-3}$, for nondegenerate semiconductors, for which $r = -1/2$ and the forbidden energy band gap, $E_g = 0.78$eV

Table (12) Values of the Seebeck coefficient, $S$, $\mu VK^{-1}$, as a function of the dimensionless thermoelectric figure of merit, $ZT$, and the scattering parameter, $r$, for degenerate and nondegenerate semiconductors

Table (13) Limits of the variation of the reduced Fermi energy, $\eta$, for nondegenerate semiconductors
Special Note

The three appendices: A1, A2 and A3, related to the Brillouin Zone Concept, Transport of Heat and Electricity in Solids and Band Theory of Solids, respectively, have been included as a tutorial aid to any prospective reader of this dissertation whose background might be deficient in those areas.

Chapter 1

Theoretical Fundamentals of Thermoelectric Energy Conversion

1.1 Introduction

In 1821, Seebeck submitted a report on his experiments to the Prussian Academy of Sciences, which showed that he had observed the first of the three thermoelectric effects. He had produced potential differences by heating the junctions between dissimilar conductors. Thirteen years later, Peltier published some results which showed that he had discovered a second thermoelectric effect. When a current is passed through a junction between two different conductors, there is absorption or generation of heat, depending on the direction of the current. This effect is superimposed upon, but quite distinct from, the Joule resistance-heating effect, usually associated with the passage of an electric current. Thomson, later Lord Kelvin, realized that a relation should exist between the Seebeck and Peltier effects, and proceeded to derive this relation from thermodynamical arguments. This led him to the conclusion that there must be a third thermoelectric effect, now called the Thomson effect; this is a heating or cooling effect in a homogeneous conductor when an electric current passes in the direction of a temperature gradient.

1.2 The Kelvin Relations

The first Kelvin relation is that between the Seebeck and Peltier effects, and is particularly important since, whereas it is the Seebeck coefficient which is most easily measured, it is the Peltier coefficient which determines the cooling capacity of a thermoelectric refrigerator.
In order to define the thermoelectric coefficients, let us consider the circuits shown in Fig. 1. In Fig. 1(a), an open circuit potential difference, $\Delta V$, is developed as a result of the temperature difference, $\Delta T$, between the junctions of conductor a to conductor b. The differential Seebeck coefficient, $S_{ab}$, is defined by

$$S_{ab} = \lim_{\Delta T \to 0} \frac{\Delta V}{\Delta T}$$

(1.2.1)

In Fig. 1(b), there is a rate of reversible heat generation or absorption, $Q$, as a result of a current, $I$, passing through a junction between the conductors, a and b. The Peltier coefficient, $\pi_{ab}$, is given by

$$\pi_{ab} = \frac{Q}{I}$$

(1.2.2)

Finally, Fig. 1(c) shows that the passage of a current, $I$, along a portion of a single homogeneous conductor, over which there is a temperature difference, $\Delta T$, leads to a rate of reversible heat generation, $\Delta Q$. The Thomson coefficient, $\gamma$, is defined by

$$\gamma = \lim_{\Delta T \to 0} \frac{\Delta Q}{I \Delta T}$$

(1.2.3)
FIG. 2. A simple Thermoelectric Circuit.

The Kelvin relations may be derived by applying the laws of thermodynamics to the simple circuit shown in Fig. 2. A current, \( I \), passes around this circuit, which consists of conductors, a and b, with junctions at temperatures, \( T_1 \) and \( T_2 \). From the principle of conservation of energy, the heat generated must be equal to the consumption of electrical energy. If the current is small enough, Joule heating may be neglected. Thus,

\[
\left( \frac{1}{2} \eta_{a} - \frac{1}{2} \eta_{b} \right) I \delta T = \int_{1}^{2} S_{ab} I dT
\]  

(1.2.4)

Differentiating, one finds that

\[
\frac{d\pi_{ab}}{dT} + \eta_{a} - \eta_{b} = S_{ab}
\]  

(1.2.5)

A second relation between the coefficients may be obtained by applying the second law of thermodynamics. In order that this step should be valid, it is essential that the process should be reversible. It is reasonable to suppose that the thermoelectric effects are reversible, but they are inevitably accompanied by the irreversible processes of Joule heating and heat conduction. Thus, the application of reversible thermodynamics is not strictly justified. However, the use of Onsager’s reciprocal relations, which are based on irreversible thermodynamics, leads to the same conclusions\(^1\), so we shall assume that there is no overall change of entropy around the
circuit of Fig. (2). Hence,

\[ \int_1^2 \text{Id} \left( \frac{\pi_{ab}}{T} \right) + \int_1^2 \frac{\gamma_a - \gamma_b}{T} \text{Id} T = 0 \quad (1.2.6) \]

Invoking Sir Isaac Newton’s differential calculus, i.e., differentiating the above equation with respect to T, one writes

\[ \frac{d}{dT} \int_1^2 \text{Id} \left( \frac{\pi_{ab}}{T} \right) + \frac{d}{dT} \int_1^2 \frac{\gamma_a - \gamma_b}{T} \text{Id} T = \frac{d}{dT} (0) \quad (1.2.7) \]

which leads to

\[ I \cdot \frac{d}{dT} \left( \frac{\pi_{ab}}{T} \right) + \frac{\gamma_a - \gamma_b}{T} I = 0 \quad (1.2.8) \]

one thus writes

\[ T \frac{d\pi_{ab}}{dT} - \frac{\pi_{ab}}{T^2} + \frac{\gamma_a - \gamma_b}{T} = 0 \quad (1.2.9) \]

consequently, one obtains

\[ \frac{1}{T} \frac{d\pi_{ab}}{dT} - \frac{\pi_{ab}}{T^2} + \frac{\gamma_a - \gamma_b}{T} = 0 \quad (1.2.10) \]

this eventually yields

\[ \frac{d\pi_{ab}}{dT} - \frac{\pi_{ab}}{T^2} + \frac{\gamma_a - \gamma_b}{T} = 0 \quad (1.2.11) \]

Combining Eqs. (1.2.5) and (1.2.11), one gets

\[ S_{ab} = \frac{\pi_{ab}}{T} \quad (1.2.12) \]

This is the first and the more important of the Kelvin relations, since it relates the Seebeck and the Peltier coefficients. The second Kelvin relation, which connects the Seebeck and Thomson coefficients, is
\[ \frac{dS_{ab}}{dT} = \frac{\gamma_a - \gamma_b}{T} \]  \hspace{1cm} (1.2.13)

Both of Kelvin’s laws have been confirmed, within experimental error, for a number of thermocouple materials\(^2\). However, it has been reported that Eq. (1.2.12) is not strictly obeyed for a germanium-copper couple\(^3\). In this case, it appears that the value of the Peltier coefficient, \(\pi\), is less than the product, \(ST\). In spite of this fact, it seems that the Kelvin relations are applicable to the materials used in thermoelectric applications, and their validity will be assumed here.

The Seebeck and Peltier coefficients are both defined for junctions between two conductors, but the Thomson coefficient is a property of a single conductor. Eq. (1.2.13) suggests a way of defining the absolute Seebeck coefficient for a single material, namely by putting

\[ \frac{dS}{dT} = \frac{\gamma}{T} \]  \hspace{1cm} (1.2.14)

It is established from the third law of thermodynamics that the Seebeck coefficient is zero, for all junctions, at the absolute zero of temperature; the absolute Seebeck coefficient of any material is, therefore, taken to be zero at this temperature. Thus,

\[ S = \int_{T_0}^{T} \frac{T}{T} dT \]  \hspace{1cm} (1.2.15)

The absolute Seebeck coefficient of a material at very low temperatures may be determined by joining it to a superconductor, the latter possessing zero thermoelectric coefficients. This procedure has been carried out for pure lead up to 18K,\(^4\) and the Thomson coefficient for lead has been measured between 20K and room temperature\(^5\). Its value in the range between 18K and 20K can be accurately extrapolated. Thus, by use of Eq. (1.2.15), the
absolute Seebeck coefficient of lead has been established. The absolute Seebeck coefficient of any other conductor may be determined by joining it to lead.

1.3 Theory of Thermoelectric Devices

In spite of the fact that thermoelectric effects have been known for a long time, practically the only devices based upon them, which have been employed until recently, are thermocouples for the measurement of temperature, and thermopiles for the detection of radiant energy. Both applications utilize the Seebeck effect; in fact, they involve the thermoelectric generation of electricity from heat. In the past this process has been extremely inefficient, but the high sensitivity of the associated instruments has enabled such devices to be employed satisfactorily. However, even inefficient thermoelectric refrigeration has been impossible up to the last several years.

The basic theory of thermoelectric generators and refrigerators was first derived satisfactorily by Altenkirch in 1909\textsuperscript{6} and 1911\textsuperscript{7}. He showed that, for both applications, materials were required with high thermoelectric coefficients, high electrical conductivities to minimize Joule heating, and low thermal conductivities, to reduce heat transfer losses. However, it was quite a different matter, knowing the favorable properties, and obtaining materials embodying them, and so long as metallic thermocouples were employed, no real progress was made. It is only since semiconductor thermocouples have been prepared that reasonably efficient thermoelectric generators and refrigerators have become possible.

1.4 Coefficient of Performance of a Thermoelectric Refrigerator

One of the virtues of thermoelectric generation and refrigeration is the fact that the efficiency is independent of the capacity of the unit. The coefficient of performance of a thermoelectric refrigerator can, therefore, be derived on the basis of the single couple shown in
Fig. (3). A current, I, is passed through a thermocouple, with one of its junctions in thermal contact with a heat sink, while the other junction removes heat from the source at a rate, Q. The subscripts, p and n, will be used to denote the two branches of the couple which should consist of conductors having positive and negative absolute Seebeck coefficients, respectively. It will be supposed that the temperature of the cold junction becomes less than the temperature of the hot junction by an amount, $\Delta T$.

The total resistance, R, of the couple is given by

$$R = \frac{l_p}{A_p \sigma_p} + \frac{l_n}{A_n \sigma_n}$$  \hspace{1cm} (1.4.1)

where $l$ represents the length, $A$ the cross-sectional area and $\sigma$ the electrical conductivity of an element. The thermal conductance, $K$, of the two arms of the couple in parallel is

$$K = \frac{A_p \kappa_p}{l_p} + \frac{A_n \kappa_n}{l_n}$$  \hspace{1cm} (1.4.2)

where $\kappa$ is the thermal conductivity.

FIG. 3. Thermocouple used as a refrigerator.
As a result of the Peltier effect, there is a rate of cooling \( Q = \pi_{pn} I \) at the cold junction. From Kelvin’s first relation, this becomes

\[
Q = \pi_{pn} I = S_{pn} T I
\]

where \( T \) is the absolute temperature at which the cooling, or the heat transfer, rate \( Q \), from the source to the cold junction, occurs. Let us investigate this a little further by doing the following transformation

\[
T = T_M - \frac{\Delta T}{2} = \frac{T_H + T_C}{2} - \frac{T_H - T_C}{2} = T_C
\]  

(1.4.3)

The above proves that the cooling rate, \( Q \), may be calculated from the temperature \( T_C \), which is that of the cold junction. One may thus write

\[
Q = \pi_{pn} I = S_{pn} T I = S_{pn} T_C I = S_{pn} \left( T_M - \frac{\Delta T}{2} \right) I
\]  

(1.4.4)

where \( T_M \) is the mean absolute temperature of the hot and cold junctions. The cooling effect is opposed by Joule heating in the branches, and by heat conducted from the hot junctions. It may be shown that half of the overall Joule heating finds its way to each junction. Thus the rate of absorption of heat from the source is

\[
Q_a = S_{pn} \left( T_M - \frac{\Delta T}{2} \right) I - \frac{1}{2} I^2 R - K\Delta T
\]  

(1.4.5)

While writing the above equation, the Thomson effect has been neglected.

Part of the potential difference applied to the couple is employed in overcoming the resistance of the branches, and part is used to balance the Seebeck voltage, resulting from the temperature difference between the junctions. One thus writes

\[
\Delta V = S_{pn} \Delta T + IR
\]  

(1.4.6)

Thus, the power, \( W \), supplied to the couple is given by
\[ W = \Delta VI = S_{pn} \Delta TI + I^2 R \]  

(1.4.7)

The coefficient of performance for refrigeration, \( \Phi \), is defined as the ratio, \( Q_a/W \). Hence,

\[
\Phi = \frac{Q_a}{W} = \frac{S_{pn} \left( T_M - \frac{\Delta T}{2} \right) I - \frac{1}{2} I^2 R - K \Delta T}{S_{pn} \Delta TI + I^2 R}
\]

(1.4.8)

For a given pair of thermoelectric materials, and for given hot and cold junction temperatures, the coefficient of performance is a function of the current, \( I \), and also of the resistance, \( R \), and the thermal conductance, \( K \). However, the latter two quantities are not independent, and only a single relation between the dimensions of the elements is really involved. For a specified cooling capacity, the ratio of length to cross-sectional area for an element should rise with the electrical and thermal conductivities. A balance is then struck between the effects of resistance heating and thermal conductance. It may be shown that \( \Phi \) reaches a maximum value when the dimensions of the elements obey the rule

\[
\frac{l_p A_n}{l_n A_p} = \left( \frac{\sigma_p \kappa_p}{\sigma_n \kappa_n} \right)^{1/2}
\]

(1.4.9)

Consequently,

\[
KR = \left\{ \left( \frac{\kappa_p}{\sigma_p} \right)^{1/2} + \left( \frac{\kappa_n}{\sigma_n} \right)^{1/2} \right\}^2
\]

(1.4.10)

and

\[
\Phi = \frac{S_{pn} \left( T_M - \frac{\Delta T}{2} \right) I R - \frac{1}{2} (IR)^2 - \Delta T \left( \frac{\kappa_p}{\sigma_p} \right)^{1/2} + \left( \frac{\kappa_n}{\sigma_n} \right)^{1/2} }{S_{pn} \Delta T (IR) + (IR)^2}
\]

(1.4.11)

By differentiation of \( \Phi \) with respect to the product \( IR \), the optimum current for a specific temperature difference may be found. It is given by

9
\[(IR)_{opt} = \frac{S_{pn}\Delta T}{\sqrt{1 + ZT_M} - 1} \quad (1.4.12)\]

where

\[Z = \frac{S_{pn}^2}{\left\{ \left( \frac{\kappa_p}{\sigma_p} \right)^{1/2} + \left( \frac{\kappa_n}{\sigma_n} \right)^{1/2} \right\}^2} \quad (1.4.13)\]

Substituting the optimum value of IR in Eq. (1.4.11), it is found that the maximum coefficient of performance is given by

\[\Phi_{max} = T_M \left( \frac{\Delta T_M - 1}{\Delta T} \right) - \frac{1}{2} \quad (1.4.14)\]

Clearly, when the values of \(T_M\) and \(\Delta T\) are given, the coefficient of performance rises with increase of \(Z\). \(Z\) is, therefore, a figure of merit for the thermocouple.

It may be noted that as \(Z\) tends towards infinity, the value of \(\Phi\), given by Eq. (1.4.14), approaches the value

\[\Phi = \lim_{Z \to \infty} \Phi_{max} = \frac{T_M}{\Delta T} - \frac{1}{2} = \frac{T_M - \Delta T/2}{\Delta T} = \frac{T_C}{\Delta T} \quad (1.4.15)\]

As expected, this is the coefficient of performance of an ideal thermodynamic refrigerating machine or an inverted Carnot engine.

A thermoelectric heat pump operates exactly the same way as a thermoelectric refrigerator, with the only difference that the desired effect, or result, is the heating rate that is produced at the hot junction, and transferred to the sink, which could be a certain environment, or enclosure. Thus, for a thermoelectric heat pump, the coefficient of performance is

\[\Psi = \frac{W + Q}{W} = 1 + \frac{Q}{W} = 1 + \Phi \quad (1.4.16)\]
where $\Phi$ is the coefficient of performance for refrigeration.

### 1.5 The Figure of Merit $Z$

The question arises as to the values of the parameters which should be used in calculating the figure of merit, $Z$, if $\Delta T$ is large, since $S$, $\sigma$ and $\kappa$ are all temperature dependent. Since the Joule heating and thermal conduction effects involve the whole of the couple, it is apparent that mean values of $\sigma$ and $\kappa$ are required. However, since the Peltier effect is localized, it might be thought that the value of $S$ at the cold junction should be employed. This conclusion is found to be false, when the Thomson effect is taken into account.

The Thomson cooling in the whole of an element amounts to $\int \gamma dT$, which, from Kelvin’s second relation, is equal to $\int T dS$. Half of this cooling appears at each junction. Thus, instead of the Peltier cooling $S_C T_C I$, we should use an overall rate of cooling $S_C T_C I + \frac{1}{2} \int_C^H T dS$, where the subscripts, $C$ and $H$, refer to the cold and hot junctions, respectively. Now, the Thomson term, in the expression for the overall cooling, is only a small fraction of the Peltier term. It is, therefore, a reasonable approximation to put

$$\int_C^H T dS = T_C \int_C^H dS = (S_H - S_C)T_C$$  \hspace{1cm} (1.5.1)

It is then found that

$$S_C T_C I + \frac{1}{2} \int_C^H T dS = \left(\frac{S_H + S_C}{2}\right)T_C I$$  \hspace{1cm} (1.5.2)

Thus, the Thomson effect can be taken into account, if the mean value of the Seebeck coefficient, $S$, rather than the value at the cold junction, is used. The figure of merit, $Z$, thus
involves the mean values of all the parameters.

In comparing different thermoelectric materials, it is rather inconvenient to deal with the figure of merit, Z, as defined by Eq. (1.4.13). We, therefore, define a figure of merit, Z, for a single material as

\[ Z = S^2 \frac{\sigma}{\kappa} \]  \hspace{1cm} (1.5.3)

The figure of merit, Z, of a thermoelectric device is strictly equal to the mean of the values of Z for the two elements, Z\(_p\) and Z\(_n\), respectively, only in exceptional circumstances. In general, Z must be regarded as a rather complicated average of Z\(_p\) and Z\(_n\). The use of an individual figure of merit, as defined by Eq. (1.5.3), is justified by the fact that nowadays the values of Z\(_p\) and Z\(_n\) are never very much different in the thermocouples which are most suitable for thermoelectric applications.

### 1.6 Efficiency of a Thermoelectric Generator

The efficiency of a thermoelectric power generator may be calculated using the simple model shown in Fig. (4)\(^9\). One junction of a thermocouple is in contact with a source of heat, while the other is in thermal contact with a heat sink. Power is delivered to a load of resistance, R\(_L\). The efficiency of the generator is defined as the ratio of the rate of useful work, W, to the rate of heat supply, Q, from the source.

The heat supplied is either conducted along the branches to the cold junction, or used to balance the Peltier effect at the cold junction. However, it must be remembered that half of the Joule heating in the thermocouple finds its way back to the source. Thus,

\[ Q_a = K\Delta T + S_p n \left( T_M + \frac{\Delta T}{2} \right) I - \frac{1}{2} I^2 R \]  \hspace{1cm} (1.6.1)

using the same symbols as in the case of thermoelectric refrigeration.
For maximum power output from a given couple, the load resistance, \( R_L \), should be made equal to the generator resistance, \( R \). The useful work is then

\[
W = \frac{S_{pn}^2 (\Delta T)^2}{4R}
\]

since half of the thermoelectric voltage, \( S_{pn} \Delta T \), appears across the load.

The efficiency, \( \eta \), is given by

\[
\eta = \frac{\Delta T}{2T_M + \frac{\Delta T}{2} + \frac{4RK}{S_{pn}^2}}
\]

The dimensions of the thermoelements or branches should obey Eq. (1.4.9), therefore

\[
\eta = \frac{\Delta T}{2T_M + \frac{\Delta T}{2} + \frac{4}{Z}}
\]

Since the efficiency rises with \( Z \), this quantity is a figure of merit for thermoelectric generation as well as for refrigeration.

If \( Z \) tends to infinity, \( \eta \), as given by Eq. (1.6.4), approaches the value

\[
\eta = \frac{\Delta T}{2T_M + \frac{\Delta T}{2}}
\]

which is only about half the efficiency of an ideal thermodynamic machine, or heat engine. The reason is that, when the maximum efficiency is required, \( R_L \) should not be made equal to \( R \). It may be shown that the maximum efficiency is obtained when

\[
r = \frac{R_L}{R} = (1 + ZT_M)^{1/2}
\]

Therefore,
\[ \eta_{\text{max}} = \frac{\Delta T}{(r+1)T_M + \frac{\Delta T}{2}} \]  

(1.6.7)

This is a more thorough, rigorous and comprehensive study of the performance of a thermoelectric power generator, than the preceding one. Let us consider a thermoelectric device,
Fig. (5), consisting of a p-type thermoelement or branch, (1), and an n-type thermoelement or branch, (2); joined by a metallic bridge, (3), which constitutes the hot junction. An external resistance, $R_L$, is connected in circuit across the cold ends of the thermoelements. This resistance acts as a load for the electrical energy generated; it thus constitutes the cold junction of the device. When the potential difference across the cold ends of the circuit, without the load, is denoted by $U$ volts, the power delivered by the thermoelements, or device, will be

$$W = \frac{U^2}{R_L}$$

(1.7.1)

whilst the current

$$I = \frac{U}{R_L}$$

(1.7.2)

Obviously, both thermoelements, (1) and (2), normally comprise either the same, or different, semiconducting materials. This is deliberately done in order to maximize the figure of merit and the performance of thermoelectric devices.

We shall denote the thermal emf, which is, in this case, equal to the sum of the thermal emf’s of the two branches, by $SVK^{-1}$. Thus,

$$S = S_{pn} = S_{12} = S_p - S_n = S_1 - S_2 = |S_p| + |S_n| = |S_1| + |S_2|$$

(1.7.3)

The hot ends, joined by the bridge, are maintained at a temperature, $T_1$, receiving heat energy from a source with a slightly higher temperature. The cold ends are at a temperature, $T_o$, slightly above the temperature of the surroundings, or of the heat exchange medium.

We shall denote the internal resistances of the two branches by $R_1$ and $R_2$, and their thermal conductances by $K_1$ and $K_2$. Denoting the electrical resistivities by $\rho$, the thermal conductivities by $\kappa$, and assuming that the lengths of both thermoelements, or rods, are equal to $l$, and their cross-sectional areas are $A_1$ and $A_2$, we have
\[ R = R_1 + R_2 = \left( \frac{\rho_1}{A_1} + \frac{\rho_2}{A_2} \right) l \]  \hfill (1.7.4)

\[ K = K_1 + K_2 = \left( \kappa_1 A_1 + \kappa_2 A_2 \right) \frac{1}{l} \]  \hfill (1.7.5)

On the basis of the general laws of thermoelectric phenomena, we can calculate the Peltier heat generated and absorbed by the thermoelements, considered together as one entity constituting the device, at its two ends or junctions, the Thomson heat generated or absorbed inside the thermoelements or rods, the heat transferred by conduction from the hot to the cold ends, the Joule heat generated by the current in the rods or thermoelements, and the useful electrical energy delivered by the thermoelements.

The amount of heat energy, \( Q_{T1} \), received by the hot junctions of both thermoelements, due to the Peltier effect, is

\[ Q_{T1} = S_{T1} I_1 T_1 \]  \hfill (1.7.6)

The power, \( Q_{To} \), delivered at the cold junction is

\[ Q_{To} = -S_{To} I_0 T_0 \]  \hfill (1.7.7)

The Thomson heat, \( Q_{Th} \), generated in each rod or thermoelement is, according to Eq. (1.2.13),

\[ Q_{Th} = \pm \int_{T_0}^{T_1} T \frac{dS}{dT} IdT \]  \hfill (1.7.8)

In the special case, when \( S \) at both ends has the same value, \( Q_{Th} = 0 \).

The heat flux, \( Q_h \), transferred from the hot junction to the cold junction through the two thermoelements is

\[ Q_h = K(T_1 - T_o) \]  \hfill (1.7.9)
The Joule heat generated in the two thermoelements, branches or rods is

\[ Q_J = I^2R \]  

(1.7.10)

The useful power, \( W \), delivered by the two thermoelements, or device, is

\[ W = I^2R_L \]  

(1.7.11)

The current is

\[ I = \frac{S(T_1 - T_o)}{R_L + R} \]  

(1.7.12)

Setting

\[ \frac{R_L}{R} = m \]  

(1.7.13)

we have

\[ Q_{T_1} = S_{T_1}^2T_1(T_1 - T_o)\frac{1}{R(m + 1)} \]  

(1.7.14)

\[ W = S^2(T_1 - T_o)^2\frac{m}{R(m + 1)^2} \]  

(1.7.15)

For the time being, we shall neglect the Thomson heat, regarding it as small in comparison with the other terms, and shall assume \( S_{T_1} = S_{T_o} = S \). When \( S_{T_1} \neq S_{T_o} \), the Thomson heat is accounted for, in the determination of the efficiency, by substituting for \( S \) the mean value for the two ends

\[ \frac{S}{S} = \frac{S_{T_1} + S_{T_o}}{2} \]  

(1.7.16)

Of the total Joule heat, \( I^2R \), generated in the thermoelements, half passes to the hot junction, returning the power
\[ \frac{1}{2} Q_j = \frac{1}{2} I^2 R \] (1.7.17)

and the rest is transferred to the cold junction.

The efficiency, \( \eta \), will be defined as the ratio of the useful electrical energy, \( I^2 R_L \), delivered to the external circuit, to the energy consumed, or received, from the heat source. The latter consists of the Peltier heat, \( Q_T \), and the heat, \( Q_h \), transferred by conduction to the cold junction, from which it is necessary to deduct the electrical energy, \( \frac{1}{2} I^2 R \), returned to the heat source.

\[
\eta = \frac{W}{Q_T + Q_h - \frac{1}{2} I^2 R} = \frac{S^2 (T_1 - T_o)^2 \frac{1}{R} \frac{m}{(m+1)^2}}{S^2 T_1 (T_1 - T_o) \frac{1}{R} \frac{1}{(m+1)} + K(T_1 - T_o) - \frac{1}{2} \frac{S^2 (T_1 - T_o)^2}{R(m+1)^2}} = \frac{T_1 - T_o}{T_1} \frac{m}{m+1} \left( \frac{1}{1 + \frac{KR}{S^2}} \right) \frac{1}{T_1} \frac{1}{m+1} T_1 \frac{1}{2} T_1 \frac{1}{m+1} \] (1.7.18)

Thus, the efficiency of the thermoelements, or thermoelectric device, is fully determined by: a) The hot and cold junction temperatures. b) The quantity, \( \frac{KR}{S^2} \), which depends on the properties of the materials, used in the manufacture of the thermoelements, and which will be denoted by, \( Z^{-1} \), so that

\[ Z = \frac{S^2}{KR} \] (1.7.19)

and, finally,

c) The selected ratio,
\[ m = \frac{R_L}{R} \] (1.7.20)
In order to achieve an efficiency as high as possible in thermoelements, or devices, at given values of S, κ and ρ, and an arbitrary ratio, m, as given by Eq. (1.61), it is necessary to find the optimum cross-sectional areas, A₁ and A₂, so that at given values of κ and ρ, the product, KR, is minimum. To find the condition for a minimum of KR, we shall differentiate

\[
KR = (\kappa_1 A_1 + \kappa_2 A_2) \left( \frac{\rho_1}{A_1} + \frac{\rho_2}{A_2} \right) = \kappa_1 \rho_1 + \kappa_2 \rho_2 + \kappa_1 \rho_2 \frac{A_1}{A_2} + \kappa_2 \rho_1 \frac{A_2}{A_1}
\]

with respect to A₁/A₂ and equate the derivative to zero. This gives

\[
\frac{\rho_1 \kappa_2}{\kappa_1 \rho_2} = \left( \frac{A_1}{A_2} \right)^2
\]

(1.7.21)

At this value of A₁/A₂

\[
KR = \left( \sqrt{\kappa_1 \rho_1} + \sqrt{\kappa_2 \rho_2} \right)^2
\]

(1.7.22)

\[
Z = \frac{S^2}{KR} = \frac{S^2}{\left( \sqrt{\kappa_1 \rho_1} + \sqrt{\kappa_2 \rho_2} \right)^2}
\]

(1.7.23)

This expression contains only the properties of the materials constituting the two branches or thermoelements of the device, but not their dimensions.

We now need to find the ratio, m = Rₐ / R, which gives the highest efficiency. The condition for delivering the maximum power to the load leads, in the case of a thermoelectric device, as with other current sources, to the requirement: m = 1, which when substituted into Eq. (1.7.18) yields

\[
\eta = \frac{1}{2} \frac{T_1 - T_0}{T_1 + \frac{2}{Z} \frac{1}{4} (T_1 - T_0)}
\]

(1.7.24)

We shall find the condition for the maximum efficiency by differentiating Eq. (1.7.18) with respect to m, and setting
which yields

$$\frac{\partial \eta}{\partial m} = 0$$  \hspace{1cm} (1.7.25)$$

where $M$ is a dimensionless number.

Substituting this optimum value of $m$, denoted here by $M$, into Eq. (1.7.18), we get

$$\left(\frac{R_L}{R}\right)_{\text{opt}} = M = \sqrt{1 + \frac{1}{2}Z(T_1 + T_0)}$$  \hspace{1cm} (1.7.26)$$

Henceforth, $\eta$ will only refer to this maximum efficiency. The first factor in Eq. (1.7.27)

represents the thermodynamic efficiency of a reversible heat, or Carnot, engine, and the second
describes the reduction in the efficiency due to losses, brought about by irreversibility, resulting
from heat conduction ($\kappa$) and Joule heat ($\rho$), entering into the expression for $Z$.

The greater is $M$, in comparison with unity, i.e., the larger the values of $Z$ and $T_1 + T_0$, the
smaller is the reduction in the efficiency due to losses brought about by irreversibility.

Therefore, an increase in the hot junction temperature, $T_1$, increases $\eta$ not only by increasing the
value of the efficiency of a reversible engine, $\frac{T_1 - T_0}{T_1}$, but also because of the simultaneous
increase in $M$, at a given $Z$.

This equation also clearly shows that, in order to achieve maximum efficiency, a material
has to satisfy only one condition, namely the maximum value of $Z$ compatible with the
maximum temperature, $T_1$, of the heat source, or, more precisely, the maximum product,

$$Z \frac{T_1 + T_0}{2},$$

attainable for the given material.
1.8 Early Developments

At the beginning of the past century, the theory of metals was based on the concept of the presence in each metal of a certain concentration of free electrons, \( n \), different for differing metals, moving at random within the metal, in a manner similar to gas molecules. According to these theories, as developed by Rieke, Drude, Lorentz and Debye, electrons were governed by the same Boltzmann statistics as gaseous molecules. As in a gas, their average translational energies were regarded as equal to \( \frac{3}{2} kT \), where \( k \) is the Boltzmann constant and \( T \) is the absolute temperature. The thermal emf, \( V \), in an open circuit, consisting of two metals, calculated according to this theory, is

\[
V = \frac{k}{e} \int_{T_1}^{T_2} \ln \left( \frac{n_2}{n_1} \right) dT
\]  

(1.8.1)

It can be readily calculated that \( S=86\mu\text{VK}^{-1} \) for \( n_2 / n_1 = 2.7 \), which is much higher than the experimentally observed value of \( S \), for cases where \( n_2 / n_1 > 2.7 \), since for the majority of metals \( S \) does not exceed a few \( \mu\text{VK}^{-1} \). According to the quantum theory of metals developed by Sommerfeld, who applied Fermi-Dirac statistics to electrons in metals, the thermal emf between two metals is, to a first approximation, equal to zero. Only the next approximation leads to a certain finite, although very small, value of \( S \). All this is not surprising. Let us consider a metallic rod, the ends of which are at different temperatures. Since the temperature does not change the concentration of electrons, \( n \), in the metal, but brings about only a slight redistribution in their thermal agitation velocities, it is obvious that a large thermal emf cannot arise in such a metal. The state of semiconductors in the quantum theory is similar to that of the classical metal, considered at the beginning of the 20th century. Some of the current-carrying particles in
semiconductors may be considered to be “free.” The concentration of free electrons in semiconductors varies, but is usually so much smaller than in metals, that the Boltzmann classical statistics are often applicable. A rise in temperature in certain regions leads to a change in both the concentration of free electrons and their kinetic energy, which, as in the case of gas molecules, may be equated to $\frac{3}{2}kT$. Let us consider a rod of semiconductor in which there is a temperature gradient, $dT/dx$. At the hot end of the rod, both the concentration and the velocities of the electrons are higher than at the cold end. Therefore, more electrons will start to diffuse in the direction of the temperature gradient, than in the opposite direction. The diffusion flux, carrying the negative charge away from the hot end, and transferring it to the cold end, sets up a potential difference between the ends. The diffusion process is increasingly retarded by the electrical field in the interior of the semiconductor, until the flux of electrons, caused by diffusion, is equal to the reverse flux, caused by the potential difference which has arisen. Under these conditions, a dynamic electron equilibrium in the semiconductor will be established, under which the temperature difference between the ends of the rods will maintain a corresponding potential difference.

The number of electrons passing through any cross-section of the conductor, in unit time, in both directions, are equal. However, the velocities of electrons proceeding from the hot end are higher than the velocities of electrons passing through the given cross-section from the cold end. This difference ensures that there is a continuous transfer of heat energy in the direction of the temperature gradient, without any actual charge transfer.

The heat transfer mechanism is substantially different when negative and positive charge carriers participate simultaneously in the current. Simultaneous transfer of equal numbers of holes and electrons does now lead to an accumulation of the charge and an increase in the
potential. Simultaneous diffusion of electrons and holes, from the hot end to the cold end, is caused not only by the difference in the carrier velocities, but also by their concentration gradient.

In the case of such a bipolar diffusion a thermal emf can also arise due to the following two factors: 1) When the concentration of one type of carrier exceeds the concentration of the opposite type of carrier, the flux of the first type of carrier will carry, towards the cold end, predominantly a charge which will retard their motion and, conversely, accelerate the carriers of the opposite sign, until the fluxes of both carriers are equal. An electric field, governed by the temperature gradient, will thus be established. 2) The difference of charge carrier mobilities forms the second source of thermal emfs. The mobility, \( \mu \), is related to the diffusion coefficient, \( D \), by the universal relationship, established by Einstein:

\[
\frac{\mu}{D} = \frac{e}{kT}
\]  

(1.8.2)

Under the effect of the concentration gradient, set up by the temperature gradient, the carriers, characterized by higher values of \( \mu \) and \( D \), would have moved forward, had they not created a space charge on becoming separated from the carriers of opposite sign. The corresponding electric field, \( E \), retards their motion, and accelerates the slower carriers of the opposite sign. The electric field, \( E \), equalizes the velocities of both types of carrier, enabling them to diffuse as a single body.

Therefore, even when thermal agitation creates an equal number of current carriers of each sign, their diffusion produces an electric field in the conductor, depending on the difference in carrier mobilities. The expression for this electric field, \( E \), will be similar to that for the diffusion of ions in an electrolyte, namely
where $E_0$ is the electric field which would have existed in the presence of current carriers of one sign only, and $\mu_1$ and $\mu_2$ are the mobilities of positive and negative charges.

The advantage of semiconductors, when used as materials for thermoelements, is due to their higher values of thermal emf. In metals, the concentration of free charge carriers is high, of the order of $10^{22}$ carriers cm$^{-3}$, and does not depend much on temperature. The kinetic energy of the highly degenerate electrons is virtually independent of temperature, and the same applies to the contact potential of the metal at its boundaries. Under such conditions, thermal emfs, as a rule, do not exceed a few $\mu$VK$^{-1}$. In semiconductors, however, temperature may have a pronounced effect on the concentration, $n$, and kinetic energy, $E$, of free charge carriers, while the absolute value of the concentration is smaller by a few orders of magnitude, namely $10^{14}$ to $10^{20}$ carriers cm$^{-3}$. The contact potential, $\Phi$, of semiconductors with respect to metals, and the corresponding chemical potential, $\mu$, are also functions of temperature. Thermal emfs amount in this case to hundreds $\mu$VK$^{-1}$.

The thermal emf per degree Kelvin, $S$, may be regarded as an entropy flux of 1 Coulomb of electrical charge. The value of $S$ depends not only on the entropy difference between two substances, but also on the conditions of the motion of the electrons. These conditions may, in turn, depend on the nature of the semiconductor, and on the mechanism of electron scattering in the transfer of electrons from one portion of the semiconductor to another. Therefore, the value of $S$ is closely related to the mobility, $\mu$, which is governed by the same scattering mechanism.

## 1.9 Materials for Semiconductor Thermoelements

Let us now consider the selection of the most suitable materials for thermoelements.
First of all, it should be noted that semiconductors, owing to their much higher values of \( S \), are definitely preferable to metals. The higher the value of \( Z = \frac{S^2}{\kappa \rho} \), for the individual branches of the thermoelement, the higher is the value of \( Z = (S_1 + S_2)^2 \left/ \left( \sqrt{\kappa_1 \rho_1} + \sqrt{\kappa_2 \rho_2} \right)^2 \right. \), which determines the efficiency of the entire thermoelement or device. The ratio of thermal conductivity to electrical conductivity, for all metals, is close to the theoretical value predicted by quantum mechanics: \( \frac{\kappa}{\sigma} = \left( \frac{\pi^2}{3} \right) \left( \frac{k}{e} \right)^2 T = 2.44 \times 10^{-8} T \). The thermal conductivity of a semiconductor, \( \kappa \), is composed of the electronic thermal conductivity, \( \kappa_{el} \), and the thermal conductivity due to thermal vibrations and the propagation of heat waves, usually denoted as phonon, or lattice, thermal conductivity: \( \kappa = \kappa_{el} + \kappa_{ph} \). The former is related to the electrical conductivity by the Wiedemann-Franz law. However, the coefficient of proportionality has a value equal to that for metals, \( \frac{\kappa_{el}}{\sigma} = \left( \frac{\pi^2}{3} \right) \left( \frac{k}{e} \right)^2 T \), only at very high concentrations of free electrons, namely more than \( 2.5 \times 10^{19} \) carriers \( \text{cm}^{-3} \), when the latter must be regarded as degenerate. At lower concentrations, usually prevailing in semiconductors, \( \frac{\kappa_{el}}{\sigma} = 2 \left( \frac{k}{e} \right)^2 T = 1.48 \times 10^{-8} TV^{-1} \). In general, the Wiedemann-Franz law for semiconductors has the following form: \( \frac{\kappa_{el}}{\sigma} = (r+2) \left( \frac{k}{e} \right)^2 T \), which for atomic lattices with \( r = 0 \), reduces to the preceding equation, where \( r \) is the electronic scattering parameter or constant that appears in the formula giving the electron mean free path length, \( l \), as a function of the kinetic energy, namely: \( l \propto E^r \).

By analogy with diffusion and thermal conduction in gases, the phonon part of thermal conductivity, \( \kappa_{ph} \), can be expressed as \( \kappa = \frac{1}{3} cv \bar{\lambda} \), where \( c \) is the specific heat, \( v \) is the sound velocity and \( \bar{\lambda} \) is the phonon mean free path length. The value of the thermal conductivity, \( \kappa_{ph} \), of a semiconductor is also related to electron mobility, \( \mu \), and therefore also to electrical conductivity. Both phonons and electrons are scattered by the same inhomogeneities of the
crystalline lattice, but the degree of scattering is different, owing to their different wavelengths: 7x10^{-7} cm for electrons and approximately 5x10^{-8} cm for phonons, and their different physical nature.

There should exist some correlation between $\bar{\lambda}$ and the mean free path length of electrons, $\bar{l}$. As pointed out by H.J. Goldsmid, the ratio, $\mu / \kappa_{ph}$, governing the value of $Z$ of thermocouples, increases with increasing atomic weight. The factors affecting the value of the phonon thermal conductivity, $\kappa_{ph}$, are as follows: $\kappa_{ph}$ decreases not only with increasing average atomic weight, but also in going over from purely valence type to ionic compounds, i.e., from elements in group IV of the periodic system towards compounds of groups III with V, II with IV and I with VII.

In order to obtain the highest values of $(S^2\sigma)_{max}$, it is desirable to select semiconductors with maximum mobility, $\mu$, and minimum effective mass, $m^*$. The phonon thermal conductivity is strongly affected by scattering. It decreases with increase of the atomic mass and with the reduction of the hardness of the material. For ionic compounds, $\kappa_{ph}$ has a lower value than for covalent compounds. When excess, or impurity, atoms become ionized in the lattice, giving up their electrons to be free, they produce electronic conductivity; on the other hand, when impurity atoms pick up electrons from the band filled with valence electrons, hole conductivity is produced.

The following are A. F. Ioffe’s qualitative criteria for minimizing the lattice thermal conductivity: 1) The thermal conductivity would be expected to be minimum in crystals, consisting of heavy atoms or ions, with a low Young’s modulus. 2) The minimum thermal conductivity can be expected to occur in crystals with a large thermal expansion coefficient. 3) In nondegenerate semiconductors, the wavelength of electronic waves, at room temperature, $\lambda_{el}$,
is tens of times larger than the interatomic distance ($\lambda_{el} \approx 10^{-6}$ cm), whereas most of the normal lattice vibrations occur at a wavelength, $\lambda_{ph}$, which is of the order of the lattice constant ($\lambda_{ph} \approx 10^{-8}$ cm). This difference between $\lambda_{el}$ and $\lambda_{ph}$ makes it possible to create inhomogeneities in the crystal lattice, which are effective in scattering phonons, but which virtually do not scatter electron waves, i.e., do not reduce the carrier mobility. Such inhomogeneities can, in particular, be produced by the introduction into the lattice of neutral impurity atoms, or by the formation of solid solutions, based on chemical compounds crystallizing in similar lattices.

The lattice thermal conductivity, $\kappa_{ph}$, depends on the phonon scattering mechanism. At temperatures above the Debye temperature: 1) $\kappa_{ph} \propto T^{-1}$ in ideal crystals, i.e., for the scattering of phonons by phonons. 2) $\kappa_{ph}$ is independent of temperature for the scattering of phonons by lattice defects. Intermediate cases may, of course, also occur, with both scattering mechanisms participating; in this case, the numbers of collisions of phonons with phonons, and with defects, i.e., the reciprocals of the mean free path lengths, are additive. The temperature dependence of the numerator, $S^2\sigma$, in the expression for $Z$, is governed by the temperature dependence of the carrier concentration and the mobility. It has been mentioned earlier that, for $S^2\sigma$ to preserve its maximum value over a wide temperature range, the carrier concentration should rise gradually with temperature: namely $n \propto T^{3/2}$. However, there are no substances in nature which follow such a law and, therefore, the materials for thermocouple arms consist, as a rule, of semimetals, in which the carrier concentration is constant. Depending on the working temperature range, a suitable number of donors or acceptors is introduced into the semimetal, so as to fulfill the condition
approximately, within this particular temperature range. Therefore, the temperature dependence of $S^2\sigma$ in semimetals is entirely governed by the temperature dependence of the mobility.
Chapter 2

Minimizing Thermal Conductivity

2.1 Atomic Weight and Melting Point

First, one considers pure elements and compounds, and one directs attention to the so-called high-temperature region, over which lattice thermal conductivity varies as $T^{-1}$. This region usually extends below room temperature and, for most materials of interest, even below liquid nitrogen temperature. It may be assumed that phonon-phonon scattering is dominant under the conditions stated.

Some success in predicting high-temperature lattice thermal conductivity was achieved by Leibfried and Schlömann\textsuperscript{8}. By using a variational method, they found that

$$\kappa_L = 3.5 \left( \frac{k_B}{h} \right)^3 \frac{1}{MV^3 \theta D^3} \frac{\gamma^2 T}{\theta} \right)$$

(2.1.1)

where $M$ is the average mass per atom, $V$ is the average atomic volume and $\gamma$ is the Grüneisen parameter, which enters into the equation of state for solids and, together with the expansion coefficient, is a measure of the anharmonicity of lattice vibrations.

One can obtain Leibfried and Schlömann’s formula, or rather one that differs from it only in the value of the numerical constant, by the following simple means: One starts by using Dugdale and MacDonald’s argument that there should be a relationship between $\kappa_L$ and the thermal expansion coefficient, $\alpha_T$, because both depend on the anharmonicity of interatomic forces. Dugdale and MacDonald\textsuperscript{19} showed that anharmonicity can be represented by the dimensionless quantity, $\alpha_T \gamma T$, and they suggested that the phonon free path length, $l_\nu$, should be
approximately equal to the lattice constant, $a$, divided by this quantity. Thus, by substitution in Eq. (A2.2.6)

$$\kappa_L = \frac{c_V a v}{3\alpha_T \gamma T} \quad (2.1.2)$$

One may eliminate $\alpha_T$ from this equation by using the Debye equation of state

$$\alpha_T = \frac{\chi \gamma c_V}{3} \quad (2.1.3)$$

where $\chi$ is the compressibility. The speed of sound, $v$, and $\chi$ are also related to the Debye temperature, $\theta_D$, through the equation

$$v = (\rho_d \chi)^{1/2} = \frac{2 k_B a \theta_D}{h} \quad (2.1.4)$$

where $\rho_d$ is the density. Then, if $a$ is set equal to the cube root of the atomic volume,

$$\kappa_L = 8 \left(\frac{k_B}{h}\right)^3 \frac{M}{\gamma^2} \frac{1}{\theta_D^3} \quad (2.1.5)$$

which may be compared with Eq. (2.1.1). Because one is not concerned with the value of the numerical constant here, one may use either Eq. (2.1.1) or (2.1.5), and we opt for the latter.

Neither Eq. (2.1.2) nor (2.1.5) is as useful as one might hope in predicting the value of $\kappa_L$. One of the relationships requires a knowledge of the expansion coefficient and the speed of sound, whereas the other involves $\theta_D$. Keyes\textsuperscript{20} showed that a more useful formula could be obtained by using further approximations. His starting point was Lawson's\textsuperscript{21} equation for thermal conductivity,

$$\kappa_L = \frac{a}{3 \gamma^2 T \chi^{3/2} \rho_d^{1/2}} \quad (2.1.6)$$

which is easily obtained from Eq. (2.1.2) or (2.1.5) by substitution of Eq. (2.1.4). Keyes then
eliminated the compressibility by use of the Lindemann melting rule,

\[ \chi = \frac{\varepsilon_m V}{RT_m} \quad (2.1.7) \]

where \( R \) is the gas constant and \( T_m \) is the melting temperature. This rule is based on the assumption that a solid melts when the amplitude of the lattice vibrations reaches a fraction, \( \varepsilon_m \), of the lattice constant, \( \varepsilon_{m, \text{m}} \) being approximately the same for all substances. Keyes found that

\[ \kappa_L T \cong B \frac{T_m^3 \rho_d^2}{A^6} \quad (2.1.8) \]

where

\[ B = \frac{3}{R^2} \frac{\gamma^2 \varepsilon_m^3 N_A^{-1/3}}{2} \quad (2.1.9) \]

with \( N_A \) being Avogadro’s number and \( A \) being the mean atomic weight. The factor, \( B \), involves only the universal constants, \( R \) and \( N_A \), and two quantities, \( \gamma \) and \( \varepsilon_m \), that are unlikely to change by much from one material to another. The variables, \( T_m, \rho_d \) and \( A \), that appear in Eq. (2.1.8) for thermal conductivity are, of course, known as soon as a material is first synthesized, so that, if the equation is valid, it is a most useful aid for predicting \( \kappa_L \).

By collecting data for a wide variety of dielectric crystals, Keyes was able to show that Eq. (2.1.8) is always satisfied to within an order of magnitude, if \( B \) is set equal to \( 3 \times 10^{-4} \) SI units. If one selects different values of \( B \) for covalent and ionic crystals, much better agreement is obtained. Thus, one may use \( B \) equal to \( 1.3 \times 10^{-3} \) SI units for covalent materials, and \( 1.5 \times 10^{-4} \) SI units for ionic materials. Goldsmid\textsuperscript{72} has found good agreement between the predictions of the Keyes rule, Eq. (2.1.8), and the experimental data for a variety of semiconductors, using a value of \( 6 \times 10^{-4} \) SI units for \( B \).
It should be noted that elements or compounds of high mean atomic weight tend to have low melting temperatures. Their densities are also usually high, but one expects that $\rho d^{2/3} A^{-7/6}$ should decrease as the atomic weight rises. Thus, the Keyes rule supports the prediction of Goldsmid and Douglas $^{22}$ that materials of high mean atomic weight should have a low thermal conductivity and should be chosen for thermoelectric applications.

Ioffe and Ioffe $^{23}$ also drew attention to the dependence of thermal conductivity on atomic weight, and demonstrated a difference between covalent and ionic crystals that is consistent with the different values of B that are preferred for the respective types of material. One should avoid inferring that ionic materials are likely to be better than covalent semiconductors in thermoelectric energy conversion, on account of their lower range of thermal conductivity. Ionic compounds invariably have very much smaller carrier mobilities, and the factor, $\mu(m^*)^{3/2}/\kappa L$, appears to be highest for semiconductors in which covalent bonding is predominant.

### 2.2 Semiconductor Solid Solutions

Ioffe et al $^{24}$ first made the interesting and useful suggestion that solid solutions or alloys might be superior to elements or compounds in their thermoelectric properties. It was argued that, in forming a solid solution between isomorphous crystals, the long-range order would be preserved, so that the charge carriers, with their comparatively long wavelengths, would not suffer a reduction in their mobility. On the other hand, phonons that contribute most to the transport of heat have a much shorter wavelength and would be effectively scattered by the short-range disorder, thus reducing the thermal conductivity.

Shortly afterwards, Airapetyants et al $^{25}$ proposed a refinement of the above principle. They considered that electrons would be more strongly scattered by disturbances in the electropositive sublattice of a compound, whereas holes would be more strongly scattered by
disturbances in the electronegative sublattice. On this basis, positive and negative thermoelectric materials should be produced from solid solutions in which there is disorder only in the electronegative and electropositive sublattices, respectively.

Although there is some experimental evidence to support these views, they cannot be regarded as firmly established. Indeed, they would seem to imply that there should be strong scattering of both electrons and holes in solid solutions between elemental semiconductors, whereas, in fact, silicon-germanium alloys have proved most useful as high-temperature thermoelectric materials\textsuperscript{26}.

Whether or not account should be taken of the principle of Airapetyants et al, it is clear that the reduction of the lattice thermal conductivity, in certain solid solutions of both elements and compounds, outweighs any fall in the mobility of the appropriate charge carriers. Therefore, it is important to consider the mechanism by which phonon scattering occurs in such materials.

According to the Rayleigh theory, the scattering cross-section, $\sigma$, for point defects can be expressed as

$$\sigma = \frac{4\pi}{9} c^6 q_L^4 \left( \frac{\Delta \chi}{\chi} + \frac{\Delta \rho_d}{\rho_d} \right)^2$$

(2.2.1)

where $c$ is the linear dimension of the defect, $q_L$ is the magnitude of the phonon wave vector, $\Delta \chi$ is the local change of compressibility and $\Delta \rho_d$ is the local density change. The applicability of Rayleigh scattering has been discussed by Klemens\textsuperscript{27}. Strictly speaking, the dimensions of the defect should be much smaller than the wavelength of the radiation. For higher frequency phonons, however, this is not true, even if $c$ is no larger than the interatomic spacing. Nevertheless, the use of the Rayleigh formula can be justified, because the short-wavelength phonons are so strongly scattered by point defects, that they contribute very little to thermal conductivity. Most of the heat transport then arises from phonons for which the Rayleigh theory
is applicable. The same argument allows us to use the Debye model for lattice vibrations of alloys with more confidence than would be the case for simple elements or compounds.

Although we might expect normal processes to have some influence on thermal conductivity, even at high temperatures (see Appendix A2), we can obtain a good idea of the effect of alloy scattering by assuming that the umklapp processes are predominant, \( k_0 \to 0 \). Eq. (A2.3.16), therefore, becomes

\[
\frac{\kappa_L}{\kappa_0} = \frac{\omega_o}{\omega_D} \tan^{-1} \left( \frac{\omega_D}{\omega_o} \right) = \frac{\omega_o}{\omega_D} \arctan \left( \frac{\omega_D}{\omega_o} \right)
\]

(2.2.2)

where \( \omega_o / \omega_D \) is defined by Eq. (A2.3.18). Here, \( \kappa_L \) is the thermal conductivity of the solid solution and \( \kappa_0 \) is that of the so-called virtual or pure crystal; \( \kappa_0 \) is the lattice thermal conductivity that the solid solution would possess if it were perfectly ordered, that is, if point-defect scattering were absent. The value of \( \kappa_0 \) can be obtained by linear interpolation between the observed thermal conductivities of the components of the solid solution, if no better estimate can be made. Note that, for group IV elements, elastic constants vary inversely with the fourth power of the lattice spacing; this enables estimating \( \theta_D \) accurately for any particular alloy, and certainly assists in predicting \( \kappa_0 \) and the speed of sound, \( v \).

Now, Eq. (2.2.1) shows that the phonons are scattered by variations of both elasticity and density. The density, or mass-fluctuation, effect is easier to determine theoretically and might, in fact, predominate in the type of alloy in which we are interested. In this case, the parameter, \( A \), in Eq. (A2.3.18) is given by

\[
A = \frac{\pi}{2 \nu^3 N} \sum_i x_i (M_i - \bar{M})^2 \bar{M}^2
\]

(2.2.3)

where \( x_i \) is the concentration of unit cells of mass \( M_i \), \( \bar{M} \) is the average mass per unit cell and \( N \)
is the number of unit cells per unit volume.

The additional effect of scattering due to elasticity fluctuations, or strain scattering, is more difficult to predict. However, we can outline the approach. A foreign atom changes local compressibility, because it has bonds different from those of the host atoms, and also because it does not fit properly into a lattice site, thus straining the crystal. Based on the elastic continuum model, an impurity atom of width, $\delta'_i$ (in its own lattice) will distort the space it occupies from one having a width, $\delta$, in the host lattice to one of width, $\delta_i$, where

$$\frac{\delta_i - \delta}{\delta} \equiv \frac{\Delta \delta_i}{\delta} = \frac{\mu}{(1 + \mu)} \frac{\delta'_i - \delta}{\delta}$$  \hspace{1cm} (2.2.4)$$

and

$$\mu = \frac{(1 + \nu_p)G_1}{2(1 + 2\nu_p)G}$$  \hspace{1cm} (2.2.5)$$

$G$ and $\nu_p$ are the bulk modulus and Poisson’s ratio for the host crystal, respectively, and $G_1$ is the bulk modulus of the impurity crystal. Klemens has shown that the value of $A$ in Eq. (A2.3.18), for strain scattering, is given by

$$A_S = \frac{\pi}{N} \sum_i x_i \left( \frac{\Delta G_1}{G} - 6.4\gamma \frac{\Delta \delta_i}{\delta} \right)^2$$  \hspace{1cm} (2.2.6)$$

where $\Delta G_1 \equiv G_1 - G$ and $\gamma$ is the Grüneisen parameter.

The relative importance of mass-fluctuation scattering and strain scattering can be judged from experimental data compared to a solid curve, calculated from Eq. (2.2.2), where $\omega_p/\omega_o$ is determined using only the mass-fluctuation parameter, $A_M$, given by Eq. (2.2.3). It is seen that, for several of the solid solution systems, the experimental data lie close to the theoretical curve. However, there is considerable disagreement in other systems. It is noteworthy that, where such disagreement occurs, the observed thermal conductivity is less than predicted theoretically on the
basis of mass-fluctuation scattering alone. Then, it is reasonable to suggest that the discrepancies can be attributed to a significant strain scattering effect.28

Chapter 3

Electron Transport in a Zero Magnetic Field

3.1 Single Bands

We are now in a position to derive general expressions for the thermogalvanomagenetic coefficients. We shall restrict ourselves, however, to the case of zero magnetic field. It will also be supposed that we are dealing with one type of carrier, electrons or holes, residing in a single energy band of the parabolic form described by Eq. (A3.1.7). It is also assumed that although the charge carriers interact with the phonons, they have no significant effect on the distribution of the latter. In other words, phonons are regarded merely as scattering centers.

The Boltzmann equation for the steady state, which represents a balance between the effects of the fields and the scattering processes, yields

\[
\frac{f(E) - f_0(E)}{\tau_e} = -\frac{e}{k} \frac{\partial f(E)}{\partial k} - \frac{e}{\tau} \frac{\partial f(E)}{\partial \vec{r}}
\]  

(3.1.1)

where \(f\) and \(f_0\) are the perturbed and unperturbed distribution functions, respectively, and \(\vec{k}\) and \(\vec{r}\) are the wave vector and the position vector of the charge carriers. We must solve this equation to obtain \(f-f_0\), in terms of the electric field and the temperature gradient. It will be assumed that the conductor is isotropic and that the fields and flows are all in the direction of the \(x\)-axis.

If \(|f-f_0| \ll f_0\), one can replace \(f\) by \(f_0\) on the right-hand side of Eq. (3.1.1). Then, bearing in mind that the Fermi distribution will be dependent on displacement through both the electric
field and the temperature gradient, we may show that

$$\frac{f(E) - f_0(E)}{\tau_e} = u \frac{\partial f_0(E)}{\partial E} \left( \frac{\partial \zeta}{\partial x} + \frac{E - \zeta}{T} \frac{\partial T}{\partial x} \right)$$

(3.1.2)

where $u$ is the velocity of the carriers in the $x$-direction. It may be noted that this equation will apply equally well for holes and electrons, provided that the carrier energy, $E$, and the Fermi energy, $\zeta$, are measured from the appropriate band edge and in the appropriate direction. Thus, electron energies are measured upwards from the bottom of the conduction band, and hole energies downwards from the top of the valence band.

The number of carriers per unit volume in the range of energy from $E$ to $E+dE$ is $f(E)g(E)dE$, with $g(E)$ given by Eq. (A3.1.7). Since the carriers, with charge $\mp e$, move in the $x$-direction with a velocity, $u$, the electric current density is

$$i = \mp \int_0^\infty euf(E)g(E) dE$$

(3.1.3)

where henceforth the upper sign in such an equation refers to electrons and the lower sign to holes. Similarly, the rate of flow of heat per unit cross-sectional area, $w$, is

$$w = \int_0^\infty u(E - \zeta)f(E)g(E) dE$$

(3.1.4)

since $E-\zeta$ represents the total energy transported by a carrier. It will be noted that the upper limit to the integrals in Eqs. (3.1.3) and (3.1.4) suggests that the bands are infinitely wide, but, in practice, the function, $f(E)$, always falls to zero before $E$ becomes very large.

Since there can be no flow of charge or heat when $f=f_0$, it is appropriate to replace $f$ by $f-f_0$ in the above two equations. Also, in any case of interest to us, the thermal velocity of the carriers will always be very much greater than any drift velocity. Thence, we may set
\[ u^2 = \frac{2E}{3m^*} \]  

(3.1.5)

since \( u^2 \) will be one third of the mean square velocity corresponding to the energy, \( E \). By inserting Eq. (3.1.5) into Eqs. (3.1.3) and (3.1.4), we find

\[
i = \mp \frac{2e}{3m^*} \int_0^\infty g(E) \tau E \frac{\partial f_0(E)}{\partial E} \left( \frac{\partial \zeta}{\partial x} + \frac{E - \zeta}{T} \frac{\partial T}{\partial x} \right) dE
\]

(3.1.6)

and

\[
w = \pm \frac{\zeta}{e} \frac{2}{3m^*} \int_0^\infty g(E) \tau E^2 \frac{\partial f_0(E)}{\partial E} \left( \frac{\partial \zeta}{\partial x} + \frac{E - \zeta}{T} \frac{\partial T}{\partial x} \right) dE
\]

(3.1.7)

To find an expression for the electrical resistivity or conductivity, \( \sigma \), we set the temperature gradient, \( \partial T / \partial x \), equal to zero. Also the electric field is given by \( \pm \left( \partial \zeta / \partial x \right) e^{-1} \).

Thence

\[
\sigma = \frac{i}{E} = -\frac{2e^2}{3m^*} \int_0^\infty g(E) \tau E \frac{\partial f_0(E)}{\partial E} dE
\]

(3.1.8)

Alternatively, if the electric current is zero, Eq. (3.1.6) shows that

\[
\frac{\partial \zeta}{\partial x} \int_0^\infty g(E) \tau E \frac{\partial f_0(E)}{\partial E} dE + \frac{1}{T} \frac{\partial T}{\partial x} \int_0^\infty g(E) \tau E (E - \zeta) \frac{\partial f_0(E)}{\partial E} dE = 0
\]

(3.1.9)

This is the specified condition for the definition of both the Seebeck coefficient and the thermal conductivity. The Seebeck coefficient is defined by

\[
S = E \left( \frac{\partial T}{\partial x} \right)^{-1} = \pm \left( \frac{\partial \zeta}{\partial x} \right) e^{-1} \left( \frac{\partial T}{\partial x} \right)^{-1} = \pm \frac{\partial \zeta}{\partial x} \left( e \frac{\partial T}{\partial x} \right)^{-1} = \pm \frac{1}{e} \frac{\partial \zeta}{\partial x} \frac{\partial T}{\partial x} = \pm \frac{1}{e} \frac{\partial \zeta}{\partial T}
\]

(3.1.10)

Now, substituting Eq. (3.1.10) into Eq. (3.1.9), we arrive at the following, more specific, definition of the thermoelectric power:
\[
S = \pm \frac{1}{eT} \left[ \zeta - \int_0^\infty g(E) \tau_e E^2 \frac{\partial f_o(E)}{\partial E} \, dE \right]
\]

(3.1.11)

It turns out that the Seebeck coefficient is always negative if the carriers are electrons, and positive if the carriers are holes.

The electronic thermal conductivity is given by \(-\nu(\partial T / \partial x)^t\) and is found from Eqs. (3.1.7) and (3.1.9). Thence

\[
\kappa_e = \frac{2}{3m^*T} \left[ \int_0^\infty g(E) \tau_e E^2 \frac{\partial f_o(E)}{\partial E} \, dE - \int_0^\infty g(E) \tau_e E^3 \frac{\partial f_o(E)}{\partial E} \, dE \right]
\]

(3.1.12)

The Peltier coefficient, \(\Pi\), can be found either from the Kelvin relation: \(\Pi = ST\), in terms of the Seebeck coefficient, \(S\), or as the ratio, \(w/i\), when \(\partial T / \partial x = 0\). The same value, of course, results from each derivation.

The integrals that are found in Eqs. (3.1.8), (3.1.11) and (3.1.12) have the same general form. They may be expressed conveniently as

\[
K_S = -\frac{2T}{3m^*} \int_0^\infty g(E) \tau_e E^{s+1} \frac{\partial f_o(E)}{\partial E} \, dE = -\frac{8\pi}{3} \left( \frac{2}{\hbar^2} \right)^{\frac{3}{2}} (m^*)^{\frac{1}{2}} T \tau_o \int_0^\infty E^{s+r+\frac{3}{2}} \frac{\partial f_o(E)}{\partial E} \, dE
\]

(3.1.13)

where \(g\) and \(\tau_e\) have been eliminated in terms of \(m^*, r\) and \(\tau_o\), using Eqs. (A3.1.7) and (A3.2.2).

On integrating by parts, one finds that

\[
\int_0^\infty E^{s+r+\frac{3}{2}} \frac{\partial f_o(E)}{\partial E} \, dE = -\left( s + r + \frac{3}{2} \right) \int_0^\infty E^{s+r+\frac{1}{2}} f_o(E) \, dE
\]

(3.1.14)
Thence we may write

\[ K_S = \frac{8\pi}{3} \left( \frac{2}{h^2} \right)^{\frac{3}{2}} (m^*)^{\frac{1}{2}} T \tau_0 \left( s + r + \frac{3}{2} \right) (k_B T)^{\left( s + r + \frac{3}{2} \right)} F_{s+r+\frac{1}{2}} \]  

(3.1.15)

where

\[ F_n(\xi) = \int_0^\infty \xi^n f_o(\xi) \, d\xi \]  

(3.1.16)

and use is made of the reduced energy, \( \xi \), equal to \( E/k_B T \), as a variable.

The expressions for the transport coefficients in terms of the integrals, \( K_S \), are

\[ \sigma = \frac{e^2}{T} K_0 \]  

(3.1.17)

\[ S = \pm \frac{1}{eT} \left( \zeta - \frac{K_1}{K_0} \right) \]  

(3.1.18)

and

\[ \kappa_e = \frac{1}{T^2} \left( K_2 - \frac{K_1^2}{K_0} \right) \]  

(3.1.19)

It is these equations that allow expressing the figure of merit, \( Z \), in terms of \( \zeta \), \( m^* \), the relaxation time parameters, \( \tau_0 \) and \( r \), and \( \kappa_L \). The lattice conductivity is involved because the total thermal conductivity, \( \kappa \), is given by

\[ \kappa = \kappa_L + \kappa_e \]  

(3.1.20)

3.2 Degenerate and Nondegenerate Conductors

In general, it is necessary to use numerical methods to determine the transport properties from Eqs. (3.1.17) to (3.1.20). However, if the Fermi energy is either much greater than or much less than zero, it is possible to use simple approximations to the Fermi distribution function.
First, we shall consider the nondegenerate approximation that is applicable when 
\( \zeta / k_B T \ll 0 \), that is, the Fermi level lies within the forbidden gap, well away from the 
appropriate band edge. The approximation is very good when \( \zeta \approx -4k_B T \), and is acceptable for 
many purposes when \( \zeta \approx -2k_B T \). It is necessary that the Fermi level should be much further 
from the opposite band edge, if minority carriers are to be neglected. From a practical point of 
view, the independent variable is the carrier concentration, because this can usually be controlled 
directly in a semiconducting material. Here, it will be more convenient to regard the Fermi 
energy as the variable quantity.

When \( \zeta / k_B T \ll 0 \), the Fermi-Dirac integrals become

\[
F_n(\eta) = \exp(\eta) \int_0^\infty \xi^n \exp(-\xi) \, d\xi = \exp(\eta) \Gamma(n+1) \tag{3.2.1}
\]

where the gamma function has the property

\[
\Gamma(n+1) = n \Gamma(n) \tag{3.2.2}
\]

When \( n \) is an integer, \( \Gamma(n+1) = n! \), and the gamma function can be calculated for half-integral 
values of \( n \) from the fact that \( \Gamma\left(\frac{1}{2}\right) = \pi^{\frac{1}{2}} \).

In terms of the gamma function, the transport integrals take the form

\[
K_S = \frac{8\pi}{3} \left(\frac{2\hbar}{m^*}\right)^\frac{3}{2} T \tau_0 (k_B T)^{\frac{s+r+\frac{3}{2}}{2}} \Gamma\left(s + r + \frac{5}{2}\right) \exp(\eta) \tag{3.2.3}
\]

Thus, the electrical conductivity of a nondegenerate semiconductor is

\[
\sigma = \frac{8\pi}{3} \left(\frac{2\hbar}{m^*}\right)^\frac{3}{2} e^2 (m^*)^\frac{1}{2} \tau_0 (k_B T)^{r+\frac{3}{2}} \Gamma\left(r + \frac{5}{2}\right) \exp(\eta) \tag{3.2.4}
\]

It is convenient to express electrical conductivity as
\[ \sigma = n e \mu \] (3.2.5)

where \( n \) is the carrier concentration and \( \mu \) is the mobility, which is the average drift speed of the carriers in a unit electric field. The concentration of the carriers is given by

\[ n = \int_0^\infty f(E)g(E)\,dE = 2\left(\frac{2\pi m^* k_B T}{\hbar^2}\right)^{\frac{3}{2}} \exp(\eta) \] (3.2.6)

so that the mobility is

\[ \mu = \frac{4}{3\pi^2} \Gamma\left(r + \frac{5}{2}\right)e\tau_o \frac{(k_B T)^f}{m^*} \] (3.2.7)

One of the consequences of the nondegenerate approximation is that the mobility in Eq. (3.2.7) does not directly depend on the Fermi energy. It must be remembered that \( \tau_o \) and \( r \) may well depend on the carrier concentration, unless lattice scattering predominates.

It is seen from Eq. (3.2.6) that the carrier concentration has the value that one would expect if there were just \( 2\left(2\pi m^* k_o T / \hbar^2\right)^{\frac{3}{2}} \) states located at the band edge. Thus, this quantity is known as the effective density of states. It has different values for different bands because it depends on \( m^* \).

The Seebeck coefficient of a nondegenerate semiconductor from Eqs. (3.1.18) and (3.2.3) is

\[ S = \pm \frac{k_B}{e} \left[ \eta - \left(r + \frac{5}{2}\right) \right] \] (3.2.8)

One may regard \(-\eta\) as the reduced potential energy of the carriers, and \( (r + 5/2) \) as the reduced kinetic energy that is transported by the current. Thus, the Peltier coefficient, equal to \( ST \), represents the total energy transport per unit charge.
It is conventional to describe $\kappa_{e}$ in terms of the Lorenz number, $L$, defined as $\kappa_{e} / \sigma \tau$.

Then, from Eqs. (3.1.17) and (3.1.19),

$$L = \frac{1}{e^{2}T^{2}} \left( \frac{K_{2} - K_{1}^{2}}{K_{0} - K_{0}^{2}} \right)$$

(3.2.9)

When the nondegenerate approximation is applicable

$$L = \left( \frac{k}{e} \right)^{2} \left( r + \frac{5}{2} \right)$$

(3.2.10)

We see that $L$ is independent of the Fermi energy, provided that the exponent, $r$, in the energy dependence of the relaxation time is constant.

Now we turn to the degenerate condition when $\zeta / k_{B}T >> 0$. This means that the Fermi level is well above the conduction-band edge for electrons or well below the valence-band edge for holes. In other words, the conductor is metallic. The Fermi-Dirac integrals are expressed in the form of a rapidly converging series

$$F_{n}(\eta) = \frac{e^{n+1}}{n+1} + n\pi^{2} / 6 + n(n-1)(n-2)\eta^{n-3} \frac{7\pi^{4}}{360} + ....$$

(3.2.11)

In the degenerate approximation, one uses only as many of the terms in the series as needed to yield a finite, or nonzero, value for the appropriate parameter.

The electrical conductivity of a degenerate conductor is found by employing only the first term in the series,

$$\sigma = \frac{8\pi}{3} \left( \frac{2}{h^{2}} \right)^{3} e^{2}(m^{*})^{1/2} \tau_{o}^{3/2} \zeta^{r - 3/2}$$

(3.2.12)

On the other hand, if only the first term in Eq. (3.2.11) were employed the Seebeck coefficient would be zero, which is consistent with the fact that most metals have exceedingly small values of $S$. To obtain a nonzero value for the Seebeck coefficient, the first two terms are used. Thus,
\[
S = \frac{\pi^2}{3} \frac{k_B}{e} \left( \frac{r + \frac{3}{2}}{\eta} \right) \tag{3.2.13}
\]

The first two terms are also needed to obtain the Lorenz number, which is given by

\[
L = \frac{\pi^2}{3} \left( \frac{k_B}{e} \right)^2 \tag{3.2.14}
\]

This shows that the Lorenz number should be the same for all metals and, in particular, it should not depend on the scattering law for the charge carriers. These features agree fully with the well-established Wiedemann-Franz-Lorenz law which states that all metals have the same ratio of thermal to electrical conductivity and that this ratio is proportional to the absolute temperature.

The only significant exceptions are for some pure metals at very low temperatures and for certain alloys in which the small value of \( \kappa_e \) means that the lattice contribution cannot be neglected.
Chapter 4

The Role of Solid-State Chemistry in the Discovery of New Thermoelectric Materials

4.1 Introduction

The challenge in any effort to discover new thermoelectric materials lies in achieving simultaneously high electronic conductivity, high thermoelectric power and low thermal conductivity in the same solid. These properties define the dimensionless thermoelectric figure of merit, $ZT = \left( \frac{S^2 \sigma}{\kappa} \right) T$, where $S$ is the thermoelectric power, $\sigma$ the electronic conductivity, $\kappa$ the thermal conductivity and $T$ the temperature. The first three quantities are determined by the details of the electronic structure and scattering of charge carriers (electrons or holes), and thus are not independently controllable parameters. The thermal conductivity, $\kappa$, has a contribution from lattice vibrations, $\kappa_L$, which is called the lattice thermal conductivity. Thus, $\kappa = \kappa_e + \kappa_L$, where $\kappa_e$ is the carrier thermal conductivity.

There seem to be several possible ways one may go about increasing $ZT$. One is to minimize $\kappa_L$ while retaining good electronic and thermoelectric power properties. Although there are several approaches to minimizing $\kappa_L$, the most intriguing is that associated with the concept of a “phonon glass electron crystal” (PGEC) that was suggested by Slack$^{30,31}$ as the limiting characteristic for a superior thermoelectric material. A material that is a PGEC features cages (or tunnels) in its crystal structure inside which reside atoms small enough to “rattle.” This situation produces a phonon damping effect that results in drastic reduction of the solid’s lattice...
thermal conductivity. In this picture, a loosely bound atom with a large thermal parameter scatters phonons much more strongly than electrons, thus permitting a glasslike thermal conductivity to coexist with the high electron mobilities found in crystals. The thermal conductivity can also be decreased by introducing large nonperiodic mass fluctuations in the crystal lattice, through solid solutions (i.e., alloy scattering), or by increasing the lattice period (i.e., large unit cell parameters), thus providing short mean path lengths for the heat-carrying phonons.

Another way to maximize ZT would be to increase the thermoelectric power and electronic conductivity without drastically increasing the total thermal conductivity. This is a difficult problem, because if the electronic conductivity is greatly increased, according to the Wiedeman-Franz (WF) law, \( \kappa \) also increases. Based on WF law, the electronic conductivity scales linearly with the carrier thermal conductivity, \( \kappa_e \), and thus very high electronic conductivities (i.e., >2000 (ohm cm)\(^{-1}\)) are ineffectual, even undesirable, in a good thermoelectric material. What is the maximum conductivity that can be tolerated for a room temperature thermoelectric material, without causing the electronic thermal conductivity, \( \kappa_e \), to be overly high? Let us first consider the total thermal conductivity at room temperature of the best currently known thermoelectric materials, that is, Bi\(_2\)Te\(_{3-x}\)Se\(_x\) and Bi\(_{2-x}\)Sb\(_x\)Te\(_3\) alloys, which is 1.5 Wm\(^{-1}\)K\(^{-1}\). About 0.7 Wm\(^{-1}\)K\(^{-1}\) of this value is due to the electronic thermal conductivity, \( \kappa_e \), that arises from an electronic, or electrical, conductivity of \( \sim 1000 \) (ohm cm)\(^{-1}\). Therefore for every 1000(ohm cm)\(^{-1}\) increase in electrical conductivity, the total thermal conductivity is burdened with an additional 0.7Wm\(^{-1}\)K\(^{-1}\) (at room temperature). This places severe constraints on the magnitude of the lattice thermal conductivity, if the total amount is to remain in the neighborhood of 1.0-2.0Wm\(^{-1}\)K\(^{-1}\). Based on these considerations, we estimate that an optimum
electronic conductivity between ~1000 and ~1400(ohm cm)$^{-1}$ is required. For a room temperature ZT~2, and a $\kappa$ value of ~1.5Wm$^{-1}$K$^{-1}$, we obtain a target range for the thermoelectric power values: $260 \leq |S| \leq 440 \muVK^{-1}$. Currently, bulk materials with these physical characteristics remain elusive. The sought-after materials almost certainly have to be novel with new compositions or structures and, therefore, they have to be discovered through exploratory chemical synthesis. By every assessment, the solution to this problem seems to be essentially a chemical one.

The least understood problem is how to increase the thermoelectric power of a material without depressing the electrical, or electronic, conductivity and how to predict precisely which materials will have very large thermoelectric power. Boltzmann transport theory describes both electronic and thermal transport in the vast majority of solids. This theory provides a general understanding of the thermoelectric power that is expressed in the Mott equation:

$$S = \frac{\pi^2 k^2 T}{3e} \left. \frac{d \ln \sigma(E)}{dE} \right|_{E = \zeta}$$

(4.1)

$\sigma(E)$ is the electronic, or electrical, conductivity determined as a function of band filling or Fermi energy, $E = E_F = \zeta$. If the electronic scattering is independent of energy, then $\sigma(E)$ is just proportional to the density of states (DOS) at $E$. In the general case, $S$ is a measure of the variation in $\sigma(E)$ above and below the Fermi surface, specifically through the logarithmic derivative of $\sigma$ with respect to $E$. Since the thermoelectric power of a material is a measure of the asymmetry in electronic structure and scattering rates near the Fermi level, we should aim to produce complexities in either or both in a small energy interval (a few $k_B T$) near $\zeta$. What insights can be gleaned from considering the implications of the Mott equation? How can the logarithmic derivative of $\sigma(E)$ (essentially DOS) be maximized in a real chemical compound?
Although there is no straightforward answer, crude insights include the search for compounds with complex structures and compositions, so that they have a good chance of possessing complex electronic structure. Another class of materials with a good probability of possessing a high $d \ln \sigma / dE$ is mixed-valent compounds, particularly those of f elements.33-37

Additional guidance to the experimenter comes from electronic band structure considerations. For example, the following parameters are critical: actual bandgap, the shape and width of the bands near the Fermi level, and the carrier effective masses and mobilities as well as the degeneracy of band extrema (i.e., the number of valleys in the conduction, or peaks in the valence, band) 38. It is often emphasized that the presence of a large number of valleys in the bands tends to significantly increase the dimensionless thermoelectric figure of merit, ZT (mainly through increasing the thermoelectric power). This is true if the carrier scattering between valleys is minimized or absent. Compounds with a large number of valleys are typically those with high symmetry crystal structures. Systems with many degenerate extrema, characterized by a degeneracy parameter $\gamma$, have higher thermoelectric power than those with a single extremum. This is because, for the same total carrier concentration, the concentration in each pocket is smaller for a larger $\gamma$. This increases the value of S associated with each pocket compared to the value obtained for the single band case, because S increases with decreasing pocket carrier concentration. The amount of increase, however, depends on $\gamma$, the temperature, bandgap and other band parameters. If the carrier mobilities associated with each pocket are the same, then the total conductivity is independent of $\gamma$, but the conductivity-weighted thermoelectric power for the multiband case is larger than for the single-band case 39-40.

For an anisotropic three-dimensional single-band case and band degeneracy of $\gamma$, when the thermal and electrical currents travel in the same direction, $x$, Hicks and Dresselhaus 39 have
shown that the figure of merit, ZT, increases with a parameter, B, defined as

$$B = \gamma \left( \frac{2k_B T}{\hbar^2} \right)^{3/2} \sqrt{\frac{m_x m_y m_z}{e \kappa_L}} \mu_x$$  \hspace{1cm} (4.2)

where $m_i$ is the effective mass of the carriers (electrons or holes) in the $i^{th}$ direction, $\mu_x$ is the carrier mobility along the transport direction, and $\kappa_L$ is the lattice contribution to the thermal conductivity. Thus, in order to increase the value of Z, large effective masses, high carrier mobility and low lattice thermal conductivity are necessary. It has been shown that semiconductors with a bandgap of approximately $10k_BT$ best satisfy this criterion, a property that can be easily determined from appropriate band-structure calculations. A detailed analysis of the electronic structure of a semiconducting compound (obtained through appropriate quantum-mechanical calculations) can give information about the gap, the degeneracies of the conduction and valence band extrema (i.e. $\gamma$), and the effective mass parameters. Equation (4.2) also suggests that while high mobility along the current flow direction is needed, high effective masses need not occur along the same crystallographic directions. Therefore, highly anisotropic structures may be very suitable for high ZT, by combining high $\mu$ along the “good” direction with very large effective masses associated with other directions.

The challenge in thermoelectrics research is to incorporate all the desirable features associated with the charge and thermal transport in a single solid-state material. The following is a review of the criteria and the types of compounds that might be suitable.

**4.2 Necessary Criteria for Thermoelectric Materials**

Based on the preceding discussion, materials suitable for thermoelectric applications fall mainly into two categories: semiconductors and mixed-valent compounds, although some semimetals may also be viable. Many research groups are also investigating whether multiple
quantum wells in semiconductors will improve thermoelectric behavior\textsuperscript{39,42-44}. The improvements for the electronic properties so far seem small, but quantum wells may have potentially a much smaller thermal conductivity than bulk counterparts, which may eventually make them useful for devices.

Metals generally have small Seebeck coefficients, but S has a large contribution from spin fluctuations near the Kondo temperature. Several intermetallic mixed valent compounds, such as CePd\textsubscript{3} and YbAl\textsubscript{3}, have the largest value of the power factor, S\textsuperscript{2}\sigma, among all known materials\textsuperscript{45,46}. Their large thermal conductivity prevents them from having a large ZT, yet mixed-valent materials have the ideal shape of the density of states for producing large values of S, according to the Mott equation given earlier.

For semiconductors, the best materials have the following properties:
1. Electronic bands near the Fermi level, with many valleys, preferably away from the Brillouin zone boundaries. This requires high symmetry.
2. Elements with large atomic number and large spin-orbit coupling\textsuperscript{40}.
3. Compositions with more than two elements (i.e., ternary, quaternary compounds).
4. Low average electronegativity differences between elements\textsuperscript{47,48}.
5. Large unit cell sizes.
6. Energy gaps equal to 10k\textsubscript{B}T, where T is the operating temperature of the thermoelectric material. For room temperature operation, this should be

\[ 0 < E_g < 0.30\text{eV}. \]

These basic criteria, if satisfied, should give rise to high carrier mobility (criteria 1 and 4), low thermal conductivity (criteria 2, 3 and 5) and large thermoelectric power (criteria 1 and 6). The last criterion, 6, suggests that low temperature thermoelectric materials (T<300K) have
very small bandgaps. For operation at higher temperatures (e.g. power generation applications), higher band gaps must be used. A more extensive discussion of criteria for thermoelectric compounds has been elaborated in the past by Slack $^{30,31}$.

### 4.3 Solid-State Chemistry as a Tool for Targeted Materials Discovery

The critical role of solid-state chemistry and physics in modern technology is firmly established. Solid-state compounds have been the foundation of the chemical industry and metallurgy, as well as the entire electronics industry for many years, and many emerging technologies, such as thermoelectrics, will hinge on developments in the discovery of new materials with new or enhanced properties. In this context, the importance of exploratory solid-state synthesis cannot be overemphasized.

Traditionally, the use of high temperatures to directly combine elements or simpler compounds into more complex ones has been quite successful in providing new materials, however, they often give rise to important synthetic limitations. For example, the reactions almost always proceed to the most thermodynamically stable products and the high energies involved often leave little room for kinetic control. These thermodynamically stable products are typically the simplest of binary or ternary compounds, and because of their high lattice stability, they become synthetic obstacles. Second, the high reaction temperatures also dictate that only the simplest chemical building blocks can be used, that is, elements on the atomic level. Attempts to synthesize using molecules of known structure are doomed because the high temperatures used sunder all bonds and reduce the system to atoms rushing to a thermodynamic minimum. Hence, multinary compounds can be more difficult to form, the preference lying with the more stable binary and ternary compounds. Being almost totally at the mercy of
thermodynamics, the solid-state chemist has traditionally relied on experience and intuition, rather than a set of predictable rules.

Over the past decade, work has been underway on many fronts to move away from the high temperatures of classical solid-state synthesis toward techniques that take advantage of lower reaction temperatures. The expectation here is that moderate temperature will favor the stabilization of more complex compounds that otherwise may be unstable with respect to disproportionation to simpler compounds. One technique is chemical vapor deposition (CVD) or molecular beam epitaxy. Here the synthesis of solid-state compounds in various technologically useful forms or shapes proceeds by the intimate gas-phase mixing of volatile precursors leading to deposition of solid-state materials on various substrates. This technique is intended for the synthesis of known solid-state compounds rather than the discovery of new ones, although there is no real reason why new ones (including metastable compounds) could not be discovered in this way. Also noteworthy is the low-temperature solid state metathesis technique, which achieves rapid synthesis of pure binary and ternary chalcogenides. The hydro(solvo)thermal technique has proven advantageous in that polyatomic building blocks have been used in the synthesis of solid-state frameworks. This method uses solvents heated in closed containers above their boiling point (but below their critical point) as reaction media. Solubility of reactants is increased by virtue of the unusually high temperature and pressure within the container. Diffusion and crystal growth are further enhanced by the inclusion of mineralizers, species that can aid in the solvation and reprecipitation of solid state reactants and products, analogous to the function of the transport agent in chemical vapor transport crystal growth. Solvothermal synthesis has been the method of choice for the synthesis of a large variety of solid-state materials and eventually may find uses in thermoelectric materials discovery.
A highly successful and most relevant approach has been performing reactions using molten salts as solvents. Such media have been employed for well over 100 years for high-temperature single crystal growth. Although many salts are high-melting species, eutectic combinations of binary salts and salts of polyatomic species often have melting points well below the temperatures of classical solid-state synthesis, making possible their use in the exploration of new chemistry at intermediate temperatures. In many cases, such salts act not only as solvents, but also as reactants, providing species and building blocks that can be incorporated into the final product. In the search for new thermoelectric materials, this molten salt method has been found to be most suitable for exploratory synthesis involving heavy elements, such as Ba, Sr, Bi, Pb, Sn, Se and Te, and alkali atoms. Since the known materials that are used in thermoelectric devices are mainly chalcogenide compounds, investigators are searching for more complex semiconducting compounds of this type, particularly for low temperature applications.

We are not merely interested in new compounds that are substitutions and variations of known structures, but in entirely new structure types. If significantly enhanced thermoelectric properties are to be found, new materials must become available. Therefore, novel types of syntheses must be explored that allow for higher ZTs. Since the electrical properties of solids are directly dependent on their crystal structure, we are motivated to look for new materials with new lattice structures.

4.4 Thermoelectric Materials Discovery

Structural and compositional complexity can result in corresponding complexities in the electronic structure that may produce the required large asymmetry in the density of states, DOS, (see Equation (4.1)) to bring about large thermoelectric power. The phonon contribution to the
thermal conductivity can also be lowered by such structural complexity, by choosing heavy elements as components of the material and by choosing combinations of elements that normally make moderate to weak chemical bonds.

The concept that certain materials can conduct electricity like a crystalline solid, but heat like a glass, is very useful. In these materials, a weakly bound atom or molecule, called a “rattler,” is used to lower the thermal conductivity of the solid without severely affecting electronic conduction. This can lead to improved thermoelectric efficiency 58-61.

The class of chalcogenide materials described here tends to satisfy this description because, as will become apparent later, they are made of three-dimensional or two-dimensional bismuth-chalcogenide frameworks, stabilized by weakly bonded alkali atoms that reside in cavities, tunnels, or galleries of the framework. These electropositive atoms almost always possess the highest thermal displacement parameters in the structure, which is evidence that a certain degree of “rattling” is present 62. This feature is very important in substantially suppressing the thermal conductivities of these materials. These materials also incorporate other beneficial features, such as large unit cells and complex compositions.

For chalcogenide materials discovery, the use of molten alkali metal polychalcogenides, of the type A$_2$Q$_x$ (A = alkali metal, Q = S, Se, Te), as solvents, is very appropriate as has been demonstrated already 63. As solvents for intermediate temperature reactions, A$_2$Q$_x$ salts are especially well suited because the melting points range between 200°C and 600°C. Most alkali polytelluride salts melt between 300°C and 500°C. Low-melting A$_2$Q$_x$ fluxes remain nonvolatile over a wide temperature range, and so, once above the melting point, reaction temperatures can be varied considerably without concern for solvent loss. Polychalcogenide fluxes are highly reactive toward metals because they are strong oxidants. Reactions between metals and molten
A$_2$Q$_x$ are performed in situ. The powdered reagents (polychalcogenide and metal or metal chalcogenide) are mixed under inert atmosphere and loaded into reaction vessels of either Pyrex or silica. Once evacuated, the tubes are sealed under vacuum and subjected to the desired heating program in a computer-controlled furnace. In a given A-M-Q system, large composition ranges are explored simultaneously (many synthesis runs at a given time), so that all possible phases can be found or identified. At the end of the experiment, and before product isolation, the tubes can be inspected with an optical microscope to assess the existence of promising reactions. In this sense, this is a “poor man’s” combinatorial method. Because this is a flux technique, it means that only the part of the phase space that is rich in flux can be explored. Metal-rich compositions will have to be investigated in other ways.

To synthesize new compounds, one or more metals are added directly to the molten A$_2$Q/Q reaction mixture and heated in a sealed pyrex or silica container. Crystalline productes either precipitate from the melt or form on slow cooling of the melt, depending on the specific stoichiometric and processing conditions. Presumably, the nucleated species are in equilibrium with the soluble intermediates, especially if the flux is present in excess, and hence a solvation/reprecipitation effect (often referred to as the mineralizer effect) occurs. This aids in the growth of single crystals, because the flux can redissolve small or poorly formed crystallites and then reprecipitate the species onto large, well-formed, crystals. The advantage of the flux method is that one allows the system to end up “where it wants,” in the kinetic or thermodynamic sense, without attempting to force upon it a certain stoichiometry or structure. Provided the temperature and time are appropriate, the reaction systems have all the ingredients and freedom to form a new phase. The benefit of this becomes apparent from the unusual compositions often found in the new materials that, most certainly, could not have been predicted “a priori.”
The molten flux approach is suitable for quick and wide explorations of phase space but is most practical when carried out in small (<2g) quantities. In this sense, it is a “discovery” method not a production method. Of course, once a new material has been discovered by such techniques, the net step is to attempt to prepare it on a large scale. Therefore, once the composition of a new material has been established, other synthesis methods amenable to scale-up must then be employed.

What are the empirical guidelines with which we have to design thermoelectric materials, and how do we go about choosing a particular system for exploration? We know that heavy atoms are desirable because they tend to vibrate slowly, giving rise to low-frequency lattice phonons that help slow down heat transfer through a material, leading to low thermal conductivity. The fact that Bi₂Te₃ is the best material known to date suggests that it combines many of the features necessary for a high figure of merit, which is true for refrigeration applications only. We note here that even Bi/Sb alloys show some promise for thermoelectric applications. Therefore, we reason that if there is anything particular, or special, about bismuth that simultaneously gives rise to high electronic conductivity and thermoelectric power, it could be manifested in other compounds of bismuth as well. Because of this, we think a sensible approach is a research direction exploring other multinary chalcogenides of bismuth and, by extension, antimony, in the hope that these elements will impart some, or all, of the key properties needed for superior thermoelectric performance. Furthermore, structurally and compositionally more complex bismuth chalcogenides would, most likely, have a low lattice thermal conductivity for the reasons outlined earlier. Based on these considerations, exploratory chemical syntheses were performed involving Bi as one of the elements, as reported in an extensive literature. The results have shown that promising new materials can be found.
other elements employed in the synthesis are chalcogens, such as S, Se and Te, as well as alkali metals.

The presence of alkali metals in the structures of ternary and quaternary bismuth chalcogenides induce the stabilization of covalently bonded Bi-chalcogen frameworks with cages or tunnels that accommodate the charge-balancing alkali atoms. The interactions of alkali metals and the Bi-chalcogen frameworks are considered to be mainly electrostatic in nature. Alkaline earth metals, such as Sr and Ba and, by extension, Eu with a +2 charge are also desirable as substitutes for alkali metals. They tend to have similar properties to alkali metals, but their greater charge tends to stabilize different structures.

4.5 Conclusions and Outlook

In conclusion, the future of thermoelectrics relies heavily on new materials and, consequently, solid-state chemistry has an important role to play in the field of thermoelectrics. The flux method is an excellent discovery technique that has produced a large number of interesting and promising new materials. This is why one of the most important long-term projects in the thermoelectrics community should be the exploratory synthesis and study of a large range of heavy element multinary compounds. We note that maximum ZT will probably not be discovered in as-made, unintentionally doped, material (barring a stroke of luck). It will come from finding a new material with promising enough thermoelectric properties to begin with, so that optimization from those levels would result in great values of ZT. These new materials have not been optimized, whereas the best samples of Bi$_2$Te$_{3-x}$Se$_x$ and Bi$_{2-x}$Sb$_x$Te$_3$ have undergone decades of development. It is believed that some of the materials discussed here may indeed emerge as competitive ones in the future, if their development is seriously pursued. The approach to performance improvement needs to be twofold: alloying and doping. To optimize a
new material, such as CsBi₄Te₆, it must be kept in mind that one is dealing with a “multidimensional parameter space.” This is because not only must solid solutions be made, but also different dopants have to be tried, and different concentrations for each dopant. If, for example, there are 5 different solid solution members to be made and 5 different dopants to be investigated in 5 concentrations each, then in the CsBi₄Te₆ system alone, we need to test at least 125 samples. At this stage, measuring these samples seems to be the slowest step in the process of evaluation. Therefore, new high-throughput thermoelectric screening methods need to be devised that are capable of handling the number of samples available.

From the chalcogenide materials described here, we know most about CsBi₄Te₆ and β-K₂Bi₈Se₁₃. The rest need further investigation, and we hope that some of them will emerge as promising thermoelectric candidates. Since it is almost certain that the “as prepared” materials are not in their optimum doping state, we can expect substantial enhancements in the power factor with doping.

Where do we go from here? What other types of compounds can be targeted for high ZT? Based on the arguments presented earlier, systems with high γ or high B parameters are needed. While high effective masses could come from carriers residing in the flat, low dispersion, regions of bands, as well as in heavy relativistic atoms, large degeneracy in the band extrema, i.e., large γ parameters, can come from compounds with highly symmetric crystal structures, ideally cubic or rhombohedral. The maximum degeneracy arises when the extrema, in the valence or conduction bands, are found in the middle of the Brillouin zone rather than at the edges. One needs not only large degeneracy in the band extrema, but also a large number of extrema of similar energy in the same band near the Fermi level. Fig. (6) shows a hypothetical band structure for such an ideal complex cubic material with the needed characteristics. None of
the materials presented earlier have a high crystal symmetry. Therefore, the $\gamma$ parameter is low. Their promising thermoelectric properties presumably come from high carrier effective masses and inherently low thermal conductivity. One direction to pursue in the future is the development of cubic compounds in the A/Pb/Bi/Q ($A = \text{alkali metal}$; $Q = S, \text{Se}, \text{Te}$) system. Inroads have already been made and the results therefrom reported on in the literature $^{64}$.

![Diagram](image)

**FIG. 6.** One possibility of a generic yet simplistic band structure of a “nearly ideal” thermoelectric material. The high number of valleys in the conduction band or peaks in the valence band, coupled with a high band degeneracy parameter, $\gamma$ (deriving from high crystal symmetry), could produce enormous power factors.

As the number of atoms increases, the structural complexity increases and makes it difficult to achieve highly symmetric, e.g., cubic, structures. Nevertheless, large cubic cells of complex compounds have been observed, and their existence suggests that many more may be
stable. One salutary example is RE$_{117}$Fe$_{52}$Ge$_{112}$ (RE = Sm, Gd, Tb, Dy, Ho, Er, Tm), which has a primitive cubic unit cell with a=29Å and space group Fm3m$^{65}$. Other examples include phases that are single crystal approximants of quasicrystals$^{66-68}$. If the elemental composition is such that a narrow-gap semiconductor is obtained, the very complex compositions and heavy, electron-rich, elements involved may well give rise to multiple band extrema and, we may hope, high ZT. Future investigations in the discovery and selection of candidate thermoelectric materials should take such considerations into account.

Interest in thermoelectric materials is not new, but the need for new materials is increasing and the next few years will prove critical in the development of this field. Taking into account the difficulty of identifying the “right” compound and optimizing to \( ZT > 1 \), long-term sustainable planning is necessary, and close collaboration among chemists, physicists and engineers is key to success. Such collaborative approaches constitute the model in which all present and future efforts should be based.
Chapter 5

General Review of the Thermoelectric Figure of Merit

The figure of merit for direct thermoelectric, or solid state, energy conversion is given by Altenkirch’s equation

\[ Z = \frac{S^2 \sigma}{\kappa} \]  \hspace{1cm} (5.1)

The dimensionless figure of merit is expressed as

\[ ZT = \frac{S^2 \sigma T}{\kappa} \]  \hspace{1cm} (5.2)

where \( S \) is the Seebeck coefficient, or thermoelectric power, \( \sigma \) is the electrical conductivity, \( \kappa \) is the thermal conductivity and \( T \) is the absolute temperature. The quantity, \( Z \), the thermoelectric figure of merit, has the dimension, \( K^{-1} \), so when it gets multiplied by the absolute temperature, \( T \), then it becomes dimensionless. When we want to increase, or maximize, \( Z \) or \( ZT \), we must either increase the product \( S^2 \sigma \), also known as the power factor, or decrease \( \kappa \), or better still, do both. If we look at the above equations more carefully, we notice that the quantity in the denominator, \( \kappa \), the thermal conductivity, comprises two components, the lattice thermal conductivity, \( \kappa_L \), and the electronic thermal conductivity, \( \kappa_e \). We may thus write

\[ ZT = \frac{S^2 \sigma T}{\kappa_L + \kappa_e} \]  \hspace{1cm} (5.3)

Now, according to the Wiedemann-Franz law, we know that

\[ \frac{\kappa_e}{\sigma} = LT \] \hspace{0.5cm} \text{or} \hspace{0.5cm} \kappa_e = \sigma LT \]  \hspace{1cm} (5.4)

Therefore,
If we can reduce the lattice thermal conductivity to such an extent that it becomes zero, or at least so much smaller than the electronic thermal conductivity, that it becomes negligible, then we obtain

\[
ZT = \frac{S^2 \sigma T}{\kappa_L + \sigma L T}
\]  

(5.5)

or

\[
ZT = \frac{S^2 \sigma T}{\sigma L T} = \frac{S^2}{L}
\]  

(5.6)

Thus, to maximize the thermoelectric figure of merit, \(ZT\), assuming that the lattice thermal conductivity is zero, or negligible, what we need to do is to increase, or maximize, the thermoelectric power, and nothing else. Eqs. (5.6) or (5.7) serve to explain why metals have such a miserable \(ZT\), because the conduction of both heat and electricity in them is substantially electronic and their thermoelectric power, \(S\), is very low. Eqs. (5.6) or (5.7) qualitatively explain the much better \(ZT\) possessed by semiconductors, in general, and that is because of their higher thermoelectric power. Ionic compounds, or insulators, do not conform to these equations, and so they again have a very low figure of merit, even though they exhibit a very high thermoelectric power.

The Seebeck coefficient \(^{69}\) of a nondegenerate semiconductor is

\[
S = \pm \frac{k_B}{e} \left[ \eta - \left( r + \frac{5}{2} \right) \right]
\]  

(5.8)
where \( \eta \) is the reduced Fermi energy, defined by
\[
\eta = \frac{E_F}{k_B T} = \frac{\zeta}{k_B T}
\]
where \( E_F \) is the Fermi energy, \( k_B \) is the Boltzmann constant and \( T \) is the absolute temperature, and \( r \) is the scattering parameter or index that appears in the equation that gives the relaxation time as a function of the energy, namely \( \tau_c = \tau_0 E^r \), where the reduced Fermi energy is defined as negative when measured downwards from the edge of the conduction band or upwards from the edge of the valence band, and where the Fermi energy is positive when measured upwards from the edge of the conduction band and downwards from the edge of the valence band. Thus both \( E_F \) and \( \eta \) acquire a zero value at both edges.

The Lorenz number \(^69\) of a nondegenerate semiconductor is

\[
L = \left(\frac{k_B}{e}\right)^2 \left( r + \frac{5}{2} \right)
\]

In Eq. (5.8), one may regard \( -\eta \) as the reduced potential energy of the charge carriers and \( (r + 5/2) \) as the reduced kinetic energy that is transported by the current. Thus, the Peltier coefficient, equal to \( ST \), represents the total energy transport per unit charge. From Eq. (5.9), we also see that \( L \) is independent of the Fermi energy, provided that the exponent, \( r \), in the energy dependence of the relaxation time is constant. All of the foregoing statements refer to nondegenerate semiconductors.

Substituting Eqs. (5.8) and (5.9) into Eq. (5.6) we obtain

\[
ZT = \frac{\left(\frac{k_B}{e}\right)^2 \left[\eta - \left( r + \frac{5}{2} \right) \right]^2}{\left(\frac{k_B}{e}\right)^2 \left( r + \frac{5}{2} \right)}
\]

thus,
Assuming either acoustic-mode lattice scattering or alloy scattering, we may set \( r = -\frac{1}{2} \). We thus have,

\[
ZT = \frac{\eta^2}{\frac{1}{2} + \frac{5}{2}} - 2\eta - \frac{1}{2} + \frac{5}{2} = \frac{\eta^2}{2} - 2\eta + 2 = \left( \frac{\eta}{\sqrt{2}} - \sqrt{2} \right)^2
\]

(5.12)

Setting \( r = \frac{1}{2} \), which is valid when the bonds are primarily ionic, when scattering is due to the optical modes, when the interaction between the electrons or holes and the optical modes is not very strong and when the temperature is well above the Debye temperature (\( T >> T_D \)), we have

\[
ZT = \frac{\eta^2}{\frac{1}{2} + \frac{5}{2}} - 2\eta + \frac{1}{2} + \frac{5}{2} = \frac{\eta^2}{3} - 2\eta + 3 = \left( \frac{\eta}{\sqrt{3}} - \sqrt{3} \right)^2
\]

(5.13)

Setting \( r = \frac{3}{2} \), which is valid for ionized-impurity scattering, we have

\[
ZT = \frac{\eta^2}{\frac{3}{2} + \frac{5}{2}} - 2\eta + \frac{3}{2} + \frac{5}{2} = \frac{\eta^2}{4} - 2\eta + 4 = \left( \frac{\eta}{2} - 2 \right)^2
\]

(5.14)

Eqs. (5.12), (5.13) and (5.14) are algebraic quadratic equations in \( \eta \). Their roots or eigenvalues have formed the basis for the construction of the right hand side of Table (2) for nondegenerate semiconductors.

Now we turn to the degenerate condition when \( E_F / k_B >> 0 \). This means that the Fermi level is well above the conduction-band edge for electrons or well below the valence-band edge.
for holes. In other words, the conductor is metallic.

The Seebeck coefficient of a degenerate semiconductor is given by

$$S = \mp \frac{\pi^2}{3} \frac{k_B}{e} \left( \frac{r + \frac{3}{2}}{\eta} \right)$$  \hspace{1cm} (5.15)

Eq. (5.15) correctly predicts that when \( \eta \gg 0 \), \( S \) becomes very small which is typical of metallic conductors. Degenerate semiconductors also behave almost the same way. In order for a degenerate semiconductor to be able to generate a reasonably high Seebeck coefficient, the doping level or the free charge carrier concentration must be kept as close as possible to \( 2.5 \times 10^{19} \) carriers cm\(^{-3} \), which is the maximum value or limit allowed for nondegenerate semiconductors.

The Lorenz number of a degenerate semiconductor is

$$L = \frac{\pi^2}{3} \left( \frac{k_B}{e} \right)^2$$  \hspace{1cm} (5.16)

Now, substituting Eqs. (5.15) and (5.16) into Eq. (5.6) yields

$$ZT = \frac{\mp \frac{\pi^2}{3} \frac{k_B}{e} \left( \frac{r + \frac{3}{2}}{\eta} \right)^2}{\frac{\pi^2}{3} \left( \frac{k_B}{e} \right)^2} = \frac{\pi^2}{3} \left( \frac{r + \frac{3}{2}}{\eta} \right)^2$$  \hspace{1cm} (5.17)

Setting \( r = -\frac{1}{2} \) yields

$$ZT = \frac{\pi^2}{3} \frac{1}{\eta^2}$$  \hspace{1cm} (5.18)

Setting \( r = \frac{1}{2} \) yields

$$ZT = \frac{4\pi^2}{3} \frac{1}{\eta^2}$$  \hspace{1cm} (5.19)
Setting \( r = 3/2 \) yields

\[
ZT = \frac{3\pi^2}{\eta^2} \tag{5.20}
\]

Now, we know that the Fermi energy has well established values for many materials, especially elements. Its value for metals and heavily doped degenerate semiconductors can be obtained from the equation

\[
E_F = \frac{\hbar^2}{2m} \left( \frac{3}{8\pi} \frac{N_e}{V} \right)^{\frac{2}{3}} \tag{5.21}
\]

where \( \hbar \) is Planck’s constant, \( m \) is the electronic mass, \( N_e \) is the number of electrons up to an energy, \( E \), and \( V \) is the molar volume. The choice of the units of \( \hbar \) determines the units of \( E_F \). For heavily doped semiconductors, Eq. (5.21) is still approximately valid. It should be modified, however, in order to include the linear dependency of \( \eta \), or \( E_F \), on \( \left( r + \frac{3}{2} \right) \) for degenerate semiconductors. We thus write

\[
E_F = \frac{\hbar^2}{2m} \left( \frac{3}{8\pi} \frac{N_e}{V} \right)^{\frac{2}{3}} \left( r + \frac{3}{2} \right) \tag{5.22}
\]

For semiconductors, of course, this assumes that \( E_F = 0 \) at \( E_C \), i.e., at the bottom or edge of the conduction band. It also assumes that \( E_F = 0 \) at \( E_V \), i.e., at the top or edge of the valence band.

Let us now investigate the Seebeck coefficient, or thermoelectric power, as given by the Mott \(^{32}\) equation

\[
S = \frac{\pi^2}{3} \frac{k_B^2 T}{e} \frac{d \ln \sigma(E)}{dE} \bigg|_{E = E_F} \tag{5.23}
\]

where \( \sigma(E) \) is the electronic conductivity determined as a function of band filling, or Fermi
energy, $E_F$. If the electronic scattering is independent of energy, then $\sigma(E)$ is just proportional to the density of states at $E$. In the general case, $S$ is a measure of the variation in $\sigma(E)$ above and below the Fermi surface, specifically through the logarithmic derivative of $\sigma$ with respect to $E$. There is no straightforward or simple way to calculate the logarithmic derivative of $\sigma(E)$. Therefore, we will follow a different path in our analysis.

Let us calculate the value of $S$ for certain predetermined values of the dimensionless thermoelectric figure of merit, $ZT$, for both degenerate and nondegenerate semiconductors, assuming that the electronic scattering is not independent of energy, i.e., $\tau_e = \tau_0 E^r$. We will thus perform the calculations for $r = -1/2, 1/2$ and $3/2$ as usual. We will also assume the validity of Eq. (5.6). Thus,

$$S^2 = (ZT)L$$  \hspace{1cm} (5.24)

or

$$S = (ZTL)^{1/2}$$  \hspace{1cm} (5.25)

We may thus proceed to determine the thermoelectric power required to achieve certain predetermined values of the dimensionless thermoelectric figure of merit for degenerate and nondegenerate semiconductors. In order to do that, we need to calculate the Lorenz number, bearing in mind that its value is constant for degenerate conductors, whereas for nondegenerate ones, it does depend on $r$.

The results of these calculations are set forth in Table (1), which gives the Seebeck coefficient for various values of the dimensionless thermoelectric figure of merit ranging from 1 to 100 for both degenerate and nondegenerate semiconductors.

The basis for performing the calculations of the thermoelectric power set forth in Table (1) is as follows:
For degenerate semiconductors, namely having a free charge carrier concentration more than 2.5x10¹⁹ carriers cm⁻³, i.e., heavily doped semiconductors, we have used a Lorenz number, \( L = 2.45 \times 10^{-8} V^2 K^{-2} \), which is equal to that for metals, and substituted that value into Eq. (5.25).

For nondegenerate semiconductors, namely having a free charge carrier concentration less than 2.5x10¹⁹ carriers cm⁻³, i.e., moderately doped and lightly doped semiconductors, we have used Lorenz number values of
\[
2 \beta \frac{k_B}{e}, 2 \beta \frac{k_B}{e}, \quad \text{and} \quad 4 \beta \frac{k_B}{e},
\]
equal to 1.48x10⁻⁸, 2.22x10⁻⁸ and 2.96x10⁻⁸ V²K⁻², corresponding to \( r = -1/2, 1/2 \) and 3/2, respectively. These values were once more substituted into Eq. (5.25). Furthermore, the basis for calculating those Lorenz numbers has been Eq. (5.9).

### TABLE 1. Values of the thermoelectric power, \( S, \mu VK^{-1} \).

<table>
<thead>
<tr>
<th>ZT</th>
<th>Degenerate Semiconductors</th>
<th>Nondegenerate Semiconductors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( r )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1/2</td>
</tr>
<tr>
<td>1</td>
<td>157</td>
<td>122</td>
</tr>
<tr>
<td>2</td>
<td>221</td>
<td>172</td>
</tr>
<tr>
<td>3</td>
<td>271</td>
<td>211</td>
</tr>
<tr>
<td>4</td>
<td>313</td>
<td>243</td>
</tr>
<tr>
<td>5</td>
<td>350</td>
<td>272</td>
</tr>
<tr>
<td>10</td>
<td>495</td>
<td>385</td>
</tr>
<tr>
<td>20</td>
<td>700</td>
<td>544</td>
</tr>
<tr>
<td>30</td>
<td>857</td>
<td>666</td>
</tr>
<tr>
<td>40</td>
<td>990</td>
<td>769</td>
</tr>
<tr>
<td>50</td>
<td>1107</td>
<td>860</td>
</tr>
<tr>
<td>60</td>
<td>1212</td>
<td>942</td>
</tr>
<tr>
<td>70</td>
<td>1310</td>
<td>1018</td>
</tr>
<tr>
<td>80</td>
<td>1400</td>
<td>1088</td>
</tr>
<tr>
<td>90</td>
<td>1485</td>
<td>1154</td>
</tr>
<tr>
<td>100</td>
<td>1565</td>
<td>1217</td>
</tr>
</tbody>
</table>

Carrying on with this analysis, we now consider the following equations:
\[ S = \pm \frac{1}{e} \frac{\partial \zeta}{\partial T} = \pm \frac{1}{e} \frac{\partial E_F}{\partial T} = \frac{\partial \zeta}{\partial x} \left[ e \frac{\partial T}{\partial x} \right]^{-1} \] (5.26)

\[ \sigma = \frac{8\pi}{3} \left( \frac{2}{\hbar^2} \right)^{\frac{3}{2}} e^2 (m^*)^{\frac{1}{2}} \tau_0 (k_B T)^{r+\frac{3}{2}} \Gamma \left( r + \frac{5}{2} \right) \exp(\eta) \] (5.27)

where \( \Gamma \) is the gamma function.

\[ m^* = \frac{\hbar^2 k^2}{8\pi^2 E} \] (5.28)

which is the same expression as for a free-electron gas, except that the electronic mass, \( m \), has been replaced by \( m^* \), the effective mass.

\[ (m^*)^{1/2} = \frac{\hbar k}{2\sqrt{2}\pi E^{1/2}} \] (5.29)

\[ S = \frac{\pi^2}{3} \frac{k_B^2 T}{e} \ln \frac{\sigma}{dE} \bigg|_{E = E_F} = \frac{\pi^2}{3} \frac{k_B^2 T}{e} \frac{d\sigma}{\sigma dE} \bigg|_{E = E_F} \] (5.30)

\[ \frac{d\sigma}{dE} \bigg|_{E = E_F} = \frac{3e}{\pi^2 k_B^2} \frac{\sigma S}{T} = \frac{3e}{\pi^2 k_B^2} \frac{\sigma(ZTL)^{1/2}}{T} \] (5.31)

Therefore,

\[ \frac{\ln \sigma}{dE} \bigg|_{E = E_F} = \frac{3e}{\pi^2 k_B^2} \frac{(ZTL)^{1/2}}{T} \frac{k_B}{e} \left( r + \frac{5}{2} \right)^{\frac{1}{2}} = \frac{3e}{\pi^2} \frac{(ZTL)^{1/2}}{T} \frac{1}{r + \frac{5}{2}} \frac{1}{k_B T} \] (5.32)

The above result applies to degenerate semiconductors and assumes the validity of Eq. (5.6).

Thus, Eq. (5.32) gives the logarithmic derivative of \( \sigma \) with respect to \( E \) at the Fermi surface in terms of a constant multiplied by the reciprocal of the product of the Boltzmann constant and the temperature. Substitution of Eq. (5.32) into Eq. (5.30) yields
The Seebeck coefficient is given by a constant multiplied by the Boltzmann constant, divided by the electronic charge. Eq. (5.33) is thus dimensionally correct, since it gives $S$ as $VK^{-1}$, the way it should be. The constants involved in both Eqs. (5.32) and (5.33), actually depend on $ZT$. The higher $S$ is, the higher $ZT$ will be.

One can now use Eq. (5.26) to determine the Fermi energy as a function of the temperature. One thus obtains

$$d\zeta = \pm eSdT$$

$$\int d\zeta = \pm \int eSdT = \pm \int (ZT)^{\frac{1}{2}} \left( r + \frac{5}{2} \right)^{\frac{1}{2}} k_BdT$$

or

$$\zeta = E_F = \pm (ZT)^{\frac{1}{2}} \left( r + \frac{5}{2} \right)^{\frac{1}{2}} k_B T + \text{constant}$$

Now we will take an alternative and different approach in order to obtain $S$, using Eqs. (5.27), (5.28), (5.29) and (5.30). We will thus proceed as follows

$$\frac{d \ln \sigma}{dE} = \frac{\frac{8\pi}{3} \left( \frac{2}{h^2} \right)^{\frac{3}{2}} e^2 \frac{\hbar k}{2\sqrt{2}\pi} \tau_o (k_B T)^{\frac{r}{2}} \Gamma \left( r + \frac{5}{2} \right) \exp(\eta)}{-\frac{1}{2} E^{-1}}$$

consequently,

$$\frac{d \ln \sigma}{dE} |_{E = E_F} = -\frac{1}{2E_F}$$
and finally

\[ S = \frac{\pi^2}{3} \frac{k_B^2 T}{e} \left( -\frac{1}{2E_F} \right) = -\frac{\pi^2}{6} \frac{k_B^2 T}{e} \frac{1}{E_F} \]  

(5.38)

or

\[ S = -\frac{\pi^2}{6} \frac{k_B}{e} \frac{1}{\eta} \]  

(5.39)

where \( \eta = E_F / k_B T \) is the reduced Fermi energy.

It should be pointed out that Eq. (5.27) gives the electrical conductivity for a nondegenerate semiconductor. Eqs. (5.38) or (5.39) thus give the thermoelectric power of a nondegenerate semiconductor, according to Eq. (5.23), established by Mott \(^{32}\).

Proceeding now to investigate the degenerate condition, we know that the electrical conductivity of a degenerate conductor is given by

\[ \sigma = \frac{8\pi}{3} \left( \frac{2}{\hbar^2} \right)^{3/2} e^2 (m^*)^{1/2} \tau \zeta \tau_{\gamma}^{r+3/2} \]  

(5.40)

which, along with Eqs. (5.28) and (5.29), yields

\[ \frac{d \ln \sigma}{dE} = \frac{8\pi}{3} \left( \frac{2}{\hbar^2} \right)^{3/2} e^2 \frac{\hbar k}{2\sqrt{2\pi} \tau \zeta \tau_{\gamma}^{r+3/2}} \left( -\frac{1}{2E^{-1}} \right) \]  

(5.41)

thus

\[ \frac{d \ln \sigma}{dE \bigg|_{E=E_F}} = -\frac{1}{2E_F} \]  

(5.42)

therefore, substituting the above result into Eq. (5.30), yields
\[ S = -\frac{\pi^2 k_B^2 T}{6 e} \frac{1}{E_F} \] (5.43)

or

\[ S = -\frac{\pi^2 k_B}{6 e} \frac{1}{\eta} \] (5.44)

where \( \eta = E_F / k_B T = \zeta / k_B T \) is the reduced Fermi energy. This is the same result as that obtained from Eqs. (5.38) and (5.39). We thus conclude that Eq. (5.30), also known as the Mott\textsuperscript{32} equation, gives the same thermoelectric power for any semiconductor condition, regardless of whether it is degenerate or nondegenerate.

We may now use Eq. (5.6) to obtain the dimensionless thermoelectric figure of merit, \( ZT \), substituting the thermoelectric power, \( S \), as given by the Mott equation, i.e., either Eq. (5.43) or (5.44). The Lorenz number, \( L \), has to be carefully evaluated, however, since we no longer have a unique expression for it, as was the case for \( S \). The only way to obtain the Lorenz number must be based on the Goldsmid analysis, which distinguishes between degenerate and nondegenerate semiconductors. Furthermore, in the case of a nondegenerate semiconductor, the value of \( L \) will also vary, depending on the scattering parameter, \( r \). Thus, for the broad-spectrum of semiconductors in general, we are going to have different values for the Mott \( ZT \), bearing the aforementioned variables in mind.

For degenerate semiconductors, we thus have

\[ ZT = \frac{\pi^4 k_B^2}{36 e^2 \eta^2} = \frac{\pi^2}{12} \frac{1}{\eta^2} \] (5.45)

The above result is the Mott dimensionless thermoelectric figure of merit for degenerate
conductors.

For nondegenerate semiconductors, we obtain the following equations, depending on the respective values of $r$. Thus, for $r = -1/2$

$$ZT = \frac{\pi^4 k_B^2}{36 e^2 \eta^2} \frac{1}{\eta} = \frac{\pi^4}{72} \frac{1}{\eta^2}$$  \hspace{1cm} (5.46)

for $r = 1/2$

$$ZT = \frac{\pi^4 k_B^2}{36 e^2 \eta^2} \frac{1}{\eta} = \frac{\pi^4}{108} \frac{1}{\eta^2}$$  \hspace{1cm} (5.47)

for $r = 3/2$

$$ZT = \frac{\pi^4 k_B^2}{4 e^2 \eta^2} \frac{1}{\eta} = \frac{\pi^4}{144} \frac{1}{\eta^2}$$  \hspace{1cm} (5.48)

We will now use Eqs. (5.12), (5.13) and (5.14), to calculate the reduced Fermi energy, $\eta$, of nondegenerate semiconductors, for appropriate values of the scattering parameter, $r$, and values of the dimensionless thermoelectric figure of merit, $ZT$, ranging from 1 to 100. We will also use Eqs. (5.18), (5.19) and (5.20) to calculate $\eta$ for degenerate semiconductors for the same values of $r$ and $ZT$. All of the aforementioned equations are based on the original formulation by Goldsmid for the Seebeck coefficient and the Lorenz number. The results of these calculations are presented in Table (2). Table (3) contains the corresponding values of $\eta$ based on the Mott formulation.
TABLE 2. Values of the reduced Fermi energy, $\eta$ (Goldsmid).

<table>
<thead>
<tr>
<th>ZT</th>
<th>$\eta$</th>
<th>Degenerate Semiconductors</th>
<th>Nondegenerate Semiconductors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$-1/2$</td>
<td>$1/2$</td>
</tr>
<tr>
<td>1</td>
<td>1.814</td>
<td>3.628</td>
<td>5.441</td>
</tr>
<tr>
<td>2</td>
<td>1.283</td>
<td>2.565</td>
<td>3.848</td>
</tr>
<tr>
<td>3</td>
<td>1.047</td>
<td>2.094</td>
<td>3.142</td>
</tr>
<tr>
<td>4</td>
<td>0.907</td>
<td>1.814</td>
<td>2.721</td>
</tr>
<tr>
<td>5</td>
<td>0.811</td>
<td>1.622</td>
<td>2.433</td>
</tr>
<tr>
<td>10</td>
<td>0.574</td>
<td>1.147</td>
<td>1.721</td>
</tr>
<tr>
<td>20</td>
<td>0.406</td>
<td>0.811</td>
<td>1.217</td>
</tr>
<tr>
<td>30</td>
<td>0.331</td>
<td>0.662</td>
<td>0.993</td>
</tr>
<tr>
<td>40</td>
<td>0.287</td>
<td>0.574</td>
<td>0.86</td>
</tr>
<tr>
<td>50</td>
<td>0.257</td>
<td>0.513</td>
<td>0.77</td>
</tr>
<tr>
<td>60</td>
<td>0.234</td>
<td>0.468</td>
<td>0.702</td>
</tr>
<tr>
<td>70</td>
<td>0.217</td>
<td>0.434</td>
<td>0.65</td>
</tr>
<tr>
<td>80</td>
<td>0.203</td>
<td>0.406</td>
<td>0.608</td>
</tr>
<tr>
<td>90</td>
<td>0.191</td>
<td>0.382</td>
<td>0.574</td>
</tr>
<tr>
<td>100</td>
<td>0.181</td>
<td>0.363</td>
<td>0.544</td>
</tr>
</tbody>
</table>

TABLE 3. Values of the reduced Fermi energy, $\eta$ (Mott).

<table>
<thead>
<tr>
<th>ZT</th>
<th>$\eta$</th>
<th>Degenerate Semiconductors</th>
<th>Nondegenerate Semiconductors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$-1/2$</td>
<td>$1/2$</td>
</tr>
<tr>
<td>1</td>
<td>0.907</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.641</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.524</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.453</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.406</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.287</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.203</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.166</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.143</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.128</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>0.117</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>0.108</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>0.096</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.091</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Now, the following comments are noteworthy regarding Tables (2) and (3). With reference to Table (2), we observe that the values of the reduced Fermi energy, \( \eta \), for degenerate semiconductors are related to one another in the ratio of 1:2:3, corresponding to \( r = -1/2 \), \( 1/2 \) and \( 3/2 \), respectively, and that is, of course, for the same ZT value. These \( \eta \) values are all positive and they start either pointing upwards from the edge of the conduction band, or pointing downwards from the edge of the valence band. The zero values of the Fermi energy coincide with either of these two edges, with the forbidden energy band gap being enclosed between them. We thus also observe that as the value of the dimensionless thermoelectric figure of merit, ZT, gets higher and higher, the reduced Fermi energy gets lower and lower. This, obviously, agrees with the fact that metals, which have a negligible ZT, also possess a Fermi energy that is much higher than that of semiconductors. Thus, the only way to get an infinite ZT is to reduce the reduced Fermi energy to zero. Thus, when performing the calculations for the reduced Fermi energy, \( \eta \), only the positive answers were accepted, while the negative ones were rejected.

Obviously, therefore, all Fermi energy levels for degenerate semiconductors must lie either within the conduction band or the valence band, they cannot be in the forbidden band gap. In contrast, the reduced Fermi energy levels for nondegenerate semiconductors are all located within the forbidden energy band gap, outside both the conduction and valence bands. These \( \eta \) values, or levels, are all negative and they either start pointing downwards from the edge of the conduction band, or pointing upwards from the edge of the valence band. Thus the zero values of the Fermi energy for nondegenerate semiconductors must coincide with either the edge of the conduction band or the edge of the valence band. For a semiconductor to be considered nondegenerate, it is necessary that \( E_F/k_B T << 0 \), i.e., the Fermi level lies within the forbidden gap well away from the appropriate band edge. The nondegenerate approximation is very good.
when $E_F < -4k_B T$ and is still acceptable for many purposes when $E_F < -2k_B T$. It is necessary that the Fermi level should be much further from the opposite band edge, if minority carriers are to be neglected. In order to assure this, it is, therefore, necessary that we choose or design materials having a sufficiently wide forbidden energy band gap. This fact is crucially important, if we want to develop materials with superior thermoelectric properties and performance.

Table (2) contains values of the reduced Fermi energy, $\eta$, for different values of $r$ and $ZT$, for both degenerate and nondegenerate semiconductors. The calculations of the $\eta$ values for degenerate semiconductors were rather straightforward, whereas those for nondegenerate semiconductors involved solving and obtaining the roots or eigenvalues of forty five algebraic quadratic equations. Only the negative eigenvalues were selected, since that agrees with the original assumption that the semiconductor is nondegenerate and, as such, its Fermi energy can only acquire a negative value.

Table (3) contains the values of $\eta$, obtained according to the Mott formulation. For degenerate semiconductors, there is a single $\eta$ value for each $ZT$ one, regardless of that of $r$. Looking at the Mott data, we clearly see that they can be reconciled with the corresponding ones given in Table (2), for $r = -1/2$, if we multiply the former by a factor of 2. And that is because the thermoelectric power, $S$, obtained according to the Mott formulation, namely Eq. (5.39), is one half that given by the Goldsmid formulation, namely Eq. (5.15), for $r = -1/2$. That would justify multiplying the original Mott equation for $S$ by a correction factor of 2. As for nondegenerate semiconductors, the Mott data leave a lot to be desired. All the $\eta$ values are in flagrant violation of the simple, but very basic, rule that $\eta$ be at least $< -2$, so that the Fermi level within the forbidden band gap be well away from the appropriate band edge. For both of the
aforementioned reasons, Mott’s data and formulation are by no means considered credible or useful. This, of course, leaves Table (2) as the only viable basis for any further research on this subject.

Regardless of the above criticism of Mott’s formulation and data, as expressed by Eq. (5.39) and Table (3), it is probably more constructive to look more closely into the reasons for the discrepancy in the data pertaining to Tables (2) and (3). We are specifically referring to the disagreement between the \( \eta \) values as presented in these two tables, for degenerate semiconductors.

As stated earlier, if the data presented in the left column of Table (3), for degenerate semiconductors, get multiplied by a factor of 2, the discrepancy between them and the corresponding ones given in Table (2), again for degenerate semiconductors, for \( r = -1/2 \), is entirely removed. The two sets of data become identical. This means that the original Mott equation for the Seebeck coefficient must be multiplied by a correction factor of 2. Doing this, Eq. (5.30) then becomes

\[
S = \frac{2\pi^2}{3} \frac{k_B T}{e} \ln \sigma \bigg|_{E = E_F} = \frac{\pi^2}{3} \frac{k_B T}{e} \frac{1}{E_F} = \frac{\pi^2}{3} \frac{k_B}{e} \frac{1}{\eta} \tag{5.49}
\]

Having introduced the above correction factor, yields

\[
S_M = S_G = \frac{\pi^2}{3} \frac{k_B}{e} \frac{1}{\eta} \tag{5.50}
\]

or

\[
\eta_M = \eta_G = \frac{\pi^2}{3} \frac{k_B}{e} \frac{1}{S} = \frac{\pi^2}{3} \frac{k_B}{e} \frac{1}{(ZTL)^{1/2}} \tag{5.51}
\]

where M and G denote Mott and Goldsmid, respectively. Applying the aforementioned correction factor implies that the earlier Eqs. (5.45), (5.46), (5.47) and (5.48) must also be
modified and rewritten. The new equations are thus expressed as follows:

For degenerate semiconductors

\[
ZT = \frac{\frac{\pi^4}{9} \frac{k_B^2}{e^2} \frac{1}{\eta^2}}{\frac{\pi^2}{3} \frac{k_B^2}{e^2}} = \frac{\pi^2}{3} \frac{1}{\eta^2} = \frac{5.52}{3\eta^2}
\]

For nondegenerate semiconductors, with \( r = -1/2 \)

\[
ZT = \frac{\frac{\pi^4}{9} \frac{k_B^2}{e^2} \frac{1}{\eta^2}}{\frac{2}{3} \frac{k_B^2}{e^2}} = \frac{\pi^4}{18} \frac{1}{\eta^2} = \frac{5.53}{18\eta^2}
\]

For nondegenerate semiconductors, with \( r = 1/2 \)

\[
ZT = \frac{\frac{\pi^4}{9} \frac{k_B^2}{e^2} \frac{1}{\eta^2}}{\frac{3}{2} \frac{k_B^2}{e^2}} = \frac{\pi^4}{27} \frac{1}{\eta^2} = \frac{5.54}{27\eta^2}
\]

For nondegenerate semiconductors, with \( r = 3/2 \)

\[
ZT = \frac{\frac{\pi^4}{9} \frac{k_B^2}{e^2} \frac{1}{\eta^2}}{\frac{4}{3} \frac{k_B^2}{e^2}} = \frac{\pi^4}{36} \frac{1}{\eta^2} = \frac{5.55}{36\eta^2}
\]

The above four equations can be further worked out to give the reduced Fermi energy, \( \eta \), directly according to the Mott formulation. We thus have

For degenerate semiconductors

\[
\eta_M = \frac{\pi}{\sqrt{3ZT}} = \frac{5.56}{\sqrt{3ZT}}
\]

For nondegenerate semiconductors, with \( r = -1/2 \)

\[
\eta = \frac{\pi}{\sqrt{2ZT}} = \frac{5.53}{\sqrt{2ZT}}
\]
\[
\eta_M = \frac{-\pi^2}{3\sqrt{2ZT}} \tag{5.57}
\]

For nondegenerate semiconductors, with \( r = 1/2 \)

\[
\eta_M = \frac{-\pi^2}{3\sqrt{3ZT}} \tag{5.58}
\]

For nondegenerate semiconductors, with \( r = 3/2 \)

\[
\eta_M = \frac{-\pi^2}{6\sqrt{ZT}} \tag{5.59}
\]

We may now proceed to introduce all of the above modifications and reconstruct Table (3), which will become Table (4), and which will be included at the end of this discussion.

Considering now Mott’s formula in its modified form, namely Eq. (5.49), we notice from Table (4) that it now gives reliable values for the reduced Fermi energy for degenerate semiconductors for which \( r = -1/2 \). For other values of \( r \), namely 1/2 and 3/2, it does not yield any answers, because it is not related to the scattering parameter, \( r \), in any way whatsoever. Furthermore, despite the modification already made, it still does not give reliable and acceptable values for the reduced Fermi energy of nondegenerate semiconductors, no matter what the value of \( r \) is. It is thus quite evident that Mott’s equation generally does not apply to nondegenerate semiconductors.

We therefore finally conclude that the Mott equation does not yield any quantitatively correct values for the Seebeck coefficient or for the reduced Fermi energy, resulting therefrom, regardless of whether the semiconductor is degenerate or nondegenerate, and regardless of what the value of the scattering parameter or index is, since Mott’s formulation does not take account of the scattering parameter. It is, nevertheless, worthwhile or useful qualitatively, since it provides a good physical picture of the nature of the Seebeck coefficient, presenting it as
fundamentally proportional to the derivative of the electrical conductivity with respect to the energy evaluated at the Fermi surface in the Brillouin space.

TABLE 4. Modified Values of the reduced Fermi energy, \( \eta \) (Mott).

<table>
<thead>
<tr>
<th>ZT</th>
<th>Degenerate Semiconductors</th>
<th>Nondegenerate Semiconductors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( r )</td>
<td>-1/2</td>
</tr>
<tr>
<td>1</td>
<td>1.814</td>
<td>-2.326</td>
</tr>
<tr>
<td>2</td>
<td>1.283</td>
<td>-1.644</td>
</tr>
<tr>
<td>3</td>
<td>1.047</td>
<td>-1.344</td>
</tr>
<tr>
<td>4</td>
<td>0.907</td>
<td>-1.164</td>
</tr>
<tr>
<td>5</td>
<td>0.811</td>
<td>-1.040</td>
</tr>
<tr>
<td>10</td>
<td>0.574</td>
<td>-0.736</td>
</tr>
<tr>
<td>20</td>
<td>0.406</td>
<td>-0.520</td>
</tr>
<tr>
<td>30</td>
<td>0.331</td>
<td>-0.424</td>
</tr>
<tr>
<td>40</td>
<td>0.287</td>
<td>-0.368</td>
</tr>
<tr>
<td>50</td>
<td>0.257</td>
<td>-0.328</td>
</tr>
<tr>
<td>60</td>
<td>0.234</td>
<td>-0.300</td>
</tr>
<tr>
<td>70</td>
<td>0.217</td>
<td>-0.278</td>
</tr>
<tr>
<td>80</td>
<td>0.203</td>
<td>-0.260</td>
</tr>
<tr>
<td>90</td>
<td>0.191</td>
<td>-0.246</td>
</tr>
<tr>
<td>100</td>
<td>0.181</td>
<td>-0.232</td>
</tr>
</tbody>
</table>
Chapter 6

Critical Analysis of the Thermoelectric Figure of Merit as Related to the Wiedemann-Franz and Eucken Laws

It is now appropriate to carry out a critical review of Eq. (5.6) with the objective of determining the conditions under which the equation may or may not be valid. Writing this equation once more, we have

$$\frac{S^2}{L} = \lim_{\kappa_L \to 0} ZT = \lim_{\kappa_e \gg \kappa_L} ZT = \lim_{\kappa_L \to 0} ZT = \lim_{\kappa_e \gg \kappa_L} ZT = (6.1)$$

Although $\kappa_L$ can always be reduced through mass-and volume-fluctuation scattering, or alloy scattering, also known as lattice point-defect scattering, the use of rattlers, quantum wells, superlattices and clathrates, or the use of Glen Slack’s concept of the phonon glass, which is essentially an amorphous material in which crystallinity is completely destroyed through the creation of a generalized disorder in the lattice, yet $\kappa_L$ no matter how small it is, or how great the disorder is, never becomes exactly equal to zero. Rewriting Eq. (6.1), we have

$$\lim_{\kappa_L \to 0} \frac{S^2 \sigma T}{\kappa_L + \kappa_e} = \lim_{\kappa_e \gg \kappa_L} \frac{S^2 \sigma T}{\kappa_L + \kappa_e} = \frac{S^2 \sigma T}{\kappa_e} = \frac{S^2 \sigma T}{\sigma L T} = \frac{S^2}{L} = (6.2)$$

Now considering only the denominator in the above equation, yields

$$\kappa_L + \kappa_e = \kappa_L \left(1 + \frac{\kappa_e}{\kappa_L}\right) = (6.3)$$

In order for $\kappa_e \gg \kappa_L$, then
If \( \frac{\kappa_e}{\kappa_L} >> 1 \), then the limiting value of Eq. (6.3) becomes

\[
\frac{\kappa_e}{\kappa_L} \to 1
\]

(6.5)

Thus, in order to ensure the validity of Eqs. (5.6) or (6.2), we have to assure that inequality (6.4) is satisfied, i.e.,

\[
\frac{\kappa_e}{\kappa_L} >> 1
\]

The above inequality, although apparently simplistic, yet provides a powerful means or criterion that enables us to investigate the domain of validity of Eq. (5.6) on a rather rigorous basis.

We will now proceed to derive a theoretical expression giving the value of the ratio between the electronic and lattice components of the thermal conductivity. In doing so, we will invoke the following two laws, which are well known in physics:

1. The Wiedemann-Franz law, already referred to several times in this dissertation.
2. The Eucken law, which states that for pure dielectric crystals the thermal conductivity varies inversely with the absolute temperature. Debye found that by taking account of the anharmonic nature of forces in real crystals, thermal conductivity takes on finite values. In effect, the elastic constants of a substance vary from point to point according to the local displacements of the atoms. Variation in either the density or the elastic properties are responsible for scattering the vibrational waves and, thus, for limiting the value of the thermal conductivity. Peierls showed how the anharmonicity of the lattice vibrations could be taken into account by introducing the concept of phonon wave packets that arise from the quantization of
vibrational waves. These wave packets are now simply referred to as phonons, and they may be regarded as the energy carriers that are responsible for heat conduction by the lattice. A limit on thermal conductivity is set by phonon collisions that do not conserve momentum. The thermal conductivity that we are referring to is actually the phonon or lattice component of the total thermal conductivity. Thus, if $\kappa_L$ is used to denote the lattice thermal conductivity, then it may be expressed as

$$\kappa_L = AT^{-1}$$  \hspace{1cm} (6.6)

where $A$ is a constant. The above equation is Eucken’s law for the lattice thermal conductivity, and holds when $T > \theta_D$. It is known to be particularly true for semiconductors in which the chemical bonding is of the covalent-ionic type, specifically alloys or solid solutions of intermetallic compounds in which mass and volume fluctuation scattering, also known as alloy scattering, prevails, and is normally used to reduce the lattice thermal conductivity.

Let us now take a look at the electronic component of the thermal conductivity. According to the Wiedemann-Franz law, it is expressed as

$$\kappa_e = \sigma LT$$  \hspace{1cm} (6.7)

The electrical conductivity, $\sigma$, for nondegenerate semiconductors is given by Eq. (5.27), which comprises a number of parameters, or quantities, that are outright constants, as well as other parameters which are not exactly constants, yet may be regarded as such, according to the way we are carrying out this analysis. We are explicitly trying to find out the degree of dependency of the electrical conductivity on the two variables that are crucially important in controlling the thermoelectric performance of semiconducting materials. These are the absolute temperature, $T$, and the free charge carrier concentration, $n$. To enable us to do so, we invoke Eq. (5.21) and use it to express the reduced Fermi energy in terms of the concentration of the free charge carriers.
and the temperature even though we know that it is only valid for metals and approximately valid for heavily doped degenerate semiconductors. This will be our preliminary, or first, attempt to establish a theoretical correlation of the electronic to lattice thermal conductivity ratio as a function of free charge carrier concentration and temperature.

Thus, the electrical conductivity for nondegenerate semiconductors can be expressed as follows

\[
\sigma = \frac{8\pi}{3} \left(\frac{2}{\hbar^2}\right)^{3/2} e^2 (m^*)^{1/2} \tau_0 (k_B T)^{r + 3/2} \Gamma \left( r + \frac{5}{2} \right) \exp \left[ \frac{h^2}{2m} \left( \frac{3}{8\pi} \right)^{2/3} \frac{n^{2/3}}{k_B T} \right] 
\]

(6.8)

\[
= \frac{8\pi}{3} \left(\frac{2}{\hbar^2}\right)^{3/2} e^2 (m^*)^{1/2} \tau_0 (k_B T)^{r + 3/2} \Gamma \left( r + \frac{5}{2} \right) \exp \left[ \frac{h^2}{8mk_B} \left( \frac{3}{\pi} \right)^{2/3} \frac{n^{2/3}}{T} \right] 
\]

(6.9)

\[
= \frac{8\pi}{3} \left(\frac{2}{\hbar^2}\right)^{3/2} e^2 (m^*)^{1/2} \tau_0 (k_B T)^{r + 3/2} \Gamma \left( r + \frac{5}{2} \right) \exp \left[ A \frac{n^{2/3}}{T} \right] 
\]

(6.10)

where

\[
A = \frac{h^2}{8mk_B} \left( \frac{3}{\pi} \right)^{2/3} = \left( \frac{6.62618 \times 10^{-34} \times 10^{31} \times 10^{23} \times 10^4}{8 \times 9.10953 \times 1.38066} \right)^{2/3} = 4.23156822 \times 10^{-11} \text{m}^2\text{K}
\]

(6.11)

The exponential term in the above Eqs. (6.8), (6.9) or (6.10) actually stands for \( e^\eta \). Its magnitude has been calculated for various values of the free charge carrier concentration, \( n \), and the temperature, \( T \), and the results are given in Table (5). The corresponding values of the reduced Fermi energy, \( \eta = \frac{E_F}{k_B T} \), have also been calculated and presented in Table (6).
TABLE 5. Values of the exponential term, $e^{-\eta}$, in Eq. (6.8) for nondegenerate semiconductors.

<table>
<thead>
<tr>
<th>$T$ (K)</th>
<th>$n_{10^{18}}$ carriers cm$^{-3}$</th>
<th>$n_{10^{19}}$ carriers cm$^{-3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>1.151485</td>
<td>1.924675</td>
</tr>
<tr>
<td>500</td>
<td>1.088316</td>
<td>1.481202</td>
</tr>
<tr>
<td>1000</td>
<td>1.043224</td>
<td>1.217047</td>
</tr>
</tbody>
</table>

TABLE 6. Reduced Fermi energy, $\eta$, according to Eqs. (6.8)-(6.10), as well as Table (5), for nondegenerate semiconductors.

<table>
<thead>
<tr>
<th>$T$ (K)</th>
<th>$n_{10^{18}}$ carriers cm$^{-3}$</th>
<th>$n_{10^{19}}$ carriers cm$^{-3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>0.1411</td>
<td>0.6548</td>
</tr>
<tr>
<td>500</td>
<td>0.0846</td>
<td>0.3929</td>
</tr>
<tr>
<td>1000</td>
<td>0.0423</td>
<td>0.1964</td>
</tr>
</tbody>
</table>

Now, a critical study of the above Eqs., (6.8)-(6.11), reveals the following:

1. There is a direct proportionality between the electrical conductivity, $\sigma$, and the free charge carrier concentration, $n$. This dependency, however, is somewhat weakened due to the fact that $n$ is raised to the power $2/3$.

2. The quantity, $n$, aside from being so reduced in value, is also included as part of the index of an exponential function, along with a constant, $A$, whose value is enormously small, of the order of $10^{-11}$, which further reduces its effect on the magnitude of the electrical conductivity.

3. Most importantly, the Fermi energy is given by Eq. (5.21) as a positive quantity. This cannot be tolerated or accepted, because we already know that for nondegenerate semiconductors the Fermi energy lies within the forbidden energy band gap, and is always measured downwards from the edge of the conduction band or upwards from the edge of the valence band, and as such it can only have negative values. If we attempt to introduce a negative sign into the argument of the exponential term in Eq. (6.8), then $e^{n}$ will have to move to the denominator of that equation.
In that case, we will have an inverse proportionality between $\sigma$ and $n$, which is completely absurd.

Based on the above arguments, we must discontinue our reliance on Eq. (5.21) to express the reduced Fermi energy as a function of free charge carrier concentration and temperature in Eq. (6.8). We will deal with this issue again further on, using another approach.

We will now proceed to investigate the quantity, $\kappa_e / \kappa_L$, which gives the ratio between the electronic and the lattice components of the total thermal conductivity. In doing so, we invoke Eq. (6.6), which we will rewrite as follows

$$\kappa_L = \left( \kappa_{L_0} T_o \right) T^{-1} \tag{6.12}$$

where $\kappa_{L_0}$ is an initial value of the lattice thermal conductivity, which may be estimated or measured at a certain reference temperature, $T_o$, like the Debye temperature or the room temperature. We will also invoke Eq. (6.7), take account of Eq. (5.27) and combine the latter with Eqs. (6.7) and (6.12). We will thus end up being able to write down definitive expressions for the ratio between the electronic and the lattice thermal conductivity. To explain more clearly what we are doing, we will substitute Eq. (5.27) into Eq. (6.7) to obtain an expression for $\kappa_e$.

Now we will substitute the value of the Lorenz number, $L$, for nondegenerate semiconductors, as given by Eq. (5.9), into the so obtained equation for $\kappa_e$, and then divide that by the corresponding expression for $\kappa_L$ as given by Eq. (6.12). We will also no longer express the index of the exponential term in these equations according to Eq. (5.21), but we will simply write it as $e^\eta$. Consequently, we end up deriving the following equations

$$\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \left( \frac{8\pi}{3} \right) \left( \frac{2}{\hbar^2} \right)^{3/2} e^2 (m^*)^{1/2} \tau_{o} (k_B T)^{r+3/2} \Gamma \left( r + \frac{5}{2} \right) e^\eta \left( \frac{k_B}{e} \right)^2 \left( r + \frac{5}{2} \right) T \tag{6.13}$$
or
\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{\text{L}_o} T_o} \right) \left( \frac{8\pi}{3} \frac{2}{\hbar^2} \right)^{3/2} (m^*)^{1/2} \tau_o (k_B T)^{r+7/2} \Gamma \left( r + \frac{5}{2} \right) e^\eta \left( r + \frac{5}{2} \right) \tag{6.14}
\]

For \( r = -1/2 \)
\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{\text{L}_o} T_o} \right) \left( \frac{16\pi}{3} \frac{2}{\hbar^2} \right)^{3/2} (m^*)^{1/2} \tau_o (k_B T)^3 e^\eta \tag{6.15}
\]

For \( r = 1/2 \)
\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{\text{L}_o} T_o} \right) \left( 64\pi \frac{2}{\hbar^2} \right)^{3/2} (m^*)^{1/2} \tau_o (k_B T)^4 e^\eta \tag{6.16}
\]

For \( r = 3/2 \)
\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{\text{L}_o} T_o} \right) \left( 64\pi \frac{2}{\hbar^2} \right)^{3/2} (m^*)^{1/2} \tau_o (k_B T)^5 e^\eta \tag{6.17}
\]

In the foregoing equations for nondegenerate semiconductors, it is worthwhile to introduce the free charge carrier concentration, \( n \), instead of \( \exp(\eta) \). We thus make use of the relationship
\[
n = \int_0^\infty f(E)g(E)dE = 2 \left( \frac{2\pi m^* k_B T}{\hbar^2} \right)^{3/2} \exp(\eta) \tag{6.18}
\]

which yields
\[
\exp(\eta) = e^\eta = \frac{n}{2} \left( \frac{2\pi m^* k_B T}{\hbar^2} \right)^{3/2} \tag{6.19}
\]

This is preferable from the practical point of view, since the free charge carrier concentration is the independent variable that can be directly controlled, at least for semiconducting materials.

We thus rewrite both Eqs. (5.27) and (6.8), as follows

87
\[ \sigma = \frac{8\pi}{3} \left( \frac{2}{\hbar^2} \right)^{3/2} e^2 (m^*)^{1/2} \tau_o \left( k_BT \right)^{r+3/2} \Gamma \left( r + \frac{5}{2} \right) n \left( \frac{2\pi m^* k_B T}{\hbar^2} \right)^{-3/2} \]  

(6.20)

or

\[ \sigma = \frac{4\pi}{3} ne^2 \tau_o (m^*)^{-1} \left( k_BT \right)^{r} \Gamma \left( r + \frac{5}{2} \right) \pi^{-3/2} \]  

(6.21)

or

\[ \sigma = \frac{4}{3\pi^{1/2}} \Gamma \left( r + \frac{5}{2} \right) ne^2 \tau_o (k_BT)^{r} \left( \frac{m^*}{e} \right)^2 \left( r + \frac{5}{2} \right) \]  

(6.22)

Now invoking Eqs. (5.9), (6.6), (6.7), (6.12) and (6.22), and combining them together, we obtain the following expression for the thermal conductivity ratio for nondegenerate semiconductors

\[ \frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0}/T_0} \right) \frac{4}{3\pi^{1/2}} \Gamma \left( r + \frac{5}{2} \right) ne^2 \tau_o (k_BT)^{r} \left( \frac{m^*}{e} \right)^2 \left( r + \frac{5}{2} \right) \]  

(6.23)

Working out all the algebra, Eq. (6.23) finally yields

\[ \frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0}/T_0} \right) \frac{4}{3\pi^{1/2}} \Gamma \left( r + \frac{5}{2} \right) n\tau_o (k_BT)^{r+2} \left( \frac{m^*}{e} \right)^2 \left( r + \frac{5}{2} \right) \]  

(6.24)

or

\[ \frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0}/T_0} \right) \frac{2(2r+5)}{3\pi^{1/2}} \Gamma \left( r + \frac{5}{2} \right) n\tau_o (k_BT)^{r+2} \left( \frac{m^*}{e} \right)^2 \]  

(6.25)

Alternatively, substituting now Eq. (6.19) into Eq. (6.13) or Eq. (6.14) gives

\[ \frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0}/T_0} \right) \frac{8\pi}{3} \left( \frac{2}{\hbar^2} \right)^{3/2} (m^*)^{1/2} \tau_o \left( k_BT \right)^{r+7/2} \Gamma \left( r + \frac{5}{2} \right) n \left( \frac{2\pi m^* k_B T}{\hbar^2} \right)^{-3/2} \left( r + \frac{5}{2} \right) \]  

(6.26)

Again, working out all the algebra, Eq. (6.26) finally yields

\[ \frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0}/T_0} \right) \frac{4}{3\pi^{1/2}} \Gamma \left( r + \frac{5}{2} \right) n\tau_o (k_BT)^{r+2} \left( \frac{m^*}{e} \right)^2 \left( r + \frac{5}{2} \right) \]  

(6.27)
or

\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \frac{2(2r + 5)}{3\pi^{1/2}} \Gamma \left( r + \frac{5}{2} \right) \frac{n \tau_0 (k_B T)^{r+2}}{m^*} \quad (6.28)
\]

Eqs. (6.25) and (6.26) are identical. We have thus derived the final expression for the thermal conductivity ratio, \( \kappa_e / \kappa_L \), for nondegenerate semiconductors. The first term between parentheses on the right hand side of the above Eqs., (6.14) to (6.17) and (6.24) to (6.28), actually represents the reciprocal of the constant, \( A \), in Eucken’s inverse temperature law, as expressed by Eq. (6.6). We now proceed to derive specific expressions for the thermal conductivity ratio for the three standard denominations of the scattering parameter, or exponent, \( r \), which yields

For \( r = -1/2 \)

\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \frac{8}{3\pi^{1/2}} \frac{n \tau_0 (k_B T)^{3/2}}{m^*} \quad (6.29)
\]

For \( r = 1/2 \)

\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \frac{8}{\pi^{1/2}} \frac{n \tau_0 (k_B T)^{5/2}}{m^*} \quad (6.30)
\]

For \( r = 3/2 \)

\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \frac{32}{\pi^{1/2}} \frac{n \tau_0 (k_B T)^{7/2}}{m^*} \quad (6.31)
\]

The terms in the above Eqs., (6.29), (6.30) and (6.31), can still be better rearranged and the equations written in their definitive forms as follows

For \( r = -1/2 \)
The electrical conductivity of degenerate semiconductors is given by Eq. (5.40).

Invoking now Eq. (5.22) and using it to express the Fermi energy in terms of the concentration of the free charge carriers, we can rewrite Eq. (5.40) as follows

\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \frac{8}{3 \pi^{1/2}} \frac{\tau_o k_B^{3/2}}{m^*} n T^{3/2} \quad (6.32)
\]

For \( r = 1/2 \)

\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \frac{8}{\pi^{1/2}} \frac{\tau_o k_B^{5/2}}{m^*} n T^{5/2} \quad (6.33)
\]

For \( r = 3/2 \)

\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \frac{32}{\pi^{1/2}} \frac{\tau_o k_B^{7/2}}{m^*} n T^{7/2} \quad (6.34)
\]

The thermal conductivity ratio can thus be expressed according to the following equation

\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \left[ \frac{8 \pi}{3} \left( \frac{2}{h^2} \right)^{3/2} e^2 (m^*)^{1/2} \tau_o \left( \frac{2}{2m \left( \frac{3}{8 \pi} \right) n} \right)^{2/3} \right]^{r+3/2} \quad (6.35)
\]

Now taking account of Eq. (5.16), which gives the Lorenz number for degenerate semiconductors, and substituting it into the above equation, and working out all the algebra, we finally obtain
\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \left[ \frac{16 \sqrt{2} \pi^3}{9} h^{2r} \left( m^* \right)^{1/2} \tau_o \left( \frac{1}{2m} \left( \frac{3n}{8\pi} \right)^{2/3} \right)^{r+1/2} \right] (k_B T)^2 \left( r + \frac{3}{2} \right)^{r+1/2} \]  

(6.37)

Proceeding as usual, we are able to derive specific forms of the above equation for the various denominations of the scattering parameter, \( r \)

For \( r = -1/2 \)

\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \left[ \frac{16 \sqrt{2} \pi^3}{9} h \left( m^* \right)^{1/2} \tau_o \left( \frac{1}{2m} \left( \frac{3n}{8\pi} \right)^{2/3} \right)^{2/3} (k_B T)^2 \right] \]

\[
= \left( \frac{1}{\kappa_{L_0} T_0} \right) \left[ \frac{16 \sqrt{2} \pi^3}{9} h \left( m^* \right)^{1/2} \tau_o \left( \frac{1}{8m} \left( \frac{3}{\pi} \right)^{2/3} \right)^{2/3} n \left( k_B T \right)^2 \right] \]  

(6.38)

For \( r = 1/2 \)

\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \left[ \frac{64 \sqrt{2} \pi^3}{9} h \left( m^* \right)^{1/2} \tau_o \left( \frac{1}{2m} \left( \frac{3n}{8\pi} \right)^{2/3} \right)^{2} \right] \]

\[
= \left( \frac{1}{\kappa_{L_0} T_0} \right) \left[ \frac{64 \sqrt{2} \pi^3}{9} h \left( m^* \right)^{1/2} \tau_o \left( \frac{1}{8m} \left( \frac{3}{\pi} \right)^{2/3} \right)^{2} n^4 \left( k_B T \right)^2 \right] \]  

(6.39)

For \( r = 3/2 \)

\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \left[ 48 \sqrt{2} \pi^3 h^3 \left( m^* \right)^{1/2} \tau_o \left( \frac{1}{2m} \left( \frac{3n}{8\pi} \right)^{2/3} \right)^{3} \right] \]

\[
= \left( \frac{1}{\kappa_{L_0} T_0} \right) \left[ 48 \sqrt{2} \pi^3 h^3 \left( m^* \right)^{1/2} \tau_o \left( \frac{1}{2m} \left( \frac{3n}{8\pi} \right)^{2/3} \right)^{3} \left( k_B T \right)^2 \right] \]

91
\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \frac{1}{2} \frac{48 \sqrt{2} \pi^3 h^3 (m^*)^{1/2} \tau_o}{8m} \left( \frac{3}{\pi} \right)^{2/3} n^2 (k_B T)^2 \] 
\]

Now the algebra in the above Eqs., (6.37), (6.38), (6.39) and (6.40), can still be worked out further and the terms better rearranged, and thus the equations are finally written as follows

\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \frac{2^{-3r}}{3-2r} \frac{2^{3r}}{3} \frac{2^{2r}}{3} \frac{6-2r}{3} \frac{h^2 r (m^*)^{1/2} \tau_o k_B^2}{n^{2r+3}} \frac{2}{3} \frac{T^2}{m^2} \left( r + \frac{3}{2} \right)^{r+3/2} \] 
\]

For \( r = -1/2 \)

\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \frac{3}{2} \frac{3}{4} \frac{3}{3} \frac{1}{3} \frac{1}{3} \frac{(m^*)^{1/2} \tau_o k_B^2}{h m} \frac{2}{n^3 T^2} \] 
\]

For \( r = 1/2 \)

\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \frac{1}{2^3} \frac{1}{3^2} \frac{1}{3} \frac{1}{3} \frac{1}{3} \frac{h (m^*)^{1/2} \tau_o k_B^2}{m^2} \frac{2}{n^3 T^2} \] 
\]

For \( r = 3/2 \)

\[
\frac{\kappa_e}{\kappa_L} = \left( \frac{1}{\kappa_{L_0} T_0} \right) \frac{3^3 \pi h^3 (m^*)^{1/2} \tau_o k_B^2}{m^3} \frac{n^2 T^2}{n^2 T^2} \] 
\]

92
Chapter 7

Correlations of the Reduced Fermi Energy and the Thermoelectric Figure of Merit as Functions of the Free Charge Carrier Concentration and the Absolute Temperature

At this point, it is naturally appropriate to derive theoretical relationships expressing the dimensionless thermoelectric figure of merit, $ZT$, as a function of the concentration of free charge carriers, $n$, and the absolute temperature, $T$. We will carry out this new investigation, as usual, first for nondegenerate, and then for degenerate, semiconductors.

For nondegenerate semiconductors, $ZT$ is given by Eq. (5.11), which we will rewrite as follows

$$ZT = \frac{\eta^2}{r + \frac{5}{2}} - 2\eta + r + \frac{5}{2}$$

(7.1)

From Eq. (6.19), we can derive the following equation for the reduced Fermi energy, $\eta$, for a nondegenerate semiconductor

$$\eta = \ln\left(\frac{n}{2}\left(\frac{2\pi m^* k_B T}{h^2}\right)^{\frac{3}{2}}\right)$$

(7.2)

or
\[
\eta = \ln \left( \frac{n}{2} \left( \frac{h^2}{2\pi m^* k_B T} \right)^{3/2} \right) \tag{7.3}
\]

or

\[
\eta = \ln \left( \frac{n}{2} \left( \frac{1}{T^{3/2}} \frac{h^3}{(2\pi m^* k_B)^{3/2}} \right) \right) \tag{7.4}
\]

or

\[
\eta = \ln \left( \frac{n}{3} \right) + \ln \left( \frac{h^3}{2(2\pi m^* k_B)^{3/2}} \right) \tag{7.5}
\]

The second term between parentheses on the right-hand side of Eq. (7.5) can be considered another constant of nature, since its value depends on Planck’s constant, Boltzmann’s constant and the effective mass of the electron. The latter can be approximately estimated on the basis of the mass of the free electron, m, multiplied by a factor which may be provisionally considered somewhat greater than 1. In carrying out our calculation of the aforementioned term, we have roughly assumed that factor to be 1.5. That is, undoubtedly, not a bad assumption, since we are merely trying to find out the order of magnitude value of that term. Our calculations run as follows

\[
A = \frac{h^3}{2(2\pi m^* k_B)^{3/2}} = \frac{(6.62618)^3 \times 10^{-102}}{2(2\pi \times 1.5 \times 9.10953 \times 1.38066 \times 10^{-4})^{3/2}} = 1.127 \times 10^{-22} \text{m}^3 \text{K}^{-2/3}
\tag{7.6}
\]

With reference to Eq. (7.6) above, we have assumed that \( m^* = 1.5m \), where m is the electronic free mass. If the effective mass, \( m^* \), of the nondegenerate semiconducting material,
under investigation, happens to be known and to conform to the relation, \( m^* = f_m \), where \( f \neq 1.5 \), then Eq. (7.6) should be modified as follows

\[
A_{fm} = A_{1.5m} \left( \frac{1.5}{f} \right)^{\frac{3}{2}} = 1.127 \left( \frac{1.5}{f} \right)^{\frac{3}{2}} \times 10^{-22}
\]  

(7.7)

Eq. (7.7) should then be used instead of Eq. (7.6) for all calculations of \( n \) as a function of \( T \) and \( \eta \), and eventually also \( ZT \) and \( S \), whenever \( f \neq 1.5 \).

Since Eqs. (7.2) to (7.5) are logarithmic, and since \( \eta \) is dimensionless, therefore, the argument of the logarithm must also be dimensionless. Thus, Eq. (7.5) should better be written as follows

\[
\eta = \ln \left( A \frac{n^{\frac{3}{2}}}{T^2} \right)
\]

(7.8)

Trying to express \( ZT \) as a function of \( n \) and \( T \), through the substitution of Eq. (7.8) into Eq. (7.1), will eventually prove not practical or successful. We will end up obtaining a quadratic, or second degree, polynomial, where the main variable is the logarithm of \( n \), \( T \) and a constant. The best bet would then be to try to optimize the dimensionless thermoelectric figure of merit with respect to either of the two variables, \( n \) or \( T \), while maintaining the other constant. Doing so, we proceed as follows, operating on Eq. (7.1)

\[
\frac{\partial (ZT)}{\partial n} = \frac{2 \eta}{r + \frac{5}{2}} \frac{\partial n}{\partial n} = \frac{2 \eta}{r + \frac{5}{2}} = 0
\]

(7.9)

or

\[
\frac{\eta}{r + \frac{5}{2}} = 1
\]

(7.10)

which yields

95
\[ \eta = r + \frac{5}{2} \quad (7.11) \]

For \( r = -1/2 \)

\[ \eta = r + \frac{5}{2} = 2 \quad (7.12) \]

For \( r = 1/2 \)

\[ \eta = r + \frac{5}{2} = 3 \quad (7.13) \]

For \( r = 3/2 \)

\[ \eta = r + \frac{5}{2} = 4 \quad (7.14) \]

The correlation between \( ZT \) and both \( n \) and \( T \) could generally be as follows, according to Eq. (7.1)

\[ ZT = \frac{\left( \ln A \frac{n}{T^{3/2}} \right)}{r + \frac{5}{2}} - 2\left( \ln \left( A \frac{n}{T^{3/2}} \right) + r + \frac{5}{2} \right) \quad (7.15) \]

Looking now more carefully into Eqs. (7.9) to (7.11), and the values of \( \eta \) that resulted therefrom, namely, Eqs. (7.12) to (7.14), we immediately realize that these results are misleading. The values of the reduced Fermi energy, \( \eta \), for nondegenerate semiconductors, must always be negative, whereas for degenerate semiconductors, these values are always positive. These are the rules established by H.J. Goldsmid\textsuperscript{73}, and we are not going to violate or disregard them in this dissertation. The veracity of these rules can be immediately established by means of the following example: We need to find the reduced Fermi energy for a nondegenerate semiconductor for which \( n = 1 \times 10^{18} \) carriers cm\(^{-3} \) and \( T = 1000 K \). Substituting the value of \( A \) given by Eq. (7.6) into Eq. (7.8) yields
\[ \eta = \ln \left( 1.127 \times 10^{-22} \times \frac{1 \times 10^{18} \times 10^6}{(1000)^{3/2}} \right) = -5.637 \]

We thus must now move in a different direction in this analysis. Instead of attempting to maximize or optimize \( ZT \), we will take the initiative to find out what happens to \( ZT \) when \( \eta \) acquires the maximum or the minimum acceptable values, respectively. Invoking Eq. (7.1), we thus obtain

For \( \eta = -2 \)

\[ ZT = \frac{(-2)^2}{r + \frac{5}{2}} - 2(-2) + r + \frac{5}{2} \quad (7.16) \]

or

\[ ZT = \frac{4}{r + \frac{5}{2}} + 4 + r + \frac{5}{2} \quad (7.17) \]

For \( r = -\frac{1}{2} \)

\[ ZT = 2 + 4 + 2 = 8 \quad (7.18) \]

For \( r = \frac{1}{2} \)

\[ ZT = \frac{4}{3} + 4 + 3 = 8.333 \quad (7.19) \]

For \( r = \frac{3}{2} \)

\[ ZT = 1 + 4 + 4 = 9 \quad (7.20) \]

For \( \eta = -4 \)

\[ ZT = \frac{(-4)^2}{r + \frac{5}{2}} - 2(-4) + r + \frac{5}{2} \quad (7.21) \]

or
\[
ZT = \frac{16}{r + \frac{5}{2}} + 8 + r + \frac{5}{2} \tag{7.22}
\]

For \( r = -\frac{1}{2} \)

\[
ZT = 8 + 8 + 2 = 18 \tag{7.23}
\]

For \( r = \frac{1}{2} \)

\[
ZT = \frac{16}{3} + 8 + 3 = 16.333 \tag{7.24}
\]

For \( r = \frac{3}{2} \)

\[
ZT = 4 + 8 + 4 = 16 \tag{7.25}
\]

The above results, as given by Eqs. (7.18) to (7.20), indicate that \( ZT \) values of from 8 to 9 are achievable for nondegenerate semiconductors for which the reduced Fermi energy equals -2. Subsequent results, yielded by Eqs. (7.23) to (7.25), likewise show that \( ZT \) values ranging from 16 to 18 are attainable for nondegenerate semiconductors, for which the reduced Fermi energy equals -4. These new results, developments or conclusions should have come as a totally unexpected surprise to us. However, there is actually no big mystery about them. They came about as a natural consequence of substituting the minimum or maximum values of the reduced Fermi energy, that a nondegenerate semiconductor is allowed to have in order to perform satisfactorily, into Eq. (7.1). Nevertheless, something very important has already happened, at least theoretically and motivationally. The above new developments show that a dimensionless thermoelectric figure of merit, \( ZT \), of much greater value than the current experimental one of about 1.5, should be achieved in practice. Nonetheless, believing that \( ZT \) values as high as 8 to 18 are achievable or possible does not necessarily assure or guarantee that such a spectacular development will come about in real life. That will depend, more than anything else, on one issue only, and that is whether a semiconducting material can be found whose thermoelectric
power, or Seebeck coefficient, conforms to the data given in Table (1). This is where the real issue or challenge in thermoelectrics lies. When we constructed that Table, we started by selecting certain values for ZT, from 1 to 100, and then calculated the corresponding Seebeck coefficients through multiplying ZT by the Lorenz number and then getting the square root of the product. What happens in real life is exactly the opposite. It is the value of the measured Seebeck coefficient of the material we select that determines how much ZT will be, and here is where the crux in thermoelectrics lies. We can now use Table (1) to find, either directly or by interpolation, the values of the Seebeck coefficient that a nondegenerate semiconducting material must have in order to produce ZT values in the range from 8 to 18, as indicated above. We are going to address this issue once more very soon. For the time being, however, let us pause here and proceed to investigate the degenerate case.

For degenerate semiconductors, ZT is given by Eq. (5.17), which we will rewrite as follows

\[
ZT = \frac{\pi^2}{3} \left( r + \frac{3}{2} \right)^2 \frac{1}{\eta^2}
\]  

(7.26)

Now, by definition, \( \eta = E_F / k_B T \), and invoking Eq. (5.22), gives

\[
\eta = \frac{\hbar^2}{2m} \left( \frac{3n}{8\pi} \right)^{\frac{2}{3}} \frac{1}{k_B T} \left( r + \frac{3}{2} \right)
\]  

(7.27)

where

\[
n = \frac{N_e}{V}
\]  

(7.28)

which is the number of free charge carriers, per unit volume, also known as the free charge carrier concentration. Eq. (7.27) can be worked further, which yields
\[
\eta = \frac{h^2}{8mk_B} \left( \frac{3}{\pi} \right)^{\frac{2}{3}} \frac{n^3}{T} \left( r + \frac{3}{2} \right)
\]  

(7.29)

From the very beginning of this investigation, when we constructed Table (1), we were able to infer, or conclude, that the Seebeck coefficient, S, for degenerate semiconductors, is not a function of the electronic scattering parameter or index, \( r \). That has happened, because the Lorenz number, \( L \), for degenerate semiconductors is not a function of \( r \). Consequently, the dimensionless thermoelectric figure of merit, \( ZT \), also is not a function of \( r \). These results or conclusions can be inferred from either Eq. (5.6) or Eq. (5.25).

When we constructed Table (2), earlier in this dissertation, we found that the values of the reduced Fermi energy, \( \eta \), for degenerate semiconductors were related to one another in the ratio of 1:2:3 corresponding to \( r = -1/2, 1/2 \) and \( 3/2 \), respectively. This evidently demonstrates a linear dependency of \( \eta \) on \( r + \frac{3}{2} \). This is also proven, if we factor \( \eta \) out of Eq. (5.17)

\[
\eta = \frac{\pi}{\sqrt{3}ZT} \left( r + \frac{3}{2} \right)
\]

This means that both \( S \) and \( ZT \) are not functions of \( r \) for degenerate semiconductors.

Substituting now Eq. (7.29) into Eq. (7.26), gives rise to

\[
ZT = \left( \frac{8mk_B}{h^2} \right)^2 \left( \frac{\pi}{3} \right)^{\frac{4}{3}} \frac{T^2}{n^3} \]

(7.30)

or

\[
ZT = B \frac{T^2}{n^3}
\]

(7.31)
where

\[ B = \frac{\pi^2}{3} \left( \frac{8\pi \mu_B}{\hbar^2} \right)^2 \left( \frac{\pi}{3} \right)^3 = \frac{\pi^2}{3} \left( \frac{1}{A} \right)^2 = 0.18372819 \times 10^{22} \text{m}^{-4} \text{K}^{-2} \]  

(7.32)

where the constant \( B = \pi^2 / 3A^2 \), and the constant, \( A \), has been defined and calculated according to Eq. (6.11).

For degenerate semiconductors, regardless of \( r \)

\[ ZT = B \frac{T^2}{n^3} \]  

(7.33)

Looking now into the above Eqs., (7.30) to (7.33), we realize that we have already succeeded in establishing a general relationship between \( ZT \) and both \( n \) and \( T \), which is Eq. (7.31). We may thus consider this part of our investigation concluded.

At this point in our research, it is probably appropriate to focus our attention and efforts on the dimensionless thermoelectric figure of merit, \( ZT \), for both nondegenerate and degenerate semiconductors, in order to find out the \( ZT \) values that can possibly be achieved, at least theoretically. We will try to determine the limits that could eventually be set or placed on those possible \( ZT \) values. The aforementioned limits could originate from the values of the free charge carrier concentration or temperature, or a combination of both, required to create certain allowed, predetermined or preferred reduced Fermi energy values which are fundamentally conducive to attaining those possible \( ZT \) values. Although less likely, additional constraints could still be placed on the possible \( ZT \) values, if we set our goal, for example, at a \( ZT \) value equal, or close, to 100. The Seebeck coefficient required in that case could be too demanding, or difficult to achieve, with any material whatsoever.

Part of the research work referred to above has already been carried out earlier in relation
to nondegenerate semiconductors. We have been able to verify, for example, that when the scattering parameter \( r = -1/2 \), the magnitude of \( ZT \) becomes 8 and 18, for \( \eta = -2 \) and \(-4\), respectively. Looking now into Table (2), we notice that we could still push \( \eta \) down to a value which is slightly below the minimum allowed value, or limit, of \(-4\), i.e., \(-4.32\), and thus get \( ZT = 20 \). This \( ZT \) value is the maximum we can achieve with any nondegenerate semiconducting material. If we still choose to ignore this result and look instead into Table (1), as an alternative, we could argue as follows: We could base our criterion on the Seebeck coefficient and assume that \( S = 1000 \mu VK^{-1} \) is attainable. This could be true for a few semiconductors, not at room temperature of course, but at higher temperatures. Looking into Table (1) for nondegenerate semiconductors, for which \( r = -1/2 \) and \( S = 1000 \mu VK^{-1} \), we interpolate a corresponding value of \( ZT = 67.6 \). Looking back into Table (2), we see that such a \( ZT \) value corresponds to \( \eta = -9.62 \), which is far below the minimum allowed value, or limit, of \(-4\), as established by Goldsmid. Thus, \( ZT = 67.6 \) cannot be achieved for nondegenerate semiconductors, regardless of the value of \( r \). Therefore, we end up reconfirming our original conclusion which runs as follows:

For nondegenerate semiconductors

\[
(ZT)_{\text{MAX}} = 20
\]

(7.34)

and is most appropriate for \( r = -1/2 \). The above result is an important finding of this dissertation.

Let us continue our analysis of nondegenerate semiconductors in order to help us construct new tables that give the free charge carrier concentration, or doping level, necessary to achieve certain predetermined values of \( ZT \) for certain constant temperature values. Those constant temperatures will be 300, 500, 750 and 1000K. We will do all the numerical
calculations with the help of Eqs. (7.1) to (7.8).

We will thus proceed as follows: From Eq. (7.8), we deduce

\[
\ln(e^n) = \ln A \frac{n}{T^{3/2}}
\]

(7.35)

therefore

\[
e^n = A \frac{n}{T^{3/2}}
\]

(7.36)

or

\[
n = \frac{e^n}{A} T^{-3/2}
\]

Considering, for example, the case in which \( \eta = -2 \), we thus proceed as follows

For \( T = 300K \):
\[
n = \frac{e^{-2}(300)^{3/2}}{1.127 \times 10^{-22}} \times 10^{-6} = 6.24 \times 10^{18} \text{ carriers cm}^{-3}
\]

For \( T = 500K \):
\[
n = \frac{e^{-2}(500)^{3/2}}{1.127 \times 10^{-22}} \times 10^{-6} = 1.34 \times 10^{19} \text{ carriers cm}^{-3}
\]

For \( T = 750K \):
\[
n = \frac{e^{-2}(750)^{3/2}}{1.127 \times 10^{-22}} \times 10^{-6} = 2.47 \times 10^{19} \text{ carriers cm}^{-3}
\]

For \( T = 1000K \):
\[
n = \frac{e^{-2}(1000)^{3/2}}{1.127 \times 10^{-22}} \times 10^{-6} = 3.80 \times 10^{19} \text{ carriers cm}^{-3}
\]

The free charge carrier concentration, \( n \), has thus generally been calculated for \( \eta = -2 \), \(-2.32\), \(-2.47\), \(-2.48\), \(-3\), \(-4\), \(-4.32\) and \(-5\). All of these results are summarized and presented in Tables (7), (8) and (9), and are related to nondegenerate semiconductors for which \( r = -1/2 \), \( 1/2 \) and \( 3/2 \), respectively. In order to render the aforementioned tables still more practical and useful, the corresponding ZT and S values have also been included at the bottom of each one of them.
All the calculated data referred to above, as summarized in Tables (7), (8) and (9), conclude our investigation in relation to nondegenerate semiconductors, and serve to demonstrate the following:

(1) When using nondegenerate semiconductors for thermoelectric energy conversion, operating in the temperature range of 300K to 1000K, and maintaining the reduced Fermi energy approximately in the range from $-2$ to $-4$ and the free charge carrier concentration within the limits of $6.13 \times 10^{17}$ to $2.47 \times 10^{19}$ carriers cm$^{-3}$, a dimensionless thermoelectric figure of merit, $ZT$, of from $8$ to $20$ can be achieved, depending on the value of the scattering index, $r$.

(2) In order for the above results, specifically $ZT$, to be realized, a minimum Seebeck coefficient must be generated or maintained by the semiconductors, as indicated in these tables.

(3) If the operating temperature can be maintained in the range from 500K to 1000K the above requirement for the Seebeck coefficient, $S$, should prove less difficult (although still not guaranteed).

(4) The material that places the least demand on $S$ is that for which $r = -1/2$, which is the material in which acoustic-mode lattice scattering or alloy scattering predominates. Therefore, such a material should be preferred or selected for that purpose.

(5) The values of $n$ for $T = 1000K$ and $\eta = -2$ and $-2.32$ are $3.80 \times 10^{19}$ and $2.76 \times 10^{19}$ carriers cm$^{-3}$, respectively. These $n$ values are untenable and represent cases of reductio ad absurdum, since the uppermost value, or limit, of $n$ for nondegenerate semiconductors is $2.5 \times 10^{19}$ carriers cm$^{-3}$. Therefore, these cases are unacceptable and must be disregarded, since they relate to degenerate semiconductors.

We now must go back and carry out a more thorough and detailed analysis of degenerate semiconductors. We will specifically concentrate our effort on Eq. (7.33) in order to find out the
ZT values that can possibly be achieved for certain values of the free charge carrier concentration, n, and the temperature, T. Or better still, we can calculate the values of n, for certain predetermined values of T, required to attain specific predetermined integral values of ZT. The latter could be taken directly from Tables (1) or (2). Those predetermined values of T could be set at 300K, 500K, 750K and 1000K. We thus select beforehand the values of both ZT and T, and end up obtaining the value of n, using Eq. (7.33). Of course, there are constraints placed on n. Its value must be higher than $2.5 \times 10^{19}$ carriers cm$^{-3}$ and must not exceed $5 \times 10^{20}$ carriers cm$^{-3}$. The latter is inconsequential, since we are more concerned or interested in keeping n as low as possible, i.e., only slightly above $2.5 \times 10^{19}$ carriers cm$^{-3}$, in order to ensure that we will get the maximum ZT possible for a certain constant, or predetermined, temperature T, while assuring that the material still remains degenerate.

Investigating degenerate semiconductors, we look into Table (1) or Table (2) and use those ZT values, set forth in the left hand column thereof, as the input to our analysis. We will use the same standard set of 300K, 500K, 750K and 1000K as our preselected temperatures, as we did with the nondegenerate semiconductors. We will now invoke Eq. (7.33) and substitute the aforementioned two inputs, i.e., the dimensionless thermoelectric figure of merit, ZT, and the temperature, T, into that equation. We will thus finally obtain the free charge carrier concentration, n. The so-obtained value of n should then be examined as to whether it falls within the range: $2.5 \times 10^{19} < n < 5 \times 10^{20}$ carriers cm$^{-3}$. If it doesn’t, then the ZT value that we used as an input must be rejected, since that means that that ZT value cannot be realized in practice at the preselected temperature, T.

We will thus have to rewrite Eq. (7.33) as follows, in order to put it into a more appropriate form.
\[ n^4 = \frac{BT^2}{ZT} \]  \hspace{1cm} (7.37)

or

\[ n = \left( \frac{BT^2}{ZT} \right)^{\frac{3}{4}} \]  \hspace{1cm} (7.38)

We may now proceed to investigate degenerate semiconductors regardless of the value of \( r \), since their \( ZT \) does not depend on \( r \).

Using, therefore, Eq. (7.38) enables us to construct Table (10), which gives values of the free charge carrier concentration, \( n \), as a function of the dimensionless thermoelectric figure of merit, \( ZT \), and the temperature, \( T \), for degenerate semiconductors.

Examining now Tables (7), (8) and (9) as well as Table (10) enables us to draw some very important conclusions concerning the magnitude of \( ZT \) that is to be expected from nondegenerate and degenerate semiconductors that constitute the p-type and n-type thermoelements or branches of devices for direct thermoelectric energy conversion.

Let us concentrate first on Tables (7), (8) and (9), related to nondegenerate semiconductors, for which \( r = -1/2, 1/2 \) and \( 3/2 \), respectively. For \( r = -1/2 \), which refers to Table (7), we notice that when \( \eta = -2 \) and \( T = 1000K \), the concentration of free charge carriers \( n = 3.80 \times 10^{19} \text{ carriers cm}^{-3} \), which is much higher than the well known limit for nondegenerate semiconductors, \( n = 2.5 \times 10^{19} \text{ carriers cm}^{-3} \). Referring to Table (7), this means that in order for that semiconductor to bring about \( ZT \) and \( S \) values of 8 and \( 344\mu\text{VK}^{-1} \), respectively, the material must actually be a degenerate semiconductor. This violates the original assumption of this analysis that the material is a nondegenerate semiconductor. This kind of contradiction is called: “reductio ad absurdum,” that is, “reduction to the absurd.” All that leads to the conclusion that
the ZT and S values of 8 and 344 μVK⁻¹, respectively, cannot be realized in practice when 
T=1000K and η = −2. The aforementioned ZT and S values are, however, possible for lower 
temperatures. In the latter case, that means that as long as the n and T values match, or 
correspond to each other, as given in the table, and as long as the material is capable of 
generating an S value as indicated, then the ZT value given in the table is guaranteed, or 
automatically achieved. The foregoing analysis applies to all three tables. That means that the 
only ZT value that is not achievable is that for which η = −2 and T=1000K, regardless of the 
value of r. Comparing the three Tables, (7), (8) and (9), we notice that the highest ZT value for a 
nondegenerate semiconductor, i.e., 24.5, is achieved through a material for which r = −1/2, 
which again refers to Table (7). In that case, η = −5 and the material must be capable of 
generating a thermoelectric power or Seebeck coefficient, S=602 μVK⁻¹. If the material fails to 
generate or bring about a Seebeck coefficient of the aforementioned magnitude, then the 
foregoing highest magnitude of ZT will not be achieved. And that is because both quantities, ZT 
and S, are tied-up together through Eq. (5.6).

Looking now into Table (2), and referring once more to a nondegenerate semiconductor 
for which r = −1/2, we infer that by reducing the value of the reduced Fermi energy, η, from 
−5 to −5.75, −6.94, −8.00, −8.95, −9.83, −10.65, −11.42 and −12.14, we could get ZT 
values of 30, 40, 50, 60, 70, 80, 90 and 100, respectively, as well as S values of 666, 769, 860, 
942, 1018, 1088, 1154 and 1217 μVK⁻¹, respectively. So we may initially fall under the 
influence of unbridled enthusiasm and think that the sky is only the limit as to how far we can go 
in thermoelectrics. Unfortunately, this is not so. While dealing with nondegenerate 
semiconductors, we should not lose sight of the fact that the reduced Fermi energy, η, values are 
all negative and start pointing downwards from the edge of the conduction band or upwards from 

107
the edge of the valence band. Furthermore, it is necessary that $\eta \ll 0$, i.e., that the Fermi level lies within the forbidden band gap well away from the appropriate band edge and also that the Fermi level should be much further from the opposite band edge, if minority carriers are to be neglected or avoided. The foregoing imposes a powerful constraint upon how much can we lower the reduced Fermi energy, $\eta$. The following rule is thus established: We should lower $\eta$ as much as we can so that $\eta < -4$, but we cannot go below the middle of the forbidden energy band gap, because the minority carriers from the opposite band edge will start adversely affecting the Seebeck coefficient, $S$, which will lead to a deterioration of the thermoelectric performance of the semiconductor.

A simple rule or relationship that could be used to calculate the minimum value of the reduced Fermi energy, $\eta$, for a nondegenerate semiconductor, for which the forbidden energy band gap, $E_g$, is known, is therefore the following: A certain confusion that may ensue due to the fact that $E_g$ is a positive, while $\eta$ is a negative, quantity, can be removed by neglecting the negative sign of $\eta$. In this case, we will actually be calculating the maximum absolute value of the reduced Fermi energy, and then we will simply apply the negative sign to it, and thus everything will be restored back to normal. Thus the relationship is expressed as follows

$$E_F = |\eta| k_B T = \frac{E_g}{2}$$  \hspace{1cm} (7.39)

or

$$|\eta| = \frac{E_g}{2k_B T}$$  \hspace{1cm} (7.40)

or

$$\eta = -\frac{E_g}{2k_B T}$$  \hspace{1cm} (7.41)
In order to avoid any possible confusion related to their future use, Eqs. (7.40) and (7.41) should be designated as referring to the maximum absolute value, and minimum value, of the reduced Fermi energy, respectively. We thus rewrite them as follows

\[ |\eta|_{\text{MAX}} = \frac{E_g}{2k_B T} \quad (7.42) \]

or

\[ \eta_{\text{MIN}} = -\frac{E_g}{2k_B T} \quad (7.43) \]

In order to be able to use Eqs. (7.42) and (7.43), both the energy band-structure diagram and the forbidden energy band gap, \( E_g \), for the specific material under investigation, must be known or available. If that is not so, then it will be impossible for us to know how far we can go in lowering the reduced Fermi energy, \( \eta \), in order to get the best possible values of \( S \) and \( ZT \). In that case then, the only alternative we would have would be to carry out experimental research work on the material in question. Let us now work out some numerical calculations using Eqs. (7.42) and (7.43), so that we may get the feel of what kind of results can be expected from them.

For a semiconductor having, for example, a forbidden energy band gap, \( E_g = 0.78 \text{ eV} \), and operating at a temperature, \( T=1000\text{K} \), we obtain

\[ |\eta|_{\text{MAX}} = \frac{E_g}{2k_B T} = \frac{0.78 \times 1.60219 \times 10^{-19}}{2 \times 1.38066 \times 10^{-23} \times 1000} = 4.53 \]

or

\[ \eta_{\text{MIN}} = -4.53 \]

The above result for \( \eta_{\text{MIN}} \) seems reasonable, and for a semiconductor for which \( r = -1/2 \) yields a dimensionless thermoelectric figure of merit, \( ZT=21.32 \). Continuing the above calculation for the operating temperatures of 750K, 500K and 300K, we obtain additional results which, along
with the foregoing one for T=1000K, are presented as shown in Table (11).

A study of the foregoing results indicates that the values of the minimum reduced Fermi energy, \( \eta_{\text{min}} \), of \(-4.53\) and, to a lesser extent, \(-6.03\) corresponding to 1000K and 750K, respectively, can be considered tentatively approved or acceptable, whereas those corresponding to 500K and 300K, respectively, i.e., \(-9.05\) and \(-15.9\), should be regarded as unreliable and suspicious. The reason they acquired such huge magnitudes is purely mathematical and emanates from Eq. (7.43), but there is no physical evidence to support it or back it up. Therefore, the latter two values are rejected. It is extremely unlikely to be able to manufacture, or come across, a semiconducting or thermoelectric material that would generate or bring about Seebeck coefficients, \( S \), of 951 and 1470 \( \mu \text{VK}^{-1} \), and figures of merit, \( ZT \), of 61 and 146 at such relatively low temperatures as 500K and 300K, respectively. Since, in this dissertation, we are essentially more interested in thermoelectric power generation than in thermoelectric refrigeration and/or thermoelectric heat pumping, we are naturally more focused on investigating the higher temperature range up to 1000K, since that would lead to the highest possible Carnot engine energy conversion efficiency, according to the second law of thermodynamics. Operating our power generating thermoelectric device at such low temperatures as 500K or 300K, considered hot junction, or hot end, temperatures, doesn’t make sense and could eventually be conducive to a very low or even zero Carnot engine efficiency, and we don’t want that.

In the above investigation or analysis, we have used Eq. (5.11) to calculate \( ZT \) from values of \( \eta \) determined on the basis of the width of the forbidden energy band gap, \( E_g \), i.e., Eqs. (7.42) and (7.43). We then used Eq. (5.6) to obtain the corresponding values of the Seebeck coefficient, \( S \). When doing that, we should always make sure that Eq. (5.6) is valid. As long as the lattice thermal conductivity has been reduced to its absolute minimum value, through mass
and volume fluctuation scattering, also known as alloy scattering, or through any other means, and even if the alloy or material is lightly doped, or even undoped, and even if the temperature is at least not below room temperature, then Eq. (5.6) should still be considered reliable and standing on a firm foundation. For that purpose, Eqs. (6.32), (6.33) and (6.34) as well as Eqs. (6.42), (6.43) and (6.44) for nondegenerate and degenerate semiconductors, respectively, should be used whenever the validity of Eq. (5.6) needs to be verified or checked. In this context, a rather suspicious situation has come about in Table (11), where we have the T and n combination of 300K and $1.29 \times 10^{13}$ carriers cm$^{-3}$, which certainly is untenable. Another situation, which is still similar but less serious, is given by $T=500K$ and $n=1.16 \times 10^{16}$ carriers cm$^{-3}$.

Comparing now Tables (7), (8) and (9) with each other, we notice that, e.g., for $\eta = -5$, the highest ZT value is given by Table (7) and the lowest is given by Table (9), i.e., 24.5 and 20.25, respectively. We also notice that the lowest value for the Seebeck coefficient, S, is given by Table (7) and the highest is given by Table (9), i.e., 602 and 774 $\mu$VK$^{-1}$, respectively. Thus, of all three tables, Table (7) is the one that yields the highest ZT value possible and it also is the one that requires the lowest value for the Seebeck coefficient, S. This leads to the conclusion that, when dealing with nondegenerate semiconductors in general, we should select one for which $r = -1/2$, that is, one in which either acoustic-mode lattice scattering or alloy scattering is predominant. The lower the demand on the Seebeck coefficient, the more likely it is that such a demand will be realized, or met, in practice. That is why, should we decide to use a nondegenerate semiconductor for direct thermoelectric energy conversion, we should always choose one for which $r = -1/2$.

We examine now Table (10) related to degenerate semiconductors. We notice that for $T=300K$, only the ZT values of 1 and 2 are possible. The rest of the ZT values, from 3 to 100,
represent cases of “reductio ad absurdum,” since the free charge carrier concentration, \( n \), has to remain below \( 2.5 \times 10^{19} \) carriers cm\(^{-3} \), i.e., the material must actually be nondegenerate. This violates the original assumption that the material is a degenerate semiconductor. Thus those ZT values from 3 to 100 cannot be achieved or realized in practice. For \( T=500K \), only the ZT values from 1 to 5 are possible, the remaining ZT values, from 10 to 100, are cases of reductio ad absurdum, and thus they cannot be realized in practice. For \( T=750K \), only the ZT values from 1 to 10 are possible, the remaining ones, from 20 to 100, are impossible to realize in practice. For \( T=1000K \), only the ZT values from 1 to 20 are possible, the remaining ones, from 30 to 100, are impossible to realize in practice. For a degenerate semiconductor, again maintaining \( T=1000K \), if we lower \( n \) down to \( 2.6 \times 10^{19} \) carriers cm\(^{-3} \) and use Eq. (7.37), we obtain ZT=23.85 which is almost identical with, or equal to, the highest ZT value obtained with a nondegenerate semiconductor for which \( r = -1/2 \) and \( \eta = -5 \) (Table (7)).

The foregoing observations, which can be easily inferred from Eq. (7.33), serve to bring into sharp focus two of the most important characteristics of degenerate semiconductors:

1. The higher the temperature, the better is the thermoelectric performance.
2. The higher the level of doping, the lower is the thermoelectric performance.

It is now time to proceed with the construction of Table (12), which is the same as Table (1) except that the Seebeck coefficient, \( S \), for degenerate semiconductors, appears to depend on the scattering parameter, \( r \), as can be seen from Eq. (5.15). Table (12) will thus contain three different values, for degenerate semiconductors, and three different values, for nondegenerate semiconductors, of the Seebeck coefficient, or thermoelectric power, \( S \), for each predetermined ZT value. We will thus start calculating the values of the Seebeck coefficient, \( S \), from the values of the reduced Fermi energy, \( \eta \), corresponding to predetermined values of the dimensionless
thermoelectric figure of merit, ZT, and appropriate values of the scattering parameter, \( r \), for both degenerate and nondegenerate semiconductors, as presented in Table (2). To do that, we will have to rely on Eqs. (5.8) and (5.15) for nondegenerate and degenerate semiconductors, respectively. Three sample calculations for each of degenerate and nondegenerate semiconductors are shown below.

Degenerate Semiconductors:

\( r = -1/2 \)

\[
ZT = 1, \quad \eta = 1.814, \quad S = \frac{\pi^2}{3} \frac{k_B}{e} \frac{1}{\eta} = \frac{\pi^2}{3} \frac{1.38066 \times 10^{-23}}{1.60219 \times 10^{-19}} \frac{10^6}{1.814} = 156 \muVK^{-1}
\]

\( r = 1/2 \)

\[
ZT = 1, \quad \eta = 3.628, \quad S = \frac{2\pi^2}{3} \frac{k_B}{e} \frac{1}{\eta} = \frac{2\pi^2}{3} \frac{1.38066 \times 10^{-23}}{1.60219 \times 10^{-19}} \frac{10^6}{3.628} = 156 \muVK^{-1}
\]

\( r = 3/2 \)

\[
ZT = 1, \quad \eta = 5.441, \quad S = \frac{3\pi^2}{3} \frac{k_B}{e} \frac{1}{\eta} = \frac{3\pi^2}{3} \frac{1.38066 \times 10^{-23}}{1.60219 \times 10^{-19}} \frac{10^6}{5.441} = 156 \muVK^{-1}
\]

Nondegenerate Semiconductors:

\( r = -1/2 \)

\[
ZT = 10, \quad \eta = -2.47, \quad S = \frac{k_B}{e} \left[ \eta - \left( r + \frac{5}{2} \right) \right] = \frac{-4.47 \times 1.38066 \times 10^{-23} \times 10^6}{1.60219 \times 10^{-19}} = 385 \muVK^{-1}
\]

\( r = 1/2 \)

\[
ZT = 10, \quad \eta = -2.48, \quad S = \frac{k_B}{e} (-2.48 - 3) = -5.48 \times \frac{1.38066 \times 10^{-23} \times 10^6}{1.60219 \times 10^{-19}} = 472 \muVK^{-1}
\]

\( r = 3/2 \)

\[
ZT = 10, \quad \eta = -2.32, \quad S = \frac{k_B}{e} (-2.32 - 4) = -6.32 \times \frac{1.38066 \times 10^{-23} \times 10^6}{1.60219 \times 10^{-19}} = 545 \muVK^{-1}
\]
In the foregoing six sample calculations for degenerate and nondegenerate semiconductors, as well as in this entire dissertation, we have ignored any negative sign that S might have, or acquire, i.e., we have reported only the absolute values of the Seebeck coefficient. The plus-or-minus signs associated with Eqs. (5.8) and (5.15) are meant to ensure that the sign of the Seebeck coefficient, S, is negative if the free charge carriers are predominantly electrons and positive if the free charge carriers are predominantly holes.

Checking first the values of the thermoelectric power, S, as obtained according to the three sample calculations, given above, for degenerate semiconductors, for \( r = -1/2, 1/2 \) and \( 3/2 \), respectively, and ZT=1, we notice that they are all the same, i.e., \( S = 156 \mu VK^{-1} \). Looking now into the data given in Table (12) for the Seebeck coefficient of degenerate semiconductors for the entire range of ZT values, from 1 to 100, we perceive that S is independent of \( r \) for any ZT value. This is contrary to what we had thought earlier, that the Seebeck coefficient, S, depends on \( r \). We were led to that conclusion, initially, because Eq. (5.15) shows a dependency of S on \( r \). We failed to realize, however, that the reduced Fermi energy, \( \eta \), also depends on the scattering parameter, \( r \). So, when the two dependencies, i.e., of S and \( \eta \) on \( r \), are included, or come to act, together, they cancel out, and S, therefore, becomes independent of \( r \). Furthermore, Table (1), which gives one single value of S, which is independent of \( r \), for every predetermined value of ZT, turns out to be correct. The data given in Table (1) are based on Eq. (5.6), through the use of Eq. (5.25). Calculation of the data given in Table (2) for the reduced Fermi energy, \( \eta \), for degenerate semiconductors, has been fundamentally carried out by means of Eq. (5.17), while substituting \( r = -1/2, 1/2 \) and \( 3/2 \) for the different denominations of the scattering parameter, which has led to the eventual use of Eqs. (5.18), (5.19) and (5.20). A thorough scrutiny of the \( \eta \) values for \( r = -1/2, 1/2 \) and \( 3/2 \) reveals that indeed those \( \eta \) values are related to one another.
in the ratio of 1:2:3, respectively. This, therefore, proves that the dependency of \( \eta \) on \( r \) is the same as that of \( S \) on \( r \), when looking at Eq. (5.15). So we end up having the factor, \((r + 3/2)\), in the numerator and the factor, \((r + 3/2)\), in the denominator, of that equation. These two factors then cancel out and \( S \) eventually becomes independent of \( r \) for degenerate semiconductors. For each predetermined \( ZT \) value, the \( S \) values are substantially identical, with insignificant numerical differences existing among them. Also there exist some minor and virtually negligible differences between the data presented in Table (12) and the corresponding ones of Table (1).

For nondegenerate semiconductors, we had to rely on Eq. (5.8) in order to convert the \( \eta \) values given in Table (2) into the \( S \) values that are presented in Table (12). Once more, the \( S \) data given in Table (12) are almost the same as those shown in Table (1), the differences being insignificant. The reduced Fermi energy, \( \eta \), data given in Table (2) were obtained, for each predetermined \( ZT \) value, and the relevant \( r \) values, through the use of Eq. (5.11), by solving forty five algebraic quadratic equations and then obtaining the roots or eigenvalues thereof. Those eigenvalues that did not have a negative sign were ignored or cancelled, since they represented cases of reductio ad absurdum, i.e., they stood in violation of the original assumption, that the semiconductor is nondegenerate and, as such, the reduced Fermi energy must be negative. Those empty spaces in Table (2) are all due to this kind of situation. There is also another restriction concerning nondegenerate semiconductors, which we chose not to implement. For nondegenerate semiconductors, the rule, as given in Table (13), is that \( \eta < -2 \) and preferably that \( \eta < -4 \). We did not apply that rule at all, and presented all the \( \eta \) values, that we calculated, in the table, as long as they were negative. The thermoelectric power or Seebeck coefficient, \( S \), of nondegenerate semiconductors, as we already know, and as proven through the data set forth in Tables (1) and (12), is indeed a function of \( r \), for any predetermined \( ZT \) value. Having created
Tables (1), (2) and (12), has had the effect that all three of them are now standing on a much firmer foundation. And that applies to both degenerate and nondegenerate semiconductors.

Focusing our attention on all the foregoing data presented in Tables (7)-(11), inclusive, as well as the lengthy statements and comments made with reference thereto, we are in a position now to draw several very significant conclusions with respect to how to select semiconducting materials, both degenerate and nondegenerate, their range of operation and the maximum performance that can be expected therefrom. Nondegenerate semiconductors are more appropriate for operation at lower temperatures, whereas degenerate ones are more suited to operate at higher temperatures. When we refer to lower and higher temperatures, what we have in mind, of course, is the entire temperature range we are dealing with, from 300K to 1000K. Generally, semiconductors, whether degenerate or nondegenerate, can still be operated at any temperature whatsoever, as long as they do not become cases of reductio ad absurdum. As far as the concentration of free charge carriers, or doping level, n, carriers cm\(^{-3}\), of degenerate semiconductors is concerned, the lower the doping level, the higher the thermoelectric figure of merit, ZT, will be. So, for degenerate semiconductors, in general, in order to maximize the thermoelectric performance, or ZT, we have to do two things simultaneously:

1. Increase the temperature as much as possible.
2. Decrease the doping level, or free charge carrier concentration, as much as possible.

We could increase the temperature, T, up to about 1000K, while trying to avoid overdoing it, for two very important reasons:

(a) The melting point of the material, T\(_m\), may be reached or approached.
(b) The state of the material may change from extrinsic to intrinsic, which normally leads to a drop in the Seebeck coefficient and a consequent deterioration or decrease of ZT. We should
thus increase the temperature way up as much as we can, while avoiding reaching both the melting and intrinsic temperatures, $T_m$ and $T_{int}$. There is no fixed upper temperature limit for degenerate semiconductors, because that may individually vary from one material to another. Lowering the doping level, $n$, of degenerate semiconductors is not going to be an easy task, because we are already hampered by the fact that their free charge carrier concentration must always remain higher than $2.5 \times 10^{19}$ carriers cm$^{-3}$, since otherwise a case of reductio ad absurdum will be created. The best we can do is maintain $n$ at about 2.6, 2.7 or $2.8 \times 10^{19}$ carriers cm$^{-3}$. That is the minimum free charge carrier concentration that a semiconductor can have while still being classified as degenerate. So, maximizing the temperature, $T$, along with minimizing the free charge carrier concentration, $n$, creates an unbeatable combination that leads to a vastly superior degenerate semiconducting material with extraordinary thermoelectric figure of merit, $ZT$. That, coupled with an enhanced Carnot engine efficiency, due to the higher operating temperature, leads to maximum overall energy conversion efficiency from heat into electricity.

For nondegenerate semiconductors, the situation is entirely different. Eqs. (5.8), (5.9) and (5.11) do give the Seebeck coefficient, $S$, the Lorenz number, $L$, and the figure of merit $ZT$, and they are all functions of $r$. Tables (1) and (12) also reflect this fact. Examining Tables (7), (8) and (9), and comparing them with each other, we observe that the highest value of the thermoelectric figure of merit, $ZT$, for a nondegenerate semiconductor, is 24.5 and is achieved through a material for which $r = -1/2$, according to Table (7). Other $r$ denominations, namely $1/2$ and $3/2$, yield $ZT$ values of 21.333 and 20.25, respectively. That fact establishes the superiority of the nondegenerate semiconductor, having $r = -1/2$, over the other two, namely those with $r = 1/2$ and $r = 3/2$. That fact is further corroborated when we look into Table (1).
We then notice that the Seebeck coefficient for any ZT value is lowest when \( r = -1/2 \). That means that for the same ZT value, the Seebeck coefficient is lesser for \( r = -1/2 \) than for \( r = 1/2 \) or \( r = 3/2 \). That makes the nondegenerate semiconductor for which \( r = -1/2 \) our material of preference among all three nondegenerate semiconductors. The lower the demand or requirement on the Seebeck coefficient, the more likely it is that such a demand will be met in practice.

Now we come to study the dimensionless thermoelectric figure of merit, ZT, for semiconducting materials, in order to draw some final conclusions about it. For nondegenerate semiconductors, we have already found out that the maximum possible value of the thermoelectric figure of merit is \( ZT = 24.5 \) achieved through a material with \( r = -1/2 \), \( \eta = -5 \) and \( S = 602\mu VK^{-1} \). Can the foregoing ZT figure be improved or exceeded? Yes, but only if we could reduce \( \eta \) to a lower value, or, in other words, increase the absolute value of \( \eta \), i.e., \(|\eta| > 5\). We are not so sure that we could do that. Actually, there is a set of 6 criteria in the recent literature\(^{41}\) on thermoelectrics, the last one of which runs as follows: “For semiconductors, the best materials should have energy band gaps equal to \( 10k_B T \), where \( T \) is the operating temperature of the thermoelectric material.” So, if we combine this criterion with Eq. (7.41), we have

\[
\eta_{\text{MIN}} = -\frac{E_g}{2k_BT} = -\frac{10k_BT}{2k_BT} = -5
\]

The above minimum value of \( \eta \) coincides with the lowest \( \eta \) values given in Tables (7), (8) and (9). In those tables, the decision to make \( \eta = -5 \) was based on pure intuition. The agreement between these values, actually a single value, and that obtained through Eq. (7.44), above, is a mere coincidence. Consequently, we are now in a position to finally wrap it up as far as
nondegenerate semiconducting thermoelectric materials are concerned:

For a nondegenerate semiconducting material, the maximum possible dimensionless thermoelectric figure of merit, $ZT=24.5$, for which $r = -1/2$, $\eta = -5$ and $S = 602 \mu VK^{-1}$.

One final word: Should nondegenerate semiconducting thermoelectric materials always be operated at rather low temperatures? The answer is: not necessarily. They can still be operated at any temperature whatsoever, as long as they conform to the $n$ and $T$ combination according to the $\eta$ value that characterizes them. This kind of $n$-$T$ relationship depends on $\eta$ only and is quite independent of $r$. So, no matter which Table we use, (7) or (8) or (9), the $n$-$T$ relationship remains exactly the same.

As far as degenerate semiconductors are concerned, their thermoelectric performance, as expressed by their dimensionless thermoelectric figure of merit, $ZT$, is not a function of $r$. Their Seebeck coefficient, $S$, as well, is not a function of $r$. Actually, $ZT$ for degenerate semiconductors does not depend on $r$, because both $S$ and $L$ are not dependent on $r$. This can be inferred from Eqs. (5.6), (5.15) and (5.16) and has been thoroughly investigated and proven earlier in this dissertation. $ZT$ for degenerate semiconductors can be increased by increasing the temperature, $T$, and decreasing the doping level, or free charge carrier concentration, $n$. The increase in $T$ should be carried out in such a way that we stay below both the melting temperature, $T_m$, and the intrinsic temperature, $T_{\text{int}}$. A maximum operating temperature of about 1000K would be reasonable. The free charge carrier concentration, $n$, should be adjusted according to the temperature, $T$, according to Table (10). Any combination of $n$ and $T$ should lead to a definite value for the dimensionless thermoelectric figure of merit, $ZT$. Maximizing the latter, requires increasing $T$ as much as possible, so that it gets close, or becomes equal, to preferably 1000K, as well as decreasing $n$ as much as possible, while assuring that it always
stays above $2.5 \times 10^{19}$ carriers cm$^{-3}$, and while conforming to Eq. (7.38). That would be our ideal temperature and doping levels for maximizing the dimensionless thermoelectric figure of merit, $ZT$, for degenerate semiconductors.

Degenerate semiconductors, therefore, can bring about a maximum thermoelectric figure of merit, $(ZT)_{\text{MAX}} = 23.85$ at 1000K, which is virtually the same as the maximum thermoelectric figure of merit $(ZT)_{\text{MAX}} = 24.5$, achieved through a nondegenerate semiconducting material for which the electronic scattering parameter, $r = -1/2$, the reduced Fermi energy, $\eta = -5$ and the Seebeck coefficient, $S = 602 \mu VK^{-1}$.

Studying now Tables (7), (8) or (9), we notice that for nondegenerate semiconductors, decreasing the concentration of the free charge carriers, $n$, i.e., lowering the doping level, while maintaining the temperature, $T$, constant, actually increases the Seebeck coefficient, $S$. Also, maintaining the concentration of the free charge carriers, $n$, constant, while increasing the temperature, $T$, for nondegenerate semiconductors, normally increases the Seebeck coefficient, $S$. This is corroborated by substituting Eq. (7.2) into Eq. (5.8), and inferring that this will actually happen. It would then be safe to assume that doing both, i.e., decreasing $n$, while increasing $T$, would tend to increase $S$ much more. This fact, related to nondegenerate semiconductors, is further supported by Ioffe’s formula (21), which is

$$S = \frac{k_B}{e} \left[ A + \ln \left( \frac{2(2\pi m^* k_B T)^{3/2}}{h^3 n} \right) \right]$$

where the constant $A = r + 5/2$ according to Goldsmid. The aforementioned fact has also been proven experimentally for semiconducting intermetallic compounds, which were substantially in the undoped state, i.e., having an extremely small free charge carrier concentration, $n$. The fact that a nondegenerate semiconducting material will actually experience a large increase in its
Seebeck coefficient, $S$, if its temperature is substantially increased, especially when it is also undoped, i.e., $n$ becomes very small, which has also been corroborated experimentally, is of paramount significance and should be taken full advantage of. This fact should still not be considered good enough. The material, furthermore, must already possess a reasonably high Seebeck coefficient of its own at temperatures equal, or close, to room temperature. Only then, could we be sure that we are dealing with a really promising and worthwhile thermoelectric material.

Referring again to nondegenerate semiconductors, we have already mentioned and applied, quite frequently, the following rules: that the reduced Fermi energy, $\eta$, should be at least less than $-2$, and better still that it be less than $-4$, i.e., $\eta < -2$ or $\eta < -4$, the former rule is considered acceptable, while the latter is regarded as highly recommended or preferable. Credit for establishing these rules undoubtedly belongs to Professor H. J. Goldsmid\textsuperscript{73}. In the course of our work on this dissertation, we have pushed those rules one step further by stipulating that $-4 > \eta \geq -5$. This is the new rule for nondegenerate semiconductors, which is simultaneously conducive to achieving the maximum possible $ZT$ value of 24.5. The aforementioned statements are also presented in a new table, namely, Table (13), for easy reference.
TABLE 7. Values of the free charge carrier concentration, $n$, carriers cm$^{-3}$, as a function of the temperature, $T$, K, and the reduced Fermi energy, $\eta$, along with the corresponding values of the dimensionless thermoelectric figure of merit, $ZT$, and the Seebeck coefficient, $S$, $\mu$VK$^{-1}$, for nondegenerate semiconductors for which $r = -1/2$.

<table>
<thead>
<tr>
<th>$T$ (K)</th>
<th>$n$ ($10^{18}$)</th>
<th>$\eta$</th>
<th>$ZT$</th>
<th>$S$ ($\mu$VK$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>300 K</td>
<td>6.24</td>
<td>3.90</td>
<td>8.44</td>
<td>6.13</td>
</tr>
<tr>
<td></td>
<td>1.34</td>
<td>0.83</td>
<td>1.82</td>
<td>1.32</td>
</tr>
<tr>
<td>750 K</td>
<td>2.47</td>
<td>1.54</td>
<td>3.34</td>
<td>2.42</td>
</tr>
<tr>
<td>1000 K</td>
<td>3.80</td>
<td>2.37</td>
<td>5.14</td>
<td>3.73</td>
</tr>
</tbody>
</table>

TABLE 8. Values of the free charge carrier concentration, $n$, carriers cm$^{-3}$, as a function of the temperature, $T$, K, and the reduced Fermi energy, $\eta$, along with the corresponding values of the dimensionless thermoelectric figure of merit, $ZT$, and the Seebeck coefficient, $S$, $\mu$VK$^{-1}$, for nondegenerate semiconductors for which $r = 1/2$.

<table>
<thead>
<tr>
<th>$T$ (K)</th>
<th>$n$ ($10^{18}$)</th>
<th>$\eta$</th>
<th>$ZT$</th>
<th>$S$ ($\mu$VK$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>300 K</td>
<td>6.24</td>
<td>3.86</td>
<td>2.30</td>
<td>8.44</td>
</tr>
<tr>
<td></td>
<td>1.34</td>
<td>0.83</td>
<td>4.94</td>
<td>1.82</td>
</tr>
<tr>
<td>750 K</td>
<td>2.47</td>
<td>1.53</td>
<td>9.07</td>
<td>3.34</td>
</tr>
<tr>
<td>1000 K</td>
<td>3.80</td>
<td>2.35</td>
<td>1.40</td>
<td>5.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T$ (K)</th>
<th>$n$ ($10^{18}$)</th>
<th>$\eta$</th>
<th>$ZT$</th>
<th>$S$ ($\mu$VK$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>10</td>
<td>18</td>
<td>20</td>
<td>24.5</td>
</tr>
<tr>
<td>344</td>
<td>385</td>
<td>516</td>
<td>544</td>
<td>602</td>
</tr>
</tbody>
</table>

| 3.833   | 10             | 12     | 16.333 | 21.333 |
| 430     | 471            | 516    | 602    | 688     |
TABLE 9. Values of the free charge carrier concentration, $n$, carriers cm$^{-3}$, as a function of the temperature, $T$, K, and the reduced Fermi energy, $\eta$, along with the corresponding values of the dimensionless thermoelectric figure of merit, $ZT$, and the Seebeck coefficient, $S$, $\mu$VK$^{-1}$, for nondegenerate semiconductors for which $\tau = 3/2$.

<table>
<thead>
<tr>
<th>$T$ (K)</th>
<th>$n$ ($\times 10^{18}$)</th>
<th>$\eta$ ($\times 10^{18}$)</th>
<th>$\eta$ ($\times 10^{19}$)</th>
<th>$\eta$ ($\times 10^{20}$)</th>
<th>$\eta$ ($\times 10^{21}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>300 K</td>
<td>6.24</td>
<td>4.53</td>
<td>2.30</td>
<td>8.44</td>
<td>3.11</td>
</tr>
<tr>
<td>500 K</td>
<td>1.34</td>
<td>0.97</td>
<td>0.49</td>
<td>1.82</td>
<td>0.66</td>
</tr>
<tr>
<td>750 K</td>
<td>2.47</td>
<td>1.79</td>
<td>0.91</td>
<td>3.34</td>
<td>1.23</td>
</tr>
<tr>
<td>1000 K</td>
<td>3.80</td>
<td>2.76</td>
<td>1.40</td>
<td>5.14</td>
<td>1.89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$ZT$</th>
<th>9</th>
<th>10</th>
<th>12.25</th>
<th>16</th>
<th>20.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S$</td>
<td>516</td>
<td>544</td>
<td>602</td>
<td>688</td>
<td>774</td>
</tr>
</tbody>
</table>

TABLE 10. Values of the free charge carrier concentration, $n$, carriers cm$^{-3}$, as a function of the dimensionless thermoelectric figure of merit, $ZT$, and the temperature, $T$, K, for degenerate semiconductors.

<table>
<thead>
<tr>
<th>$ZT$</th>
<th>$n$ ($\times 10^{19}$)</th>
<th>$n$ ($\times 10^{20}$)</th>
<th>$n$ ($\times 10^{21}$)</th>
<th>$n$ ($\times 10^{22}$)</th>
<th>$n$ ($\times 10^{23}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.61</td>
<td>9.92</td>
<td>1.82</td>
<td>2.81</td>
<td>1.67</td>
</tr>
<tr>
<td>2</td>
<td>2.74</td>
<td>5.90</td>
<td>1.08</td>
<td>1.67</td>
<td>1.23</td>
</tr>
<tr>
<td>3</td>
<td>2.02</td>
<td>4.35</td>
<td>0.80</td>
<td>1.23</td>
<td>0.99</td>
</tr>
<tr>
<td>4</td>
<td>1.63</td>
<td>3.51</td>
<td>0.64</td>
<td>0.99</td>
<td>0.76</td>
</tr>
<tr>
<td>5</td>
<td>1.38</td>
<td>2.97</td>
<td>0.54</td>
<td>0.83</td>
<td>0.63</td>
</tr>
<tr>
<td>10</td>
<td>8.20</td>
<td>1.76</td>
<td>0.32</td>
<td>0.49</td>
<td>0.39</td>
</tr>
<tr>
<td>20</td>
<td>4.88</td>
<td>1.05</td>
<td>0.19</td>
<td>0.29</td>
<td>0.21</td>
</tr>
<tr>
<td>30</td>
<td>3.60</td>
<td>0.74</td>
<td>0.14</td>
<td>0.21</td>
<td>0.16</td>
</tr>
<tr>
<td>40</td>
<td>2.90</td>
<td>0.62</td>
<td>0.11</td>
<td>0.17</td>
<td>0.13</td>
</tr>
<tr>
<td>50</td>
<td>2.45</td>
<td>0.53</td>
<td>0.09</td>
<td>0.14</td>
<td>0.10</td>
</tr>
<tr>
<td>60</td>
<td>2.14</td>
<td>0.46</td>
<td>0.08</td>
<td>0.13</td>
<td>0.10</td>
</tr>
<tr>
<td>70</td>
<td>1.91</td>
<td>0.41</td>
<td>0.07</td>
<td>0.12</td>
<td>0.10</td>
</tr>
<tr>
<td>80</td>
<td>1.72</td>
<td>0.37</td>
<td>0.06</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td>90</td>
<td>1.58</td>
<td>0.34</td>
<td>0.05</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>100</td>
<td>1.46</td>
<td>0.31</td>
<td>0.05</td>
<td>0.08</td>
<td>0.07</td>
</tr>
</tbody>
</table>
TABLE 11. Values of the minimum reduced Fermi energy, $\eta_{\text{min}}$, as a function of the temperature, $T$, K, along with the corresponding values of the dimensionless thermoelectric figure of merit, $ZT$, the Seebeck coefficient, $S$, $\mu V K^{-1}$, and the concentration of the free charge carriers, $n$, carriers cm$^{-3}$, for nondegenerate semiconductors, for which

$$r = -1/2$$

and the forbidden energy band gap, $E_g = 0.78 eV$.

<table>
<thead>
<tr>
<th>T</th>
<th>300</th>
<th>500</th>
<th>750</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_{\text{min}}$</td>
<td>-15.09</td>
<td>-9.05</td>
<td>-6.03</td>
<td>-4.53</td>
</tr>
<tr>
<td>$ZT$</td>
<td>146.03</td>
<td>61.05</td>
<td>32.24</td>
<td>21.32</td>
</tr>
<tr>
<td>$S$</td>
<td>1470</td>
<td>951</td>
<td>691</td>
<td>562</td>
</tr>
<tr>
<td>$n$</td>
<td>$1.29 \times 10^{13}$</td>
<td>$1.16 \times 10^{16}$</td>
<td>$4.38 \times 10^{17}$</td>
<td>$3.02 \times 10^{18}$</td>
</tr>
</tbody>
</table>

TABLE 12. Values of the Seebeck coefficient, $S$, $\mu V K^{-1}$, as a function of the dimensionless thermoelectric figure of merit, $ZT$, and the scattering parameter, $r$, for degenerate and nondegenerate semiconductors.

<table>
<thead>
<tr>
<th>$ZT$</th>
<th>-1/2</th>
<th>1/2</th>
<th>3/2</th>
<th>-1/2</th>
<th>1/2</th>
<th>3/2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>156</td>
<td>156</td>
<td>156</td>
<td>221</td>
<td>156</td>
<td>221</td>
</tr>
<tr>
<td>2</td>
<td>221</td>
<td>221</td>
<td>221</td>
<td>271</td>
<td>221</td>
<td>271</td>
</tr>
<tr>
<td>3</td>
<td>271</td>
<td>271</td>
<td>271</td>
<td>313</td>
<td>271</td>
<td>271</td>
</tr>
<tr>
<td>4</td>
<td>313</td>
<td>313</td>
<td>313</td>
<td>350</td>
<td>313</td>
<td>350</td>
</tr>
<tr>
<td>5</td>
<td>350</td>
<td>350</td>
<td>350</td>
<td>394</td>
<td>350</td>
<td>394</td>
</tr>
<tr>
<td>10</td>
<td>494</td>
<td>494</td>
<td>494</td>
<td>698</td>
<td>494</td>
<td>698</td>
</tr>
<tr>
<td>20</td>
<td>698</td>
<td>698</td>
<td>698</td>
<td>856</td>
<td>698</td>
<td>856</td>
</tr>
<tr>
<td>30</td>
<td>856</td>
<td>856</td>
<td>856</td>
<td>988</td>
<td>856</td>
<td>988</td>
</tr>
<tr>
<td>40</td>
<td>988</td>
<td>988</td>
<td>988</td>
<td>1103</td>
<td>988</td>
<td>1103</td>
</tr>
<tr>
<td>50</td>
<td>1103</td>
<td>1103</td>
<td>1103</td>
<td>1212</td>
<td>1103</td>
<td>1212</td>
</tr>
<tr>
<td>60</td>
<td>1212</td>
<td>1212</td>
<td>1212</td>
<td>1306</td>
<td>1212</td>
<td>1306</td>
</tr>
<tr>
<td>70</td>
<td>1306</td>
<td>1306</td>
<td>1306</td>
<td>1397</td>
<td>1306</td>
<td>1397</td>
</tr>
<tr>
<td>80</td>
<td>1397</td>
<td>1397</td>
<td>1397</td>
<td>1484</td>
<td>1397</td>
<td>1484</td>
</tr>
<tr>
<td>90</td>
<td>1484</td>
<td>1484</td>
<td>1484</td>
<td>1566</td>
<td>1484</td>
<td>1566</td>
</tr>
<tr>
<td>100</td>
<td>1566</td>
<td>1566</td>
<td>1566</td>
<td>1566</td>
<td>1566</td>
<td>1566</td>
</tr>
</tbody>
</table>

TABLE 13. Limits of the variation of the reduced Fermi energy, $\eta$, for nondegenerate semiconductors.

<table>
<thead>
<tr>
<th>Goldsmid</th>
<th>This Dissertation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta &lt; -2$ acceptable</td>
<td>$-4 &gt; \eta \geq -5$</td>
</tr>
<tr>
<td>$\eta &lt; -4$ preferable</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 8

Comparison of the Results and Findings of this Dissertation with Those of Other Investigators

Dwelling on the topic of the upper limit or bound on the Seebeck coefficient, S, and the thermoelectric figure of merit, ZT, Professor H.J. Goldsmid\textsuperscript{74} made the following comments in his pioneering book on thermoelectric refrigeration:

“It is unreasonable to suppose that a single semiconductor could have the highest known values of both carrier mobility and effective mass, since these two parameters are inter-related. However, in estimating an upper limit of ZT, it will be supposed that a single semiconductor does combine the very low lattice thermal conductivity derived above, which is about 0.002Wcm\textsuperscript{-1}K\textsuperscript{-1}, with the best known value of $\mu(m^*/m)^{3/2}$, i.e., that for bismuth at 300K, which is about 750cm\textsuperscript{2}V\textsuperscript{-1}sec\textsuperscript{-2}. It is also assumed that the energy gap is large enough for the conduction to be extrinsic, when the Seebeck coefficient is optimized ($E_g > 10k_BT$).

The theory of Section 3.2.1 shows that the optimum Seebeck coefficient for the hypothetical material would be $\pm 350\mu$VK\textsuperscript{-1} and the corresponding value of ZT is 4.1. Since a most unlikely combination of parameters has been assumed in this estimation of the upper limit of ZT, the situation may be summarized:

(i) Present trends indicate that ZT cannot be much more than unity.
(ii) It seems unlikely that ZT will exceed two.
(iii) Values of ZT greater than four seem out of the question.”

Furthermore, we now refer to earlier research work done by Rittner\textsuperscript{75} and Rittner and
Neumark\textsuperscript{76}. The former\textsuperscript{75} carried out the first computation of an upper bound on $Z$ for a thermocouple with n- and p-type branches of the same material. This calculation was based upon essentially a single-band semiconductor model with spherical energy surfaces, a carrier mean free path, $l$, determined by elastic scattering and dependent on carrier energy, $E$, in accordance with $l = l_0 E^r$ and neglect of the lattice component of heat conductivity relative to the electronic component. Rittner\textsuperscript{75} reported $ZT$ values ranging from 0.2 to 4.4. In a strictly single-band model, the Seebeck coefficient can approach infinity as the Fermi level is allowed to diverge without limit from the band edge. This difficulty was avoided by the latter authors \textsuperscript{76} by constraining the maximum value of the Seebeck coefficient into correspondence with approach of the Fermi level to its position in the intrinsic semiconductor, thus to this extent, introducing a second band into the model. The upper bound proved

$$Z_{\text{MAX}} = \frac{(r + 2 + E_g / 2k_B T)^2}{(r + 2)T}$$

(8.1)

where $E_g$ is the intrinsic band gap. For bismuth-telluride based thermocouples operating in the neighborhood of room temperature, Eq. (8.1) leads to an upper bound, about an order of magnitude higher than the best devices presently existing. A somewhat smaller upper bound on $Z$ has been derived by Donahoe\textsuperscript{77} based on very similar reasoning, but using for the maximum possible Seebeck coefficient the one obtained by Price\textsuperscript{78} for a two band model. In this treatment\textsuperscript{78}, the figure of merit was calculated for a thermocouple with the two branches of the couple having equal values of a material parameter, $\gamma$, involving effective mass and mobility ratios. Litttman and Davidson\textsuperscript{79} have concluded, from arguments of irreversible thermodynamics, that there is no theoretical upper bound on $Z$ imposed by the second law. However, it has been pointed out by Rittner\textsuperscript{80} that thermodynamics is not a proper discipline to
deal with questions of this kind and that it is necessary to employ statistical or kinetic methods in connection with a specific physical model of the solid, in order to optimize the figure of merit, as has been done in references 75 and 77. It is also emphasized in reference 80 that there is no unique upper bound on $Z$, as the value of $Z_{\text{max}}$ depends upon the particular model employed to represent the material.

In another paper, a general two-band semiconductor model, treating each branch of the couple separately, and allowing for unequal values of the parameter, $\gamma$, in both branches, was examined by Simon$^{81}$, using statistical methods. He was unable to set a finite upper bound on $Z$, but pointed out that in order to demonstrate the existence of such a bound, his mathematical analysis must be supplemented by the establishment of a sufficient number of limiting values, or relationships, among the various physical quantities that enter into the definition of the important material parameters. The Simon$^{81}$ model was discussed further from this point of view. It was shown that the limited adjustability of the Fermi level position, because of dopant solubility limitations, together with other reasonable physical restrictions associated with the semiconductor model, lead to an upper bound on $Z$ only somewhat higher than Eq. (8.1).

It has been shown, in conjunction with the Simon$^{81}$ two-band semiconductor model, that consideration of existing constraints on the Fermi level position, together with use of reasonable physical values for the band parameters, lead to a finite upper bound on the thermoelectric figure of merit, which is of the same order of magnitude as that previously found by Rittner$^{75}$ for a one-band model. However, the extra degree of freedom in the Simon model represented by the parameter, $\gamma$, permits a higher value for $Z_{\text{max}}T$ than the upper bounds derived from earlier models$^{75,77}$, provided that a sufficiently large ratio for $m^*_h / m^*_e$ can be realized concomitantly with a sufficiently small absolute value of $m^*_e$. As in the earlier models, the contradictory
conditions persist that the lattice component of the heat conduction be as small as possible and the band gap as large as possible. Peculiar to the Simon two-band model is the requirement that acoustic lattice scattering be the predominant process determining the mobility in the negative branch of the couple and that impurity scattering be the dominant one in the positive branch of the couple. Also, in the negative branch, it is desirable that there be a large number of equivalent extrema in the conduction band and a small number in the valence band, that there be a large mass anisotropy for the electrons but not for the holes, and that the deformation potential for the valence band edge be much larger than that for the conduction band edge. The first two of these three factors (with conduction and valence band reversing roles) are also somewhat useful in the positive branch.

It is abundantly evident that most, if not all, of the assumptions and arguments made by the aforementioned authors\textsuperscript{75-81}, particularly those indicated by Simon\textsuperscript{81}, have a very remote chance of eventually being realized in practice. Not surprisingly, they ended up being attacked and challenged by Wood\textsuperscript{82}. The latter’s main argument runs as follows: Rittner and Neumark\textsuperscript{83} and Donahoe\textsuperscript{77} have argued that the maximum $S$, and hence maximum $Z$, will occur when the Fermi level is at the middle of the band gap, $E_g$. They assumed extrinsic n-type material, with $m^*_p > m^*_n$ and $\mu_p << \mu_n$, since contributions from charge carriers of the opposite sign (holes) will otherwise reduce the value of $S$ in the expression for the total Seebeck coefficient, which is

$$S = \frac{S_n\sigma_n + S_p\sigma_p}{\sigma_n + \sigma_p}$$

Hence they obtain

$$Z = \frac{(A + E_g / 2k_B T)^2}{AT}$$

Since there is no theoretical limit to the band gap, obviously there is no theoretical limit to the
magnitude of $Z$. Similar, but more refined, arguments, by other investigators, considering bipolar conduction, have essentially concluded that the upper limit on $Z$ is governed by the value of the energy gap and the mobility ratio, i.e., in every case, it was found that no theoretical upper bound could be placed on $Z$. This conclusion is supported by the work of Littman and Davidson\textsuperscript{79} who derived $Z$ using irreversible thermodynamics and found a theoretical upper bound of infinity. In lieu of a theoretical limit, likely values of the relevant parameters have been substituted in the equations of Rittner and Neumark\textsuperscript{83} and Donahoe\textsuperscript{77} to obtain typical $Z_{\text{max}}$ values\textsuperscript{82} of about $17 - 45 \times 10^{-3} \text{K}^{-1}$. These $Z$ values are far in excess of those found experimentally to date.

Now, before closing this section on the comparison of our results with those of other investigators, we thought it worthwhile to try to derive, or prove the validity of, Eq. (8.1), which corresponds to, or is identical with, Eq. (21) in Rittner’s\textsuperscript{75} paper, Eq. (1) in Rittner and Neumark’s\textsuperscript{76} paper and Eq. (3.29) in Dr. C. Wood’s\textsuperscript{82} research report. We will rewrite the latter equation below in order to make our discussion easier and more clearly understood

$$ (ZT)_{\text{MAX}} = \left( \frac{A + E_g}{2k_B T} \right)^2 \quad (8.4) $$

Now there should be no doubt whatsoever that this whole analysis deals exclusively with nondegenerate semiconductors. Also, earlier in this dissertation, when we decided to lower the reduced Fermi energy, $\eta$, until the middle of the forbidden energy band gap, also known as the intrinsic Fermi energy level, in order to maximize the value of $ZT$, we ended up deriving Eqs. (7.39-7.41). Now invoking Eqs. (5.8) and (5.9), substituting the value of $\eta$, according to Eq. (7.41), into Eq. (5.8) and then substituting both Eqs., (5.8) and (5.9), into Eq. (5.6), we finally obtain
(ZT)_{\text{MAX}} = \frac{\left(\frac{k_B}{e}\right)^2 \left[-\frac{E_g}{2k_B T} - \left(r + \frac{5}{2}\right)\right]^2}{\left(\frac{k_B}{e}\right)^2 \left(r + \frac{5}{2}\right)} = \frac{\left[\frac{r + 5}{2} + \frac{E_g}{2k_B T}\right]^2}{\left(r + \frac{5}{2}\right)} = \frac{(A + \frac{E_g}{2k_B T})^2}{A} \quad (8.5)

where

\[ A = r + \frac{5}{2} \quad (8.6) \]

according to Goldsmid\textsuperscript{72}. And that is the proof, which all of the referred to authors\textsuperscript{75, 76, 82} failed to provide. Our evaluation of the constant, A, is more accurate than theirs, which was based on \( A = r + 2 \), with the r denominations being 0, 1/2 and 2, which reflects the intellectual world, or school, of semiconductors and thermoelectrics of the thirties, forties and fifties of the twentieth century, dominated by A.F.Ioffe. Beyond that era, leadership of the thermoelectrics field went to H.J.Goldsmid.

With respect to our foregoing statements concerning the need, or decision, to lower the reduced Fermi energy until the middle of the forbidden energy band gap, we would like to emphasize that that is not absolutely necessary. The only thing that is definitely unavoidable is to reduce the reduced Fermi energy to the value \( \eta = -5 \), because that assures that the maximum dimensionless figure of merit, \( ZT = 24.5 \), along with, of course, \( r = -1/2 \), will be achieved. If the reduced Fermi energy happens to reach or touch the middle of the forbidden energy band gap and if the reduced energy band gap \( \frac{E_g}{k_B T} \leq 10 \), then the intrinsic condition may come about or exist. If that is the case, then it may be advisable for the Fermi energy to stay short of, or higher than, the middle of the band gap. In that case, the maximum figure of merit, referred to earlier, will not be achieved. That is why it is always advisable to lower the reduced Fermi energy and bring it as close to the middle of the forbidden energy band gap as possible. There is
no other alternative.

In our foregoing analysis and review of the research reports of a number of other investigators\textsuperscript{74-83}, we have been exposed to a large number of arguments and counterarguments as to whether there exists an upper bound or limit on $E_g$, $S$ and, more importantly, on $Z$ and/or $ZT$. Both the opponents and proponents of the upper bound on $ZT$ are making wonderful expositions in support of their respective points of view. Some of them are going as far as to suggest

$$E_g = 16k_B\, T$$  \hfill (8.7)

as their preferred or recommended value for the forbidden energy band gap, which, obviously, applies only to nondegenerate semiconductors and, therefore, has nothing to do with degenerate ones. Of course, the higher the value of $E_g$, the higher will $ZT$ be according to any of Eqs. (8.1) and (8.3–8.5). Such high values as 16 for the reduced energy band gap, nevertheless, are no longer acceptable nowadays. The maximum $\eta$ value recommended at present for good thermoelectric materials is 10, so that Eq. (8.7) becomes

$$E_g = 10k_B\, T$$  \hfill (8.8)

In thermoelectrics, as in many other fields, more is not necessarily always better. A very large energy band gap should be regarded always with suspicion, because that implies a greater difficulty for free electrons to be generated and move around, which ends up lowering the electrical conductivity until it becomes zero. For best results, energy band gaps must be neither too narrow nor too wide. That is why semiconductors, with their moderately-sized energy band gaps, are the materials most indicated and recommended for thermoelectric energy conversion. According to Pierre Aigrain\textsuperscript{84}, a reasonable energy band gap for power generation purposes should be about $0.6eV$. For refrigeration, that value goes down to about $0.3eV$. 
Let us now try to see and analyze what happens if we move away from semiconductors to either metals or dielectric and ionic materials:

For metals, \( \kappa_L \approx 0 \) and \( S \) is very small, thus

\[
ZT = \frac{S^2 \sigma T}{\kappa_e} = \frac{S^2 \sigma T}{\sigma LT} = \frac{S^2}{L}
\]

and because \( S \) is very small, metals are not appropriate for thermoelectric energy conversion, but they can still be used as thermocouples for temperature measurement.

For dielectric and ionic materials, \( \sigma \approx 0 \) and \( S \) is very high, or enormous, therefore

\[
ZT = \frac{S^2 \sigma T}{\kappa_L} = 0
\]

because \( \sigma \approx 0 \), as a consequence to the fact that there are no free electrons available to carry the electric charge and thus conduct electricity. Thus, despite their very high \( S \), dielectric and ionic materials are not suitable for thermoelectric energy conversion.

Thus, only semiconductors with their moderately wide energy band gaps are the materials most suited for direct thermoelectric energy conversion.

Referring now to the possible upper bounds or limits on \( ZT \) that we discussed earlier, what we must say here is that yes, these limits or bounds do exist. We actually did impose those limits on the maximum possible \( ZT \) values for degenerate and nondegenerate semiconductors, while at the same time, we were totally unaware that we were doing so. These bounds or limits actually came about as a natural consequence to, or outcome of, our entire research work on this dissertation, not because we were looking for ways to impose them in order to satisfy, or be responsive to, any preconceived concepts or notions we might have had in relation thereto. These bounds or limits became known to us only after most of the research work on this dissertation had already been carried out. The limits or bounds that we have found out were that
for nondegenerate semiconductors, the maximum possible figure of merit $ZT=24.5$, or about 25, and that for degenerate semiconductors, the maximum possible $ZT=23.85$, or about 24. So, we many end up writing a generic equation for both types of semiconductors

$$ (ZT)_{\text{MAX}} \begin{array}{c} \text{degenerate} \\ \text{nondegenerate} \end{array} \leq 25 \quad (8.11) $$

For degenerate semiconductors, the limit was imposed quite naturally by the fact that in order to achieve a $ZT$ value greater than 23.85, the material would have had to behave as a nondegenerate semiconductor, i.e., the rules and equations of degenerate semiconductors were not applicable any more beyond that $ZT$ value of 23.85. We thus had a case of reductio ad absurdum. For nondegenerate semiconductors, the situation was completely different, since the limit was imposed when we tried to lower the Fermi energy down to the middle of the forbidden energy band gap, which is also known as the intrinsic Fermi level. And since we knew that we couldn’t lower or push the Fermi level any further below the intrinsic level, therefore, that imposed the upper bound or limit on $ZT$ for nondegenerate semiconductors which, as we already know, is $(ZT)_{\text{MAX}} = 24.5$.

When we began our research work on this dissertation, we reviewed the theoretical fundamentals of thermoelectric energy conversion and the transport of heat and electricity in solids. We then quickly derived Eq. (5.6) and a number of other relationships, up to Eq. (5.25), and ended up constructing Table (1). As soon as we had done that, we remained with the secret inner notion and hope that the maximum $ZT$ value that would possibly be achieved would lie somewhere in the middle of the entire spectrum of $ZT$ values presented, maybe 40 and probably even 50. Unfortunately, however, our initial inner secret hope or belief concerning the maximum $ZT$ value achievable with either nondegenerate, or degenerate, semiconductors will
not be realized. What we are going to actually get will be about 50% of that, that is, only one fourth of the entire extension or interval of the ZT spectrum, i.e.,

\[(ZT)_{\text{MAX}} = \frac{100}{4} = 25\]  \hspace{1cm} (8.12)

The generic inequality (8.11) expresses the same result. These are the new facts of life in thermoelectrics. Having arrived at the above result, justifies our having embarked on and carried out the entire research work related to this dissertation. Should the above result be realized, or proven true, experimentally by some extraordinary new material, then that would make us very happy indeed, since that would add more value to what we have achieved and accomplished while working on this dissertation.

We will now focus our attention once more on the arguments and counterarguments made by several research workers\textsuperscript{75-83} as to whether an upper bound or limit on \(E_g\) exists. To those who argue that there is no upper bound on \(E_g\), we would like to say that if that is true, then as \(E_g\) increases, the material will eventually become, and behave like, an ionic or dielectric one with an electrical conductivity, \(\sigma \approx 0\). Although it will acquire a very high Seebeck coefficient, its thermoelectric figure of merit, ZT, will vanish, according to Eq. (8.10). All of the foregoing events will happen long before \(E_g\) reaches infinity. So, an upper bound, or limit, on \(E_g\) must exist. Its value must be maintained neither too low nor too high. Eq. (8.8) provides a valuable clue. Therefore, statements by some authors\textsuperscript{79, 82} about the nonexistence of an upper bound on \(E_g\) must now be rejected. Furthermore, these authors\textsuperscript{82} argue, since there is no theoretical limit to the energy band gap, obviously there is no theoretical limit to the magnitude of \(Z\). Both parts of the foregoing statement are mistaken, since we definitely know already that neither the energy band gap, \(E_g\), nor the figure of merit, \(Z\), and the dimensionless version thereof, ZT, are unbounded or limitless. The big blunder that was committed here by all these authors\textsuperscript{75-82} was
their introduction of $E_g$ into the equations for $Z$ or $ZT$. Therefore, if $E_g$ were considered, or assumed to be, unbounded, then also $ZT$ would consequently become unbounded or limitless. That again is another case of reductio ad absurdum. These authors have considered only the mathematics of the problem and ignored the physics thereof. That is where they went wrong. In this dissertation, we have established that the thermoelectric figure of merit, $ZT$, is indeed bounded for both nondegenerate and degenerate semiconductors, with the maximum value or limit being 24.5 and 23.85, respectively. These two limits were not obtained as a result of any superficial assumption or quick and careless evaluation, but were the outcome of a thorough, persistent and prolonged research effort that lasted more than two years. Consequently, we reject the range of values for $(ZT)_{\text{MAX}}$ of 17-45 ($T = 1000K$), reported by C. Wood$^{82}$, and mentioned earlier in this section, as unreliable.

Lin, Sun and Dresselhaus$^{85}$ have recently developed a theoretical model for the transport properties of cylindrical Bi nanowires. Based on the band structure of Bi nanowires and the semiclassical transport model, the thermoelectric figure of merit, $ZT$, was calculated for Bi nanowires, with various wire diameters and wire orientations. Their results show that the trigonal axis is the most favorable wire orientation for thermoelectric applications, and $ZT > 1$ is predicted for n-type trigonal wires with diameters $d_w < 10 \text{nm}$. The effect of T-point holes on the thermoelectric performance has also been discussed, and it was found that $ZT$ can be significantly enhanced, especially for p-type nanowires, if T-point holes are removed or suppressed, which can possibly be achieved by Sb doping. The entire investigation was carried out at only one temperature, namely 77K. $ZT$ values of about 2 and 6 were obtained or calculated for n-type Bi nanowires having diameters of 10 and 5 nm, respectively. It was found that the thermoelectric performance substantially decreases as the diameter of the Bi nanowires...
increases. A similar thermoelectric performance was obtained, or calculated, for p-type Bi nanowires.

Lin et al.’s results are encouraging and appear to be useful for temperatures well below room temperature, i.e., for refrigeration. It is unlikely that they can be extrapolated or extended to temperatures above room temperature. It is therefore, highly unlikely that they can be used for power generation applications. At temperatures well above room temperature, the Seebeck coefficient tends to increase substantially above its room temperature value, and since

\[ ZT = \frac{S^2}{L} \]

at temperatures higher than room temperature, there is no way for Lin’s results to be applicable or competitive here.

In a recent report or article, Professor H. J. Goldsmid\textsuperscript{86}, who in 1964\textsuperscript{74} had stated that ZT cannot be much more than 1, that it seems unlikely that ZT will exceed 2 and that values of ZT greater than 4 seem out of the question, now comes and tries to convince us that the figure of merit of new bulk thermoelectric materials cannot exceed the limit of ZT = 4. In establishing this new limit, Professor Goldsmid is basing his argument on the assumption that any promising figure of merit of new bulk thermoelectric materials is due entirely to their very low thermal conductivity, which already has its minimum value. We strongly disagree with Professor Goldsmid on this issue and we believe that the limit on the figure of merit of these materials \((ZT)_{\text{MAX}} = 25\), that is, 6 times greater than the limit established by him. In reaching that conclusion, we have removed the lattice thermal conductivity entirely from any equations we might have used to calculate ZT for semiconductors. To this effect, we used the very simple, elegant and very powerful equation that represents the cornerstone of this investigation, which is

\[ ZT = \frac{S^2}{L} \]
Professor Goldsmid, obviously encouraged by Lin et al.’s results on Bi-based nanowires, or quantum wires, has suggested that a minimum lattice conductivity $\kappa_L$ of 0.1 W m$^{-1}$ K$^{-1}$, coupled with a power factor, $S^2\sigma T$, of 103.9 W m$^{-1}$ K$^{-1}$ at 77 K and a Seebeck coefficient, $S$, of 630 $\mu$V K$^{-1}$, would yield a ZT value of about 20. Professor Goldsmid suggested, furthermore, that the aforementioned value, ZT=20, will remain the same, even if the temperature is increased from 77 K to 300 K. Professor Goldsmid pushed his ideas and concepts much further, saying: “We might go further and propose that ZT might have a limiting value of about 20 at all temperatures of practical interest.” Professor Goldsmid finally concluded his report making the following statements: “It has been shown that there is no reason to raise the upper bound for the dimensionless figure of merit above four for bulk thermoelectric devices but there are good grounds for supposing that a value as high as 20 might be reached using low-dimensional structures. It is possible that a still greater value might be reached if one can realize the situation in which the electron and phonon flows are independent of one another. If, indeed, ZT were to reach a value of about 20, then the performance at ordinary temperatures would be similar to that of a vacuum diode with a cathode having a work function of 0.3 eV.”

We strongly disagree with many of the statements of Professor Goldsmid. His reference to the value of the minimum lattice thermal conductivity as being no more than about 0.1 to 0.2 W m$^{-1}$ K$^{-1}$ is not true. The research reports on that matter that came to our attention, thus far, say otherwise. They indicate that the minimum lattice thermal conductivity for bulk materials is about 0.3 W m$^{-1}$ K$^{-1}$. Also the value of the lattice thermal conductivity $\kappa_L$ remains important as long as $\kappa_L \gg \kappa_e$ which is true for low temperatures and for applications where doping is moderate, i.e., refrigeration. But when we go to temperatures substantially higher than room temperature, coupled with mass and volume fluctuation scattering of phonons, along with
moderate to heavy doping, what we get is $\kappa_L \ll \kappa_e$, so $\kappa_L$ becomes negligible or unimportant. We don’t care how much its value is, we just neglect it. That is the main difference between the low temperature and the high temperature domains, or ranges, in thermoelectrics. They are governed by different laws. So the fact that the highest observed value for $\mu(m^*/m)^{3/2}$ is 0.075 $m^2V^{-1}sec^{-1}$ for electrons in Bi is of no interest to us in the temperature range above room temperature. Also the Seebeck coefficient value, $S=630\mu VK^{-1}$, as well as the lattice thermal conductivity value, $\kappa_L=0.1\text{Wm}^{-1}\text{K}^{-1}$, are only assumed values. They have not been corroborated experimentally.

In response to Professor Goldsmid’s assertions, setting a limit, or bound, of four on the dimensionless thermoelectric figure of merit of new bulk thermoelectric materials, i.e.,

$$(ZT)_{\text{max}}^{\text{bulk}} = 4$$

we would like to carry out an experimental research program aiming at producing and evaluating the thermoelectric performance of the following four component material:

Mg-Si-Pb-Ba

Production of the aforementioned material through any appropriate process or method, such as melting or by using the powder metallurgy technique, or any other method, will end up bringing about a quaternary alloy or solid solution of intermetallic compounds containing magnesium silicide, magnesium plumbide, barium silicide and barium plumbide, wherein the alloy or solid solution is defined by the following chemical constitutional formula:

$$\text{Ba}_{2r}\text{Mg}_{2(1-r)}\text{Si}_{(1-x)}\text{Pb}_x$$

wherein $r$, $(1-r)$, $(1-x)$ and $x$ represent the atomic or molecular proportion of each of barium, magnesium, silicon and lead in the alloy, respectively, and wherein $r$ varies from 0.1 to 0.4 and $x$
The objective would be to prepare doped and undoped samples of this material by any appropriate method and then measure the following three properties:

1. Seebeck Coefficient
2. Electrical Conductivity
3. Thermal Conductivity

We would then calculate the dimensionless figure of merit as follows:

\[
ZT = \frac{S^2 \sigma T}{\kappa}
\]

All the measurements should be carried out at various temperatures ranging from 300K to 1000K. In the above expression for \(ZT\), \(\sigma\) and \(\kappa\) will tend to cancel out, because \(\kappa_L\) is very small or negligible. So \(S\) becomes the crucial parameter that determines how good the material is thermoelectrically. This fact can be expressed by the following equation:

\[
ZT = \frac{S^2}{L}
\]

where \(L\) is the Lorenz number, whose value depends on the concentration of the free charge carriers, whether it is equal to or less than \(2.5 \times 10^{19}\) carriers \(\text{cm}^{-3}\), in which case we have a nondegenerate semiconductor, which implies that \(L\) depends on the value of the scattering parameter or index, \(r\), whether it is \(-1/2\), \(1/2\) or \(3/2\). If the concentration of the free charge carriers is greater than \(2.5 \times 10^{19}\) carriers \(\text{cm}^{-3}\), we then have a degenerate semiconductor and the value of \(L\) in that case will be the same as that for metals, as well as independent of \(r\).

In 1995, Slack\textsuperscript{30} wrote that there has long been an interest in the question of the ultimate performance limits for thermoelectric materials, both for generators and for coolers. Since \(Z\) depends inversely on the thermal conductivity of the material from which the device is made, it
is clear that low thermal conductivity values are desirable. The highest $Z$ values are likely to be found in materials with very low thermal conductivities. It has been pointed out $^{31}$ that for each material there is a well defined value for its minimum lattice thermal conductivity, $\kappa_{\text{MIN}}$. This value is reached when all phonons have a mean free path equal to their wavelength. An alternative model $^{87}$ sets this minimum at the point where the limit is one-half wavelength. These minimum values differ by a factor of two, if acoustic phonons carry the heat. The effect of such a $\kappa_{\text{MIN}}$ concept on predicting the ultimate performance of Si-Ge thermoelectric generators has been demonstrated by Slack and Hussain$^{88}$. The highest possible values of $ZT$ in this system are for n-type material with $ZT = 1.75$ at 1050K. The highest values achieved to date are $ZT = 1.0$ at 1050K.

For thermoelectric cooling materials, the analysis must be broader than it was for Si-Ge because we need to consider not only the lower limits on the thermal conductivity but also the upper limits on the power factor, $P$. Therefore, an exploration has been made of a large number of known elemental and binary compound semiconductors in order to fix reasonable limits on both $P_{\text{max}}$ and $\kappa_{\text{min}}$. The conclusion is that in the temperature range between 77K and 300K, one might obtain $ZT$ values as high as 4, but that substantially larger values are highly unlikely. Slack$^{30}$ eventually arrived at the following nine conclusions:

1. There are no currently known materials that have $R$ values sufficiently high to produce $ZT=1$, over the temperature range from 77K to 300K, where $R = \frac{U}{\kappa_L} = \mu \left( \frac{m^*}{m_o} \right) ^{3/2} / \kappa_L$, where $\mu$ is the electronic mobility, $\left( \frac{m^*}{m_o} \right)$ is the ratio of the effective to the free electronic mass and $\kappa_L$ is the lattice thermal conductivity.

2. The model for PGEC (phonon glass electron crystal) indicates that a $ZT = 4$ is possible from
77K to 300K if a material with $U = 1800 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ and $\kappa_L = 2.5 \text{ mWcm}^{-1}\text{K}^{-1}$ can be found. The decisive ratio is $R = U / \kappa_L = 7.20 \text{ cm}^2\text{s}^2\text{Kg}^{-1}\text{V}^{-1}$.

3. It seems unlikely that cooling materials with $ZT > 4$ will ever be found.

4. The goal of $ZT=1$ can be obtained with $R=0.60 \text{ cm}^2\text{s}^2\text{Kg}^{-1}\text{V}^{-1}$.

5. The place to look for such materials is in small-band-gap, heavy element compound semiconductors, with electronegativity differences of $|\Delta X| \leq 0.5$.

6. For doping levels less than $10^{19}\text{cm}^{-3}$ charged impurity scattering that might lower the mobility will be screened out by the high dielectric constants.

7. For $ZT = 4$ materials, the traditional use of mixed crystal formation to enhance the ZT values will fail because alloy scattering of the charge carriers increases as $|\Delta X|^2$. There are very few pairs of elements with small $|\Delta X|$ values available.

8. For $ZT = 1$ materials, mixed crystals might help in a few special cases.

9. Minimum thermal conductivity values can sometimes be found in crystals that contain “rattling” atoms or molecular groups.

   In his research report, Slack\textsuperscript{30} wrote that a model thermoelectric material is “a phonon glass and an electron crystal.” Its ZT values are calculated as a function of carrier concentration and temperature. He also stated that the figure of merit, $Z$, for a thermoelectric material is given by

   $Z = \frac{S^2\sigma}{\kappa_{TOT}}$ \hspace{1cm} (8.13)

   where

   $\kappa_{TOT} = \kappa_L + L\sigma T$ \hspace{1cm} (8.14)
where \( S \) = Seebeck coefficient, \( \sigma \) = electrical conductivity, \( \kappa_L \) = lattice thermal conductivity, \( L \) = Lorenz number and \( T \) = absolute temperature. This can be rewritten as:

\[
ZT = \frac{\left(\frac{S}{\sqrt{\kappa_L}}\right)^2}{1 + \frac{\kappa_L}{L\sigma T}} \tag{8.15}
\]

If \( \kappa_L \ll L\sigma T = \kappa_e \), then Eq. (8.15) becomes

\[
ZT = \frac{S^2}{L} \tag{8.16}
\]

which is identical with Eqs. (5.6) or (5.7), and which is mainly applicable in the high temperature range, above room temperature, that is, for thermoelectric power generation. For temperatures well below room temperature, namely refrigeration, and especially if \( \kappa_L \) has not also been substantially reduced through mass and volume fluctuation scattering of the phonons, or through any other means, then Eq. (8.15) will have to be used rather than Eq. (8.16). And in that case, the value of \( ZT \) will be lower than that calculated according to Eq. (8.16). That is the major difference between low temperature and high temperature applications, as far as the dimensionless thermoelectric figure of merit, \( ZT \), is concerned. Slack\textsuperscript{30} correctly stated that his results were mainly related to thermoelectric refrigeration.
Chapter 9

General Summary and Conclusions

In this dissertation, we have started our research by deriving an equation for calculating an equation for calculating an effective dimensionless thermoelectric figure of merit, ZT, which constitutes a departure from the earlier equation established by Altenkirch. The traditional Altenkirch equation has been used by almost all research workers in thermoelectrics since the beginning of the twentieth century. In order to enable the derivation of the new equation for the effective ZT, and make it stand on a firm foundation with little or negligible error, the phonon or lattice component of the thermal conductivity, of the material used in the manufacture of the thermoelectric energy conversion device, must be reduced to a negligible value compared to the electronic component thereof. One way of achieving this, for temperatures higher than room temperature, is to use alloys or solid solutions, comprising at least three or four different elements, where mass and volume fluctuation scattering, or alloy scattering, of phonons is predominant. The foregoing is obviously related to thermoelectric power generation. For low temperature applications of thermoelectrics, notably refrigeration, rattlers, quantum wells, superlattices, clathrates, etc., are currently often used to reduce the lattice thermal conductivity. Values from 0.001 to 0.002 Wcm\textsuperscript{-1}K\textsuperscript{-1} for the minimum phonon thermal conductivity have been reported in the literature. In the opinion of most experts in the field, the minimum phonon thermal conductivity should be no greater than 0.003 Wcm\textsuperscript{-1}K\textsuperscript{-1} at room temperature. For thermoelectric energy conversion devices operating in the temperature range of 300-1000K, the phonon thermal conductivity should be further lowered to 0.0009 Wcm\textsuperscript{-1}K\textsuperscript{-1} at the higher temperature limit of 1000K, according to the Eucken inverse temperature law for semiconductor
alloys and solid solutions. This already should give us an enormous headway in the design of future thermoelectric devices for power generation. Since the phonon thermal conductivity of the semiconducting materials that we use for thermoelectric energy conversion does not go to zero, we have to make sure that the ratio of the electronic thermal conductivity to the phonon thermal conductivity is much greater than unity, in order to assure the validity of the equation, $ZT = S^2/L$, that we derived for the effective dimensionless thermoelectric figure of merit, where $S$ is the Seebeck coefficient and $L$ is the Lorenz number. For that purpose, several equations have been derived expressing the ratio of the electronic to the phonon, or lattice, thermal conductivity for the various values of the electron scattering parameter, $r$, namely: -1/2, 1/2 and 3/2, for both nondegenerate and degenerate semiconductors.

When one compares the effective $ZT$ equation with the earlier equation due to Altenkirch, one notices that out of the three properties involved in the latter equation for $ZT$, two are missing in the effective $ZT$ equation, which are the electrical conductivity and the thermal conductivity. We have thus ended up retaining only one property, namely the Seebeck coefficient, in the newer equation. Over the years, the best way to improve or optimize the performance of thermoelectric devices has been to maximize the product, $S^2\sigma$, known as the power factor, in the numerator of the Altenkirch equation, by taking its derivative and equating it to zero, i.e., $\partial(S^2\sigma)/\partial\sigma = 0$. This has led, or contributed, to the idea or concept of optimum doping, which leads to drastically increased $\sigma$ with virtually little variation in $S$, and thus $(S^2\sigma)$ is maximized. However, thanks to the Wiedemann-Franz law, the electronic component of the overall thermal conductivity, which is in the denominator of the Altenkirch equation, also increases in proportion to the $\sigma$ increase. This leads to an increase in the total thermal conductivity, which is always undesirable, because it tends to reduce or eliminate the effect of
increasing the power factor, \((S^2\sigma)\). This kind of situation is called: “Conservation of nastiness.” We indeed do have conservation of nastiness in thermoelectrics, because, while on the one hand we increase the numerator through maximizing the power factor, yet on the other hand, the denominator of the Altenkirch equation also increases, owing to the Wiedemann-Franz law. In this dissertation, through our derivation and use of the effective ZT, we have put an end to this conservation of nastiness and eliminated that vicious circle, once and for all, particularly in the temperature range above room temperature, notably for power generation applications.

Next, we moved forward to the most important part of this investigation. New relationships have been worked out for the effective ZT as a function of the reduced Fermi energy, \(\eta\), and the scattering parameter, \(r\), for nondegenerate semiconductors. We elaborated correlations of the reduced Fermi energy, \(\eta\), and the effective ZT as functions of the free charge carrier concentration, \(n\), and the absolute temperature, \(T\). This elaboration was done for values of the scattering parameter of \(-1/2\), \(1/2\) and \(3/2\). For nondegenerate semiconductors, we calculated the values of ZT for certain values of \(\eta\) ranging from -2 to -5, and we also calculated the corresponding values of \(S\). We also developed an independent correlation for \(\eta\) as a function of \(n\) and \(T\), which we converted into another relationship giving \(n\) as a function of \(\eta\) and \(T\). We then used all of the foregoing to construct three tables giving the values of \(n\), ZT and \(S\) for the aforementioned values of \(\eta\) and values of \(T\) from 300K to 1000K. We carried out these calculations for \(r = -1/2, 1/2\) and \(3/2\). For degenerate semiconductors, we developed an independent relationship for ZT as a function of \(T\) and \(n\). This relationship was then converted into a new one giving \(n\) as a function of \(T\) and ZT. We then used this new equation to construct one additional table giving values of \(n\) for certain values of ZT ranging from 1 to 100, and the predetermined values of \(T\) of 300K, 500K, 750K and 1000K.

145
On the basis of the numerical data presented in the aforementioned four tables, we were able to draw the following conclusions:

1. The highest ZT value possible for a nondegenerate semiconductor is 24.5 and is achieved through a material for which \( r = -1/2, \eta = -5 \) and \( S = 602 \mu \text{VK}^{-1} \).

2. The highest ZT value possible for a degenerate semiconductor is 23.9 and is achieved through a material for which \( T = 1000 \text{K}, n = 2.6 \times 10^{19} \text{ cm}^{-3} \) and \( S = 764 \mu \text{VK}^{-1} \).

3. Comparing the foregoing value for \( S \), namely, \( S = 764 \mu \text{VK}^{-1} \), with the earlier one for the nondegenerate semiconductor having the highest ZT value of 24.5 and \( S = 602 \mu \text{VK}^{-1} \), it may be noticed that the nondegenerate material places a substantially lesser requirement on \( S \) than the degenerate one. The higher the demand on the Seebeck coefficient, the less likely it is that such a demand will be realized in practice. For that reason, and bearing in mind that the two ZT values are almost identical, we end up adopting the nondegenerate semiconductor, as the material of preference.

More generally, we should select nondegenerate in preference to degenerate semiconductors for thermoelectric energy conversion applications. Nondegenerate semiconductors require far lesser doping than degenerate ones. As mentioned earlier, they also require a lower Seebeck coefficient than the degenerate semiconductors for the same ZT value, especially those for which \( r = -1/2 \). Therefore, nondegenerate semiconductors for which \( r = -1/2 \) and, if at all possible, \( \eta = -5 \), that is, with relatively little doping, should be our materials of preference for thermoelectric energy conversion, particularly for power generation applications. That is perhaps the most important conclusion of this dissertation.

The key to success in this research program, and the results and conclusions originating therefrom is to assure the validity of the equation, \( ZT = S^2/L \). This definitely comes about
through the near elimination of the lattice thermal conductivity. The latter has been the main culprit hindering the progress of thermoelectricity throughout the twentieth century.
Appendix 1

General Review of the Brillouin Zone Concept

The face-centered and body-centered cubic lattices have been proven to be mutually reciprocal. A face-centered cubic lattice has a body-centered cubic lattice as its reciprocal. Wave propagation in a periodic medium for three dimensions has a general solution given by the equation $\Psi = A(r) \exp[2\pi i (\mathbf{v} \cdot \mathbf{r} - \omega t)]$ where $A(r)$ is the amplitude, $\mathbf{v}$ is the frequency, $a$ is the reciprocal of the wavelength ($1/\lambda$) and $r$ is the distance, from some point in a cell $(n_1, n_2, n_3)$ to the origin of the first cell $r = r_0 + r' + n_1d_1 + n_2d_2 + n_3d_3$. This is the basis of the Wigner-Seitz method of the theory of solids. Thus the wave motion of particles in a discontinuous lattice is given by

$$\Psi_j = A_j \exp[i \left( \omega t - n_1k_1 - n_2k_2 - n_3k_3 \right)]$$

where $k_j = \frac{2\pi a_j}{\lambda}$ is the wave vector. This is the wave function inside the cell $(n_1, n_2, n_3)$, or

$$\Psi = A''(r') \exp[i \left( \omega t - n_1k_1 - n_2k_2 - n_3k_3 \right)]$$. Wigner and Seitz choose their elementary cell in a manner similar to that in which zones are constructed for the reciprocal lattice, i.e. by constructing a polyhedron about a lattice point. The faces of the polyhedron are planes that are perpendicular bisectors of the lines joining the points taken as the origin with neighboring lattice points. The motion of electrons in a potential field $(x, y, z)$, e.g. free electrons in a metal, is obtained from the solution of the Schrödinger equation

$$\nabla^2 \Psi + k^2 (E - V) \Psi = 0$$

where

$$k = \frac{8\pi^2 m}{\hbar^2} = \frac{2m}{\hbar^2}$$. This theory applies directly to the motion of free electrons in a metal and yields the usual rules about zone structure. In fact, zone structure is completely independent of the special meaning of the waves considered and must be the same for elastic, electromagnetic and Schrödinger electronic waves.
In the analysis of wave propagation in a one-dimensional lattice, it is found that the frequency is a periodic function of the wave number, \( a = 1/\lambda \), or the wave vector, \( k = 2\pi/\lambda \). Hence, for a given frequency there is ambiguity in the wave length and the direction of propagation. So an interval containing one period of the frequency is chosen and taken symmetrically about the origin, and then the wave number is restricted to values in this interval. This interval is called the first Brillouin zone. The second Brillouin zone, in one dimension, would consist of two intervals, containing half a period each, one on each side of the first zone, and similarly for higher order zones. Analogues of these zones can be carried over and extended to the two- and three-dimensional lattices. These new zones will be regions in the reciprocal lattice, since this is the lattice describing the periodicity of frequency, as a function of wave number. It can be shown that the frequency \( \nu \) of the wave in a two-dimensional lattice is a periodic function of the wave number \( a=1/\lambda \), in the reciprocal lattice, with basis vectors \( b_1 \) and \( b_2 \). We must now establish a rule for choosing the area to which the wave number \( a \) is to be confined. We may start at the origin of the basis system, and confine all vectors \( a = 1/\lambda \) to the first elementary cell of the reciprocal lattice. To overcome a number of objections, we try to construct a zone that will be analogous to the first zone in the one-dimensional case. This means, first, that we must place the origin in the center of the zone. The remainder of the construction is accomplished by drawing perpendicular bisectors of the lines joining the origin to each of the other points in the reciprocal lattice. The smallest closed polygon formed by these perpendicular bisectors is taken as the first zone. It is independent of the basis system chosen, and allows propagation in all directions. Furthermore, it requires the longest wave length, describing a given disturbance, to be used, since a complete period for each direction of propagation is included in the zone. The first zone has the same area as the first elementary cell.
of the reciprocal lattice.

The construction of the second zone, and higher order zones, is more complicated. They must all be bounded by perpendicular bisectors of lines joining the origin with other lattice points, and there can be no perpendicular bisectors passing through the interior of a zone. If a wave propagates through a continuous medium, with small periodic variations, in such a manner that its wave number, $a=1/\lambda$, measured from the origin, terminates on a perpendicular bisector of a line, joining the origin with a lattice point, a discontinuity occurs in the $\nu$ vs. $|a|$ curve. The object of introducing zones is to eliminate discontinuities in the $\nu$ vs. $|a|$ curve, except at the boundaries. This is exactly analogous to the one-dimensional case.

To construct the second zone, we draw the second smallest closed figure about the origin, bounded by perpendicular bisectors. The second zone is the area enclosed between the boundary of the first zone and the boundary of this second figure. Similarly, one may construct a third zone, that is the area enclosed between the boundary of the second zone, and the third smallest closed figure, bounded by perpendicular bisectors. Higher order zones are constructed in just the same manner. It should be noted that the $(n+1)^{st}$ zone consists of figures having at least one side in common with one side of the $n^{th}$ zone and all vertices in common with the $(n-1)^{st}$ zone.

Each zone has the same area as the elementary cell in the reciprocal lattice. This is shown by taking sections of the zone under consideration, and noting their position relative to some lattice point. Now the wave vectors, terminating in the corresponding section of any other cell, will give the same value for $\Psi$, since the only change in $\Psi$ is the addition of $2m\pi$ in the exponent of the exponential. Therefore, we may consider the two sections equivalent. Furthermore, the area of each of the zones is equal to the area of the elementary cell in the reciprocal lattice, and this is true for all cases that have been worked out.
An analysis of the propagation of waves in a continuous two-dimensional medium with periodic perturbation leads to a first-order correction in the frequency

$$\omega^2 = \omega_0^2 + \varepsilon k_1$$  \hspace{1cm} (A1.1)

and a wave whose general shape is still represented by

$$u_1 = \sum B_{m1_m2} \exp\left(-2\pi i(a'_{m1_m2} \cdot r)\right)$$  \hspace{1cm} (A1.2)

In general, the perturbation is small when $\tilde{\alpha}$ is not too close to any $a'_{n1n2}$. When it is, however, the perturbation becomes larger and when $\tilde{\alpha} = a'_{n1n2}$, the frequency has two possible values, one for each of the two possible values of $k_1$. Thus, for the unperturbed wave, the frequency is a linear function of $|\tilde{\alpha}|$, as shown by the solid line in Fig. (A1.1). The dotted line shows $\omega$ as a function of $\alpha$ for the perturbed wave. For a certain value of $\tilde{\alpha}$, the curve splits up, and the perturbation becomes less as $\tilde{\alpha}$ gets farther away from $a'_{n1n2}$.

![FIG. A1.1. Perturbed and Unperturbed Wave Frequency.](image1)

![FIG. A1.2. Bragg’s Reflection.](image2)

The forgoing discussion shows that discontinuities in the function $\nu(a)$ appear only when the $\tilde{\alpha}$
vector has its extremity on the perpendicular bisector of the vectors in the reciprocal lattice. This justifies the rule given earlier for the construction of successive zones.

The analysis just developed contains the principle of X-ray reflection from crystal lattices. X-rays propagate through the crystal with the velocity of light in vacuum, and atoms or molecules may just slightly perturb the propagation. The perturbation is practically proportional to the electronic density and the following equation of propagation is obtained

$$\nabla^2 u + \frac{\omega_0^2}{V_0^2} u = 0 = \nabla^2 u_0 + \frac{\omega_0^2}{V_0^2} u_0 + \varepsilon \left[ \nabla^2 u_1 + \frac{\omega_0^2}{V_0^2} u_1 + \left( \frac{k_1}{V_0} + \omega_0^2 f \right) u_0 \right]$$  \hspace{1cm} (A1.3)

with a perturbation term, f, proportional to the electronic density. In a crystal lattice, atoms are regularly distributed along a direct lattice, each atomic nucleus being surrounded by a cloud of electrons that may partly overlap that of its neighbors. The electronic density is a periodic function, with the periodicity of the lattice, and can be expanded in a multiple Fourier series, like e.g.

$$f = \Sigma C_{m_1m_2} \exp\left(2\pi i [m_1(b_1 \cdot r) + m_2(b_2 \cdot r)]\right)$$  \hspace{1cm} (A1.4)

In the final results, given by either of the following equations

$$\omega^2 = \omega_0^2 + \varepsilon k_1 = \omega_0^2 \pm \varepsilon \omega_0^2 V_0^2 \left| C_{n_1n_2} \right|$$  \hspace{1cm} (A1.5)

or

$$\frac{k_1}{V_0} = \pm \sqrt{\omega_0^4 \left| C_{n_1n_2} \right|^2 + \left[ 8\pi^2 h(\tilde{a}_0 \cdot \tilde{B}) \right]^2}$$  \hspace{1cm} (A1.6)

only the absolute value $\left| C_{m_1m_2} \right|$ of the coefficients of the Fourier terms appear. Hence, any experiment on wave propagation, through the crystal lattice, will give only the absolute value, and not the phase angle. This latter may often be obtained from symmetry considerations and

152
some general knowledge of the lattice structure, and then the whole Fourier expansion of the electronic space charge is found, from which the distribution of the space charge, at any point can be computed.

The relation between the conditions yielding discontinuities in the \( \nu(a) \) curve and Bragg’s reflection condition can be explained in the same way as for the relation between the passing bands and stopping bands in the one-dimensional lattices. Let a certain direction of propagation be given (direction of \( \vec{a} \)), and the frequency \( \nu \) of the wave be varied. If the frequency corresponds to a passing band, then the wave falling upon the lattice, with this direction of propagation and frequency, may be propagated through the crystal. Surface conditions at the boundary of the crystal will give only partial reflection from the surface. If, however, the frequency, \( \nu \), falls inside one of the stopping bands, like the intervals obtained in Fig. (A1.1), then the corresponding wave cannot be propagated through the crystal, and must be totally reflected from the surface.

The elementary Bragg theory, as sketched in Fig. (A1.2), predicts reflection for just one frequency, while a more comprehensive treatment yields reflection for the whole stopping \( \Delta \nu \). For X-rays, the perturbation of the wave by the crystal lattice is extremely small and the \( \Delta \nu \) bands for ideal crystals are very narrow. The elementary theory of Bragg reflection is as follows:

1. \( \Phi = \Phi' \) and thus \( \theta = \theta' \). This ensures a uniform reflection from each lattice row.
2. Waves reflected from two successive lattice rows must be in phase; hence

\[
\text{AOB} = 2\delta \cos \Phi = 2\delta \sin \theta = m\lambda
\]  

(A1.7)

This is Bragg’s formula, which is conducive to

\[
\lambda = 2\delta \cos \Phi = 2\delta \sin \theta
\]  

(A1.8)
when m=1. The latter equation has been derived by L. Brillouin from the following intermediate result

\[ |\vec{B}|^2 = 2(\vec{a}_o \cdot \vec{B}) = -2(\vec{a}_o' \cdot \vec{B}) \]  
(A1.9)

or

\[ |\vec{a}_o|\cos(\vec{a}_o, \vec{B}) = -|\vec{a}_o'|\cos(\vec{a}_o', \vec{B}) = \frac{1}{2}|\vec{B}| \]  
(A1.10)

and since \( \vec{a}_o \) and \( \vec{a}_o' \) differ only in direction, but have the same length

\[ \cos(\vec{a}_o, \vec{B}) = -\cos(\vec{a}_o', \vec{B}) \]  
(A1.11)

or \( \vec{a}_o \) makes the same angle with \( \vec{B} \) that \( \vec{a}_o' \) makes with \(-\vec{B} \). From the second equation above, it now follows at once that \( \vec{a}_o \) terminates on the perpendicular bisector of \( \vec{B} \), while \( \vec{a}_o' \) terminates on the perpendicular bisector of \(-\vec{B} \). This means that the projection of \( \vec{a} \) on \( \vec{B} \) has magnitude \( |\vec{B}|/2 \), or

\[ |\vec{a}| \cos \Phi = \frac{1}{\lambda} \cos \Phi = \frac{|\vec{B}|}{2} = \frac{1}{2\delta} \]  
(A1.12)

or

\[ \lambda = 2\delta \cos \Phi = 2\delta \sin \theta \]

where \( \Phi \) is the angle between \( \vec{a} \) and \( \vec{B} \), and \( \delta \) is the distance between successive rows passing through the lattice points in the direct lattice, and perpendicular to the basis vector conjugate to \( \vec{B} \), i.e.

\[ \delta = \frac{|c|}{|h|} = \frac{1}{|h|} \]  
(A1.13)

since \(|c|=1\). Thus, we have the following statement: A point \((h_1, h_2)\) in the reciprocal lattice
defines a set of lattice rows in the direct lattice. These straight rows are perpendicular to the vector \( \mathbf{h} \) and are spaced at a distance of \( 1/|\mathbf{h}| \) from one another. This is known as the Bravais notation for crystal lattices. According to the physical properties to be analyzed or investigated, sometimes the direct, and at other times the reciprocal, lattice will yield the better description of the periodic structure. Now, referring again to Fig. (A1.2), we note that it shows a set of parallel rows in the direct lattice, with a distance of separation \( \delta \), with the incident beam \( a_o \), and with the reflected beam \( a'_o \).

All of the previous discussion can be used without any change for three-dimensional problems. The areas are replaced by volumes, and atomic planes in the direct lattice replace the atomic rows occurring in two dimensions. Triply periodic functions are expanded in triple Fourier series.

A point in the direct lattice is given by

\[
\mathbf{R} = l_1 \mathbf{d}_1 + l_2 \mathbf{d}_2 + l_3 \mathbf{d}_3
\]  

where \( l_1, l_2 \) and \( l_3 \) are integers and \( \mathbf{d}_1, \mathbf{d}_2 \) and \( \mathbf{d}_3 \) are the lattice vectors. An arbitrary vector in the direct lattice is given by

\[
\mathbf{r} = \xi_1 \mathbf{d}_1 + \xi_2 \mathbf{d}_2 + \xi_3 \mathbf{d}_3
\]  

where \( \xi_1, \xi_2 \) and \( \xi_3 \) are the components of \( \mathbf{r} \) in the directions \( \mathbf{d}_1, \mathbf{d}_2 \) and \( \mathbf{d}_3 \), respectively. If a wave is propagating through the lattice, and is observed only at the lattice points \( \mathbf{R} \), we may express the disturbance \( \Psi \) by

\[
\Psi = A \exp \left( 2\pi i \left[ \mathbf{t} - (\mathbf{a} \cdot \mathbf{R}) \right] \right)
\]

where \( \mathbf{a} \) is the wave vector (actually wave number), and is regarded as a vector of the reciprocal
lattice. The function $\Psi$ is, evidently, periodic in $\mathbf{a}$ and $\mathbf{R}$. The vector

$$\mathbf{a}_h = \mathbf{a} + h_1 \mathbf{b}_1 + h_2 \mathbf{b}_2 + h_3 \mathbf{b}_3$$  \hspace{1cm} (A1.17)

where $h_1$, $h_2$ and $h_3$ are integers, will describe the motion at the lattice points just as well as $\mathbf{a}$; for

$$\left(\mathbf{a}_h \cdot \mathbf{R}\right) = \left(\mathbf{a} \cdot \mathbf{R}\right) + h_1 \mathbf{b}_1 \cdot \mathbf{R} + h_2 \mathbf{b}_2 \cdot \mathbf{R} + h_3 \mathbf{b}_3 \cdot \mathbf{R}$$

$$= \left(\mathbf{a} \cdot \mathbf{R}\right) + h_1 l_1 + h_2 l_2 + h_3 l_3 = \left(\mathbf{a} \cdot \mathbf{R}\right) + \text{integer}$$  \hspace{1cm} (A1.18)

Since the same motion can be represented by $\mathbf{a}$ or any arbitrary $\mathbf{a}_h$, the frequency must be a periodic function of $\mathbf{a}$ also, with the periodicity $\mathbf{b}_1$, $\mathbf{b}_2$ and $\mathbf{b}_3$ of the reciprocal lattice.

We wish to set up conventions for eliminating all but one of the values for $\mathbf{a}$, i.e., define zones to which $\mathbf{a}$ is to be restricted, as was done for the two-dimensional case. The method is exactly the same. We take the first zone to be a volume centered upon the origin $0$ of the reciprocal lattice, and limited by plane perpendicular bisectors of vectors in the reciprocal lattice. The first zone has a volume $V_b$ equal to the volume of the elementary cell in the reciprocal lattice, and any point in space can be brought back into the first zone by $\mathbf{H}$ translations in the reciprocal lattice.

$$\mathbf{H} = h_1 \mathbf{b}_1 + h_2 \mathbf{b}_2 + h_3 \mathbf{b}_3$$  \hspace{1cm} (A1.19)

Higher zones surrounding the first one are built in a similar way.

The choice of these particular planes, bisecting and perpendicular to the vectors of the reciprocal lattice, is based on the earlier analysis for two dimensions. The theory of waves, propagating through a continuous medium, with a small periodic perturbation, is identical with the two-dimensional theory, and one must distinguish between
Case 1, where all \(|\vec{a}_h' - \vec{a}| \neq 0\)

Case 2, where one certain \(\vec{a}_h', \) say \(\vec{a}'_n, \) is \(|\vec{a}_n'| = |\vec{a}|\) where \(\vec{a}_h'\) is a vector given by

\[
\vec{a}_h' = \vec{a} + h_1\vec{b}_1 + h_2\vec{b}_2 + h_3\vec{b}_3
\]

(A1.20)

where \(h_1, h_2\) and \(h_3\) are integers. Case 2 arises when two vectors \(\vec{a}_n'\) and \(\vec{a}\), having almost the same length, can be found, such that their difference is a vector of the reciprocal lattice

\[
\vec{a}_n' - \vec{a} = \vec{H} = h_1\vec{b}_1 + h_2\vec{b}_2 + h_3\vec{b}_3
\]

(A1.21)

A geometrical interpretation of this condition was presented earlier in Fig. (A1.2), where it was proven that the circumstances leading to case 2 were identical with those yielding Bragg’s reflections in the crystal. This is always true if atomic rows are replaced by atomic planes in the direct lattice.

Another geometrical interpretation was given by P.P.Ewald. From point 0 of the reciprocal lattice, he draws a vector \(\vec{a}\) to point P. A sphere is drawn with radius \(|\vec{a}|\), about the point P, at the end of the vector \(\vec{a}\). If it happens that this sphere passes through, or near to, a second point B of the reciprocal lattice, we obtain a discontinuity in the frequency as a function of \(\vec{a}\). The reason for this is that there is some \(\vec{a}_n'\) that just fulfills the above equation. There is no way of telling, from the motion of the lattice points, which of vectors \(\vec{a}\) and \(\vec{a}_n'\) should be preferred. Since the two vectors have different directions, we must assume that the motion is given by a superposition of the two waves. One of these is an incident wave, and the other is a reflected wave (Bragg wave). There will be two values for the frequency, for this single value of \(|\vec{a}|\), and hence a discontinuity in the \(\nu\) vs. \(\vec{a}\) curve; see Fig. (A1.3).

We note that when Bragg reflection occurs, the vector terminates on a plane, that is a perpendicular bisector of the line \(\vec{H}\), joining two lattice points of the reciprocal lattice. This
justifies the rule for constructing zones. We construct planes that are perpendicular bisectors of lines joining lattice points with the origin. The smallest polyhedron bounded by such planes is the first zone. The area bounded by the first zone and the second smallest polyhedron is the second zone, and similarly for higher order zones.

The results obtained from investigating the problem of wave propagation in a two-dimensional continuous medium, with a small periodic perturbation, can be immediately extended to three dimensions without any difficulty. In order to discuss the general problem of wave propagation in a periodic medium, without restricting the discussion to the case of a small periodic perturbation, we now consider the three-dimensional wave equation

\[ \nabla^2 \Psi + \frac{\omega^2}{V^2} \Psi = 0 \]  \hspace{1cm} (A1.22)

We confine ourselves to waves in a periodic medium and, therefore, we may assume

\[ \frac{1}{V^2} = F(\bar{r}) = F(\bar{r} + n_1\bar{d}_1 + n_2\bar{d}_2 + n_3\bar{d}_3) \]  \hspace{1cm} (A1.23)

where \( n_1, n_2 \) and \( n_3 \) are integers, and \( \bar{d}_1, \bar{d}_2 \) and \( \bar{d}_3 \) are lattice vectors, which represent the three directions of the periodic function \( F(\bar{r}) \). The basis vectors of the reciprocal lattice are \( \bar{b}_1, \bar{b}_2 \) and \( \bar{b}_3 \), where

\[ (\bar{b}_i \cdot \bar{d}_j) = \delta_{ij} \]  \hspace{1cm} (A1.24)

\( F(\bar{r}) \) may be expanded in a triple Fourier sum

\[ F = \Sigma C_{m_1m_2m_3} \exp[2\pi i (m_1(\bar{b}_1 \cdot \bar{r}) + m_2(\bar{b}_2 \cdot \bar{r}) + m_3(\bar{b}_3 \cdot \bar{r}))] \]  \hspace{1cm} (A1.25)

The solution of the three-dimensional wave equation may also be expressed as a Fourier sum

\[ \Psi = A(\bar{r})\exp[2\pi i \{t - (\bar{a} \cdot \bar{r})\}] \]  \hspace{1cm} (A1.26)
where $A$ is to be periodic in $\mathbf{r}$, and may be written

$$A(\mathbf{r}) = \Sigma A_{m_1m_2m_3} \exp\left[2\pi i \left( m_1(\mathbf{b}_1 \cdot \mathbf{r}) + m_2(\mathbf{b}_2 \cdot \mathbf{r}) + m_3(\mathbf{b}_3 \cdot \mathbf{r}) \right) \right]$$  \hspace{0.5cm} (A1.27)

The form taken for the $\Psi$ solution above is generally known as F. Bloch’s theorem.

A lattice with basis for three dimensions is based on three vectors $\mathbf{d}_1$, $\mathbf{d}_2$ and $\mathbf{d}_3$, as before, and inside each $\mathbf{d}$ cell, there is a certain number $N$ of particles, whose positions with respect to the origin of the cell are

$$\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N$$

and whose masses are taken as $M_1, M_2, ..., M_N$, respectively. If $N = 1$, we obtain the monatomic lattice, with one particle per cell, and it is convenient to take $r_1 = 0$, and have this particle at the origin of the cell. In the general case, an arbitrary particle $k$ anywhere in the lattice is located at

$$\mathbf{r} = \mathbf{r}_j + n_1\mathbf{d}_1 + n_2\mathbf{d}_2 + n_3\mathbf{d}_3$$ \hspace{0.5cm} (A1.28)

where particle $k$ is the $j^{th}$ particle in cell $(n_1, n_2, n_3)$. A general scheme for propagation of waves through such a lattice could be developed as follows: For each $j$ type of particle in the lattice, we could write a wave equation

$$\Psi_j = A_j \exp\left\{ 2\pi i \left[ \nu t - (\mathbf{a} \cdot \mathbf{r}) \right] \right\}$$ \hspace{0.5cm} (A1.29)

but it is more convenient to include the exponential

$$\exp\left[ -2\pi i \left( \mathbf{a} \cdot \mathbf{r}_j \right) \right]$$

in the complex amplitude $A_j$ and to keep the same imaginary exponential in all the $j$ waves

$$\exp\left[ i \left( \omega t - n_1k_1 - n_2k_2 - n_3k_3 \right) \right]$$

where

$$k_1 = 2\pi(\mathbf{a} \cdot \mathbf{d}_1), \quad k_2 = 2\pi(\mathbf{a} \cdot \mathbf{d}_2), \quad k_3 = 2\pi(\mathbf{a} \cdot \mathbf{d}_3)$$ \hspace{0.5cm} (A1.30)

The wave motion of particles $j$ is then described by
\[ \Psi_j = A_j \exp[i(\omega t - n_1k_1 - n_2k_2 - n_3k_3)], \quad A_j = A'_j \exp[-2\pi i (\tilde{a} \cdot \tilde{r}_j)] \quad (A1.31) \]

There will be \( N \) constants: \( A_1, A_2, \ldots, A_N \), in general complex, to include the phase of particle \( j \) with particle 1.

As in the one-dimensional case, we will have \( N \) equations of motion, and substitution of the foregoing equation in the equations of motion will yield \( N \) linear equations in the \( A_i \). Each \( A_i \) is a vector function and has, therefore, three components, thus yielding altogether \( 3N \) linear homogeneous equations to be solved simultaneously. As before, we equate the determinant of the coefficients of the \( A_j \) to zero, which yields an equation of degree \( 3N \) in \( \omega^2 \). This means that for each value of \( a \) in the first zone of the reciprocal lattice, there will be \( 3N \) values for the frequency. This is shown schematically in Fig. (A1.4). The number of branches for the function \( \omega(\tilde{a}) \) is thus equal to three times the number \( N \) of particles per cell. A continuous periodic medium may be regarded as the limit \( N \to \infty \). Uncertainty in the wave vector \( \tilde{a} \) is best avoided by restricting \( \tilde{a} \) to the interior of the first zone.

For any kind of waves, elastic, electromagnetic, or wave mechanical de Broglie, waves, propagating in a medium with periodic structure, discontinuities are found in the relation
between the frequency \( \nu \) and the wave vector \( \vec{a} \) (a vector of length \( |\vec{a}| = 1/\lambda \) and pointing in the direction of propagation). These discontinuities are obtained whenever the vector \( \vec{a} \) terminates on a plane that is a perpendicular bisector of one of the vectors of the reciprocal lattice. These planes play the same role for all waves, whatever their particular nature may be. For instance, in electromagnetic waves, the discontinuities in \( \nu(\vec{a}) \) are directly responsible for the selective reflection of X-rays (Bragg-Laue spots). This general property is the reason for choosing these special planes as limits of the zones defined earlier, and the zone structure, in a given crystal lattice, is the same for all waves. These zones are generally known in the literature as Brillouin zones.

Elastic vibrations, for instance, are propagated through a crystal lattice as elastic waves. Their properties result from the earlier analyses and their generalizations in three dimensions. Assuming \( N \) atoms per lattice cell (lattice with basis), the best representation of the waves is obtained by restricting the \( \vec{a} \) vector to the first zone. Each \( \vec{a} \) vector yields 3\( N \) different wave motions, of which three are of the acoustic type and 3(\( N-1 \)) of the optical type, as Fig. (A1.4) shows schematically for the case \( N=3 \). The three acoustic waves correspond to the well-known waves in an isotropic solid: one longitudinal and two transverse vibrations. In an ideal continuous anisotropic medium, three vibrations at right angles to each other are obtained, none of which is exactly longitudinal, or transverse. In the crystal lattice with discontinuous structure, there are still three different acoustic waves for each \( \vec{a} \) vector, but their properties are much more complicated than for a continuous medium. We may, for the sake of visualization, call the waves longitudinal, \( L \), and transverse, \( t_1, t_2 \), but these names do not correspond exactly to the properties of waves in a lattice structure.

The hexagonal lattice, although it contains only one type of atom, yet it is a lattice with
basis, and yields three acoustic branches, and three branches of higher frequency, which may be called optical. A real monatomic crystal lattice, such as the face- or body-centered cubic structure, is one without basis, and contains only one atom per cell. It yields only three acoustic branches and no optical branches.
Appendix 2

Transport of Heat and Electricity in Solids

A2.1 Crystalline Solids

Condensed matter exists in many different forms: organic, inorganic, liquids, crystalline solids, amorphous solids, ionic conductors, covalent conductors, etc. However, the materials presently employed in thermoelectric devices are inorganic semiconductor crystalline solids. We suppose, therefore, that such solid bodies consist of one or more crystals; the atoms in each of which are regularly arranged in the form of a lattice. The crystal lattice is a framework along which the carriers of electric charge and thermal energy move. The vibrations of the lattice provide one of the principal mechanisms of heat transfer. There is a wide variety of crystal structures that may be assumed by an element or compound. The particular structure that is observed is governed, to a certain extent, by the nature of the binding between the atoms. There are four types of atomic bonds, namely, the ionic, covalent, metallic and van der Waals bonds.

The simplest bonds arise from coulombic attraction between ionized atoms that bear charges of opposite sign. Sodium chloride constitutes the most typical example of an ionic bond. In this compound, each of the sodium atoms loses one electron, to become positively charged. The chlorine atoms become negatively charged, as each takes up an electron. The arrangement in the sodium chloride is stable when each positive or negative ion is surrounded symmetrically by six ions of the opposite sign. If all the atoms were identical, we would have a simple cubic lattice. Taking account of the two different types of atom, we see that the lattice is of the face centered cubic type. We note that a repulsive force that becomes very strong as the spacing becomes smaller opposes the coulombic attraction between oppositely charged ions. At equilibrium
atomic spacing, the attractive and repulsive forces balance. If one examines the periodic table, one notices that Group IV contains the elements silicon and germanium, which, like one form of carbon and gray tin, have a diamond structure. In this structure, each atom shares the four electrons in its outer shell with its four nearest neighbors. Thus the binding between each pair of atoms involves two electrons, and is termed covalent. Unlike an ionic bond, a covalent bond is strongly directional, so that, in the diamond structure, the four nearest neighbors of each atom lie at the corners of a tetrahedron. In covalently bonded crystals, the electrons are concentrated near the lines that join neighboring atoms, whereas in ionic crystals, the electrons form almost spherical clouds around the negative ions. The atoms to the left of Group IV are termed electropositive, because they tend to donate, or give up, electrons, thereby becoming positive ions. Likewise, the atoms to the right of Group IV are termed electronegative, because they tend to accept, or take, electrons. The alkali halides exemplify the principle that compounds between strongly electropositive and strongly electronegative elements contain ionic bonds, whereas when the electronegativity difference is small, covalent bonding predominates. The nature of the bonding has a considerable bearing on the transport properties of solids.

In a metallic bond, as in a covalent bond, atoms are bound together by shared electrons. In this case, however, the participating electrons are less influenced by s-p hybridization, and the electron density is more uniform. The number of nearest neighbors is no longer determined by considerations of valence, and many metals form structures in which the atoms are as near to one another as possible. These are face centered cubic and close packed hexagonal structures.

The fourth type of bond is much weaker than the aforementioned three ones. This bond is particularly important in condensed inert gases. It is also of some significance in certain thermoelectric materials, notably those that have layered structures.
A2.2 Heat Conduction by the Lattice

The fact that metals are good conductors of heat can be attributed to the transport of heat by the charge carriers. Of course, heat can also be conducted through electrical insulators. The thermal conductivity of diamond, for example, exceeds that of any metal. It is, in fact, the transfer of vibrational energy, from one atom to the next, that constitutes this process.

The atoms in a crystal do not vibrate independently. Rather, the vibrations are such that elastic waves continually pass backward and forward through the crystal. An increase in temperature should lead to an increase in the amplitude of a given wave, as well as redistribution of the relative amplitudes for different wavelengths. This concept was used by Debye to determine the specific heat of a solid. Although Debye’s theory is now known to have shortcomings, it is still regarded as a first step toward the complete description of the thermal properties of solids.

Debye represented the crystal as an elastic continuum. When boundary conditions are imposed, only certain modes of vibration are possible. For example, if the surfaces of the crystal are to correspond to nodes, this condition can be realized only for waves in a particular direction, if they have particular wavelengths. In a true continuum, the total number of modes of vibration would be infinite. Debye showed that the discrete nature of atoms sets a lower limit to the wavelength. Therefore, one should take account of only the 3N modes of lowest frequency, where N is the number of atoms per unit volume. This restriction on the number of vibrational modes gives a result at high temperatures that is consistent with the Dulong and Petit law, which states that the specific heat per atom is equal to 3k_B, where k_B is the Boltzmann constant. The number of modes, dw, per unit volume, that have frequencies between \( \nu \) and \( \nu + d\nu \), is given by
where \( v \) is an appropriate average value for the speed of sound. In a simple continuum, the speed of sound is independent of frequency, but, even in an isotropic solid, it will be different for longitudinal and transverse waves.

The plot of frequency, \( \nu \), against wave number, \( a = 1/\lambda \), or the dispersion curve, is linear through the origin for a continuum, but becomes nonlinear at high frequencies in real materials. If more than one atom is associated with each lattice point, the dispersion curve also has separate branches that have one or more optical branches to each acoustic branch. In three dimensions, there will be three acoustic branches (one for the longitudinal vibrations and one for each polarization of the transverse vibrations) and 3(n-1) optical branches, where \( n \) is the number of atoms per primitive cell. The group velocity, determined by the slope of the dispersion curve, is of particular interest, because it is related to the rate of energy transport. It clearly depends on frequency or wave number, and also differs considerably for acoustic and optical modes.

The limit of 3N on the number of modes means that there is an upper frequency limit, or cut off frequency, \( \nu_D \), which may be obtained by integrating Eq. (A2.2.1). \( \nu_D \) is given by

\[
\int_0^N dw = \int_{-\nu_D}^{\nu_D} \frac{2\pi\nu^2}{\nu^3} d\nu
\]

and therefore

\[
[w]_0^N = \frac{2\pi}{3\nu^3} \nu^3 \Bigg|_{-\nu_D}^{\nu_D} = \frac{2\pi}{3\nu^3} \left(\nu_D^3 - (-\nu_D)^3\right)
\]

thus
The above equation forms the basis for the definition of the so-called Debye temperature, which is

\[ \theta_D = \frac{\hbar \nu_D}{k_B} \]  

(A2.2.3)

Debye’s success in explaining the general behavior of specific heat as a function of temperature was due to his use of quantum theory. He showed that the average energy, \( W \), in a mode of frequency, \( \nu \), is

\[ W = \hbar \nu \left( \exp \left( \frac{\hbar \nu}{k_B T} \right) - 1 \right)^{-1} \]  

(A2.2.4)

where \( \hbar \) is Planck’s constant. With the frequency limit set by Eq. (A2.2.2), Eqs. (A2.2.1) and (A2.2.4) allow one to determine the internal energy of the crystal. Then, by differentiating the internal energy with respect to temperature, the specific heat at constant volume is

\[ c_v = 9Nk_B \left( \frac{T}{\theta_D} \right)^3 \int_0^{\theta_D/T} \frac{x^4 \exp(x)}{(\exp(x) - 1)^5} dx \]  

(A2.2.5)

Debye’s theory has been so successful that it is usual to fit Eq. (A2.2.5) to the experimental data, by regarding \( \theta_D \) as a temperature-dependent parameter.

The relative accuracy of Eq. (A2.2.5) does not indicate that Debye’s theory gives a particularly good description of the vibrations in a real crystal. Therefore, specific heat studies are not a good way to determine the distribution of the vibrational modes. The vibrational spectra can, however, be found by the neutron scattering method of Brockhouse\(^{A1}\) and others.

The first requirement of any theory of lattice heat transport is that it should explain the observation of Eucken\(^{A6}\) that, for pure dielectric crystals, thermal conductivity varies inversely
with absolute temperature. Eucken’s rule has been confirmed for many materials, provided that the temperature is not much less than $\theta_D$.

Debye attempted to use his continuum model to explain Eucken’s $1/T$ law, but found out that the thermal conductivity is infinite. It is only by taking account of the anharmonic nature of the forces in real crystals that thermal conductivity takes on finite values. If an atom is displaced from its equilibrium position by an infinitesimal amount, the restoring force is proportional to the displacement, i.e., it is harmonic. As the displacement becomes larger, the expression for the restoring force includes higher order terms. Then the force is partially anharmonic. In effect, the elastic constants of the substance vary from point to point according to the local displacements of the atoms. Variations in either the density or the elastic properties are responsible for scattering the vibrational waves and, thus, for limiting the value of thermal conductivity.

Peierls showed how the anharmonicity of the lattice vibrations could be taken into account. He introduced the concept of phonon wave packets that arise from the quantization of vibrational waves. One refers to the wave packets simply as phonons, and they may be regarded as the energy carriers that are responsible for heat conduction by the lattice. A limit on thermal conductivity is set by phonon collisions that do not conserve momentum. One may draw on the kinetic theory of gases to express the lattice thermal conductivity, $\kappa_L$, in terms of the mean free path of the phonons. Thus,

$$\kappa_L = \frac{1}{3} c_v v l_t$$  \hspace{1cm} (A2.2.6)

where the specific heat, $C_v$, is defined for unit volume, $v$ is the speed of sound and $l_t$ is the mean free path length.

In pure crystals at high temperatures, phonons are scattered predominantly by other
phonons. The most important processes involve three phonons and can be of two kinds. In the normal, or N-processes, both energy and momentum are conserved. In Umklapp or U-processes, momentum, or more strictly, wave number, is not conserved. The N-processes lead to a redistribution of phonons, but no thermal resistance to first order. Therefore, the U-processes directly control high-temperature thermal conductivity.

The difference between the N- and U-processes is indicated in Fig. (A2.2.1). Two-dimensional representations of wave-vector space are shown: the squares correspond to all values of the phonon wave vector set by the frequency limit, \( \nu_D \). The maximum x or y component of the wave vector is equal to half the side of the square. The region of wave-vector space that contains the allowed values for the wave vector is called the Brillouin zone.

In the N-process, two phonons of wave vector, \( \vec{q}_1 \) and \( \vec{q}_2 \), interact to form a third phonon of wave vector, \( \vec{q}_3 \), such that

\[
\vec{q}_1 + \vec{q}_2 = \vec{q}_3
\]  

(A2.2.7)

On the other hand, in the U-process, the sum of the wave vectors, \( \vec{q}_1 \) and \( \vec{q}_2 \), would produce a resultant outside the Brillouin zone, if Eq. (A2.2.7) were applied. Thus one subtracts the vector, \( \vec{G} \), the reciprocal lattice vector, to produce a resultant wave vector, \( \vec{q}_3 \), that lies within the allowed zone. The rule for the conservation of a wave vector for the U-process is

\[
\vec{q}_1 + \vec{q}_2 = \vec{q}_3 + \vec{G}
\]  

(A2.2.8)

Peierls showed that the increasing probability of U-processes with a rise in temperature explains Eucken’s 1/T law. However, at low temperatures, it becomes exceedingly difficult for any U-processes to occur. When \( T << \theta_D \), there are virtually no phonons where wave vectors are comparable to half-width of the Brillouin zone, and very few with even half that value. The only
real possibility for finding a U-process is when two phonons have wave vectors marginally greater than a quarter of the Brillouin zone width, provided that they happen to have, more or less, the same direction. The number of such phonons is proportional to \( \exp(-\theta_D/2T) \), so one might expect that \( l_t \) at low temperatures is proportional to \( [\exp(-\theta_D/aT)]^{-1} \), where \( a \approx 2 \). The exponential variation of \( l_t \) should have a major influence on the dependence of thermal conductivity on temperature.

Although there is no doubt that the inverse temperature dependence of thermal conductivity, when \( T > \theta_D \), will always change to an exponential dependence for \( T \ll \theta_D \) in a sufficiently pure and perfect large crystal. In practice, it is often difficult to observe such behavior. It is usually found that scattering by all kinds of internal defects, and by the external crystal boundaries, masks phonon-phonon scattering at low temperatures. In fact, we can sometimes take advantage of these other phonon-scattering processes to improve thermoelectric materials. Therefore, it is important for one to consider lattice conductivity in less than perfect crystals.

\[ \begin{align*}
\text{FIG. A2.2.1. Two-dimensional representation of: (a) N-process and (b) U-process.}
\end{align*} \]
A2.3 Scattering of Phonons by Defects

As the temperature of a dielectric crystal is lowered, its thermal conductivity rises to a maximum value, and then falls so as to approach zero at 0 K. At the lowest temperatures, thermal conductivity is proportional to $T^3$, which is the same variation with temperature as that of specific heat. This implies that $l_l$ must have reached a limiting value. Casimir showed that this value is comparable to the dimensions of the crystal, and can be attributed to the scattering by phonons at the crystal boundaries. In a polycrystalline sample, one expects that the maximum free path length of the phonons is set by the grain size. It is even found that amorphous solids behave as if heat conduction arises from phonons whose free path lengths are comparable to the inter-atomic spacing. One might expect boundary scattering to have an effect on the lattice conductivity of crystalline materials only when the temperature is low enough for $l_l$ of phonons to be very large.

Now, one discusses scattering by the various point defects that are found in real crystals. Examples of such defects are the local variations of elasticity and density, that occur in solid solutions between isomorphous crystals, e.g., silicon-germanium alloys, or in an element or compound that contains foreign impurities or vacancies. It is even possible to observe scattering of phonons due to the density variations associated with the different isotopes in naturally occurring elements.

The treatment of the scattering of phonons by point defects presents us with a difficulty. The relaxation time, $\tau_1$, for this form of scattering should be proportional to $1/\nu^4$. This means that point defect scattering does not have a significant effect on low-frequency phonons. The thermal conductivity at low temperatures then becomes exceedingly large because U-processes also have little effect in this region. That this phenomenon does not occur can be explained if
one takes account of the redistribution of phonons by N-processes into modes for which the scattering is stronger.

Callaway$^{A2}$ developed a satisfactory method of including N-processes. In Callaway’s treatment, one uses the relaxation time, $\tau$, rather than $l$; the two quantities are related through $l = \nu \tau$. The basic principle is that non-momentum-conserving processes cause the perturbed phonons to relax toward equilibrium distribution, whereas N-processes cause them to relax toward a non-equilibrium distribution. If all of the scattering processes were non-momentum-conserving, the overall relaxation time could be obtained simply as the reciprocal of the sum of the reciprocal relaxation times of the individual processes. Callaway successfully obtained a multiplying factor that corrects for the momentum-conserving behavior of N-processes.

Suppose that there are two different relaxation times, $\tau_R$, for the processes that change the momentum, and $\tau_N$ for the N-processes. The rate at which the phonon distribution function, $N$, changes through the various scattering effects is

$$\left( \frac{dN}{dt} \right)_{scatter} = \frac{N_0 - N}{\tau_R} + \frac{N_N - N}{\tau_N} \quad \text{(A2.3.1)}$$

where $N_0$ is the equilibrium distribution function and $N_N$ is the distribution function to which N-processes on their own would lead. At the same time, the temperature gradient, $\nabla T$, changes the distribution function according to

$$\left( \frac{dN}{dt} \right)_{\text{diffusion}} = -\bar{\nabla} T \frac{\partial N}{\partial T} \quad \text{(A2.3.2)}$$

where $\bar{\nabla}$ is the sound velocity in the direction of the phonon wave vector, $\bar{q}_L$. Because the diffusion and scattering processes must balance,

$$\frac{N_0 - N}{\tau_R} + \frac{N_N - N}{\tau_N} - \bar{\nabla} T \frac{\partial N}{\partial T} = 0 \quad \text{(A2.3.3)}$$
Now, the N-processes must lead to a distribution that carries momentum against the temperature gradient. A suitable distribution is

\[ N_N = \left[ \exp \left( \frac{\hbar \omega - \mathbf{q}_L \cdot \mathbf{l}}{k_B T} \right) - 1 \right]^{-1} \]  

(A2.3.4)

instead of the usual Bose-Einstein distribution function

\[ N_o = \left[ \exp \left( \frac{\hbar \omega}{k_B T} \right) - 1 \right]^{-1} \]  

(A2.3.5)

It is convenient to use the angular frequency, \( \omega \), which is equal to \( 2\pi \nu \), and \( \hbar = \hbar / 2\pi \).

In Eq. (A2.3.4), \( \mathbf{l} \) is a constant vector in the direction of \( \nabla \mathbf{T} \), where \( \mathbf{q}_L \cdot \mathbf{l} \ll \hbar \omega \). In effect, Eq. (A2.3.4) differs from Eq. (A2.3.5) in that the frequency, \( \omega \), is changed by \( \mathbf{q}_L \cdot \mathbf{l} / \hbar \).

Consequently, one finds that

\[ N_N - N_o = \frac{\mathbf{q}_L \cdot \mathbf{l}}{k_B T} \exp(x) \left( \frac{\exp(x)}{(\exp(x) - 1)^2} \right) \]  

(A2.3.6)

where \( x = \hbar \omega / k_B T \). Also

\[ \frac{\partial N}{\partial T} = \frac{\hbar \omega}{k_B T^2} \exp(x) \left( \frac{\exp(x)}{(\exp(x) - 1)^2} \right) \]  

(A2.3.7)

so that

\[ N_N - N_o = \frac{\mathbf{q}_L \cdot \mathbf{l} T}{\hbar \omega} \frac{\partial N}{\partial T} \]  

(A2.3.8)

Substituting Eq. (A2.3.8) into Eq. (A2.3.3) gives

\[ \left( N_o - N \right) \left[ \frac{1}{\tau_R} + \frac{1}{\tau_N} \right] - \left[ \mathbf{v} \cdot \nabla \mathbf{V} - \frac{\mathbf{q}_L \cdot \mathbf{l} T}{\hbar \omega \tau_N} \right] \frac{\partial N}{\partial T} = 0 \]  

(A2.3.9)

Because \( \mathbf{l} \) should be proportional to the temperature gradient, one writes
\[ \vec{I} = -\frac{\hbar}{T} \beta \nu^2 \vec{\nabla} T \]  

(A2.3.10)

where \( \beta \) is a constant whose dimension is time. Also, using the Debye model to express \( \nu \) as \( \omega / q_L \), Eq. (A2.3.9) becomes

\[
(N_o - N) \left[ \frac{1}{\tau_R} + \frac{1}{\tau_N} \right] \left[ 1 + \frac{\beta}{\tau_N} \right]^{-1} - \vec{\nu} \cdot \vec{\nabla} T \frac{\partial N}{\partial T} = 0
\]

(A2.3.11)

This means that the behavior is exactly the same as it would be if there were an effective relaxation time, \( \tau_{\text{eff}} \), given by

\[
\frac{1}{\tau_{\text{eff}}} = \left[ \frac{1}{\tau_R} + \frac{1}{\tau_N} \right] \left[ 1 + \frac{\beta}{\tau_N} \right]^{-1} = \frac{1}{\tau_C} \left[ 1 + \frac{\beta}{\tau_N} \right]^{-1}
\]

(A2.3.12)

Here, \( \tau_C \) is the relaxation time that we would expect if the N-processes did not conserve wave vector, because then we would merely add the reciprocal relaxation times for all of the scattering processes. We thus see that \( \tau_C \) must be multiplied by \( (1 + \beta / \tau_C) \) to yield \( \tau_{\text{eff}} \).

The constant, \( \beta \), is found by using the fact that N-processes conserve the wave vector. In other words,

\[
\int \frac{N_N - N}{\tau_N} q_m^3 \, d^3 q_m = \int \frac{4 \pi \omega^2}{v^3} \frac{N_N - N}{\tau_N} q_L^3 \, d\omega = 0
\]

(A2.3.13)

Because \( N_N - N = (N_N - N_o) + (N_o - N) \) is proportional to (\( \beta - \tau_{\text{eff}} \)) \( \partial N / \partial T \), Eq. (A2.3.13) can be written as

\[
\int_0^{\theta/\nu} \left[ \frac{\beta}{\tau_N} - \frac{\tau_C}{\tau_N} - \frac{\beta \tau_C}{\tau_N^2} \right] \omega^4 \frac{\exp(x)}{(\exp(x) - 1)^2} \, dx = 0
\]

(A2.3.14)

Thus, we find that
\[
\beta = \frac{\int_0^{\theta_D/T} \frac{\tau_C x^4}{\tau_N} \frac{\exp(x)}{(\exp(x) - 1)^2} \, dx}{\int_0^{\theta_D/T} \frac{1}{\tau_N} \left[ 1 - \frac{\tau_C}{\tau_N} \right] x^4 \frac{\exp(x)}{(\exp(x) - 1)^2} \, dx} \tag{A2.3.15}
\]

The general determination of \( \beta \) is difficult, but Callaway showed that a simplification of the expression is possible in certain cases. For example, if the scattering by the imperfections is very strong, and the relaxation time \( \tau_I \ll \tau_N \), it is valid to use the approximation, \( 1/\tau_{\text{eff}} = 1/\tau_I + 1/\tau_N \). Another approximation may be made at low temperatures for fairly pure crystals, when \( \tau_I \gg \tau_N \) and U-processes can be neglected. However, when we are dealing with thermoelectric materials, the high temperature approximation used by Parrott\textsuperscript{A7} is of interest. The relaxation times for both U- and N-processes are expected to be proportional to \( \omega^{-2} \), whereas for point defects, \( \tau_I \) is proportional to \( \omega^{-4} \). Suppose, then that the relaxation times are expressed as \( 1/\tau_I = A\omega^4 \), \( 1/\tau_U = B\omega^2 \), \( 1/\tau_N = C\omega^2 \), where A, B and C are constants for a given sample. Also, when \( T \gg \theta_D, x << 1 \) for the whole phonon spectrum, so that \( x^2 \exp(x) / (\exp(x)-1)^2 \approx 1 \). Then it is possible to obtain analytical solutions for the integrals involved in \( \beta \) and to express, \( \kappa_L \), for the crystal, with point imperfections, in terms of the value, \( \kappa_o \), that would be found for a perfectly pure crystal. The expression is

\[
\frac{\kappa_L}{\kappa_o} = \left( 1 + \frac{5k_o}{9} \right)^{-1} \left[ \frac{\tan^{-1} y}{y} + \left( 1 - \frac{\tan^{-1} y}{y} \right) \left( \frac{y^4(1 + k_o)}{5k_o} - \frac{y^2}{3} - \frac{\tan^{-1} y}{y} \right) \right]^{-1} \tag{A2.3.16}
\]

where \( k_o \), equal to \( C/B \), is a measure of the relative strengths of the N- and U-processes. The quantity, \( y \), is defined by

\[
y^2 = \left( \frac{\omega_D}{\omega_o} \right)^2 \left( 1 + \frac{5k_o}{9} \right)^{-1} \tag{A2.3.17}
\]

where \( \omega_o \) is given by

175
\[
\left( \frac{\omega_0}{\omega_D} \right)^2 = \frac{k_B}{2\pi^2 v\kappa_0 \omega_D a}
\]  \hspace{1cm} (A2.3.18)

The value of \( k_0 \) may be determined experimentally by measuring the thermal conductivity of one pure crystal, and one that contains imperfections. Thereafter, the same value of \( k_0 \) can be used for crystals with other concentrations of defects.

One notes that the approximations that hold strictly for \( T \gg \theta_D \) are often used with reasonable success for the less stringent condition, \( T \geq \theta_D \).
Appendix 3

Band Theory of Solids

A3.1 Energy Bands in Metals, Insulators and Semiconductors

In the classical free electron theory of metals, it was supposed that the atomic nuclei and all, except the outer, electrons are in fixed positions, whereas the outer electrons form as “gas” that can be described by the familiar kinetic theory. In particular, a mean thermal energy of $3k_BT/2$ was ascribed to each of the free electrons; the energy being distributed according to the Maxwell-Boltzmann law.

Although the classical free-electron theory accounted for Ohm’s law, predicted the right order of magnitude for the electrical resistivity and also explained the observed relationship between the electrical and thermal conductivity in metals (the Wiedemann-Franz law), yet it failed badly in certain other respects. A most obvious shortcoming was its prediction of a substantial component of the specific heat from the electrons, whereas, at ordinary temperatures, it is observed that the electronic specific heat is negligible. This and other failures of the classical theory can be eliminated, however, if we take account of the fact that the free-electron gas should obey Fermi-Dirac statistics rather than Maxwell-Boltzmann statistics. That is to say, classical theory must be replaced by quantum theory. The following are the rules that must be complied with: (1) Electrons can have spin quantum numbers of $+1/2$ or $-1/2$ but are otherwise indistinguishable from one another. (2) Electrons have discrete values for their energy and momentum. (3) No more than one electron of a given spin can reside in one of these discrete states (Pauli Exclusion Principle).

Sommerfeld showed that the number of permitted electronic states with energies in the
range from $E$ to $E + dE$, per unit volume, is given by

$$g(E)dE = \frac{4\pi(2m)^{3/2}E^{1/2}dE}{\hbar^3} \quad (A3.1.1)$$

where $m$ is the mass of a free electron. The probability that a particular state will be filled is given by the Fermi-Dirac distribution function

$$f_o(E) = \left[ \exp \left( \frac{E - \xi}{k_BT} \right) + 1 \right]^{-1} \quad (A3.1.2)$$

where $\xi$ is the Fermi energy, which has a value such that

$$\int_0^\infty f_o(E)g(E)dE = n \quad (A3.1.3)$$

where $n$ is the total number of free electrons per unit volume. At the absolute zero of temperature, the Fermi energy has the value

$$\xi_o = \frac{\hbar^2}{2m} \left( \frac{3n}{8\pi} \right)^{2/3} \quad (A3.1.4)$$

All the states with $E < \xi_o$ are filled with electrons, whereas all the states with $E > \xi_o$ are empty. This, of course, is in contrast to the result from classical theory that all electrons have zero energy at 0 K. It should be noted that, at ordinary temperatures, $\xi >> k_BT$ for all metals, which means that $\xi \approx \xi_o$.

Eq. (A3.1.4) shows that the value of the Fermi energy at the absolute zero of temperature can become small if the free-electron concentration, $n$, is much less than is normally found in a metal. It is then possible to achieve the condition, $k_BT >> \xi_o$, at ordinary temperatures, whence Eq. (A3.1.2) reduces to

$$f_o(E) = \exp \left( -\frac{E - \xi}{k_BT} \right), \quad k_BT >> \xi_o \quad (A3.1.5)$$
This is, in fact, the Maxwell-Boltzmann distribution function that is used in the classical kinetic theory of gases. When \( n \) is small enough for Eq. (A3.1.5) to apply, the electron gas is said to be nondegenerate. On the other hand, when \( n \) is so large that \( k_{\text{B}}T \ll \zeta_0 \), we have what is called degenerate electron gas. It turns out that thermoelectric materials tend to lie between these two extremes.

In a free-electron gas, the energy, \( E \), varies as the square of the wave vector, \( \vec{k} \). This clearly cannot be generally true for energy bands, although, close to the maxima and minima, the bands are parabolic in shape. This means that, if we measure the energy and wave vector from the band edge, then \( E \propto k^2 \), for small values of \( E \) and \( k \). Thus, we may write

\[
E = \frac{\hbar^2 k^2}{8\pi^2 m^*}
\]

which is the same expression as for a free-electron gas, except that the mass, \( m \), of an electron has been replaced by \( m^* \), known as the effective mass. It should be noted that the same effective mass may be inserted into Eq. (A3.1.1) to yield a density of states, near the band maximum or minimum, of

\[
g(E)\,dE = \frac{4\pi(2m^*)^{3/2}|E|^{1/2}\,d|E|}{\hbar^3}
\]

In these cases, the interaction between the electrons and the crystal lattice is fully taken into account by using the effective mass.

The effective mass may be defined quite generally by the expression

\[
\frac{1}{m^*} = \frac{1}{\hbar^2} \frac{d^2E}{dk^2}
\]

or

\[
\text{(A3.1.8)}
\]
but it is only near band edges that Eqs. (A3.1.6) and (A3.1.7) can usefully be employed. Of special interest is the situation near a band maximum, when \( \frac{d^2E}{dk^2} \) becomes negative. This means that the crystal behaves as if it contains negative carriers of negative mass, but it is convenient to regard these as positive carriers of positive mass, known as positive holes (usually just ‘holes’). It is this feature of band theory that explains why the values for the thermoelectric and Hall coefficients are sometimes positive. When we are dealing with positive holes, the energy, \( E \), must be measured downwards from the band edge, if, for example, Eq. (A3.1.6) is to be used. Also, the Fermi energy, \( \zeta \), must be measured downwards from the same band edge, and the function, \( f(E) \), then represents the probability of a state being unoccupied by an electron, that is, occupied by a hole.

Energy-band theory has been particularly successful in accounting for the difference between metals, semiconductors and insulators. In a metal, the Fermi energy lies within one, or more, of the allowed bands, and well away from the band edge. The density of electron states is large in such a region, and the electrical conductivity is high in such a partly filled band. On the other hand, in an electrical insulator, the Fermi level lies within a forbidden gap. The bands below the gap are all completely full, so there are no free states into which electrons can move. The bands above are completely empty, so they cannot contribute to the conduction either.

In fact, this last statement is not quite true except at the absolute zero of temperature. At any temperature above 0 K, some electrons will be excited from the highest full band to the lowest empty band, so that there will be a finite, albeit small, electrical conductivity. If the energy gap, \( E_g \), is small enough, substantial numbers of electrons in the upper band, and equal
numbers of holes in the lower band, will be excited in this way. The substance is then said to be an intrinsic semiconductor.

It has been found possible to make a great many materials that are essentially insulators in their pure state into electrical conductors, by adding impurities. Sometimes the impurities give up electrons to the upper band and are said to be donors. Alternatively, the impurities may take up electrons from the lower bands, so as to leave behind positive holes; such impurities are called acceptors. In both cases, the states associated with the impurities usually lie very close to the appropriate band edge, so that excitation is virtually complete at ordinary temperatures. The substances in which the charge carriers originate from the addition of impurities are known as extrinsic semiconductors.

Energy diagrams rather than band-structure diagrams are useful, when we have no particular interest in the value of the wave vector. It will be observed that the upper and lower allowed bands are called the conduction band and valence band, respectively.

The essential difference between an insulator and an intrinsic semiconductor is just the width of the energy gap in terms of $k_B T$. Thus, an insulator at ordinary temperatures may become an intrinsic semiconductor at an elevated temperature. Likewise, as the temperature of an extrinsic semiconductor is raised, the charge carriers excited across the energy gap will eventually predominate over the carriers due to the impurities. When this happens, the extrinsic semiconductor becomes intrinsic.

There are certain materials in which the valence and conduction bands just overlap one another. They are called semimetals, and they do not differ substantially in their properties from narrow-gap semiconductors. Their carrier concentration can, for example, be changed by adding donor or acceptor impurities. This is in contrast with the situation in ordinary metals, in which
the changes of electrical resistivity, on adding impurities, are the result of a reduction in the free path length, rather than the result of any alteration of the carrier concentration.

### A3.2 Electron-Scattering Processes

The transport properties of a conductor depend not only on the concentration of the charge carriers but also on their free path length between collisions. This would have an infinite value if the electrons or holes were moving in a perfectly periodic potential, since they would then never be scattered. However, in any real crystal, the free path length is limited by scattering due to local distortions of the potential electronic function.

Transport coefficients can be derived from Boltzmann’s equation, and the scattering processes can be described in terms of a relaxation time, \( \tau_e \). If the electron distribution function is disturbed from its equilibrium value, \( f_0 \), to \( f \), where \( |f-f_0| \ll f_0 \), the system relaxes according to the equation

\[
\left( \frac{df(E)}{dt} \right)_{\text{scatter}} = -\frac{f(E) - f_0(E)}{\tau_e}
\]

(A3.2.1)

The relaxation time can, hopefully, be expressed in the form

\[
\tau_e = \tau_0 E^r
\]

(A3.2.2)

where \( \tau_0 \) and \( r \) are constants. There may well be several scattering processes, with relaxation times, \( \tau_{e1}, \tau_{e2}, \ldots \), that should all be taken into account, whereupon

\[
\frac{1}{\tau_e} = \frac{1}{\tau_{e1}} + \frac{1}{\tau_{e2}} + \ldots
\]

(A3.2.3)

but it is usually supposed that one process predominates.

In a pure and perfect crystal, the carriers are scattered by the thermal vibrations of the crystal lattice. Both the acoustic and the optical modes can upset the periodicity of the potential.
The acoustic modes perturb the lattice potential in two ways. The local changes of lattice parameter directly influence the energy-band structure, leading to so-called deformation potentials. Alternatively, if the atoms in the crystal are partly ionized, piezoelectric potentials result from their displacement. It is usually assumed that the deformation-potential scattering is much stronger than the piezoelectric scattering.

Similarly, optical vibrations can cause scattering in more or less the same two ways. In this case, it is usual to refer to the mechanisms as nonpolar and polar. The polar process that creates dipole moments, which vary in time and space, may completely outweigh nonpolar processes.

When electrons and holes with a given energy are located at different points in the Brillouin zone, as in a multivalley semiconductor, lattice scattering may involve large changes of wave vector (intervalley scattering), as well as small changes (intravalley scattering). The two processes have different effects at low temperatures, but at high temperatures, when there are substantial numbers of phonons with large wave vectors, intervalley scattering becomes similar to intravalley scattering.

A number of other scattering processes become possible in imperfect and impure crystals. Impurity atoms act as scattering centers, particularly if they are ionized. Crystal defects, such as dislocations, can also produce a significant effect. Often, mixed crystals or solid solutions are used in thermoelectric materials, in which case alloy scattering can become important. Finally, one should remember that the charge carriers can scatter one another, though it is doubtful that carrier-carrier scattering is significant at the concentrations in which we are interested.
References

1. H.B. Callen, Phys. Rev. 73, 1349 (1948).

Appendix References

A5. P. Debye, Vorträge über die kinetische Theorie, Teubner (1914).