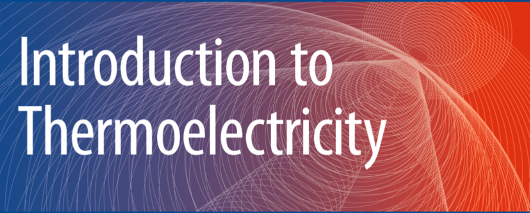
H. Julian Goldsmid

SPRINGER SERIES IN MATERIALS SCIENCE 121



Introduction to Thermoelectricity

Springer

Springer Series in
MATERIALS SCIENCE 121

Springer Series in
MATERIALS SCIENCE

The Springer Series in Materials Science covers the complete spectrum of materials physics, including fundamental principles, physical properties, materials theory and design. Recognizing the increasing importance of materials science in future device technologies, the book titles in this series reflect the state-of-the-art in understanding and controlling the structure and properties of all important classes of materials.

Please view available titles in *Springer Series in Materials Science*
on series homepage http://www.springer.com/series/856

H. Julian Goldsmid

Introduction to Thermoelectricity

With 140 Figures

Springer

Professor H. Julian Goldsmid
University of New South Wales, School of Physics
2052 Sydney, Australia
E-mail: hjgoldsmid@phys.unsw.edu.au

Series Editors:

Professor Robert Hull
University of Virginia
Dept. of Materials Science and Engineering
Thornton Hall
Charlottesville, VA 22903-2442, USA

Professor Jürgen Parisi
Universität Oldenburg, Fachbereich Physik
Abt. Energie- und Halbleiterforschung
Carl-von-Ossietzky-Straße 9–11
26129 Oldenburg, Germany

Professor R. M. Osgood, Jr.
Microelectronics Science Laboratory
Department of Electrical Engineering
Columbia University
Seeley W. Mudd Building
New York, NY 10027, USA

Professor Hans Warlimont
DSL Dresden Material-Innovation GmbH
Pirnaer Landstr. 176
01257 Dresden, Germany

Springer Series in Material Science ISSN 0933-033X
ISBN 978-3-642-00715-6 e-ISBN 978-3-642-00716-3
DOI 10.1007/978-3-642-00716-3
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2009929170

© Springer-Verlag Berlin Heidelberg 2010
This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.
The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book has been written at a time when thermoelectric energy conversion is showing great promise. It was in 1953 that I first carried out the experiments on bismuth telluride that demonstrated the potential of thermoelectric refrigeration. The present-day thermoelectric modules are based on the work that was carried out during the late 1950s and the early 1960s on bismuth telluride and its alloys. Since that time, there have been significant advances in materials for thermoelectric generation, but at all temperatures the efficiency of energy conversion using thermocouples has fallen far short of that expected for an ideal thermodynamic machine. At last, with the advent of nanostructured thermoelements, there is the promise that substantial advances will be made.

The basic principles of thermoelectric devices have not changed over the years and the theory presented in the first few chapters will always be applicable as new materials are discovered. A review of existing thermoelectric materials is presented with a chapter devoted to bismuth telluride showing how improvements in its synthesis and composition have led to the present-day performance. It is not always appreciated that the behaviour of a specific alloy is strongly dependent on the manner in which it is prepared and a chapter is devoted to the production of materials, the stress being on principles rather than on experimental detail.

The assessment of the transport properties of thermoelectric materials presents special problems. The chapter on measurement techniques includes a discussion of the errors that can arise when the so-called figure of merit is determined for non-uniform specimens. Indeed, I myself was led astray in the interpretation of experimental observations on polycrystalline samples of anisotropic material before I realised the extent of the problem.

It is usual to make use of modules rather than simple thermocouples. There is an outline of the method of selecting commercial modules for any particular application and a discussion of the problems that arise from attempts to miniaturise the size of modules so as to economise on space and material. Throughout the book, I have tried to emphasise practical considerations.

A full understanding of the behaviour of nanostuctured thermoelectric materials requires the mastery of difficult theoretical concepts but it is hoped that the elementary treatment in this book will allow the reader to comprehend the basic principles. It is expected that the so-called bulk nanostructures will find their way into commercial production in the very near future.

It is only during the past 2 or 3 years that I have appreciated the potential of the synthetic transverse thermoelement and I have included a chapter that reviews this unusual configuration. I have also included discussions of energy conversion using the transverse thermomagnetic effects and the thermionic effects in solids and in vacuum. The latter, in particular, will lead to greatly improved efficiencies if they live up to their theoretical promise.

This book draws on my experience of thermoelectricity and its applications over the past 55 years. During that time I have been supported by many people and I acknowledge with gratitude the help that I have received from all of them.

In 1953, as a very junior scientist at the Research Laboratories of the General Electric Company, I was encouraged by my group leader, R.W. Douglas, to look into the possibility of using the Peltier effect in semiconductors as a practical means of refrigeration. He continued to support the project, in spite of scepticism from some of his senior colleagues, and the success of bismuth telluride as a thermoelectric material stems from his foresight. I received support from many others in the Solid Physics Group over the next few years and should mention particularly D.A. Wright, who supervised my Ph.D. studies, and Ray Drabble, who helped me to understand transport theory.

In my academic life between 1964 and 1988, first as Reader in Solid State Physics at the University of Bath and then as Professor of Physics at the University of New South Wales, I was fortunate to be working in institutions that had been founded to promote applied science. I was encouraged to continue my research on thermoelectricity and was joined by some excellent students. I am sure that I learned much more from them than they did from me.

I acknowledge the support that I have received over much of my career from Marlow Industries. Raymond Marlow enabled me to work closely with his company and kept me in touch with practical developments. In recent years, I have been stimulated by my contact with George Nolas and Ted Volckmann and I appreciate the fact that I am still able to work with Jeff Sharp and Jim Bierschenk.

Perhaps, my greatest inspiration has been the work of Abram Ioffe and I greatly valued the opportunity, in 2005, to join in the celebration of the 125th anniversary of his birth in the town of Romny in Ukraine. This was made possible through an invitation from Professor L.I. Anatychuk and I am most appreciative of his encouragement for me to continue with my research.

Over the whole of my career I have received enthusiastic support from my wife Joan and it is to her that I dedicate this book.

Kingston Beach
Tasmania, Australia
June 2009

Julian Goldsmid

Acknowledgments

All the diagrams in this book are original but several are based on material that has been published elsewhere. The permission of the following publishers to use this material is gratefully acknowledged.

American Physical Society.
CRC Taylor and Francis.
Elsevier.
Institute of Electrical and Electronics Engineers.
Institute of Thermoelectricity.
Wiley VCH.

Contents

List of Symbols

A	Cross-section area, mean atomic weight, parameter for point-defect scattering
A_{M}	Parameter for mass-defect scattering
A_0	Richardson constant
a	Lattice constant
a_{H}	Scattering law dependent parameter in Hall coefficient
B	Magnetic field, parameter for umklapp scattering
B_{K}	Parameter in Keyes relation
C	Parameter for scattering by normal processes, concentration
c	Diameter of defect
c_{V}	Specific heat per unit volume
D	Diffusion coefficient in liquid
D^*	Specific detectivity
d	Width, electrode spacing, barrier width
d_{t}	Tunneling width
E	Electric field
E_F	Fermi energy
E_{g}	Energy gap
e	Electron charge
F_{n}	Fermi–Dirac integral
F_{NE}	Function proportional to thermomagnetic figure of merit
f	Fermi distribution function, measure of reduction of lattice conductivity in calculations for Si–Ge
f_0	Equilibrium Fermi distribution function
$\mathbf{G}$	Reciprocal lattice vector
G	Bulk modulus
g	Density of electron states, ratio of space occupied by insulation to that of thermoelements
h	Planck's constant, $h/2\pi$
I	Electric current
I_{q}	Current for maximum cooling power
I_ϕ	Current for maximum COP
i	Electric current density

i_1	Electric current density in a thermionic device
j	Heat flux density
K	Thermal conductance
K_c	Thermal conductance of end plates
K_s	Transport integral
k	Boltzmann's constant, segregation coefficient
$\boldsymbol{k}$	Wave vector for charge carriers
k_0	Parameter in Callaway's theory
L	Length, Lorenz number, latent heat
$\boldsymbol{l}$	Vector parallel to temperature gradient
l_e	Mean free path of charge carriers
l_t	Mean free path of phonons
M	$(1 + ZT_\mathrm{m})^{1/2}$, mean atomic mass, Average mass of unit cell
m	Mass of free electron, slope of liquidus
$m*$	Density-of-states effective mass
m_I	Inertial mass
m_N	Density-of-states mass for a single valley
N	Nernst coefficient, total number of modes of vibration, number of unit cells per unit volume, number of couples in a module
N_0	Bose–Einstein function
N_A	Avogadro's number
N_v	Number of valleys in an energy band
n	Subscript for electrons
n	Electron concentration, ratio of layer thicknesses in a synthetic transverse thermoelement
n_L	Number of vibrational modes per unit volume and frequency
P	Ettingshausen coefficient, Poisson's ratio
p	Porosity factor, proportion of specular reflection of phonons
p	Subscript for positive holes
$\boldsymbol{p}$	Phonon momentum
q	Rate of heat flow
$\boldsymbol{q}_\mathrm{L}$	Phonon wave vector
q_max	Maximum cooling power
q_1	Rate of heat flow from source
R	Electrical resistance, gas constant, responsivity
R_H	Hall coefficient
R_L	Load resistance
r	Scattering law parameter
r_c	Electrical contact resistance for unit area
S	Righi–Leduc coefficient
s	Compatibility factor
T	Temperature
T_1	Temperature of heat source
T_2	Temperature of heat sink
T_m	Mean temperature, melting point

ΔT	Temperature difference, difference between liquidus and solidus temperatures
ΔT^*	Temperature difference between sink and source
ΔT_{max}	Maximum temperature difference
t	Time
u	Velocity of carriers
V	Voltage, mean atomic volume
V_{q}	Voltage for maximum cooling power
v	Speed of sound, speed of zone
W	Energy in a mode of vibration, thermal resistance
w	Electrical power
x	$\hbar\omega/kT$
y	Parameter in Callaway's theory
Z	Thermoelectric figure of merit for couple
Z_{NE}	Thermomagnetic or Nernst–Ettingshausen figure of merit,
$Z_{\mathrm{NE}}{}^{\mathrm{i}}$	Isothermal thermomagnetic figure of merit
Z_{ϕ}	Transverse figure of merit
z	Figure of merit for single material
z_{d}	Phonon drag figure of merit
z_{1D}	One-dimensional figure of merit
z_{2D}	Two-dimensional figure of merit
α	Seebeck coefficient
α_{d}	Phonon drag Seebeck coefficient
α_{I}	Thermionic parameter replacing Seebeck coefficient
α_{T}	Thermal expansion coefficient
β	Chasmar and Stratton's materials parameter
β^*	$\left(1-\rho\Delta\rho/R_H^2B^2\right)$
β'	Materials parameter for a 2D conductor
β''	Materials parameter for a 1D conductor
β_{I}	Materials parameter for a solid-state thermionic device
Γ	Gamma function
γ	Grüneisen's parameter
δ	Atomic diameter
ε	Energy, emissivity, surface roughness
ε_{m}	Parameter in melting rule
η	Efficiency, reduced Fermi energy
η_{g}	Reduced energy gap
η_{r}	Reduced efficiency
Θ_{D}	Debye temperature
κ	Thermal diffusivity
Λ	Wavelength of phonons
Λ_0	Smallest phonon wavelength
λ	Thermal conductivity
λ_{e}	Electronic thermal conductivity

λ_{I}	Thermal conductivity of insulation, thermionic quantity replacing thermal conductivity
λ_{L}	Lattice conductivity
μ	Carrier mobility
ν	Frequency
ξ	Reduced energy
π	Peltier coefficient
π_{d}	Phonon drag Peltier coefficient
ρ	Electrical resistivity
ρ_{d}	Density
σ	Electrical conductivity, Stefan–Boltzmann constant
σ_{I}	Thermionic quantity replacing electrical conductivity
σ_0	Parameter that depends on mobility and effective mass
τ	Thomson coefficient, relaxation time
τ_0	Scattering law constant
τ_d	Relaxation time for phonon drag
τ_{e}	Relaxation time for charge carriers
τ_{eff}	Effective relaxation time for charge carriers
τ_{N}	Relaxation time for normal processes
τ_{R}	Relaxation time for umklapp processes
$\varPhi$	Work function
ϕ	Coefficient of performance, angle of transverse thermoelement to normal to layers
ϕ_{q}	Coefficient of performance at maximum cooling power
ϕ_{s}	Coefficient of performance of each stage of a cascade
ϕ_{ϕ}	Maximum coefficient of performance
χ	Compressibility
ω	Angular frequency
ω_{D}	Debye angular frequency

Chapter 1
The Thermoelectric and Related Effects

1.1 Introduction

The first of the thermoelectric effects was discovered, in 1821, by T. J. Seebeck. He showed that an electromotive force could be produced by heating the junction between two different electrical conductors. The Seebeck effect can be demonstrated by making a connection between wires of different metals (e.g., copper and iron). The other ends of the wires should be applied to the terminals of a galvanometer or sensitive voltmeter. If the junction between the wires is heated, it is found that the meter records a small voltage. The arrangement is shown in Fig. 1.1. The two wires are said to form a thermocouple. It is found that the magnitude of the thermoelectric voltage is proportional to the difference between the temperature at the thermocouple junction and that at the connections to the meter.

Thirteen years after Seebeck made his discovery, J. Peltier, a French watchmaker, observed the second of the thermoelectric effects. He found that the passage of an electric current through a thermocouple produces a small heating or cooling effect depending on its direction. The Peltier effect is quite difficult to demonstrate using metallic thermocouples since it is always accompanied by the Joule heating effect. Sometimes, one can do no better than show that there is less heating when the current is passed in one direction rather than the other. If one uses the arrangement shown in Fig. 1.1, the Peltier effect can be demonstrated, in principle, by replacing the meter with a direct current source and by placing a small thermometer on the thermocouple junction.

It seems that it was not immediately realised that the Seebeck and Peltier phenomena are dependent on one another. However, this interdependency was recognised by W. Thomson (who later became Lord Kelvin), in 1855. By applying the theory of thermodynamics to the problem, he was able to establish a relationship between the coefficients that describe the Seebeck and Peltier effects. His theory also showed that there must be a third thermoelectric effect, which exists in a homogeneous conductor. This effect, now known as the Thomson effect, consists of reversible heating or cooling when there is both a flow of electric current and a temperature gradient.

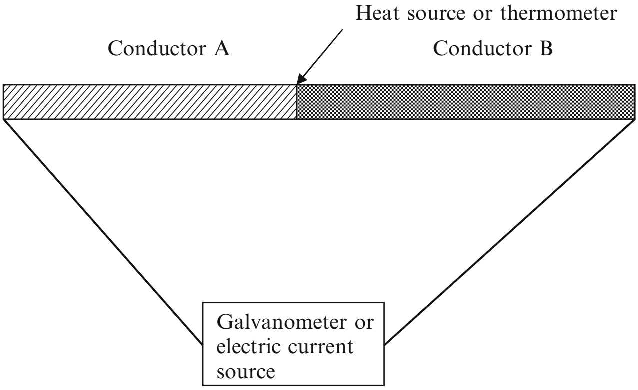


Fig. 1.1 Experiment to demonstrate the Seebeck and Peltier effects

The fact that the Seebeck and Peltier effects occur only at junctions between dissimilar conductors might suggest that they are interfacial phenomena but they are really dependent on the bulk properties of the materials involved. Nowadays, we understand that electric current is carried through a conductor by means of electrons that can possess different energies in different materials. When a current passes from one material to another, the energy transported by the electrons is altered, the difference appearing as heating or cooling at the junction, that is as the Peltier effect. Likewise, when the junction is heated, electrons are enabled to pass from the material in which the electrons have the lower energy into that in which their energy is higher, giving rise to an electromotive force.

Thomson's work showed that a thermocouple is a type of heat engine and that it might, in principle, be used either as a device for generating electricity from heat or, alternatively, as a heat pump or refrigerator. However, because the reversible thermoelectric effects are always accompanied by the irreversible phenomena of Joule heating and thermal conduction, thermocouples are generally rather inefficient.

The problem of energy conversion using thermocouples was analysed by Altenkirch [1], in 1911. He showed that the performance of a thermocouple could be improved by increasing the magnitude of the differential Seebeck coefficient, by increasing the electrical conductivities of the two branches and by reducing their thermal conductivities. Unfortunately, at that time, there were no thermocouples available in which the combination of properties was good enough for reasonably efficient energy conversion, although the Seebeck effect has long been used for the measurement of temperature and for the detection of thermal radiation. It was only in the 1950s that the introduction of semiconductors as thermoelectric materials allowed practical Peltier refrigerators to be made. Work on semiconductor thermocouples also led to the construction of thermoelectric generators with a high enough efficiency for special applications. Nevertheless, the performance of thermoelectric energy convertors has always remained inferior to that of the best conventional machines.

In fact, there was little improvement in thermoelectric materials from the time of the introduction of semiconductor thermoelements until the end of the twentieth century. However, in recent years, several new ideas for the improvement of materials have been put forward and, at last, it seems that significant advances are being made, at least on a laboratory scale. It is reasonable to expect that this work will soon lead to much wider application of the thermoelectric effects.

1.2 Relations Between the Thermoelectric Coefficients

We now define the Seebeck and Peltier coefficients and show how they are related to one another. For the time being, we assume that the conductors are isotropic. We refer to the simple thermocouple shown in Fig. 1.2. Conductor A is joined at both ends to conductor B, the latter being divided into two parts so that, for example, a voltmeter can be inserted in the gap.

Suppose that a temperature difference ΔT is established between the two junctions and that the two free ends of conductor B are maintained at the same temperature, it will then generally be found that a potential difference V will appear between the free ends. The differential Seebeck coefficient, α_{AB}, is defined as the ratio of V to ΔT. Thus,

$$\alpha_{\mathrm{AB}} = \frac{V}{\Delta T} \tag{1.1}$$

α_{AB} is deemed to be positive if the electromotive force tends to drive an electric current through conductor A from the hot junction to the cold junction. It is noted that, particularly in older texts, the quantity that is now known as the Seebeck coefficient has often been called the thermoelectric power or the thermal EMF coefficient.

We define the differential Peltier coefficient, π_{AB}, for the same thermocouple by supposing that a source of EMF is connected across the gap in conductor B so as to drive a current around the circuit in a clockwise direction. The Peltier coefficient is regarded as positive if the junction at which the current enters A is heated and the junction at which it leaves A is cooled. π_{AB} is equal to the ratio of the rate q of heating or cooling at each junction to the electric current I,

$$\pi_{\mathrm{AB}} = \frac{q}{I}. \tag{1.2}$$

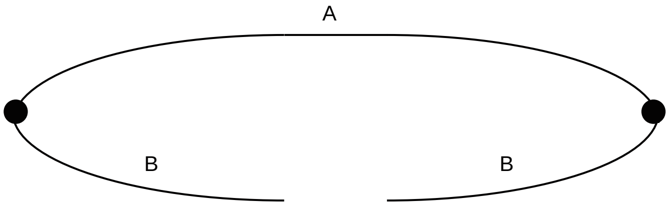


Fig. 1.2 Simple thermocouple

We note that it is much simpler to measure the Seebeck coefficient than the Peltier coefficient. Thus, while both quantities enter into the theory of thermoelectric energy conversion, it would be preferable if only one of them had to be specified. In fact, one of the Kelvin relations allows us express the Peltier coefficient in terms of the Seebeck coefficient. The relevant equation is

$$\pi_{\mathrm{AB}} = \alpha_{\mathrm{AB}} T. \tag{1.3}$$

The other Kelvin relation connects the Seebeck coefficient and the Thomson coefficient, τ, or, rather, the difference between the Thomson coefficients of the two conductors. The Thomson coefficient is defined as the rate of heating per unit length that results from the passage of unit current along a conductor in which there is unit temperature gradient. The appropriate Kelvin relation is

$$\tau_{\mathrm{A}} - \tau_{\mathrm{B}} = T\frac{\mathrm{d}\alpha_{\mathrm{AB}}}{\mathrm{d}T}. \tag{1.4}$$

The Seebeck and Peltier coefficients are defined above for a pair of conductors whereas it would be much more convenient if their values could be given for a single material. In fact, the absolute Seebeck or Peltier coefficient becomes equal to the differential coefficient if the second material can be regarded as having zero absolute coefficients. This concept can be realised, in practice, by using a superconductor as the second material. It is reasonable to assign zero Seebeck or Peltier coefficients to a superconductor since the differential coefficients between all pairs of superconductors are zero.

Of course, there is no material that remains in the superconducting state at ordinary temperatures, so it might be thought that the absolute Seebeck coefficients of other materials can be obtained only at low temperatures. However, this is not the case. It is reasonable to write (1.4) in the form

$$\tau = T\frac{\mathrm{d}\alpha}{\mathrm{d}T} \tag{1.5}$$

for a single conductor. Thus, if the absolute Seebeck coefficient of a material at low temperatures is determined by connecting it to superconductor, one can then use (1.5) to find the value at higher temperatures after measuring the Thomson coefficient [2, 3]. This procedure has been carried out for the metal lead, which may be used as a reference material when determining the absolute coefficients for other substances.

In actual fact, most metals, like lead, have very small absolute values of the Seebeck coefficient compared with practical thermoelectric materials that are almost invariably semiconductors.

1.3 Effects in a Magnetic Field

Electric charges are subject to transverse forces when they travel in a magnetic field. Thus, the thermoelectric effects, like the other transport properties, become changed when a magnetic field is applied, and there also appear some new phenomena. We need to discuss these so-called thermogalvanomagnetic effects since they can affect the performance of thermoelectric devices and can even lead to new methods of energy conversion.

As we shall demonstrate in the next chapter, the electric and thermal conductivities are properties that are of importance when we are calculating the performance of devices based on the Seebeck and Peltier effects. Both quantities become less on the application of a magnetic field, though the changes are very small unless the field is very strong and the mobility of the charge carriers is high. The Seebeck and Peltier coefficients, too, will change under the influence of a magnetic field, B.

Usually, the value of the Seebeck coefficient will be the same when the direction of the magnetic field is reversed but this is not always the case. Any difference between the values of the Seebeck coefficient upon reversal of the field is called the Umkehr effect. The Umkehr effect has been shown [4] to be very large for certain orientations of the semimetal bismuth.

Another consequence of the action of a magnetic field [5] is the need to modify the Kelvin relation (1.3). The modified equation is

$$\pi(B) = T\alpha(-B). \tag{1.6}$$

When a transverse magnetic field is applied to a current carrying conductor, an electric field appears in a direction perpendicular to both the current and B. This is the well-known Hall effect. The Hall effect is not immediately relevant to energy conversion but is a useful tool in explaining the behaviour of the charge carriers. Of more direct significance for energy conversion are the transverse Nernst and Ettingshausen effects.

The Nernst effect, like the Hall effect, manifests itself as a transverse voltage in a magnetic field but it depends on the longitudinal temperature gradient or heat flow rather than on a longitudinal electric current. The Nernst coefficient, N, is defined by the relation

$$|N| = \frac{\mathrm{d}V/\mathrm{d}y}{B_z\,\mathrm{d}T/\mathrm{d}x}. \tag{1.7}$$

Here $\mathrm{d}V/\mathrm{d}y$ is the transverse electric field. The sign of the Nernst effect is given in Fig. 1.3, which illustrates all the transverse thermogalvanomagnetic phenomena. The sign of the Nernst effect does not depend on whether the charge carriers are positive or negative and, in this respect, it differs from the Hall effect.

The Ettingshausen and Nernst effects are related to one another in the same way as the Peltier and Seebeck effects. The Ettingshausen effect is a transverse temperature gradient that is the result of a transverse magnetic field and a longitudinal flow of electric current. The Ettingshausen coefficient, P, is defined by

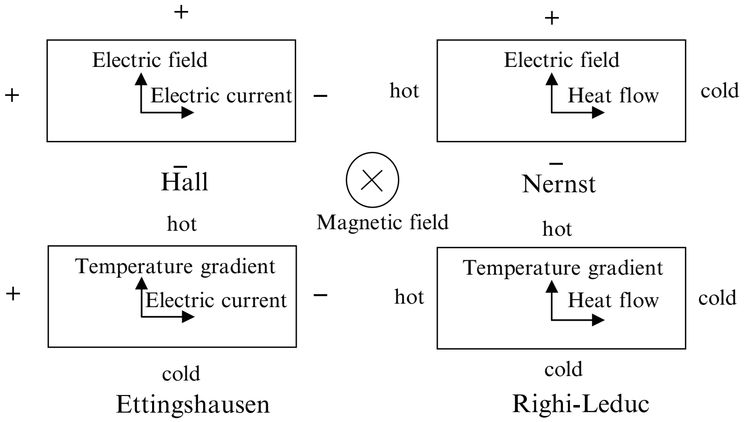


Fig. 1.3 The transverse thermogalvanomagnetic effects. When the effects are in the direction shown in the diagram, the coefficients are positive

$$|P| = \frac{\mathrm{d}T/\mathrm{d}y}{i_x B_z}, \tag{1.8}$$

where i_x is the longitudinal current density. As one might have expected, there is a thermodynamic relationship between the Nernst and Ettingshausen coefficients,

$$P\lambda = NT, \tag{1.9}$$

where λ is the thermal conductivity, which is included since the Ettingshausen coefficient is defined in terms of a temperature gradient rather than a heat flow.

To complete the transverse phenomena, there exists the Righi–Leduc effect, which is a transverse temperature gradient arising from a longitudinal heat flow. The Righi–Leduc coefficient, S, is given by

$$|S| = \frac{\mathrm{d}T/\mathrm{d}y}{B_z \mathrm{d}T/\mathrm{d}x}. \tag{1.10}$$

References

1. E. Altenkirch, Phys. Z. **12**, 920 (1911)
2. G. Borelius, W.H. Keesom, C.H. Johansson, J.O. Linde, Proc. Ned. Akad. Wet. **35**, 10 (1932)
3. J.W. Christian, J.P. Jan, W.B. Pearson, I.M. Templeton, Can. J. Phys. **36**, 627 (1958)
4. G.E. Smith, R. Wolfe, S.E. Haszko, in *Proceedings of the International Conference on the Physics of Semiconductors, Paris* (Dunod, Paris, 1964), p. 399
5. R. Wolfe, G.E. Smith, Phys. Rev. **129**, 1086 (1963)

Chapter 2
Theory of Thermoelectric Refrigeration and Generation

2.1 The Transport Effects

The thermoelectric phenomena are reversible in the sense that they do not of themselves give rise to thermodynamic losses. However, they are always, in practice, accompanied by the irreversible effects of electrical resistance and thermal conduction. It turns out that the performance of any thermocouple as an energy convertor can be expressed in terms of the differential Seebeck coefficient and the thermal and electrical resistances of the two branches. These resistances depend on the thermal and electrical resistivities and the ratios of length to cross-sectional area. Again we shall, in the first instance, assume that all the properties are independent of orientation.

The electrical resistivity, ρ, is the reciprocal of the electrical conductivity, σ, which is defined by the relation

$$I = \frac{\sigma V A}{L} \tag{2.1}$$

where I is the electric current through a specimen of constant cross-sectional area A and length L when a voltage V is applied. Likewise, the thermal conductivity, λ, is defined by the equation

$$q = -\frac{\lambda A \, \Delta T}{L} \tag{2.2}$$

where q is the rate of heat flow through a similar specimen that has a temperature difference ΔT between its two ends.

We shall refer to the thermoelectric coefficients and the electrical and thermal conductivities of a given material as its transport properties. All these properties will generally be temperature-dependent and this should be taken into account in any rigorous theory. The variation with temperature of the transport properties may not be too serious a matter in some applications of the Peltier effect when the temperature differences across the thermocouples are small but it can be very important in thermoelectric generation. However, in order to determine the relative importance of the different parameters we shall, for the time being, suppose that the conductivities and the Seebeck coefficient are all independent of temperature. The Kelvin relation (1.3) implies that, even when α is constant, π will be proportional to T.

In fact, we shall invariably express the Peltier coefficient, π, in terms of the more easily measured Seebeck coefficient, α.

2.2 Thermoelectric Refrigerators and Heat Pumps

We shall determine the performance of thermoelectric refrigerators and heat pumps using as our model the single thermocouple shown in Fig. 2.1. Practical devices usually make use of modules that contain a number of thermocouples connected electrically in series and thermally in parallel. This enables the cooler or heat pump to be operated from a power source that delivers a manageable current with a reasonable voltage drop. It is a simple matter to extend the equations for a single couple to a multi-couple arrangement.

In the elementary theory that is outlined in this chapter, it will be supposed that there is no thermal resistance between the thermocouple and the heat source or sink. It will also be assumed that all the heat flow between the source and sink takes place within the thermocouple. Thus, it will be supposed that thermal radiation and losses by conduction and convection through the surrounding medium are negligible. The two thermocouple branches in our model have constant cross-sectional areas. There have been suggestions [1] that tapered thermoelements might improve the performance but it is not difficult to show that they give no theoretical advantage. The thermoelements need not be of the same length but the ratio of length to cross-sectional area (the form factor) is of importance and, as we shall see, there is a preferred relationship between the form factors of the two branches.

The quantity of greatest importance for a refrigerator is the coefficient of performance (COP), which is defined as the ratio of the heat extracted from the source to the expenditure of electrical energy. If the thermocouple were free of losses associated with heat conduction and electrical resistance, the COP would reach the ideal value; that is, the value for a Carnot cycle. The ideal COP can be much greater than

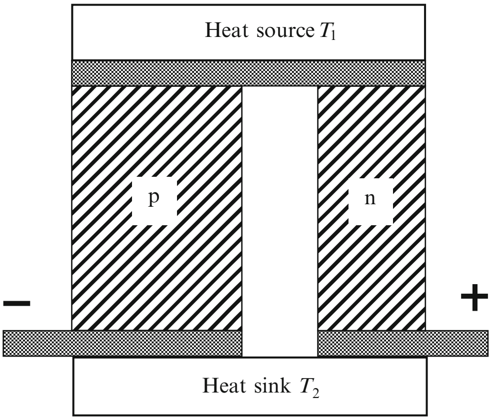


Fig. 2.1 Simple refrigerator or heat pump

unity as it is given by $T_1/(T_2 - T_1)$, where T_1 and T_2 are the absolute temperatures of the source and sink, respectively. We shall also be interested in the cooling power; i.e., the rate at which heat is extracted from the source.

A detailed derivation of the COP and cooling power may be found elsewhere [2]. Here, we shall outline the theory.

When a current, I, is passed through the couple, there is Peltier cooling at the source equal to $(\alpha_p - \alpha_n)IT_1$, where we have used the Kelvin relation (1.3) to eliminate the Peltier coefficient. α_p and α_n are the Seebeck coefficients of the two branches which, of course, should have opposite signs. This cooling effect is opposed by heat conduction at the rate $(T_2 - T_1)(K_p + K_n)$, where K_p and K_n are the thermal conductances of the branches. The cooling is also opposed by Joule heating within the thermoelements. It is easily shown that half of the Joule heating passes to the sink and half to the source, each half being equal to $I^2(R_p + R_n)/2$, where R_p and R_n are the thermal resistances of the branches.

The expression for the cooling power is

$$q_1 = (\alpha_p - \alpha_n)IT_1 - (T_2 - T_1)(K_p + K_n) - I^2(R_p + R_n)/2. \tag{2.3}$$

Also, the rate of expenditure of electrical energy is

$$w = (\alpha_p - \alpha_n)I(T_2 - T_1) + I^2(R_p + R_n) \tag{2.4}$$

where the first term is the rate of working to overcome the thermoelectric voltage whereas the second term is the resistive loss. The COP, ϕ, is then given by

$$\phi = \frac{(\alpha_p - \alpha_n)IT_1 - (T_2 - T_1)(K_p + K_n) - I^2(R_p + R_n)/2}{(\alpha_p - \alpha_n)I(T_2 - T_1) + I^2(R_p + R_n)}. \tag{2.5}$$

Equation (2.5) shows us that the COP depends on the current, as does the cooling power. As the current is increased, the Peltier cooling rises linearly but the Joule heating depends on I^2. Thus, a plot of cooling power against current has the parabolic form shown in Fig. 2.2. This plot represents, schematically, the situation in which T_1 is significantly smaller than T_2. Provided that the temperature difference is not too great, the cooling power will become positive at a certain value of the current. However, as the current is increased further, there will come a point at which the difference between the Peltier cooling and the Joule heating begins to diminish. In other words, there is a particular current at which the cooling power reaches its maximum value. Equation (2.3) shows that the maximum is reached when the Peltier cooling is twice that part of the Joule heating that reaches the cold junction.

There are two values of the current that are of special interest. The current, I_q, that yields the maximum cooling power is given by

$$I_q = (\alpha_p - \alpha_n)T_1/(R_p + R_n). \tag{2.6}$$

At this current, the COP is given by

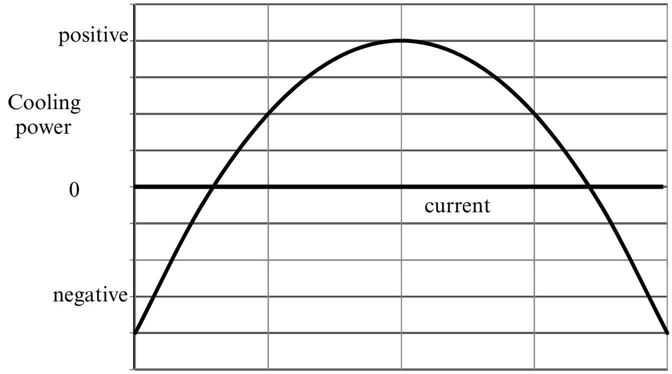


Fig. 2.2 Schematic plot of cooling power against current for a thermoelectric cooler. The cooling power is negative until the Peltier effect is great enough to counteract both heat conduction and Joule heating

$$\phi_q = \frac{ZT_1^2/2 - (T_2 - T_1)}{ZT_2T_1} \tag{2.7}$$

where Z is equal to $(\alpha_p - \alpha_n)^2/\{(K_p + K_n)(R_p + R_n)\}$. Equation (2.7) shows that the COP under the condition of maximum cooling power depends solely on Z and the temperatures of the source and sink. As we shall see, the optimum COP also depends only on these quantities and Z is therefore known as the figure of merit of the thermocouple. Z has the dimensions of inverse temperature and it is more usual nowadays to specify the dimensionless figure of merit, which is equal to ZT at a given temperature.

The other condition of particular interest is that of maximum COP. The current I_ϕ that satisfies this condition is specified by

$$I_\phi = \frac{(\alpha_p - \alpha_n)(T_2 - T_1)}{(R_p + R_n)\{(1 + ZT_m)^{1/2} - 1\}} \tag{2.8}$$

where T_m is the mean temperature. The optimum COP is

$$\phi_{max} = \frac{T_1\{(1 + ZT_m)^{1/2} - (T_2/T_1)\}}{(T_2 - T_1)\{(1 + ZT_m)^{1/2} + 1\}}. \tag{2.9}$$

It might be thought that one would wish to operate a thermoelectric refrigerator as close to the condition of optimum COP as possible. However, this is sometimes not practical. The cooling power under this condition can be much less than the maximum value, particularly, when the temperature difference between the source and sink is small. Thus, while the optimum COP condition may be economical in use of electrical energy, it may be uneconomical in the use of thermoelectric material. Generally speaking, the preferred current will lie somewhere between that for maximum cooling power and optimum COP.

The figure of merit Z itself may be optimised for a given pair of thermoelectric materials. The aim should be to make the product $(K_p + K_n)(R_p + R_n)$ as small as possible. This result is obtained when the form factors for the two branches satisfy the relation

$$\frac{L_n A_p}{L_p A_n} = \left(\frac{\rho_p \lambda_n}{\rho_n \lambda_p}\right)^{1/2}. \tag{2.10}$$

When (2.10) is satisfied the figure of merit is given by

$$Z = \frac{(\alpha_p - \alpha_n)^2}{\{(\lambda_p \rho_p)^{1/2} + (\lambda_n \rho_n)^{1/2}\}^2}. \tag{2.11}$$

It is usually the figure of merit defined by (2.11) that is meant when one discusses Z for a pair of materials.

In the search for improved thermocouples, it is uncommon to investigate a pair of substances at the same time. It would, therefore, be convenient if one could define a figure of merit for a single material. In practice, one makes use of the figure of merit z, which is given by $\alpha^2\sigma/\lambda$ or $\alpha^2/\lambda\rho$. The parameters used in defining z refer to the positive and negative thermoelements separately. It is important to realise that z cannot be used to calculate the performance of a thermocouple even if its value is known for both branches. For this purpose, the figure of merit Z must be used. However, it often turns out that Z lies close to the average of z_p and z_n so that it is meaningful to select materials on the basis of the single-material figure of merit.

One of the most important characteristics of a thermocouple is the maximum depression of temperature that can be reached through the Peltier effect using a single stage. This quantity, ΔT_{max}, can be calculated from (2.7). The maximum temperature depression is reached when the cooling power and, thus, the COP fall to zero. We see that this occurs when $(T_2 - T_1)$ becomes equal to $ZT_1^2/2$ so that

$$\Delta T_{max} = \frac{1}{2} Z T_1^2. \tag{2.12}$$

Figure 2.3 shows how ΔT_{max} varies with ZT_m when the heat sink is kept at 300 K. Practical thermoelectric refrigeration has stemmed from the development since the 1950s of thermocouples with ZT_m of the order of unity. Temperature depressions of 100° or more require values of ZT_m significantly greater than unity.

In Fig. 2.4, we show the optimum COP plotted against the dimensionless figure of merit for various heat source temperatures when the heat sink is at 300 K. It is noted that COPs that are considerably greater than unity can be achieved when ZT_m is equal to about 1 provided that the heat source is not at a too low temperature. For example, a COP value of 2 is reached when ZT_m is equal to 1, if the temperature difference between the source and sink is 20°. This is, of course, much smaller than the COP of an ideal refrigerator, which would be no less than 14 under the same conditions. A high COP is desirable, not only because it reduces the expenditure of electrical energy, but also because it allows a heat sink of smaller capacity to be

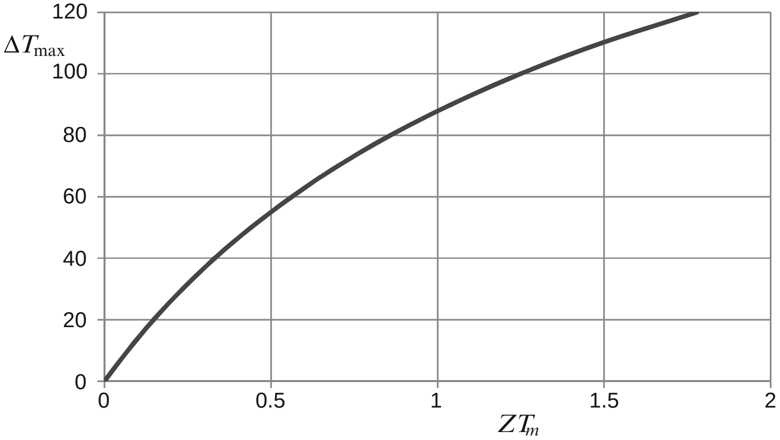


Fig. 2.3 Plot of maximum temperature depression against ZT_{m} for the heat sink at 300 K

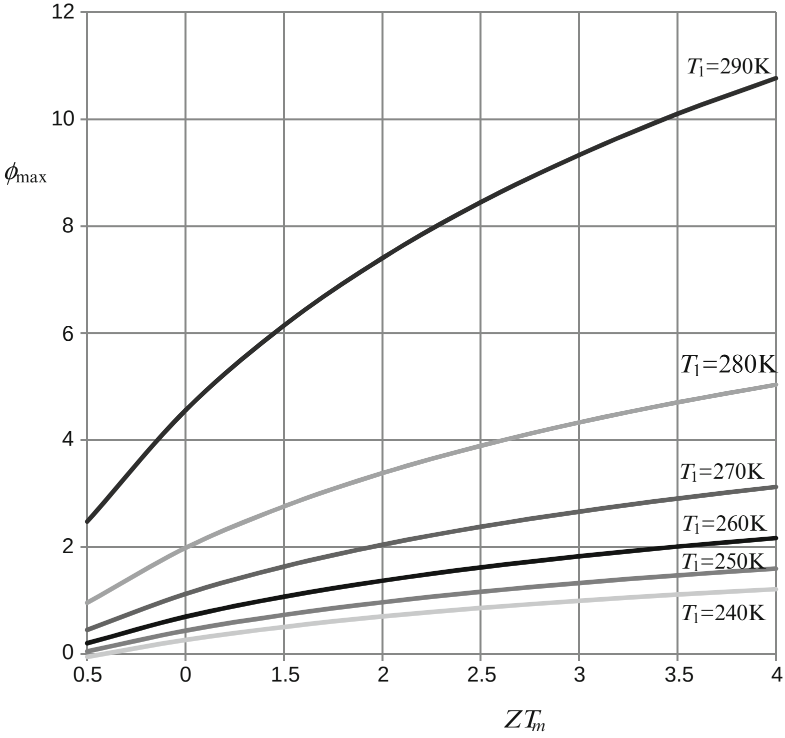


Fig. 2.4 Plot of optimum COP against ZT_{m} for the heat sink at 300 K

used. Thus, although many applications are possible with ZT_m equal to about unity, it is important that we should aim at much higher values.

We now turn briefly to thermoelectric heat pumps. Here, we are more interested in the rate that heat is delivered to the sink rather than the cooling power at the source. The difference between the two quantities will be equal to the electrical power consumed by the thermocouple. The rate at which the sink is heated is, thus, the sum of q_1 and w from (2.3) and (2.4), and is given by

$$q_2 = q_1 + w = w(\phi + 1). \tag{2.13}$$

The rate of heat delivery is normally greater than the electrical power w since ϕ is usually positive.

2.3 Thermoelectric Generators

We, now, consider the application of the Seebeck effect in the generation of electrical power. Again, we obtain the performance using the model of a single thermocouple and we assume that no heat arrives at the sink other than through the two branches. The arrangement is shown in Fig. 2.5 in which the thermocouple is connected to a load R_L that can be varied. The efficiency of the generator will depend on the value of R_L as well as on the properties of the thermocouple. We shall outline the theory of thermoelectric generation, a full treatment having been given elsewhere [3].

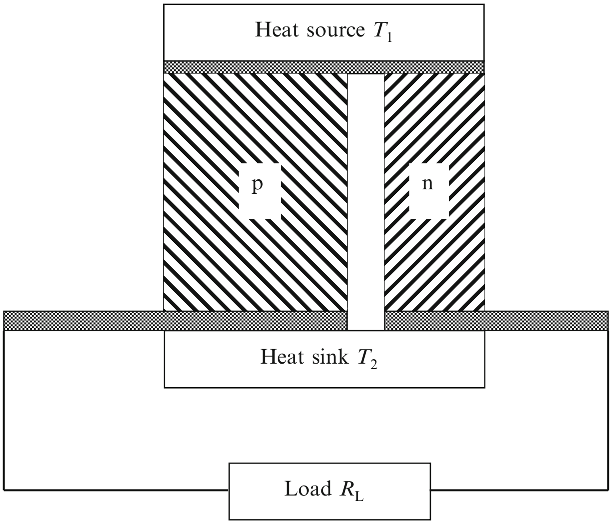


Fig. 2.5 Simple thermocouple used as a generator

We are interested in the electrical power that is delivered to the load and in the efficiency, η, that is the ratio of the output power to the rate at which heat is drawn from the source.

The thermal EMF is equal to $(\alpha_p - \alpha_n)(T_1 - T_2)$ and this gives rise to a current I that may be expressed as

$$I = \frac{(\alpha_p - \alpha_n)(T_1 - T_2)}{R_p + R_n + R_L}. \tag{2.14}$$

Thence, the power delivered to the load is

$$w = I^2 R_L = \left\{ \frac{(\alpha_p - \alpha_n)(T_1 - T_2)}{R_p + R_n + R_L} \right\}^2 R_L. \tag{2.15}$$

Part of the heat drawn from the source is used to balance the Peltier cooling associated with the flow of current. In addition, there is the flow of heat due to thermal conduction along the branches. Thus, the total rate of heat flow from the source is

$$q_1 = (\alpha_p - \alpha_n) I T_1 + (K_p + K_n)(T_1 - T_2). \tag{2.16}$$

The efficiency is equal to the ratio w/q_1.

The useful power reaches its maximum value when the load resistance is equal to the generator resistance. However, even if there were no loss of heat through thermal conduction, the efficiency could then never exceed 50%. An increase in the load resistance reduces the power output but increases the efficiency. It may be shown that the efficiency becomes a maximum when the ratio, M, of the resistance of the load to that of the generator is given by

$$M = \frac{R_L}{R_p + R_n} = (1 + Z T_m)^{1/2}. \tag{2.17}$$

As one might have expected, the same figure of merit, Z, applies for refrigeration and generation as is apparent from the expression for the efficiency

$$\eta = \frac{(T_1 - T_2)(M - 1)}{T_1(M + T_2/T_1)}. \tag{2.18}$$

If $Z T_m$ were much greater than unity, M would also be very large and the efficiency would approach $(T_1 - T_2)/T_1$, which is the value for the Carnot cycle.

In Fig. 2.6, we show the variation of the efficiency with the dimensionless figure of merit for a thermoelectric generator in which the source and sink are at 400 and 360 K, respectively. The Carnot cycle efficiency for these temperatures would be 10%.

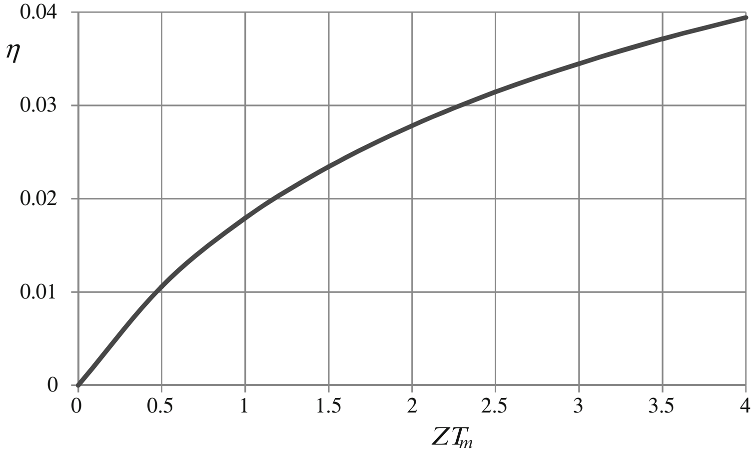


Fig. 2.6 Plot of efficiency against dimensionless figure of merit for the heat source at 400 K and the heat sink at 360 K

2.4 Multi-Stage Devices

If a thermoelectric unit has to operate between a source and sink that are at widely different temperatures, it is unlikely that a single pair of thermocouple materials will suffice. Thus, the two legs are often made up of segmented thermoelements [4]. Here, however, we concern ourselves with multi-stage devices, which provide a means for extending the maximum temperature difference for a thermoelectric refrigerator beyond the limit set by (2.12).

We shall suppose that heat can be transferred from one stage to another without losses associated with thermal resistance. In a cascade, as a multi-stage cooler is often called, each stage, as one proceeds from the heat source to the sink, has to have a greater cooling capacity than the one before. This is because every stage rejects not only the heat that it extracts from the previous stage, but also the Joule heat that is generated within it.

Suppose that there are N stages in the cascade and that the COP of the nth stage is equal to ϕ_n, then, if q_N is the rate of cooling of the Nth stage, in contact with the heat source, the rate of cooling for the nth stage is given by

$$q_n = q_N \left(1 + \frac{1}{\phi_N}\right)\left(1 + \frac{1}{\phi_{N-1}}\right)\cdots\left(1 + \frac{1}{\phi_{N-n}}\right). \tag{2.19}$$

The rate at which heat is delivered to the sink by the first stage is

$$q_n = q_N \left(1 + \frac{1}{\phi_N}\right)\left(1 + \frac{1}{\phi_{N-1}}\right)\cdots\left(1 + \frac{1}{\phi_1}\right), \tag{2.20}$$

and the overall COP is given by

$$\phi = \left[\left(1+\frac{1}{\phi_N}\right)\left(1+\frac{1}{\phi_{N-1}}\right)\cdots\left(1+\frac{1}{\phi_1}\right)-1\right]^{-1}. \tag{2.21}$$

In order to simplify the calculations, we shall assume that each stage has the same COP, ϕ_s. Then, the overall COP is

$$\phi = \left[\left(1+\frac{1}{\phi_s}\right)^N - 1\right]^{-1}. \tag{2.22}$$

It is also reasonable to assume that each stage operates with the maximum COP given by (2.9). We can then use (2.22) to determine the COP of the cascade. It will necessarily be an approximation if the temperature difference between the source and sink is large, since it is then most unlikely that we could arrange for the COPs for all the stages to be equal.

As we move from the Nth stage towards the first stage, the cooling power has to increase. Thus, a thermoelectric cascade has a pyramidal form, as shown schematically in Fig. 2.7. It is supposed that all the thermocouples are similar to one another and that extra cooling is attained by increasing the number of couples.

We have calculated the overall COP for up to four stages assuming that ZT has the same value of 0.7 throughout the device. The results are shown in Fig. 2.8, in which the COP is plotted against the heat source temperature with the heat sink at 300 K. The value of ZT is probably underestimated for commercially available thermoelectric modules at the upper end of the temperature range but overestimated at the lower end. When the coefficient becomes very small, it may be assumed that the cooling limit has been reached. Thus, the single-stage cooler has a minimum cold junction temperature of about 230 K. The minimum temperatures for the two-, three-, and four-stage cascades are of the order of 180, 160, and 140 K, respectively. Commercial multi-stage coolers do not behave quite as well as indicated by these theoretical curves. Thus, although a single-stage module supplied by

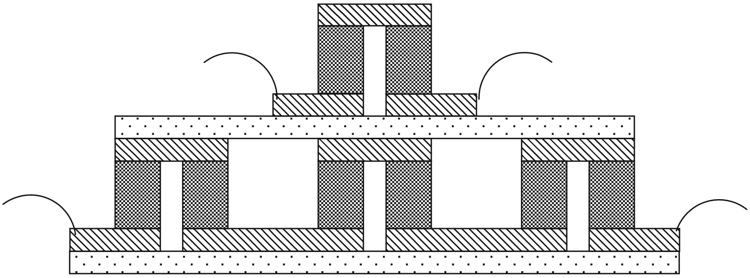


Fig. 2.7 Schematic arrangement of a two-stage thermoelectric cascade. The stages are electrically insulated from one another but in good thermal contact

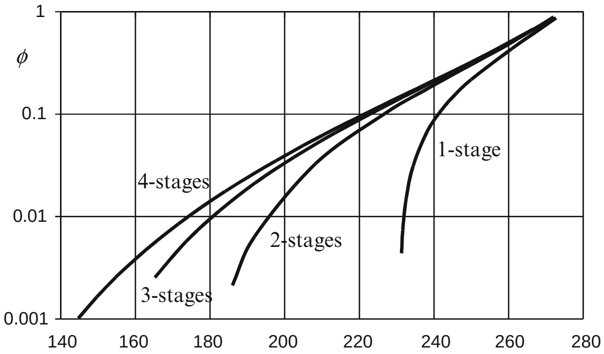


Fig. 2.8 Overall coefficient of performance plotted against heat source temperature for 1-, 2-, 3-, and 4-stage coolers. The heat sink is at 300 K and $ZT = 0.7$

Marlow Industries Inc. was found to give a minimum temperature of about 230 K, in agreement with theory, a two-stage cascade yielded only about 200 K. Even a six-stage cascade could not reach a temperature below about 170 K. This is, primarily, due to the fact that too high a value for ZT has been assumed for the calculations at the lower temperatures. Nevertheless, it is remarkable that temperatures substantially below 200 K can be achieved in practice using multi-stage coolers with a heat sink at 300 K.

2.5 Application of the Thermomagnetic Effects

Although the transverse thermomagnetic effects have not yet found many practical applications, the Ettingshausen effect is potentially superior to the Peltier effect for refrigeration at low temperatures. The Nernst effect also offers some advantages over the Seebeck effect in the detection of thermal radiation.

There is a close correspondence between the equations that describe the cooling power and COP for Ettingshausen and Peltier coolers [5]. As we shall see, there is a figure of merit that can be used for transverse thermomagnetic energy conversion that is similar to the figure of merit Z that is used for thermocouples.

An Ettingshausen cooler might take the form of a rectangular bar, as shown in Fig. 2.9. A current is passed along the bar and a magnetic field is applied in a perpendicular direction. There is then a transverse flow of heat normal to both the current and the magnetic field. The heat source and sink are thermally, but not electrically, attached to the thermomagnetic material.

The equipotential surfaces near the centre of the bar will be inclined to the yz plane because of the Hall effect, but near the ends of the specimen these surfaces will normally lie in such a plane. We shall suppose that the specimen is much longer in the x direction than in the y direction and shall neglect the end effects. We realise, however, that the presence of end effects will always be a disadvantage of any

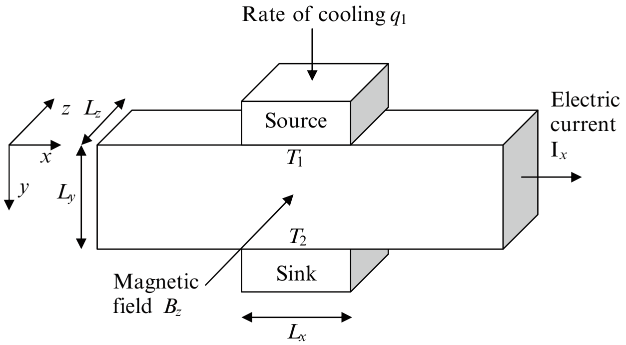


Fig. 2.9 Element of an Ettingshausen refrigerator. The heat source is at $y = 0$ and the heat sink at $y = L_y$

transverse cooling device. We shall assume the Nernst coefficient and the electrical and thermal conductivity to be independent of temperature. This is likely to be a reasonable assumption since the temperature differences in most thermomagnetic devices will probably be rather small.

We use the thermodynamic relation (1.9) to eliminate the Ettingshausen coefficient, P, in favour of the Nernst coefficient, N. Then, the current I_x in the magnetic field B_z gives rise to an Ettingshausen heat flow $NB_zI_xTL_x/L_y$. This heat flow will be opposed by thermal conduction at the rate $\lambda L_x L_z \, \mathrm{d}T/\mathrm{d}y$. Thus, for any particular value of y

$$q = \frac{NB_zI_xTL_x}{L_y} - \lambda L_x L_z \frac{\mathrm{d}T}{\mathrm{d}y}. \tag{2.23}$$

At the same time, there will be Joule heating in the bar that will disturb the linearity of the temperature gradient according to the relation

$$\frac{I_x^2 \rho L_x}{L_z L_y^2} = -\lambda L_x L_z \frac{\mathrm{d}^2 T}{\mathrm{d}y^2}, \tag{2.24}$$

since the heat generation per unit length is equal to $I_x^2 \rho L_x \,/\, (L_z L_y^2)$.

We now apply the boundary conditions that the temperature is equal to T_1 when $y = 0$ and T_2 when $y = L_x$. From (2.23) and (2.24) we find that

$$\lambda \frac{\mathrm{d}T}{\mathrm{d}y} = -\frac{I_x^2 \rho \left(y - \frac{1}{2} L_y\right)}{L_z^2 L_y^2} + \frac{(T_2 - T_1)\lambda}{L_y}, \tag{2.25}$$

and the cooling power at the source is

$$q_1 = \frac{NB_zI_xT_1L_x}{L_y} - \frac{\lambda L_x L_z (T_2 - T_1)}{L_y} - \frac{I_x^2 \rho L_x}{2L_z L_y}. \tag{2.26}$$

which is of the same form as (2.3) for a thermoelectric refrigerator.

Comparing (2.3) and (2.26), we find that the latter has NB_zL_x/L_y in place of $(\alpha_\mathrm{p}-\alpha_\mathrm{n})$, $\lambda L_xL_z/L_y$ in place of K, and $\rho L_x/(L_zL_y)$ in place of R. This means that we can make use of the equations for the COP that we derived for a thermoelectric refrigerator, if we make the appropriate substitutions, for an Ettingshausen cooler. The thermoelectric figure of merit, Z, is replaced by a thermomagnetic, or Nernst–Ettingshausen, figure of merit given by

$$Z_\mathrm{NE} = \frac{(NB_z)^2}{\rho\lambda}. \tag{2.27}$$

The differential Seebeck coefficient, $(\alpha_\mathrm{p} - \alpha_\mathrm{n})$, is replaced by NB_z, which has been called the thermomagnetic power [6], just as the Seebeck coefficient is sometimes known as the thermoelectric power.

At this point, we should draw attention to the importance of defining more precisely the thermomagnetic figure of merit and the quantities on which it depends. The question of definition has been discussed by Delves [7] and Horst [8]. Briefly, the quantity Z_NE, which is so closely related to the thermoelectric figure of merit, Z, should more properly be defined as the adiabatic thermomagnetic figure of merit as distinct from the isothermal thermomagnetic figure of merit Z_NE^i. Z_NE^i lies between the theoretical limits 0 and 1 whereas Z_NE has the limits 0 and ∞.

The isothermal Nernst coefficient is defined for zero electric current and zero transverse temperature gradient. The isothermal electrical resistivity requires there to be no transverse electric current and no temperature gradients. It is this resistivity that is used in the definition of the isothermal Nernst–Ettingshausen figure of merit. The adiabatic figure of merit Z_NE, on the other hand, is defined in terms of the adiabatic resistivity, which requires zero transverse temperature gradient and current and zero longitudinal heat flow.

The similarities between the formal equations for refrigeration using the thermoelectric and thermomagnetic effects do not mean that there are no significant differences between the two techniques. When one uses the thermomagnetic effects, there is only one material so the optimisation of the relative dimensions of two elements is no longer a requirement. However, a major difference lies in the separation of the directions of the electric current and the flow of heat. There are usually a large number of couples in a thermoelectric cooling module, if its cooling power is to be substantial, since it would be impractical to use a single couple with a large cross-sectional area. Such a couple would have to draw a very large current at a very small voltage. However, in a thermomagnetic cooler, a single bar of material might suffice. It would be possible to have a small cross-sectional area in the direction of the electric current flow and a large cross-sectional area in the direction of heat flow.

The advantage of the separation of the electric and thermal flows would certainly be apparent, if the Nernst effect were used in a broadband thermal detector. Thermocouples are often used as thermal detectors but it is difficult to achieve a rapid response time for such devices. The response time is proportional to the square of the thickness in the direction of the heat flow. However, if one is to decrease the length of the legs of a thermocouple, the output voltage for a given heat flow becomes smaller. In a thermomagnetic detector, the output voltage depends on the temperature gradient rather than the temperature difference. It is, therefore, possible to use a thin film of thermomagnetic material that combines both a rapid response and a high sensitivity [9, 10].

It is, particularly, simple to make a cascade based on the transverse thermomagnetic effects. For a given current, I, the cooling power is inversely proportional to the thickness in the y direction. Thus, instead of changing the number of elements from one stage to another, as in a thermoelectric cascade, one can use a number of bars of different thickness, as shown in Fig. 2.10. All the bars have the same length in the x direction and the same width in the direction of the magnetic field.

However, there is a method of obtaining an infinite-staged cascade using a single piece of thermomagnetic material. The sample is shaped so that it is much wider at the heat sink than it is at the source, as shown in Fig. 2.11. The potential difference, V, between the ends remains the same for all values of y.

Consider the section of thickness Δy that is bounded by the broken lines in Fig. 2.11. If we regard this section as one of the stages in the cascade its optimum COP is

$$\phi_y = \frac{T}{\Delta T}\frac{(1 + Z_{\mathrm{NE}}T)^{1/2} - 1}{(1 + Z_{\mathrm{NE}}T)^{1/2} + 1}. \tag{2.28}$$

Then, the ratio of the heat leaving the stage at $y + \Delta y$ to that entering at y is

$$\frac{q_y + \Delta y}{q_y} = 1 + \frac{\Delta T}{T}\frac{(1 + Z_{\mathrm{NE}}T)^{1/2} + 1}{(1 + Z_{\mathrm{NE}}T)^{1/2} - 1}. \tag{2.29}$$

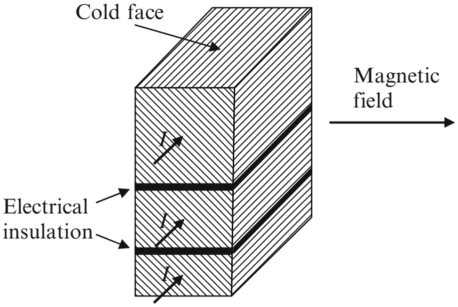


Fig. 2.10 Three-stage Ettinghausen cooler. The same current, I, passes through each stage but the stages are of different thickness

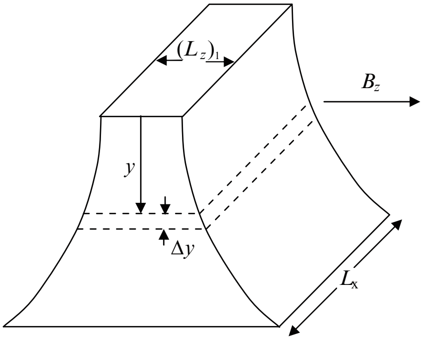


Fig. 2.11 Exponentially shaped Ettingshausen cascade. Heat is extracted from the source at the upper face

This is the ratio of the width in the z direction at $y + \Delta y$ to that at y, that is $(L_z + \Delta L_z)/L_z$.

If we denote $(1 + Z_{\mathrm{NE}}T)^{1/2}$ by M_{NE}, we may write

$$\frac{\Delta L_z}{L_z} = \frac{M_{\mathrm{NE}} + 1}{M_{\mathrm{NE}} - 1}\frac{\Delta T}{T}. \tag{2.30}$$

This leads to an integral equation which we solve after making the assumption that $[(M_{\mathrm{NE}} + 1)/\{T(M_{\mathrm{NE}} - 1)\}(\mathrm{d}T/\mathrm{d}y)]$ is approximately constant. The solution is

$$L_z = (L_z)_1 \exp\left(\frac{M_{\mathrm{NE}} + 1}{T(M_{\mathrm{NE}} - 1)} y \frac{\mathrm{d}T}{\mathrm{d}y}\right). \tag{2.31}$$

Equation (2.31) shows us that we can make an infinite-staged cascade by using a thermomagnetic element of exponential shape.

It may not be easy to cut a sample of material to the optimum shape but it has been shown [11] that even a simple trapezoidal section yields a substantial improvement in performance over that of a rectangular bar.

References

1. K. Landecker, in *Proceedings of the 1st International Conference on Thermoelectric Energy Conversion, Arlington, Texas* (IEEE, New York, 1976), p. 150
2. H.J. Goldsmid, *Electronic Refrigeration* (Pion, London, 1986), p. 7
3. A.F. Ioffe, *Semiconductor Thermoelements and Thermoelectric Cooling* (Infosearch, London, 1957), p. 36
4. T.P. Hogan, T. Smith, in *Thermoelectrics Handbook: Macro to Nano*, ed. by D.M. Rowe (CRC Taylor and Francis, Boca Raton, 2006), p. 12–5
5. B.J. O'Brien, C.S. Wallace, J. Appl. Phys. **29**, 1010 (1958)
6. J.L. Weaver, K.F. Cuff, G.M. Enslow, S.R. Hawkins, R.B. Horst, C.F. Kooi, Bull. Am. Phys. Soc. **7**, 495 (1962)
7. R.T. Delves, Br. J. Appl. Phys. **15**, 105 (1964)
8. R.B. Horst, J. Appl. Phys. **34**, 3246 (1963)
9. B. Paul, H. Weiss, Solid State Electron. **11**, 979 (1968)
10. E.R. Washwell, S.R. Hawkins, K.F. Cuff, Appl. Phys. Lett. **17**, 164 (1970)
11. R.B. Horst, L.R. Williams, in *Proceedings of the 3rd International Conference on Thermoelectric Energy Conversion, Arlington, Texas* (IEEE, New York, 1980), p. 183

Chapter 3
Thermoelectric Properties of Metals and Semiconductors

3.1 Transport by Electrons

Thermoelements are usually made from crystalline solids. They are not commonly single crystals, but their crystalline nature can be observed under a microscope. Sometimes, the transport properties vary with crystalline orientation but, until we deal with specific examples, we shall suppose that the properties are uniform in all directions.

The transport of electric charge is due to quasi-free electrons in the solid. The solids of interest to us are metals and semiconductors. In such materials, the electrons carry not only the charge but also the thermal energy. In other words, there is an electronic component of the thermal conductivity. As we shall see later, heat can also be carried by the thermal vibrations of the atoms in a crystal but, for the moment, we confine ourselves to the electronic effects.

The idea of conduction by electrons was proposed by Drude and Lorentz using the principles of classical physics. The classical free electron theory predicted that the specific heat should be much larger for a metal than for an electrical insulator but, in reality, there is very little difference. This discrepancy disappeared when Sommerfeld [1] took account of the newly developed quantum theory but neither the classical nor the quantum mechanical-free electron theories were able to explain why some solids are metallic conductors and others are insulators. It was only when notice was taken of the interaction of the electrons with the periodic potential that exists in a crystal lattice that further progress could be made. It was shown that, through this interaction, the energy of the electrons must lie in discrete bands that are separated by forbidden regions or energy gaps. Interestingly, Sommerfeld's theory can still be applied to the current-carrying electrons, if they are assigned an effective mass rather than the mass of a free electron.

According to quantum theory, the probability that an electron state of energy, E, will be occupied is given by the Fermi distribution function

$$f_0(E) = \left[\exp\left(\frac{E - E_F}{kT}\right) + 1\right]^{-1}. \tag{3.1}$$

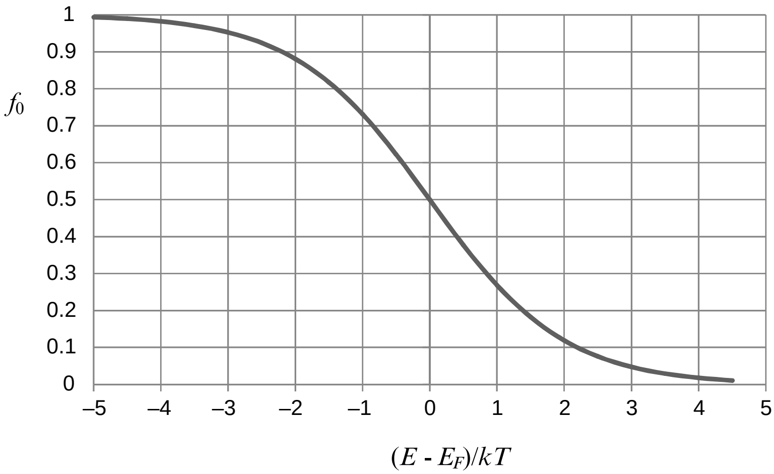


Fig. 3.1 Plot of Fermi distribution function against $(\mathrm{E}-\mathrm{E_F})/\mathrm{kT}$

The quantity E_F has a value that depends on the total number of electrons that have to be accommodated, and k is Boltzmann's constant. The Fermi distribution function has the property that it is equal to zero when $(E - E_F) \gg kT$ and equal to unity when $(E - E_F) \ll kT$. The transition from zero to unity takes place over a very narrow range of energy as shown in Fig. 3.1. The energy, E_F, at which the Fermi distribution function is equal to ½ is known as the Fermi level. If the number of permitted electron states in the energy range between E and $E + \mathrm{d}E$ is represented by $g(E)\mathrm{d}E$, the total number of electrons is

$$n = \int_0^\infty f_0(E)g(E)dE. \tag{3.2}$$

Each electron state is defined not only by its energy but also by its momentum or, more strictly, by its wave vector. Thus, although there may be gaps between the allowed bands of energy at a given position in wave-vector space, the bands might still overlap in a simple energy band diagram. Such a diagram is shown schematically in Fig. 3.2 for the situation where this complication does not arise. It must be understood that the energy bands shown in this diagram are only those in the neighbourhood of the Fermi level. There are also completely filled lower bands that do not contribute to the transport processes.

It must be appreciated that electrical conduction can take place only when the electrons in a band can move from one energy state to another. This, of course, cannot happen in an empty band for there are then no electrons at all. It also cannot happen when a band is completely full, since there are then no free states into which an electron can move. Conduction is, in fact, due entirely to those electrons whose energy is such that the states are partially filled.

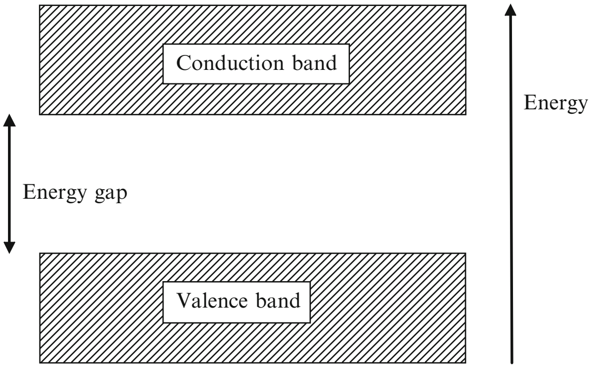


Fig. 3.2 Simple energy band diagram

The density of electron states, $g(E)$, is small near the conduction band edge but rises rapidly within the band. Thus, if the Fermi level lies well within the conduction band, there will be a large number of electrons located near vacant states and the solid will be highly conducting; that is, it will be a metal. On the other hand, if the Fermi level lies well within the energy gap, there will be virtually no electrons in the conduction band and the material will be an electrical insulator. When the Fermi level is close to the conduction band edge, there will be only a few electrons but they will all be able to contribute to the conduction process. Because of their relatively small number, the conductivity will not be large and the material is called a semiconductor. The density of electron states near the band edge is given by the expression

$$g\,(E)\,\mathrm{d}E = \frac{4\pi\,(2m^*)^{3/2}\,\mathrm{d}E}{h^3}. \tag{3.3}$$

This is the same density of states that is to be found in Sommerfeld's theory with the exception that the free electron mass m is replaced by an effective mass m^*.

Interesting effects occur when the Fermi level is close to the edge of the valence band. Then there will be some empty states in this band allowing conduction to take place. It turns out that, although the process is still due to negatively charged electrons, the effective mass is, now, negative. The behaviour can best be described in terms of what are called positive holes, which have a positive effective mass. Equation (3.3) can still be used for the positive holes.

A glance at Fig. 3.1 reveals that it is possible for the Fermi function to be significantly greater than zero at the edge of the conduction band and, at the same time, be significantly less than unity at the edge of the valence band. The Fermi level itself would be somewhere near the middle of the forbidden gap. In this case, there would be simultaneous conduction by electrons and holes. If the energy gap is small enough for this to occur, the material is what is known as an intrinsic semiconductor. Conduction by either electrons in the conduction band or holes in the valence band can be induced by adding donor or acceptor impurities to a semiconductor. In such extrinsic conduction, the Fermi level is close to the edge of either the conduc-

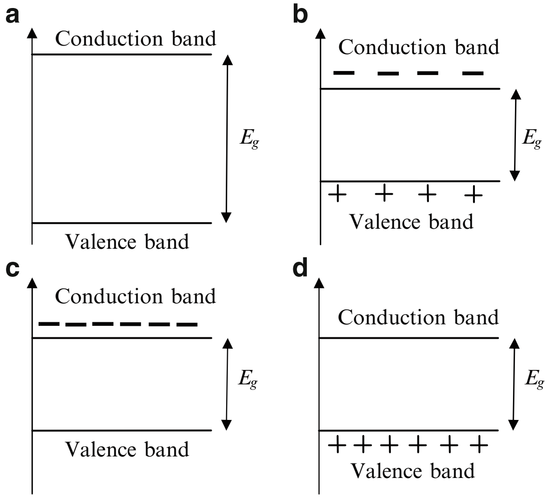


Fig. 3.3 Energy band diagrams for (**a**) an insulator, (**b**) an intrinsic semiconductor, (**c**) an n-type semiconductor, and (**d**) a p-type semiconductor

tion band or valence band. Figure 3.3 shows energy diagrams for an insulator, an intrinsic semiconductor, and for n-type and p-type extrinsic semiconductors.

We shall discuss the transport effects in a semiconductor with specific reference to electrons. The same equations will apply for hole conduction but, when dealing with holes, we must measure the energy downwards from the Fermi level. The subscripts "n" and "p" will be used for the electrons and holes respectively.

We shall make use of the concept of a relaxation time, τ_e, for the charge carriers. Then, if the distribution function, f, is disturbed from its equilibrium value, f_0, it will relax towards f_0 according to

$$\frac{\mathrm{d}f\,(E)}{\mathrm{d}t} = -\frac{f\,(E) - f_0\,(E)}{\tau_\mathrm{e}}. \tag{3.4}$$

As an approximation, we shall assume that the relaxation time can be expressed in the form $\tau_0 E^r$ where τ_0 and r are constants for a given scattering process.

In many potential thermoelectric materials, it seems that the predominant scattering of the charge carriers is due to the acoustic-mode lattice vibrations, in which case the parameter r is equal to $-1/2$. Also, for scattering by ionised impurities, r is equal to 3/2. When more than one scattering process is operative, the properties can sometimes be determined by interpolation. Generally speaking, reciprocal relaxation times are additive but, hopefully, one process may outweigh the others for any particular electron energy.

Our theory is based on the Boltzmann equation that relates the effects of the applied fields and the scattering of the carriers. If we suppose that the disturbance to the distribution is relatively small we find that

$$\frac{f\,(E) - f_0\,(E)}{\tau_e} = u\frac{\mathrm{d}f_0\,(E)}{dE}\left(\frac{\mathrm{d}E_F}{\mathrm{d}x} + \frac{(E - E_F)}{T}\frac{\mathrm{d}T}{\mathrm{d}x}\right). \tag{3.5}$$

Here, u is the velocity of the carriers in the x direction and E_F is the Fermi energy. The two terms in the brackets are associated with the electric field and the temperature gradient, respectively.

We may use (3.5) to obtain the electric current density, i, and the heat flux density, j. The equation for the electric current density is

$$i = \mp \int_0^\infty euf\,(E)\,g\,(E)\,\mathrm{d}E, \tag{3.6}$$

where e is the magnitude of the electronic charge. The upper and lower signs apply for electrons and holes respectively. The heat flux density is

$$j = \int_0^\infty u\,(E - E_F)\,f\,(E)\,g\,(E)\,\mathrm{d}E, \tag{3.7}$$

where $(E - E_F)$ is the energy transported by each carrier. The upper limit of the integrals in (3.4) and (3.5) has been arbitrarily set at infinity but, in fact, this is unimportant since $f(E)$ becomes zero before E becomes at all large.

In using these equations to determine the transport coefficients, we can replace f by $(f - f_0)$ since there is no flow of any kind when $f = f_0$. Also, since the drift velocity of the carriers is a small part of the total velocity, it is satisfactory to replace u by $2E/3\,m^*$. This enables us to write (3.6) and (3.7) in the forms

$$i = \mp \frac{2e}{3m^*} \int_0^\infty g\,(E)\,\tau_e E \frac{\mathrm{d}f_0\,(E)}{\mathrm{d}E} \left\{ \frac{\mathrm{d}E_F}{\mathrm{d}x} + \frac{(E - E_F)}{T} \frac{\mathrm{d}T}{\mathrm{d}x} \right\} \mathrm{d}E, \tag{3.8}$$

$$j = \pm \frac{E_F}{e} i + \frac{2}{3m^*} \int_0^\infty g\,(E)\,\tau_e E^2 \frac{\mathrm{d}f_0\,(E)}{\mathrm{d}E} \left\{ \frac{\mathrm{d}E_F}{\mathrm{d}x} + \frac{(E - E_F)}{T} \frac{\mathrm{d}T}{\mathrm{d}x} \right\} \mathrm{d}E. \tag{3.9}$$

In order to determine the transport parameters, we must insert the appropriate boundary conditions. Thus, the electrical conductivity is given by the ratio of i to the electric field when the temperature gradient $\mathrm{d}T/\mathrm{d}x$ is zero. The electronic contribution, λ_e, to the thermal conductivity is equal to the ratio of j to $-\mathrm{d}T/\mathrm{d}x$ when the electric current is zero. Also, the Seebeck coefficient is equal to the ratio of the electric field to the temperature gradient under the same condition. Thence, the three quantities that appear (with the lattice conductivity) in the thermoelectric figure of merit are

$$\sigma = \frac{1}{\rho} = -\frac{2e^2}{3m^*} \int_0^\infty g\,(E)\,\tau_e \frac{\mathrm{d}f_0\,(E)}{\mathrm{d}E} \mathrm{d}E, \tag{3.10}$$

$$\begin{aligned} \lambda_\mathrm{e} = \frac{2}{3m^* T} \langle & \left\{ \left[\int_0^\infty g\,(E)\,\tau_e E^2 \frac{\mathrm{d}f_0\,(E)}{\mathrm{d}E} \mathrm{d}E \right]^2 \Big/ \int_0^\infty g\,(E)\,\tau_e \frac{\mathrm{d}f_0\,(E)}{\mathrm{d}E} \mathrm{d}E \right\} \\ & - \int_0^\infty g\,(E)\,\tau_e E^3 \frac{\mathrm{d}f_0\,(E)}{\mathrm{d}E} \mathrm{d}E \rangle \end{aligned} \tag{3.11}$$

and

$$\alpha = \pm\frac{1}{eT}\left[E_F - \int_0^\infty g(E)\,\tau_e E^2 \frac{\mathrm{d}f_0(E)}{\mathrm{d}E}\mathrm{d}E \Big/ \int_0^\infty g(E)\,\tau_e E \frac{\mathrm{d}f_0(E)}{\mathrm{d}E}\mathrm{d}E\right]. \quad (3.12)$$

It is convenient to express the integrals that are included in (3.10)–(3.12) in the form

$$K_\mathrm{s} = -\frac{2T}{3m^*}\int_0^\infty g(E)\,\tau_e E^{s+1}\frac{\mathrm{d}f_0(E)}{\mathrm{d}E}\mathrm{d}E. \quad (3.13)$$

One can then eliminate g and τ_e in favour of m^*, r, and τ_0. Eventually, one then finds that

$$K_\mathrm{s} = \frac{8\pi}{3}\left(\frac{2}{h^2}\right)^{3/2}\left(m^*\right)^{1/2} T\tau_0\,(s + r + 3/2)\,(kT)^{s+r+3/2}\,F_{s+r+1/2}, \quad (3.14)$$

where

$$F_n(\xi) = \int_0^\infty \xi^n f_0(\xi)\,d\xi. \quad (3.15)$$

Here, the reduced energy, ξ, has been used as a variable in place of E/kT. The derivation of (3.14) is given more fully elsewhere [2]. The functions F_n are known as the Fermi-Dirac integrals and, for convenience, are shown in Tables 3.1 and 3.2 for different values of the reduced Fermi energy, η, which is equal to E_F/kT.

The expressions for the transport coefficients in terms of the integrals K_s are

$$\sigma = \frac{e^2}{T}K_1, \quad (3.16)$$

$$\lambda_\mathrm{e} = \frac{1}{T^2}\left(K_2 - \frac{K_1^2}{K_0}\right), \quad (3.17)$$

and

$$\alpha = \pm\frac{1}{eT}\left(E_F - \frac{K_1}{K_0}\right). \quad (3.18)$$

Equations (3.16)–(3.18) allow us to relate the thermoelectric figure of merit to the scattering parameters, the effective mass of the charge carriers, and the Fermi energy. It is also necessary to express the total thermal conductivity, λ, as the sum of the electronic component, given by (3.17) and the lattice component, λ_L. Thus,

$$\lambda = \lambda_\mathrm{e} + \lambda_\mathrm{L}. \quad (3.19)$$

Table 3.1 Fermi-Dirac integrals, F_n, as defined by (3.15), for $n = -1/2$, 0, 1/2, 1, and 3/2. η is the reduced Fermi energy and is equal to E_F/kT

η	$F_{-1/2}$	F_0	$F_{1/2}$	F_1	$F_{3/2}$
−2.0	0.21919	0.12693	0.11459	0.13101	0.1758
−1.8	0.26278	0.15298	0.13863	0.15893	0.21367
−1.6	0.31393	0.1839	0.1674	0.19253	0.25945
−1.4	0.37352	0.22042	0.2017	0.23286	0.31467
−1.2	0.44235	0.26328	0.24241	0.28112	0.38111
−1.0	0.52114	0.31326	0.2905	0.33865	0.46085
−0.8	0.61308	0.3711	0.34699	0.40695	0.55625
−0.6	0.71033	0.43749	0.41294	0.48766	0.66999
−0.4	0.82094	0.51302	0.48941	0.58255	0.80506
−0.2	0.94179	0.59814	0.57747	0.6935	0.96479
0.0	1.07213	0.69315	0.67809	0.82247	1.1528
0.2	1.21086	0.79814	0.79218	0.97143	1.373
0.4	1.35662	0.91302	0.92051	1.14238	1.62954
0.6	1.50787	1.03749	1.06309	1.33727	1.92679
0.8	1.66299	1.1711	1.22221	1.55798	2.26928
1.0	1.82037	1.31326	1.39637	1.80628	2.66167
1.2	1.97851	1.46328	1.58632	2.0838	3.10867
1.4	2.13609	1.62041	1.79206	2.39205	3.61502
1.6	2.29197	1.7839	2.01348	2.73236	4.18544
1.8	2.44526	1.95297	2.25036	3.10594	4.82462
2.0	2.59528	2.12692	2.50241	3.51383	5.53714
2.2	2.74154	2.30507	2.76928	3.95693	6.32752
2.4	2.88374	2.48681	3.05058	4.43604	7.20015
2.6	3.0217	2.67161	3.34589	4.95181	8.15931
2.8	3.15539	2.859	3.65479	5.50483	9.20915
3.0	3.28485	3.04855	3.97687	6.09556	10.3573
3.2	3.41017	3.23992	4.3117	6.72441	11.59684
3.4	3.5315	3.4328	4.65888	7.39172	12.9423
3.6	3.64903	3.62693	5.01803	8.09775	14.39367
3.8	3.76293	3.82211	5.38877	8.8427	15.95437
4.0	3.87341	4.01815	5.77074	9.62671	17.62761

3.2 Metals and Semiconductors

In the general case, we have to evaluate the transport parameters using tabulated values for the Fermi-Dirac integrals such as those in Tables 3.1 and 3.2. However, these integrals can be obtained as simple approximations when the Fermi energy, as measured from the band edge, is either very much greater than *kT* or very much less than –*kT*.

When $E_F \gg kT$, the conductor is a metal and the degenerate approximation may be employed. This approximation is usually regarded as acceptable when

Table 3.2 Fermi-Dirac integrals, F_n, as defined by (3.15), for $n = 2, 5/2, 3, 7/2$ and 4. η is the reduced Fermi energy and is equal to E_F/kT

η	F_2	$F_{5/2}$	F_3	$F_{7/2}$	F_4
−2.0	0.26627	0.44455	0.8053	1.565	3.2345
−1.8	0.32408	0.54162	0.9819	1.909	3.9471
−1.6	0.39416	0.65954	1.1967	2.3281	4.8157
−1.4	0.479	0.80264	1.4578	2.8383	5.8741
−1.2	0.58151	0.97608	1.775	3.4589	7.1631
−1.0	0.70513	1.18597	2.1598	4.2133	8.7321
−0.8	0.85386	1.43954	2.6262	5.1294	10.6406
−0.6	1.03234	1.74527	3.1904	6.2408	12.9601
−0.4	1.24588	2.11308	3.872	7.5873	15.7765
−0.2	1.50052	2.55444	4.6937	9.2162	19.1926
0.0	1.80309	3.08259	5.6822	11.1837	23.3309
0.2	2.16116	3.71261	6.8685	13.5556	28.3369
0.4	2.58316	4.46164	8.2884	16.4092	34.383
0.6	3.07826	5.34894	9.983	19.8344	41.6723
0.8	3.65642	6.396	11.9991	23.9356	50.4432
1.0	4.32833	7.62661	14.3898	28.8329	60.9745
1.2	5.10535	9.06691	17.2148	34.6643	73.5903
1.4	5.99949	10.74543	20.5409	41.5868	88.6657
1.6	7.02331	12.69308	24.4423	49.7786	106.6324
1.8	8.1899	14.94311	29.0009	59.4401	127.9849
2.0	9.5128	17.53113	34.3067	70.7962	153.2857
2.2	11.00594	20.49501	40.4578	84.097	183.1721
2.4	12.68359	23.87484	47.5608	99.6198	218.3621
2.6	14.56031	27.71286	55.7307	117.6701	259.6599
2.8	16.65089	32.05334	65.0912	138.5827	307.9621
3.0	18.97031	36.9425	75.7744	162.7227	364.2625
3.2	21.53368	42.42834	87.9208	190.486	429.6565
3.4	24.35616	48.56049	101.6791	222.2993	505.344
3.6	27.45291	55.38993	117.2055	258.6194	592.6296
3.8	30.83899	62.06875	134.6629	299.9315	692.9207
4.0	34.52922	71.34963	154.2198	346.7463	807.7205

$E_F > 4kT$. It is found that the Fermi-Dirac integrals for a degenerate conductor may be expressed as

$$F_n(\eta) = \frac{\eta^{n+1}}{n+1} + n\eta^{n-1}\frac{\pi^2}{6} + n(n-1)(n-2)\eta^{n-3}\frac{7\pi^4}{360} + \ldots \tag{3.20}$$

The series converges rapidly and one only needs to use as many terms as yield a non-zero value for the parameter in question.

The evaluation of the electrical conductivity requires only the first term in the series. Thus,

$$\sigma = \frac{8\pi}{3}\left(\frac{2}{h^2}\right)^{3/2} e^2 \left(m^*\right)^{1/2} \tau_0 E_F^{r+3/2}. \tag{3.21}$$

To determine the electronic thermal conductivity, the first two terms in the series are needed. We make use of the quantity L, known as the Lorenz number, which is defined as the ratio $\lambda_e/\sigma T$, and is given by

$$L = \frac{\pi^2}{3}\left(\frac{k}{e}\right)^2 . \tag{3.22}$$

In a metal, the electronic thermal conductivity is usually much larger than the lattice component. Thus, (3.22) is consistent with the Wiedemann–Franz law, which states that the ratio of the thermal to the electrical conductivity is the same for all metals at any particular temperature.

The Seebeck coefficient also requires the first two terms of the series in (3.20). It is found to be

$$\alpha = \mp \frac{\pi^2}{3}\frac{k}{e}\frac{(r + 3/2)}{\eta} . \tag{3.23}$$

As η becomes large, the Seebeck coefficient becomes very much less than k/e. This is borne out in practice as most metals have Seebeck coefficients of the order of only a few microvolts per degree.

We turn now to the classical or non-degenerate approximation that applies to most semiconductors when they do not contain too many impurities. This approximation is usually acceptable when $\eta < -2kT$. Under this condition, the Fermi-Dirac integrals may be written as

$$F_n(\eta) = \exp(\eta) \int_0^\infty \xi^n \exp(-\xi)\, d\xi = \exp(\eta)\, \Gamma(n+1) , \tag{3.24}$$

where the gamma function is such that

$$\Gamma(n+1) = n\Gamma(n) . \tag{3.25}$$

When n is an integer, $\Gamma(n+1)$ is equal to $n!$. Also, one can calculate the gamma function for half-integral values of n using (3.25) and the relation $\Gamma(1/2) = \pi^{1/2}$.

Under the non-degenerate condition, then, one can express the Fermi-Dirac integrals as

$$K_s = \frac{8\pi}{3}\left(\frac{2}{h^3}\right)^{3/2} (m^*)^{1/2}\, T\tau_0\, (kT)^{s+r+3/2}\, \Gamma(s + r + 5/2) \exp(\eta) . \tag{3.26}$$

The electrical conductivity of a non-degenerate conductor is

$$\sigma = \frac{8\pi}{3}\left(\frac{2}{h^3}\right)^{3/2} e^2 (m^*)^{1/2}\, T\tau_0\, (kT)^{r+3/2}\, \Gamma\left(r + 5/2\right) \exp(\eta) . \tag{3.27}$$

It is convenient to express the electrical conductivity as

$$\sigma = ne\mu, \tag{3.28}$$

where n is the carrier concentration and μ is the carrier mobility, which does not depend on the Fermi energy under classical conditions. n and μ are given by

$$n = 2\left(\frac{2\pi m^* kT}{h^2}\right)^{3/2} \exp(\eta), \tag{3.29}$$

and

$$\mu = \frac{4}{3\pi^{1/2}}\Gamma\left(r + \frac{5}{2}\right)\frac{e\tau_0 (kT)^r}{m^*}. \tag{3.30}$$

Equation (3.29) shows that the carrier concentration is that which would be found if there were $2(2\pi m^* kT/h^2)^{3/2}$ energy levels located at the band edge, so this quantity is known as the effective density of states.

The Seebeck coefficient of a non-degenerate conductor is related to the reduced Fermi energy η through the relation

$$\alpha = \mp\frac{k}{e}\left[\eta - \left(r + \frac{5}{2}\right)\right]. \tag{3.31}$$

Remembering that the Peltier coefficient is equal to αT, we see that the energy carried by each electron or hole is made up of a potential component $-\eta kT/e$ and a kinetic component equal to $(r + 5/2)kT/e$. In a typical extrinsic semiconductor, $\pm\eta$ may be much greater than unity and the magnitude of the Seebeck coefficient may be several hundred microvolts per degree.

Although the Seebeck coefficient is very much different in semiconductors and metals, the Lorenz numbers usually differ by a factor of less than 2. For a non-degenerate conductor, the Lorenz number is

$$L = \left(\frac{k}{e}\right)^2\left(r + \frac{5}{2}\right). \tag{3.32}$$

It is independent of the Fermi energy throughout the non-degenerate range. Although the Lorenz number is so close to that for a metal, semiconductors do not usually satisfy the Wiedemann–Franz law. This is because the electronic component of the thermal conductivity is so much smaller and the lattice component is, therefore, dominant.

If the Wiedemann–Franz law applied for both semiconductors and metals, there would be no doubt that the former class of material should be used in thermoelectric energy conversion in view of the larger magnitude of the Seebeck coefficient. However, as we shall see, the increase in the thermoelectric coefficients must be balanced against a decrease in the ratio of the electrical to thermal conductivity. It turns out that most of the materials of interest for practical applications are semiconductors in which the Fermi energy lies close to the edge of the conduction or valence band. That means that neither the degenerate nor the non-degenerate approximations are applicable and we must use the tabulated Fermi-Dirac integrals given in Tables 3.1 and 3.2. In Figs. 3.4–3.6, we show how the transport properties of interest vary with the reduced Fermi energy for the extreme values of r that are likely to be encountered, namely $-1/2$ and $3/2$.

In Fig. 3.4, the electrical conductivity is given as a ratio of its value at a given reduced Fermi energy, η, to its value when the Fermi level lies at the band edge. Because of the rapid variation of the conductivity with Fermi energy, a logarithmic

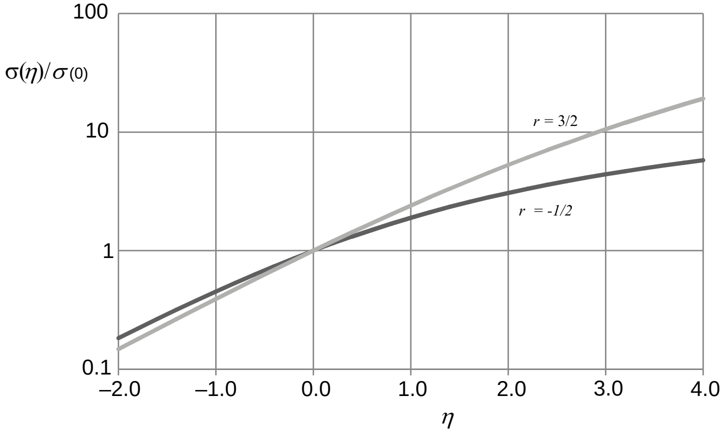


Fig. 3.4 Plot of electrical conductivity against reduced Fermi energy for r equal to $-1/2$ and 3/2. The electrical conductivity is given as a fraction of its value at $\eta = 0$ $\sigma(\eta)/\sigma(0)$

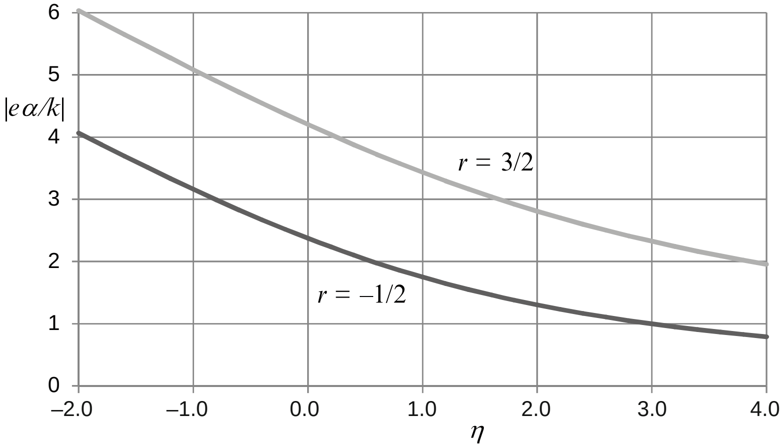


Fig. 3.5 Plot of Seebeck coefficient in units of k/e against reduced Fermi energy $r = 3/2$

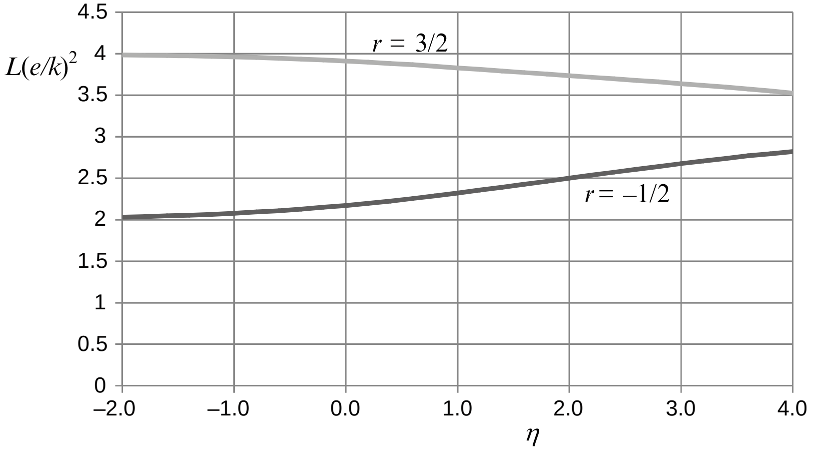


Fig. 3.6 Plot of Lorenz number in units of $(k/e)^2$ against reduced Fermi energy

scale is used. It is noted that the value of the scattering parameter, τ_0, has been supposed to be independent of η. While this is a reasonable assumption for lattice scattering of the carriers ($r = -1/2$), it is unlikely to be appropriate for ionised-impurity scattering ($r = 3/2$). The decrease in the relaxation time as the impurity concentration rises should be taken into account, if practical use is to be made of Fig. 3.4.

When the Seebeck coefficient is plotted against η, as in Fig. 3.5, a linear scale can be used since the variation is relatively small. The Seebeck coefficient will be negative for an n-type semiconductor in which the Fermi energy is measured upwards from the edge of the conduction band. It will be positive for a p-type semiconductor, the Fermi energy then being measured downwards from the edge of the valence band edge.

As shown in Fig. 3.6, the Lorenz number is only weakly dependent on the Fermi energy. It becomes smaller than the metallic value when $r = -1/2$ and larger than this value when $r = 3/2$.

3.3 Bipolar Effects

We, now, consider a conductor in which there are both electrons and positive holes. Such a material would be a wide-gap semiconductor at elevated temperatures or a narrow-gap semiconductor or semimetal at ordinary temperatures. When both types of carrier are present, it is possible for them to move in the same direction, transporting energy without an electric current. This leads to interesting effects that have a bearing on thermoelectric performance.

Let us denote the separate current densities for the electrons and holes as i_{n} and i_{p}. The expression for i_{n} or i_{p} when there is a temperature gradient as well as an electric field is

$$i_{\mathrm{n,p}} = \sigma_{\mathrm{n,p}} \left(\frac{\mathrm{d}V}{\mathrm{d}x} - \alpha_{\mathrm{n,p}} \frac{\mathrm{d}T}{\mathrm{d}x} \right). \tag{3.33}$$

The electrical conductivity is obtained by setting the temperature gradient equal to zero. Thus,

$$\sigma = \frac{i_{\mathrm{n}} + i_{\mathrm{p}}}{\mathrm{d}V/\mathrm{d}x} = \sigma_{\mathrm{n}} + \sigma_{\mathrm{p}}. \tag{3.34}$$

It is, of course, not surprising that the electrical conductivity is the sum of the conductivities of the separate carriers. Likewise, by setting the electric current equal to zero, we find that the Seebeck coefficient is a weighted average of the Seebeck coefficients associated with the two types of carrier. Thus,

$$\alpha = \frac{\mathrm{d}V/\mathrm{d}x}{\mathrm{d}T/\mathrm{d}x} = \frac{\alpha_{\mathrm{n}}\sigma_{\mathrm{n}} + \alpha_{\mathrm{p}}\sigma_{\mathrm{p}}}{\sigma_{\mathrm{n}} + \sigma_{\mathrm{p}}}. \tag{3.35}$$

It must be remembered that the Seebeck coefficients for the two carriers are of opposite sign, so the magnitude of α given by (3.35) can be quite small.

Let us now consider the flow of heat in the two-carrier system. The heat flux densities for the two types of carrier are found from

$$j_{\mathrm{n,p}} = \alpha_{\mathrm{n,p}} T i_{\mathrm{n,p}} - \lambda_{\mathrm{n,p}} \frac{\mathrm{d}T}{\mathrm{d}x}. \tag{3.36}$$

Also, the thermal conductivity is defined for the condition of zero electric current. Thence,

$$i_{\mathrm{n}} = -i_{\mathrm{p}} = \frac{\sigma_{\mathrm{n}}\sigma_{\mathrm{p}}}{\sigma_{\mathrm{n}} + \sigma_{\mathrm{p}}} \left(\alpha_{\mathrm{n}} - \alpha_{\mathrm{p}} \right) \frac{\mathrm{d}T}{\mathrm{d}x}. \tag{3.37}$$

By combining (3.36) and (3.37) to find the total heat flux density and dividing the result by the temperature gradient, we find that

$$\lambda_e = \lambda_{\mathrm{n}} + \lambda_{\mathrm{p}} + \frac{\sigma_{\mathrm{n}}\sigma_{\mathrm{p}}}{\sigma_{\mathrm{n}} + \sigma_{\mathrm{p}}} \left(\alpha_{\mathrm{n}} - \alpha_{\mathrm{p}} \right)^2 T. \tag{3.38}$$

The remarkable feature of (3.38) is that the total electronic thermal conductivity is not merely the sum of the thermal conductivities of the separate carriers. There is an additional term associated with the bipolar flow [3]. This additional term can be much larger than either λ_{n} or λ_{p}. Although the relative effect of the bipolar contribution is greatest for an intrinsic wide-gap semiconductor, the phenomenon is most easily observed for a narrow-gap material since there are then many more carriers of both types. Thus, it was first observed for intrinsic bismuth telluride [4], which has an energy gap at ordinary temperatures of about $6kT$. The Lorenz number was found to be about $25(k/e)^2$ whereas, when there is only a single type of carrier, L is no more than about $2(k/e)^2$.

3.4 Phonon Conduction

We have already discussed the conduction of heat by the charge carriers. Now, we discuss the heat conducted by the lattice vibrations. In any solid, each atom is bonded to its neighbours and, thus, even displacement of any one atom will give rise to a disturbance that is passed on to the rest of the specimen. The atoms are in continual vibration and the overall motion can be represented by waves, which may be either longitudinal or transverse in their nature. At low frequencies, the vibrations are familiar to us as sound waves but, in considering heat conduction, we are more interested in much higher frequencies.

Before proceeding further, we need to discuss the nature of the vibrational spectrum. This problem was first tackled by Debye [5] who supposed that a crystal could be represented by an elastic continuum. He showed that the boundary conditions allowed only certain wavelengths to occur. A lower limit to the permissible wavelengths is set by the atomic nature of matter. The total number of vibrational modes is equal to $3\,N$, where N is the number of atoms. According to Debye's theory, the number of modes per unit volume that have frequencies between ν and $\nu + \mathrm{d}\nu$ is

$$n_\mathrm{L} = \frac{2\pi \nu^3 \mathrm{d}\nu}{\mathrm{v}^3}, \tag{3.39}$$

where v is the speed of sound. In order that the total number of modes be equal to $3N$

$$\frac{4\pi \nu_\mathrm{D}^3}{\mathrm{v}^3} = 3N. \tag{3.40}$$

Debye used the newly developed quantum theory to determine the energy, W, in a mode of frequency ν. Quantised vibrations satisfy Bose–Einstein statistics rather than the Fermi-Dirac statistics that apply to the charge carriers. The expression for W is then

$$W = h\nu \left[\exp\left(\frac{h\nu}{kT}\right) - 1\right]^{-1}. \tag{3.41}$$

The specific heat at constant volume, c_V, is found by differentiating the integrated internal energy with respect to temperature with the result that

$$c_V = 9Nk \left(\frac{T}{\Theta_\mathrm{D}}\right)^3 f_D \left(\frac{\Theta_\mathrm{D}}{T}\right), \tag{3.42}$$

where Θ_D, which is known as the Debye temperature, is defined as

$$\Theta_\mathrm{D} = \frac{h\nu_\mathrm{D}}{k}, \tag{3.43}$$

and

$$f_D\left(\frac{\Theta_D}{T}\right) = \int_0^{\Theta_D/T} \frac{x^4 \exp(x)}{(\exp(x) - 1)^2} \mathrm{d}x. \tag{3.44}$$

The general behaviour of real solids is quite close to that predicted by Debye in spite of the very elementary model that he used. It turns out that the specific heat is not, particularly, sensitive to the details of the vibrational spectrum. The fact that Debye's specific heat theory was so successful has allowed the use of the concept of a Debye temperature to remain to this day. Discrepancies between the experimental observations and the theoretical curve are accounted for by allowing Θ_D to be temperature dependent.

A more advanced discussion of lattice vibrations requires us to differentiate between transverse and longitudinal waves. It is also necessary to distinguish between the group velocity, defined as $2\pi \mathrm{d}\nu/dq_L$, where q_L is the wave number, and the phase velocity, which is equal to $2\pi\nu/q_L$. The two velocities have the same value for the acoustic vibrations at low frequencies but become different at the other end of the spectrum. This is apparent from the schematic dispersion curve shown in Fig. 3.7. This diagram also shows acoustic and optic branches which both exist when there is more than one atom per unit cell. If there are n atoms per unit cell, there will be three acoustic branches (one longitudinal and two transverse) and $3(n-1)$ optic branches.

Debye attempted to use his elastic continuum model to explain Eucken's observation [6] that the thermal conductivity of a pure insulating crystal varies inversely with the absolute temperature. However, he was unsuccessful in accounting for the fact that the thermal conductivity is not infinite. The thermal conductivity becomes finite only if the thermal vibrations are anharmonic. Peierls [7] was able to take this anharmonicity into account and, thus, explain the phenomenon of thermal resistance in the absence of imperfections, such as impurities.

It was Peierls who first introduced the idea of phonons or quantised vibrational wave packets. He showed that phonons could interact with one another in two ways. In normal or N-processes, the momentum of the phonons is conserved, whereas in Umklapp or U-processes, momentum is not conserved. The law of conservation

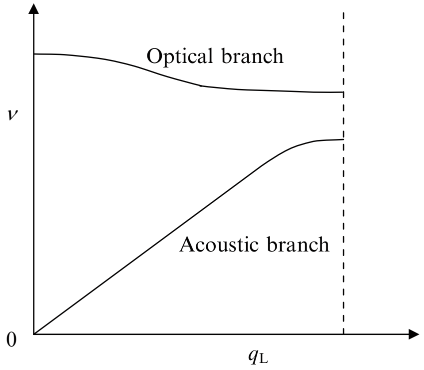


Fig. 3.7 Schematic dispersion curve for the acoustic branches in a diatomic lattice

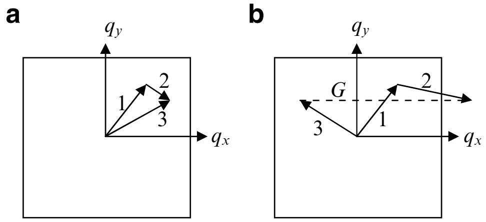


Fig. 3.8 Representation in two dimensions of (**a**) N-processes and (**b**) U-processes

of momentum is satisfied for U-processes, if the translation of the whole crystal is taken into account. N-processes are important in redistributing momentum but only the U-processes lead to thermal resistance. The two types of process are illustrated in Fig. 3.8.

In Fig. 3.8, we, more properly, use wave vector, q, rather than momentum to represent the motion of the phonons. The squares are unit cells in two-dimensional wave-vector space. In the diagram, the phonons 1 and 2 interact to produce a third phonon 3. In the N-process, the third phonon is obtained simply by vector addition of 1 and 2, but in the U-process, a reciprocal lattice vector $\boldsymbol{G}$ has to be added to bring the third phonon within the cell. We may represent the two processes by

$$q_3 = q_1 + q_2, \text{ (N-process)} \tag{3.45}$$

and

$$q_3 = q_1 + q_2 + G. \text{ (U-process)} \tag{3.46}$$

Peierls showed that U-processes become more probable as the temperature rises since they can occur only when there are reasonable numbers of phonons with wave vectors that are sufficiently large to produce a resultant outside the unit cell in wave-vector space. At low temperatures, it is predicted that the mean free path of the phonons should be proportional to $[\exp(-\Theta_D/aT)]^{-1}$, where a is a constant equal to about 2. This exponential term should dominate the temperature variation of the thermal conductivity. It was, therefore, something of a mystery that the $1/T$ variation of the thermal conductivity should persist to quite low temperatures in many materials. Eventually, it was realised that scattering of phonons on various types of point defect was masking the exponential behaviour. This behaviour was observed only after pure and perfect single crystals could be studied.

We may express the lattice thermal conductivity in terms of the specific heat, the speed of sound, and the mean free path, l_t, of the phonons using the relation

$$\lambda_L = c_V v l_t / 3. \tag{3.47}$$

We require this quantity to be small in a thermoelectric material and we shall discuss the factors that control the magnitude of the phonon free path in later chapters.

3.5 Phonon Drag

The theoretical approach that has been used in the previous sections is based on the assumption that the flows of the charge carriers and the phonons can be treated independently. However, under some circumstance, this assumption becomes invalid. When the two flows become linked, there appear what are known as the phonon drag effects.

Generally, the phonon drag effects become stronger as the temperature is reduced. Thus, their influence on the Seebeck coefficient of germanium was observed [8, 9] at temperatures below 100 K. Herring [10] showed that the higher than expected thermoelectric effect could be explained if it were supposed that electrons were scattered preferentially by the phonons in the direction of the flow of heat. Gurevich [11, 12] had earlier proposed that such an effect might occur and phonon drag is sometimes called the Gurevich effect.

We discuss the origin of phonon drag with reference to the Peltier effect. Even when phonon drag occurs, the Seebeck and Peltier coefficients still satisfy the Kelvin relation.

Phonon drag is expected to be strongest when the carrier concentration, n, is low. Under the influence of an electric field, E, these carriers accept momentum at the rate neE per unit volume. This momentum may be lost in various ways. It can be passed on to impurities and other defects and thence lost in random thermal vibrations. Alternatively, it can be given to the phonon system and retained until non-momentum-conserving collisions occur. We suppose that the fraction of collisions that involve phonons is x and the relaxation time for the loss of momentum from the phonons is τ_d. Then the excess momentum carried by the phonons is

$$\Delta p = \mp x n e \tau_d E. \tag{3.48}$$

The key to understanding phonon drag is the realisation that the time τ_d can be much greater than the relaxation time that controls the heat conduction process. The charge carriers in a semiconductor are scattered primarily by phonons of much lower energy than the heat-conduction phonons.

The electric current density is

$$i = ne\mu E, \tag{3.49}$$

and the rate of flow of heat per unit cross-sectional area is

$$w = v^2 \Delta \boldsymbol{p}. \tag{3.50}$$

The phonon drag contribution to the Peltier coefficient is the ratio of the rate of heat flow to the electric current, and is thus

$$\pi_{\mathrm{d}} = \mp \frac{x v^2 \tau_d}{\mu}, \tag{3.51}$$

and the phonon drag Seebeck coefficient is

$$\alpha_{\mathrm{d}} = \frac{\pi_{\mathrm{d}}}{T} = \mp \frac{x v^2 \tau_d}{\mu T}. \tag{3.52}$$

It is noted that the usual thermoelectric coefficients are reinforced by the phonon-drag coefficients as they have the same sign.

The phonon drag effects are strongly dependent on temperature. Typically, τ_d is proportional to T^{-5} and μ is proportional to $T^{-3/2}$, so we expect α_{d} to vary as $T^{-9/2}$. It is unusual to find a significant phonon drag effect at room temperature, though it has been observed [13] for semiconducting diamond above 300 K. Phonon drag has certainly been found in bismuth [14, 15] at low temperatures and is, generally, larger in semimetals than metals [16].

It is noteworthy that (3.51) and (3.52) do not contain the carrier concentration. In fact, the phonon drag effects become smaller as n increases. This is due in part to the scattering of charge carriers on the donor or acceptor impurities but, more significantly, to the so-called saturation effect. When the carrier concentration is high, momentum is increasingly transferred back to the electrons from the phonons. Herring [10] showed that (3.51) and (3.52) are modified due to the saturation effect. Equation (3.52) becomes

$$\alpha_{\mathrm{d}} = \mp \left(\frac{\mu T}{x v^2 \tau_d} + \frac{3 n e x v^2 \tau_d}{N_{\mathrm{d}} k \mu T} \right)^{-1}, \tag{3.53}$$

where N_{d} is the number of phonon modes that interact with the charge carriers. It is the saturation effect that will probably prevent phonon drag being exploited as a means of improving the figure of merit.

Keyes [17] showed that the value of z that can be reached using phonon drag is rather low. We suppose that the ordinary contribution to the Seebeck coefficient is small compared with the drag component. The optimum carrier concentration can be found from the expression for the figure of merit with the Seebeck coefficient given by (3.53) and the electrical conductivity equal to $ne\mu$. Since we are seeking an upper limit for the phonon drag figure of merit, we ignore the electronic component of the thermal conductivity. Thence,

$$n_{\mathrm{opt}} = \frac{N_{\mathrm{d}} k T \mu}{3 e x \tau_d v^2}. \tag{3.54}$$

We then find that the figure of merit is

$$z_{\mathrm{d}} = \frac{N_{\mathrm{d}} k x \tau_d v^2}{12 \lambda_{\mathrm{L}} T}. \tag{3.55}$$

Keyes drew attention to the fact that $N_\mathrm{d}kv^2\tau_d/3$ is the contribution, λ_d, of the low energy phonons to the thermal conductivity and is, therefore, less than the total lattice conductivity. The phonon drag dimensionless figure of merit is

$$z_\mathrm{d}T = \frac{x\lambda_\mathrm{d}}{4\lambda_\mathrm{L}}. \tag{3.56}$$

Since x is no greater than 1 and $\lambda_\mathrm{d} < \lambda_\mathrm{L}$, we see that $z_\mathrm{d}T$ must be less than $1/4$. We conclude that phonon drag cannot assist us with our aim of finding values of zT in excess of unity.

Our discussion has been restricted to bulk thermoelectric materials. Ivanov [18] has more recently considered phonon drag in low-dimensional structures and has come to the conclusion that the effect cannot improve the performance in this case either.

References

1. A. Sommerfeld, Z. Phys. **47**(1), 43 (1928)
2. H.J. Goldsmid, *Electronic Refrigeration* (Pion, London, 1986), p. 39
3. P.J. Price, Philos. Mag. **46**, 1252 (1955)
4. H.J. Goldsmid, Proc. Phys. Soc. B **69**, 203 (1956)
5. P. Debye, *Vorträge über die kinetische Theorie*, (Teubner) (1914)
6. A. Eucken, Ann. Phys. **34**, 185 (1911)
7. R.E. Peierls, Ann. Phys. **3**, 1055 (1929)
8. H.P.R. Frederikse, Phys. Rev. **92**, 248 (1953)
9. T.H. Geballe, G.W. Hull, Phys. Rev. **94**, 1134 (1954)
10. C. Herring, Phys. Rev. **96**, 1163 (1954)
11. L. Gurevich, J. Phys. USSR. **9**, 477 (1945)
12. L. Gurevich, J. Phys. USSR. **10**, 67 (1946)
13. H.J. Goldsmid, C.C. Jenns, D.A. Wright, Proc. Phys. Soc. **73**, 393 (1959)
14. V.N. Kopylov, L.P. Mezhov-Deglin, Sov. Phys. JETP. **38**, 357 (1974)
15. J. Boxus, J.P. Issi, J. Phys. C. **10**, L397 (1977)
16. J.P. Issi, in *Thermoelectrics Handbook: Macro to Nano*, ed. by D.M. Rowe (CRC Taylor and Francis, Boca Raton, 2006), p. 30–10
17. R.W. Keyes, in *Thermoelectricity: Science and Engineering*, ed. by R.R. Heikes, R.W. Ure (Interscience, New York, 1961), p. 389
18. Y.U.V. Ivanov, in *Thermoelectrics Handbook: Macro to Nano*, ed. by D.M. Rowe (CRC Taylor and Francis, Boca Raton, 2006), p. 18–19

Chapter 4

Optimisation and Selection of Semiconductor Thermoelements

4.1 Power Factor

In recent years, it has been found convenient to introduce a quantity known as the power factor that contains both the Seebeck coefficient and the electrical conductivity. The power factor is defined as $\alpha^2\sigma$ and is useful because α and σ are the parameters that are most strongly dependent on the carrier concentration. The other quantity that is involved in the definition of the figure of merit is the thermal conductivity, λ. λ is less dependent on the concentration of the charge carriers since it is often dominated by the lattice contribution. Thus, the carrier concentration that yields the maximum power factor for a given material is usually close to that which gives the highest figure of merit.

It is easy to see that the power factor will become small when the Fermi level moves too far into the forbidden gap of a semiconductor. The Seebeck coefficient of a non-degenerate material varies linearly with the Fermi energy whereas the electrical conductivity varies in an exponential fashion. Thus, even though it is the square of the Seebeck coefficient that enters into the expressions for the power factor or figure of merit, the rapid variation of the electrical conductivity will become the dominating factor when the reduced Fermi energy, η, is much less than zero. On the other hand, when we enter the metallic region, with η much greater than zero, the ratio of the thermal conductivity to the electrical conductivity eventually approaches the value given by the Wiedemann-Franz law. Further increase of η leads to a decrease in the Seebeck coefficient without any compensating increase in the ratio of electrical to thermal conductivity. Thus, there will always be an optimum value for the Fermi energy, which implies an optimum value of the carrier concentration. However, it is usual to optimise the Seebeck coefficient rather than the carrier concentration since α and η are so closely related to one another.

In Fig. 4.1, we show how the power factor varies with reduced Fermi energy when the scattering parameter $r = -1/2$, that is its value for acoustic-mode lattice scattering of the charge carriers. There is a maximum value when the Fermi level lies just inside the conduction or valence band. We would expect the optimum Fermi energy to become more negative when we consider the figure of merit rather than

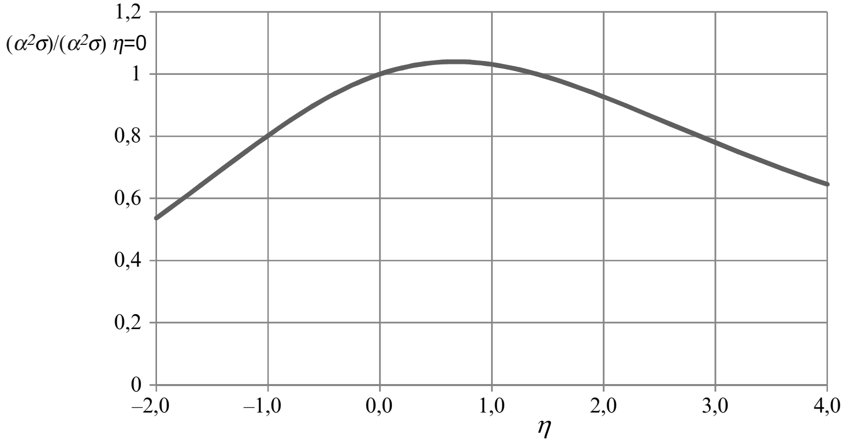


Fig. 4.1 Plot of power factor against reduced Fermi energy for $r = -1/2$

the power factor but for present day materials η would not be much less than zero. Reference to Fig. 3.5 indicates that the preferred value for the Seebeck coefficient will be close to $\pm 200\,\mu\mathrm{V\,K}^{-1}$.

In principle, it is possible for the power factor to increase continually with the Fermi energy. If the plot in Fig. 4.1 is drawn for $r = 3/2$, that is for ionised-impurity scattering, we find that the power factor continues to rise with η. However, that does not take account of the fact that the relaxation time for such scattering will become less as impurities are added to produce the charge carriers that are needed to increase the Fermi energy. Thus, in practice, we would expect the optimum Fermi energy to be close to zero whatever form of scattering is dominant.

4.2 The Materials Parameter, *β*

Although the magnitude of the optimum Seebeck coefficient does not vary very much from one material to another, the values of the power factor and of the figure of merit can change by orders of magnitude. Thus, we shall now determine the properties on which the maximum value of z will depend.

It is instructive to express the figure of merit in terms of the reduced Fermi energy for the region in which non-degenerate statistics can be employed. We then find that

$$zT = \frac{[\eta - (r + 5/2)]^2}{(\beta \exp(\eta))^{-1} + (r + 5/2)}, \tag{4.1}$$

where β is a materials parameter that was first introduced by Chasmar and Stratton [1]. This parameter is defined by the relation

$$\beta = \left(\frac{k}{e}\right)^2 \frac{\sigma_0 T}{\lambda_L} \tag{4.2}$$

and σ_0 is a quantity that depends on the carrier mobility and the effective mass according to

$$\sigma_0 = 2e\mu \left(\frac{2\pi m^* kT}{h^2}\right)^{3/2}. \tag{4.3}$$

Equation (4.1) shows us that, for a given scattering parameter r, the dimensionless figure of merit for any particular Fermi energy is a function solely of the parameter β. The greater the value of β, the greater is the value of zT.

Although (4.1) holds only for a non-degenerate semiconductor, the parameter β remains useful when the material is partly or completely degenerate. zT is still a function only of η, r and β for all values of the Fermi energy. If we ignore the fundamental constants in (4.2) and (4.3), we find that β is proportional to $(\mu/\lambda_L)(m^*/m)^{3/2}$ where m is the mass of a free electron. We see, then, that we require materials that possess high values of the mobility and effective mass for the charge carriers and low values of the lattice thermal conductivity.

In Fig. 4.2, we show how the dimensionless figure of merit, zT, varies with the reduced Fermi energy for different values of the parameter β. We have supposed that the scattering parameter r has the value of $-1/2$, as for acoustic-mode lattice scattering. We see that, as β becomes larger, the optimum value for η becomes more negative. Thus, if β were large enough, we could use classical statistics in our calculations. However, for the best materials that are used in today's thermoelectric

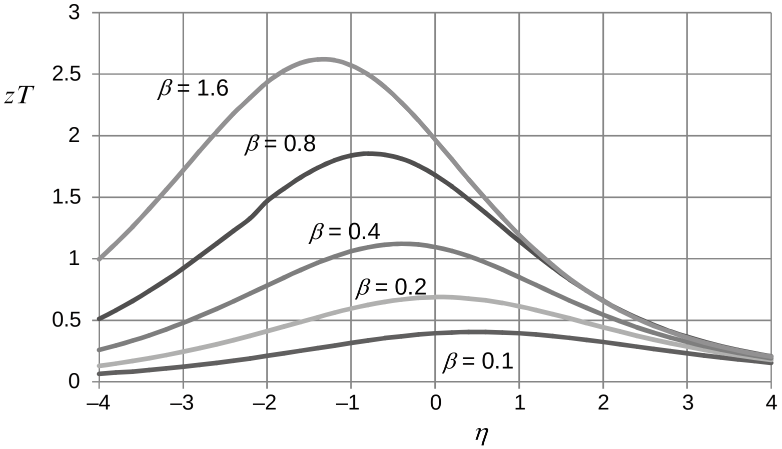


Fig. 4.2 The dimensionless figure of merit plotted against the reduced Fermi energy for different values of the parameter β. The scattering parameter $r = -1/2$

modules, β is a little less than 0.4 and we hardly expect it ever to approach the highest value in Fig. 4.2, i.e., 1.6.

Referring to Fig. 4.2, it can be seen that the optimum Fermi energy varies by little more than kT for a wide range of values for the parameter β. In fact, even if the optimum value of η were as small as -2, the Seebeck coefficient would be only about $\pm 350\,\mu\text{V K}^{-1}$. It is possible to find semiconducting materials with Seebeck coefficients of the order of $\pm 1\,\text{mV K}^{-1}$ or more, but we would not expect them to have high figures of merit. We could improve such materials by doping them with donor or acceptor impurities so as to increase the electrical conductivity and this would be more than sufficient compensation for the fall in the Seebeck coefficient.

4.3 Mobility and Effective Mass

The parameter β is proportional to $\mu m^{*3/2}$ but the selection of a material that satisfies the requirement that this product should be high is not straightforward. In general, when the effective mass of the carriers is high, the mobility is low. However, we have not yet specified exactly what we mean by the effective mass in (4.3).

In a cubic semiconductor with a simple valence or conduction band, having a maximum or minimum energy at the centre of the unit cell in wave-vector space (i.e., at the centre of the Brillouin zone), there is only one value for the effective mass. On the other hand, when the band edges are to be found at other points in the Brillouin zone, symmetry requires that the bands be of the so-called multi-valley type. The effective mass that appears in (3.3) and, thence, in the expression for β is the density-of-states mass. In a band that has N_v valleys, the density-of-states mass is $N_v^{2/3}$ times the value that it would have for a single valley.

The effective mass for each valley, i.e., the inertial mass, is still important as it is related to the carrier mobility. In fact, the inertial mass will generally be different in different directions and this may have to be taken into account, as well. Thus, the density-of-states mass m^* should really be written as

$$m^* = N_v^{2/3} \left(m_1 m_2 m_3\right)^{1/3}, \tag{4.4}$$

where m_1, m_2, and m_3 are the effective masses along the axes of symmetry within each valley. We may write the density-of-states mass, m_N, for a single valley as $\left(m_1 m_2 m_3\right)^{1/3}$ while an appropriate value for the inertial mass, m_I, is $3/\left(1/m_1 + 1/m_2 + 1/m_3\right)$. Exactly how the mobility varies with the effective mass depends on the nature of the scattering process. If acoustic-mode lattice scattering is the predominant mechanism, the mobility is proportional to $m_N^{-3/2} m_I^{-1}$ so $\mu \left(m^*\right)^{3/2}$ is proportional to N_v/m_I. This suggests that we require a multi-valley semiconductor with a low inertial mass for the carriers but a large number of valleys.

4.4 The Lattice Thermal Conductivity in Pure Crystals

The basis for one of the first guidelines in the selection of thermoelectric materials was the observation by Ioffe and Ioffe [2] that the lattice conductivity in a group of materials with similar structure and bonding falls as the atomic weight becomes larger. The results that they obtained for a number of materials are included in Fig. 4.3 in which the lattice conductivity at room temperature is plotted against the mean atomic weight.

Figure 4.3 shows that, within any group of materials, the lattice conductivity falls as the mean atomic weight rises. The thermal conductivity of the alkali halides is an order of magnitude lower than that of the diamond-type elements and the III–V compounds. Since a low lattice conductivity is desirable for thermoelectric materials, one might think that the ionic compounds would be a good choice. However, these materials have very small carrier mobilities with very low values for the power factor. The bonding in most of the useful thermoelectric materials is, in fact, largely covalent, at least for operation at ordinary temperatures.

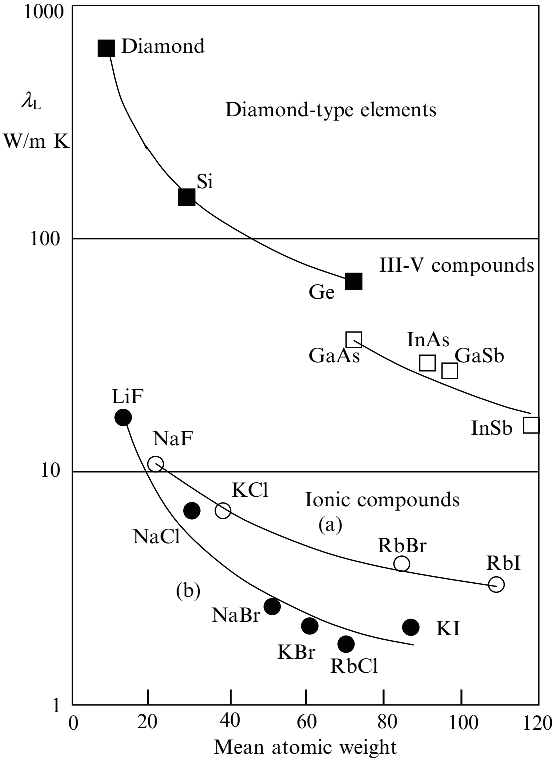


Fig. 4.3 Plot of lattice conductivity against mean atomic weight for certain covalent and ionic crystals. The plot is based on the observations of Ioffe and Ioffe [2]. The ionic compounds are divided into those (**a**) with the atomic weight ratio less than 1.5 and those (**b**) with the atomic weight ratio greater than 1.5

One feature of the behaviour of the ionic compounds is the fact that the lattice conductivity is lower for those with a high ratio of the atomic weights of the constituent elements than for those with a lower ratio. This suggests that, when using compounds, one should select those with large differences between the atomic weights of the elements of which they are composed.

A precise prediction of the lattice conductivity for a new material is not easy. However, an approximate theory that relates λ_L to the melting temperature, T_m, has been developed by Keyes [3]. Keyes based his predictions on earlier approximate theories presented by other workers.

A good starting point is the formula given by Leibfried and Schlömann [4]. By using the variational method, they were able to show that the anharmonicity of the lattice waves leads to the formula

$$\lambda_L = 3.5 \left(\frac{k}{h}\right)^3 \frac{M V^{1/3} \Theta_D^3}{\gamma^2 T}, \tag{4.5}$$

where M is the mean atomic mass and V is the mean atomic volume. γ is the Grüneisen parameter that, with the thermal expansion coefficient, is a measure of the anharmonicity.

A formula that is close to (4.5) can be obtained by simple arguments. Firstly, one makes use of a relationship given by Dugdale and MacDonald [5] who proposed that there should be a link between the lattice conductivity of a pure crystal and its thermal expansion coefficient, α_T, since both depend on the anharmonicity. Dugdale and MacDonald represented the anharmonicity by the dimensionless quantity $\alpha_T \gamma T$ and suggested that the mean free path, l_t, of the phonons is equal to $a/\alpha_T \gamma T$, where a is the lattice constant. Equation (3.47) then leads to the formula

$$\lambda_L = \frac{c_v a v}{3\alpha_T \gamma T}. \tag{4.6}$$

We may then make use of the Debye equation of state

$$\alpha_T = \frac{\chi \gamma c_v}{3}, \tag{4.7}$$

where χ is the compressibility. The speed of sound, v, is related to the Debye temperature by the equation

$$v = (\rho_d \chi)^{-1/2} = \frac{2ka\Theta_D}{h}, \tag{4.8}$$

where ρ_d is the density. Thence, if the lattice constant is set equal to the cube root of the atomic volume,

$$\lambda_L = 8 \left(\frac{k}{h}\right)^3 \frac{M V^{1/3} \Theta_D^3}{\gamma^2 T}. \tag{4.9}$$

Equation (4.9) differs from Leibfried and Schlömann's equation (4.5) only in the value of the numerical constant. For convenience, we shall use (4.9) in our treatment.

Keyes had the intention of relating the lattice conductivity to properties of a substance that would be known immediately after its synthesis. From this point of view, (4.6) and (4.9) are not really adequate. Equation (4.6) requires the knowledge of the expansion coefficient, while to use (4.9) we need the speed of sound.

As an alternative starting point, Keyes made use of Lawson's relation [6]

$$\lambda_{\mathrm{L}} = \frac{a}{3\gamma^2 T \chi^{3/2} \rho_{\mathrm{d}}^{1/2}}, \tag{4.10}$$

which can be obtained from (4.8) and (4.9). He also used the Lindemann melting rule

$$T_{\mathrm{m}} = \frac{\varepsilon_{\mathrm{m}} V}{R\chi}, \tag{4.11}$$

where R is the gas constant and T_{m} is the melting temperature. The rule is based on the principle that a solid will melt when the atomic vibrations reach a fraction ε_{m} of the lattice constant, ε_{m} being approximately the same for all substances. Thence, the Keyes relation is

$$\lambda_{\mathrm{L}} T = B_{\mathrm{K}} \frac{T_{\mathrm{m}}^{3/2} \rho_{\mathrm{d}}^{2/3}}{A^{7/6}}, \tag{4.12}$$

where

$$B_{\mathrm{K}} = \frac{R^{3/2}}{3\gamma^2 \varepsilon_m^3 N_A^{1/3}}. \tag{4.13}$$

In (4.13), N_A is Avogadro's number and A is the mean atomic weight. Thus, B_{K} involves only universal constants and the two parameters ε_{m} and γ that do not vary much from one material to another. The three variables, T_{m}, ρ_{d} and A, that appear in (4.12) will be known as soon as a material is prepared. Thus, this equation should be a valuable guide to the prediction of the lattice conductivity.

Keyes was able to compare the thermal conductivity, as predicted by (4.12), with the experimental data from a large range of dielectric crystals. He found that the data invariably agreed with this equation, to within an order of magnitude, provided that B_{K} is given the value of 3×10^{-4} SI units. He was able to improve the agreement if he selected values for B_{K} equal to 1.3×10^{-3} SI units for covalently-bonded materials and 1.5×10^{-4} SI units for ionic crystals. We are, particularly, interested in semiconductors [7] and we find that there is reasonable agreement with (4.12) for the lattice conductivity of these materials if B_{K} is given the value 6×10^{-4} SI units, as is shown in Fig. 4.4 in which $\lambda_{\mathrm{L}} T$ is plotted against $T_{\mathrm{m}}^{3/2} \rho d^{2/3} A^{-7/6}$. The data were not all obtained at room temperature but were restricted to the region in which λ_{L} satisfies Eucken's law.

The Keyes formula is consistent with the observations of Ioffe and Ioffe but shifts our attention towards the melting temperature as well as the mean atomic weight. It must be emphasised that it is an approximation and, as shown by the spread of values in Fig. 4.4, it merely serves as a useful guideline.

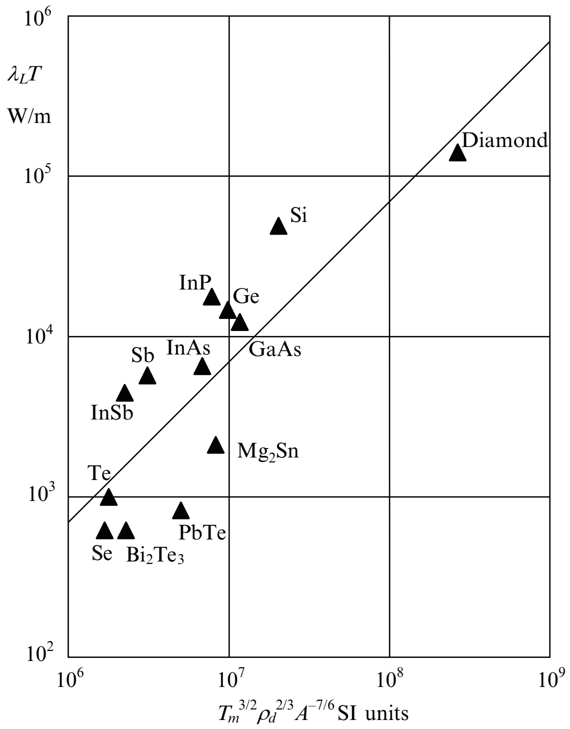


Fig. 4.4 Plot of $\lambda_L T$ against $T_m^{3/2}\rho_d^{2/3}A^{-7/6}$ for a number of semiconductors. The line represents (4.12) with the parameter B_K equal to 6×10^{-4} SI units

As we shall see later, there are ways in which the lattice conductivity can be reduced below its value for a large pure and perfect crystal. Nevertheless, the selection of a material that has a low value of λ_L when only phonon–phonon scattering exists is a good starting point in our search for good thermoelectric materials.

4.5 The Effect of Temperature

It must be realised that there is no single material that can be used in thermoelectric energy conversion at all temperatures. As we shall see, the compound bismuth telluride has been the basis of the materials for a thriving industry that has produced modules for Peltier coolers over the past half-century. However, one cannot use bismuth telluride in thermoelectric generation from a high-temperature heat source since its properties deteriorate as T becomes greater. As the temperature reaches a few hundred degrees Celsius, the compound becomes chemically unstable and, eventually, melts. This, however, is not the only problem. As the temperature is raised, electron-hole pairs are produced by excitation across the energy gap and this reduces the Seebeck coefficient while, at the same time, increasing the thermal conductivity through the bipolar effect. To some extent, the onset of mixed conduction can be minimised by adding donor or acceptor impurities but the size of the energy gap sets a limit on what can be done by this means. At some higher temperature, then, bismuth telluride must be replaced by a material that has a higher melting point and larger energy gap. The replacement material, in turn, will be overtaken by a semiconductor with an even larger melting point and wider band gap as the temperature rises still further.

Over the range of temperature for which minority carrier conduction is negligible, the dimensionless figure of merit increases with T. This, of course, occurs even if the figure of merit z remains constant, but there are, in fact, reasons for hoping that z itself will increase.

The optimum value of the reduced Fermi energy does not change very much with temperature so the preferred Seebeck coefficient is more-or-less constant. The optimum carrier concentration is then approximately proportional to $T^{3/2}$ since this quantity appears in expressions, such as (3.29), for the carrier concentration. On the other hand, if lattice scattering with $r = -1/2$ is predominant, the mobility, μ, of the carriers varies as $T^{-3/2}$. This implies that the optimum electrical conductivity should be almost independent of temperature. Any temperature dependence of the figure of merit should, thus, be associated with the thermal conductivity.

For a given value of the Lorenz number and electrical conductivity, the electronic component of the thermal conductivity is proportional to the temperature. However, in most thermoelectric materials, the lattice component is significantly larger than the electronic component and, as we have seen, it is inversely proportional to the absolute temperature for a pure crystal. Even when the phonons are partly scattered by impurities or lattice defects, we might still expect the lattice conductivity to fall as the temperature rises. Thus, z will probably rise with temperature and zT will certainly do so.

For many years, the best dimensionless figure of merit at 300 K remained at about unity but substantially higher values were being found for generator materials at elevated temperatures. This was not because of higher values of z but, rather, because of the factor T in zT. For the same reason, although large values of z have been found [8] for Bi–Sb alloys at around liquid nitrogen temperature, the corresponding values of zT at 80 K have remained less than that of alloys based on bismuth telluride at room temperature.

4.6 The Importance of the Energy Gap

For various reasons, the energy gap of most of the semiconductors that are used in thermoelectric energy convertors is rather small, in spite of the fact that this leads to the possibility of minority carrier conduction. The maximum Seebeck coefficient for a given semiconductor is closely related to the energy gap and it is necessary that this maximum should be somewhat greater than the optimum value, as derived for a single type of carrier. It is, in fact, possible to use the observed maximum Seebeck coefficient to estimate the energy gap if this quantity is not already known [9].

The Seebeck coefficient of a mixed semiconductor is given by (3.35). In order to calculate its value, we need to know the relative carrier concentrations and mobilities for the two types of carrier. We relate the carrier concentrations to the reduced Fermi energies through the expressions

$$n = 2\left(\frac{2\pi m_{\mathrm{n}}^{*} kT}{h^{2}}\right)^{3/2} \exp\left(\eta_{\mathrm{n}}\right), \tag{4.14}$$

and

$$p = 2\left(\frac{2\pi m_{\mathrm{p}}^{*} kT}{h^{2}}\right)^{3/2} \exp\left(\eta_{\mathrm{p}}\right), \tag{4.15}$$

where η_{n} and η_{p} are linked to one another and the energy gap, E_{g} by

$$\eta_{\mathrm{n}} + \eta_{\mathrm{p}} = -\frac{E_{\mathrm{g}}}{kT} \equiv -\eta_{\mathrm{g}}. \tag{4.16}$$

We shall make use of a parameter C that we define as

$$C = \left(\frac{\mu_{\mathrm{n}}}{\mu_{\mathrm{p}}}\right)\left(\frac{m_{\mathrm{n}}^{*}}{m_{\mathrm{p}}^{*}}\right)^{3/2}, \tag{4.17}$$

where m_{n}^{*} and m_{p}^{*} are the density-of-states masses for the electrons and holes. Then

$$\frac{\sigma_{\mathrm{n}}}{\sigma_{\mathrm{p}}} = C \exp\left(\eta_{\mathrm{n}} - \eta_{\mathrm{p}}\right). \tag{4.18}$$

It is likely that the maximum Seebeck coefficient will be found when the Fermi level lies sufficiently far from both the valence and conduction bands for classical statistics to be applicable. The exceptions will occur when the energy gap is very small or when the parameter C is either very large or very small [10]. When C is close to unity, the maximum Seebeck coefficient lies close to the value $E_{\mathrm{g}}/2eT$ and it remains reasonably close to this value even for $C \gg 1$ or $C \ll 1$.

Figure 4.5 shows a plot of $2eT\alpha_{\max}/E_{\mathrm{g}}$ against C in a typical case of $\eta_{\mathrm{g}} = 10$ where it has been assumed that $r = -1/2$ for both types of carrier. As C becomes much larger than 1, the maximum Seebeck coefficient becomes somewhat larger for an n-type sample and somewhat smaller for a p-type sample. However, if the maximum Seebeck coefficient is found for both types of material, the average is almost independent of C.

The easiest way to determine the energy gap from Seebeck measurements is to increase the temperature until mixed and then intrinsic conduction is observed. The maximum Seebeck coefficient for a plot against temperature at a fixed doping level is close to the maximum for a plot against electrical conductivity at a fixed temperature. When the Seebeck coefficient is measured over a range of temperature, only one sample (or, perhaps, one of each conductivity type) is needed.

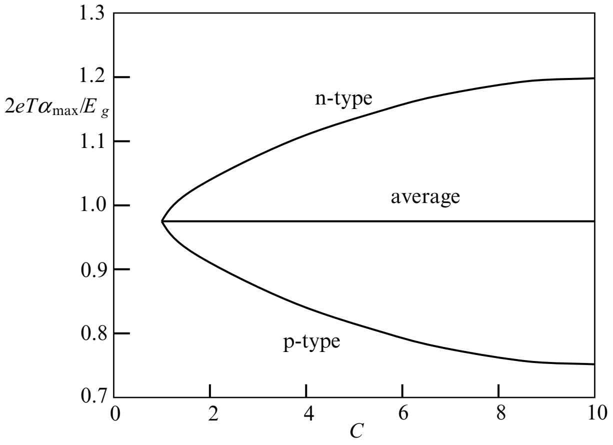


Fig. 4.5 $2eT\alpha_{\max}/E_g$ plotted against the parameter *C* for n-type and p-type material. The central line represents the average for the maximum Seebeck coefficients of both types

We can now make a reasonable estimate of the minimum energy gap that is needed for a thermoelectric material to be operated at a certain temperature. We know that the optimum Seebeck coefficient for an extrinsic semiconductor is about $\pm 200\,\mu\text{VK}^{-1}$. The maximum Seebeck coefficient should be significantly larger than this if mixed conduction is to be avoided. Let us suppose that the maximum Seebeck coefficient must not be less than, say, $\pm 260\,\mu\text{V K}^{-1}$. This means that the energy gap must not be smaller than about 6*kT*. If the energy gap is smaller than this, for an otherwise suitable material, it may be necessary to operate with a Seebeck coefficient that is smaller than the optimum value for a single carrier.

Empirically, it seems that the best material at any given temperature barely satisfies the condition $E_g > 6kT$. This is because in a given series of semiconductors, the ones with the highest mobility, or, rather, the highest product $\mu\,(m^*/m)^{3/2}$, appear to have the smallest energy gaps. Thus, many of the well-known high-mobility semiconductors, such as indium antimonide, indium arsenide, and mercury telluride have small energy gaps. This is associated with the fact that the relation between the energy and wave vector is different near the band edge from that deep within the band.

4.7 Non-Parabolic Bands

Up to this point, it has been supposed that the energy bands are parabolic; that is, the energy of the carriers is proportional to $(\boldsymbol{k} - \boldsymbol{k}_0)^2$ where $\boldsymbol{k}$ is the wave vector and $\boldsymbol{k}_0$ its value at the band extremum. This supposition breaks down if the direct gap (i.e., the gap at a given position in wave vector space) between the valence and conduction band is small. The parabolic approximation is, generally, satisfactory only close to the band edge. The problem of dealing with non-parabolic bands is difficult though a somewhat simplified approach has been presented by Kolodziejczak and Zhukotynski [11].

It was proposed by Kane [12] that the energy–wave vector relation for indium antimonide should have the form

$$E = \frac{\hbar^2 k^2}{2m} - \frac{1}{2}E_g + \frac{1}{2}\left(E_g^2 + \frac{8Q^2k^2}{3}\right)^{1/2}, \tag{4.19}$$

where the energy gap E_g is supposed to be of the direct type and Q is a parameter that is characteristic of a particular band. If one neglects the first term on the right-hand side of (4.19), the result is

$$E = \frac{1}{2}\left[\left(E_g^2 + \frac{8Q^2k^2}{3}\right)^{1/2} - E_g\right]. \tag{4.20}$$

We notice that, when the energy gap is large compared with Qk, the usual parabolic relationship between energy and wave vector is the result. In terms of Q the relation is

$$E = \frac{2Q^2}{3E_g}k^2. \tag{4.21}$$

However, when the energy gap is much smaller than Qk, we find that

$$E = \left(\frac{2}{3}\right)^{1/2} Qk, \tag{4.22}$$

and the relation between energy and wave vector is linear. The result is an increase in the effective mass of the carriers and a decrease in the mobility.

One might expect all the semiconductors in a given series to have more-or-less the same value of Q. If this is so, then (4.21) shows that effective mass should be inversely proportional to the energy gap. This explains the high mobility of compounds like InSb. It is still possible for the density-of-states mass to be reasonably large if the bands are of the many-valley type.

In view of the fact that the semiconductors that are used in thermoelectric conversion are very impure compared with those to be found in the microelectronic industry, the charge carriers are not likely to be confined to the band edges. Thus, non-parabolicity of the energy–wave vector relation is a real possibility. Hopefully, the carriers near the band edge will be dominant so that non-parabolicity may have only a minor influence on the transport properties. It may, however, account for anomalies in the behaviour of thermoelectric materials that are sometimes observed.

It is noted that Harman and Honig [13] have discussed non-parabolic bands in the context of thermoelectric and thermomagnetic devices.

4.8 Thermomagnetic Materials

It can be shown that an extrinsic conductor is best used in the Seebeck-Peltier mode rather than as a Nernst-Ettingshausen device. The Nernst coefficient for one type of charge carrier is given by the approximation [14]

$$N = \frac{\mu}{eT\left(1 + \mu^2 B_z^2\right)} \left[\frac{\langle \tau_e^2 e \rangle}{\langle \tau_e^2 \rangle} - \frac{\langle \tau_e e \rangle}{\langle \tau_e \rangle} \right], \tag{4.23}$$

where the angular brackets denote averages for all values of the energy ε. The quantity in the square brackets varies from zero for a completely degenerate conductor to rkT for non-degeneracy. The thermomagnetic power, NB_z, has a maximum value, when $\mu B = 1$, of no more than $rk/2e$. The highest value for NB_z is $3k/4e$, that is about $65 \mu\,\mathrm{V\,K^{-1}}$, if we assume that r can be no greater than its value of 3/2 for ionised-impurity scattering. This is clearly much less than the optimum Seebeck coefficient for thermoelectric energy conversion. It should be noted that the Nernst coefficient takes the same sign as r and can be used in the determination of the scattering law.

Although mixed and intrinsic conduction should be avoided in a thermoelectric device, it turns out that the bipolar effects are an advantage when using the Nernst and Ettingshausen effects. While the electrons and holes act in opposition in the Peltier and Seebeck effects, they assist one another in the Nernst and Ettingshausen effects. The difference between the Ettingshausen effects in extrinsic and intrinsic conductors is illustrated in Fig. 4.6. In the extrinsic case, we have assumed that the scattering parameter r is less than zero.

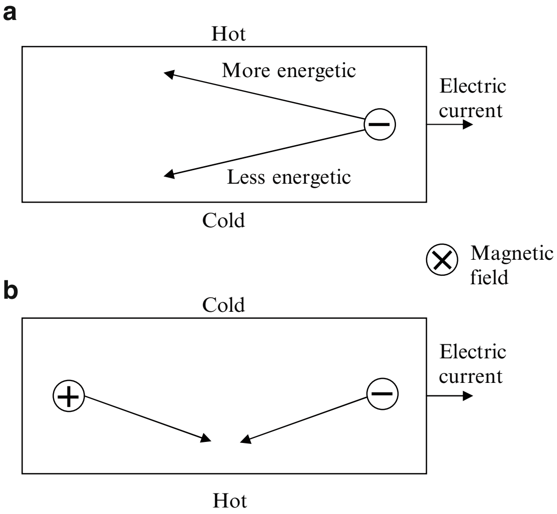


Fig. 4.6 Origin of the Ettingshausen effect in (**a**) an extrinsic conductor and (**b**) an intrinsic conductor

In the extrinsic conductor, (a), the magnetic field tries to drive the carriers downwards but the boundaries prevent this flow from taking place. There has to be zero transverse electric current. However, it is still possible for the carriers with the longer relaxation time to move down if their partial current is balanced by an upwards flow of the carriers with a shorter relaxation time. This means that the top of the sample becomes hot and the bottom becomes cold.

There is again a downwards force on the charge carriers in the intrinsic conductor, (b), and in this case, they can move downwards without any overall charge flow. Thus, the bottom of the sample becomes heated and the top is cooled. The Ettingshausen effect is clearly stronger when both positive and negative carriers are present.

We shall now derive an expression for the Nernst coefficient of a mixed conductor. We have to set the longitudinal electric current equal to zero, so, from (3.37),

$$i_{\mathrm{n,x}} = -i_{\mathrm{p,x}} = \frac{\sigma_{\mathrm{n}}\sigma_{\mathrm{p}}}{\sigma_{\mathrm{n}} + \sigma_{\mathrm{p}}} \left(\alpha_{\mathrm{p}} - \alpha_{\mathrm{n}}\right) \frac{\mathrm{d}T}{\mathrm{d}x}. \tag{4.24}$$

The magnetic field produces transverse currents $i_{\mathrm{n,y}}$ and $i_{\mathrm{p,y}}$ that must be equal and opposite. Then, if E_{y} is the electric field due to the Nernst effect and $R_{\mathrm{H,n}}$ and $R_{\mathrm{H,p}}$ are the partial Hall coefficients

$$i_{\mathrm{n,y}} = \left(E_{\mathrm{y}} + R_{\mathrm{H,n}} i_{\mathrm{n,x}} B_{\mathrm{z}}\right) \sigma_{\mathrm{n}}, \tag{4.25}$$

and

$$i_{\mathrm{p,y}} = \left(E_{\mathrm{y}} + R_{\mathrm{H,p}} i_{\mathrm{p,x}} B_{\mathrm{z}}\right) \sigma_{\mathrm{p}}. \tag{4.26}$$

Setting the transverse current equal to zero and eliminating the partial currents using (4.24),

$$E_{\mathrm{y}} = \frac{\left(R_{\mathrm{H,p}}\sigma_{\mathrm{p}} - R_{\mathrm{H,n}}\sigma_{\mathrm{n}}\right) \sigma_{\mathrm{n}}\sigma_{\mathrm{p}}}{\left(\sigma_{\mathrm{n}} + \sigma_{\mathrm{p}}\right)^2} (\alpha_{\mathrm{p}} - \alpha_{\mathrm{n}}) B_z \frac{\mathrm{d}T}{\mathrm{d}x}. \tag{4.27}$$

The Nernst coefficient is

$$N = \frac{\left(R_{\mathrm{H,p}}\sigma_{\mathrm{p}} - R_{\mathrm{H,n}}\sigma_{\mathrm{n}}\right) \sigma_{\mathrm{n}}\sigma_{\mathrm{p}}}{\left(\sigma_{\mathrm{n}} + \sigma_{\mathrm{p}}\right)^2} \left(\alpha_{\mathrm{p}} - \alpha_{\mathrm{n}}\right). \tag{4.28}$$

$|R_{\mathrm{H,n}}\sigma_{\mathrm{n}}|$ and $R_{\mathrm{H,p}}\sigma_{\mathrm{p}}$ are the Hall mobilities, $\mu_{\mathrm{H,n}}$ and $\mu_{\mathrm{H,p}}$, of the two types of carrier, which become equal to the mobilities μ_{n} and μ_{p}, as defined by (3.30) in a high magnetic field. In terms of the Hall mobilities

$$N = \frac{\left(\mu_{\mathrm{H,p}} + \mu_{\mathrm{H,n}}\right) \sigma_{\mathrm{n}}\sigma_{\mathrm{p}}}{\left(\sigma_{\mathrm{n}} + \sigma_{\mathrm{p}}\right)^2} \left(\alpha_{\mathrm{p}} - \alpha_{\mathrm{n}}\right). \tag{4.29}$$

We also need to use expressions for the electrical and thermal conductivities in a magnetic field. The electrical conductivity is given by

$$\sigma = \frac{\sigma_{\mathrm{n}}}{1 + \mu_{\mathrm{n}}^2 B_{\mathrm{z}}^2} + \frac{\sigma_{\mathrm{p}}}{1 + \mu_{\mathrm{p}}^2 B_{\mathrm{z}}^2}. \tag{4.30}$$

The conductivity is defined as the ratio of the current density in the x direction to the electric field that has no transverse component. The isothermal electrical resistivity, ρ^i, and the conductivity are related through [15]

$$\sigma = \frac{\rho^i}{\left(\rho^i\right)^2 + R_{\mathrm{H}}^2 B_{\mathrm{z}}^2}, \tag{4.31}$$

where R_{H} is the overall Hall coefficient. At the high magnetic field limit, R_{H} becomes equal to $\left(1/R_{\mathrm{H,p}} - 1/R_{\mathrm{H,n}}\right)^{-1}$.

The behaviour of the thermal conductivity in a high magnetic field has been discussed by Tsidil'kovskii [16]. The theory is complex and the exact expressions are cumbersome but we are justified in using an approximation for the electronic thermal conductivity since the lattice conductivity is likely to be predominant. In this context, a good approximation for the electronic thermal conductivity is

$$\lambda_{\mathrm{e}} = \frac{\lambda_{\mathrm{e,n}}}{1 + \mu_{\mathrm{H,n}}^2 B_{\mathrm{z}}^2} + \frac{\lambda_{\mathrm{e,p}}}{1 + \mu_{\mathrm{H,p}}^2 B_{\mathrm{z}}^2} + \frac{\sigma_{\mathrm{n}} \sigma_{\mathrm{p}} \left(\alpha_{\mathrm{p}} - \alpha_{\mathrm{n}}\right)^2 T}{\sigma_{\mathrm{n}} \left(1 + \mu_{\mathrm{H,n}}^2 B_{\mathrm{z}}^2\right) + \sigma_{\mathrm{p}} \left(1 + \mu_{\mathrm{H,p}}^2 B_{\mathrm{z}}^2\right)}, \tag{4.32}$$

where $\lambda_{\mathrm{e,n}}$ and $\lambda_{\mathrm{e,p}}$ are the partial electronic thermal conductivities in zero magnetic field.

When the transverse electric field is zero, the electronic thermal conductivity tends towards zero at very high magnetic fields. However, under the more usual experimental conditions [17], it is the transverse electric current that is zero and the total thermal conductivity then becomes

$$\lambda = \lambda_{\mathrm{L}} \left(1 + Z_{\mathrm{NE}} T\right). \tag{4.33}$$

We shall discuss this effect later when we deal with the thermal conductivity of bismuth.

We are now in a position to obtain an expression for the thermomagnetic figure of merit of an intrinsic conductor. We shall assume that the magnetic field is high enough for $\mu_{\mathrm{n}}^2 B^2 \gg 1 \ll \mu_{\mathrm{p}}^2 B^2$ and that there are equal numbers of mobile electrons and holes. Under these conditions (4.29) becomes

$$N = \frac{\mu_{\mathrm{n}} \mu_{\mathrm{p}}}{\mu_{\mathrm{n}} + \mu_{\mathrm{p}}} \left(\alpha_{\mathrm{p}} - \alpha_{\mathrm{n}}\right). \tag{4.34}$$

In a high magnetic field, the Hall coefficient of an intrinsic conductor becomes equal to zero. Thus, the isothermal electrical resistivity is given by

$$\rho^{i} = \frac{1}{\sigma} = \left(\frac{\sigma_{\mathrm{n}}(0)}{1 + \mu_{\mathrm{n}}^{2} B_{\mathrm{z}}^{2}} + \frac{\sigma_{\mathrm{p}}(0)}{1 + \mu_{\mathrm{p}}^{2} B_{\mathrm{z}}^{2}} \right)^{-1}. \tag{4.35}$$

We are interested in the isothermal thermal conductivity, which, unlike the adiabatic quantity in (4.33), does tend to λ_{L} for a very high magnetic field. Thus, the thermomagnetic figure of merit is given by

$$Z_{\mathrm{NE}} = \frac{n_{\mathrm{i}} e \mu_{\mathrm{n}} \mu_{\mathrm{p}} \left(\alpha_{\mathrm{p}} - \alpha_{\mathrm{n}} \right)}{\left(\mu_{\mathrm{n}} + \mu_{\mathrm{p}} \right) \lambda_{\mathrm{L}}}, \tag{4.36}$$

where n_{i} is the number of electrons or holes.

One cannot optimise a thermomagnetic material in the same way as a thermoelement. However, in a range of alloys, the energy gap may vary while the carrier mobilities and the lattice conductivity may remain more or less constant. Suppose that we consider first what happens in a non-degenerate conductor as the energy gap changes, the concentration of each type of carrier is then given by

$$n_{\mathrm{i}} = 2 \left(m_{\mathrm{n}}^{*} m_{\mathrm{p}}^{*} \right)^{3/2} \left(\frac{2\pi m k T}{h^{2}} \right)^{3/2} \exp\left(-\frac{E_{\mathrm{g}}}{2kT} \right). \tag{4.37}$$

and at the high magnetic field limit

$$\left(\alpha_{\mathrm{p}} - \alpha_{\mathrm{n}} \right) = \frac{E_{\mathrm{g}} + 5kT}{eT}. \tag{4.38}$$

Thus, in the classical region

$$Z_{\mathrm{NE}} \propto \left(E_{\mathrm{g}} + 5kT \right)^{2} \exp\left(-\frac{E_{\mathrm{g}}}{2kT} \right). \tag{4.39}$$

In order for classical statistics to be reasonably accurate, it is necessary that the energy gap should be greater than about $4kT$. If this condition is satisfied, Z_{NE} continually increases as the energy gap decreases so we know that the optimum gap must be less than $4kT$. We, therefore, have to use Fermi-Dirac statistics.

Let us suppose that the effective masses of the electrons and holes are equal. When this condition holds, we can obtain the following proportionality:

$$Z_{\mathrm{NE}} \propto F_{\mathrm{NE}} = F_{r+1/2}(\eta) \left[\frac{5\mathrm{F}_{3/2}(\eta)}{3\mathrm{F}_{1/2}(\eta)} - \eta \right]^{2}, \tag{4.40}$$

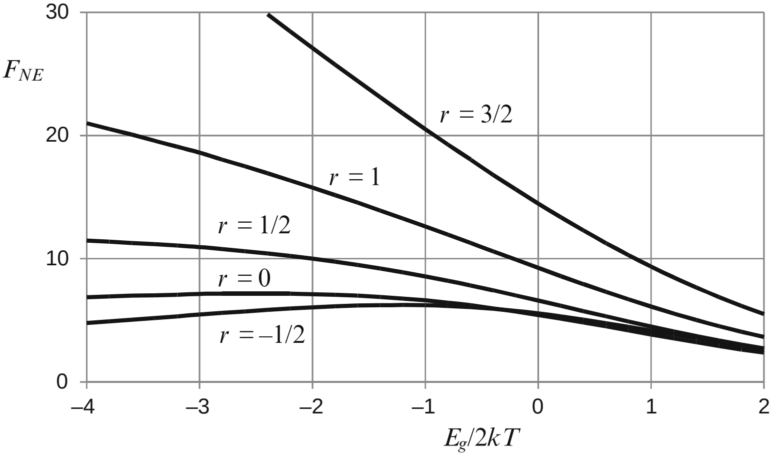


Fig. 4.7 Plot of the function F_{NE} against $E_{\mathrm{g}}/2kT$ for different values of the scattering parameter r. The Nernst-Ettingshausen figure of merit is proportional to F_{NE}. F_{NE} is defined as the function on the right hand side of the relation (4.40)

where the reduced Fermi energy, η, is equal to $-E_{\mathrm{g}}/2kT$. In Fig. 4.7, we show the function on the right hand side of the proportionality (4.40) plotted against $E_{\mathrm{g}}/2kT$.

Figure 4.7 shows us that the thermomagnetic figure of merit is highest when the bands overlap by about 2*kT*. In other words, the ideal thermomagnetic material would seem to be a semimetal with slightly overlapping bands. We note that, once *r* becomes positive, it appears that the greater the band overlap the better. However, such values of *r* imply a substantial amount of impurity scattering, with a reduced carrier mobility. This would make it very difficult to achieve the condition $(\mu B)^2 \gg 1$. Thus, it is reasonable to conclude that the best thermomagnetic materials will have only a small overlap between the conduction and valence bands.

If the energy gap has a certain value, the high field thermomagnetic figure of merit satisfies the proportionality

$$Z_{\mathrm{NE}} \propto \left(\frac{m_{\mathrm{n}}^* m_{\mathrm{p}}^*}{m^2}\right)^{3/4} \left[\left(\frac{1}{\mu_{\mathrm{n}}} + \frac{1}{\mu_{\mathrm{p}}}\right)\lambda_{\mathrm{L}}\right]^{-1}. \tag{4.41}$$

The quantity on the right-hand side of this proportionality resembles the factor $(\mu/\lambda_{\mathrm{L}})(m^*/m)^{3/2}$ that enters into the parameter β for thermoelectric materials. If one of the carriers is much more mobile than the other, say $\mu_{\mathrm{n}} \gg \mu_{\mathrm{p}}$, then Z_{NE} is proportional to μ_{p} and only the mobility of the less mobile carrier is important. Thus, we really need a material in which both the carriers are highly mobile.

If the mobilities of both carriers are equal, $\mu_{\mathrm{p}} = \mu_{\mathrm{n}} = \mu$, then (4.36) becomes

$$Z_{\mathrm{NE}} = \frac{2n_{\mathrm{i}} e \mu \alpha^2}{\lambda_{\mathrm{L}}}, \tag{4.42}$$

where $\alpha = \alpha_p = -\alpha_n$. This equation reveals some of the advantages that thermomagnetic energy convertors might have compared with thermoelectric devices. The factor of 2 in the numerator arises from the fact that the electrons and holes share a common lattice. Furthermore, the partial Seebeck coefficients usually become larger when a magnetic field is applied. Also, only the lattice contribution to the thermal conductivity remains as the magnetic field becomes very large. However, one must not forget that the thermoelectric figure of merit might itself be improved in a magnetic field.

4.9 Superconductors as Passive Thermoelements

The circumstances can arise in which there is a good thermoelectric material of one conductivity type but no equally good material of the other type. For example, bismuth is an excellent n-type material at low temperatures but not particularly good when holes are the majority carriers. When only a single material is available, the thermocouple can be completed using a normal metal as the second branch. The metal will contribute little or nothing to the Seebeck coefficient but, hopefully, it will have a much higher ratio of electrical to thermal conductivity than exists in the active material. In practice, this approach does not really work since a good thermoelectric material will have a ratio of electrical to thermal conductivity that is not much less than the value for a metal, that is the Wiedemann-Franz ratio. However, if the metal is a superconductor, it can be a truly passive branch. It will not contribute to the thermoelectric effects but will have an infinite ratio of electrical to thermal conductivity, provided that the critical current is not exceeded.

The use of superconductors as passive thermoelements was first discussed by Goldsmid et al. [18] and has been investigated more thoroughly by Vedernikov and Kuznetsov [19]. These authors discussed specifically a Bi–Sb alloy as the active component in conjunction with a high-temperature superconductor. At present, this type of combination is restricted to temperatures that are not much greater than that of liquid nitrogen, i.e., 77 K.

Although the superconducting leg need not add to the electrical resistance, it will have a finite thermal conductance. Goldsmid and his colleagues noted that a typical Bi–Sb alloy has a thermal conductivity of $3\,\mathrm{W\,m^{-1}\,K^{-1}}$. They proposed the use of $\mathrm{YBa_2Cu_3O_{9-\delta}}$, which has a thermal conductivity of $0.6\,\mathrm{W\,m^{-1}\,K^{-1}}$ at 80 K, as the superconductor. They assumed that the active thermoelement would be 10 mm in length with a current density of $6 \times 10^5\ \mathrm{A/m^2}$. Thus, so that the heat loss through the superconductor would not exceed 10% of that through the active branch, it should have a current density of $12 \times 10^5\ \mathrm{A/m^2}$. The critical current density in the sample of $\mathrm{YBa_2Cu_3O_{9-\delta}}$ studied by Goldsmid et al. was only $1.9 \times 10^4\ \mathrm{A/m^2}$ even in zero magnetic field and, in fact, a magnetic field is really needed to enhance the thermoelectric properties of the Bi–Sb.

Vedernikov and Kuznetsov discussed not only $\mathrm{YBa_2Cu_3O_{9-\delta}}$ but also superconductors from the Bi–Sr–Ca–Cu–O system. The requirement for the superconductor

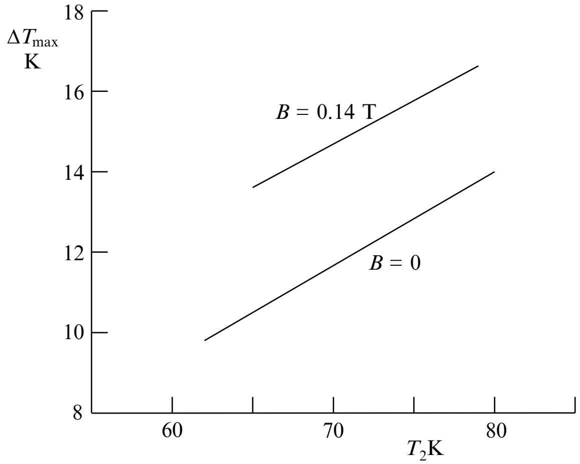


Fig. 4.8 Maximum temperature depression plotted against hot junction temperature for a thermocouple made from $Bi_{0.85}Sb_{0.15}$ and the superconductor $BiSrCaCu_2O_x$. Data of Vedernikov and Kuznetsov [19]

is that it should have a low thermal conductivity and a high critical current density. It is also necessary that its electrical contacts should have negligible resistance and, preferably, it should be capable of being used in a magnetic field if, indeed, it is to be operated in conjunction with Bi–Sb. The material chosen for the experimental work was $BiSrCaCu_2O_x$, which has a critical temperature of 87 K and a critical current density of $12 \times 10^5\ A/m^2$ at 77 K. It was provided with low-resistance electrical contacts by the electrodeposition of silver with subsequent annealing. Although this superconductor is barely adequate for the purpose, it was possible to obtain a significant temperature depression from a thermocouple in which it was combined with $Bi_{0.85}Sb_{0.15}$. It was even possible to use this couple in a magnetic field to obtain an enhanced cooling effect as shown in Fig. 4.8. It is apparent that the use of a superconducting branch is a viable option under certain conditions.

References

1. R.P. Chasmar, R. Stratton, J. Electron. Control **7**, 52 (1959)
2. A.V. Ioffe, A.F. Ioffe, Dokl. Akad. Nauk. SSSR **97**, 821 (1954)
3. R.W. Keyes, Phys. Rev. **115**, 564 (1959)
4. G. Leibfried, E. Schlömann, Nachr. Akad. Wiss., Göttingen, Kl. 2, Math-Physik **2**, 71 (1954)
5. J.S. Dugdale, D.K.C. MacDonald, Phys. Rev. **98**, 1751 (1955)
6. A.W. Lawson, J. Phys. Chem. Solids **3**, 154 (1957)
7. H.J. Goldsmid, *Thermoelectric Refrigeration* (Plenum, New York, 1964), p. 68
8. G.E. Smith, R. Wolfe, J. Appl. Phys. **33**, 841 (1962)
9. E. Müller, W. Heiliger, P. Reinshaus, H. Süssmann, in *Proceedings of the 15th International Conference on Thermoelectrics* (IEEE, Pasadena, 1996), p. 412
10. H.J. Goldsmid, J.W. Sharp, J. Electronic Materials **28**, 869 (1999)

11. J. Kolodziejczak, S. Zhukotynski, Phys. Stat. Solidi **5**, 145 (1964)
12. E.O. Kane, J. Phys. Chem. Solids **1**, 249 (1957)
13. T.C. Harman, J.M. Honig, *Thermoelectric and Thermomagnetic Effects and Applications* (McGraw-Hill, New York, 1967)
14. R.T. Delves, Br. J. Appl. Phys. **13**, 440 (1962)
15. A.C. Beer, *Galvanomagnetic Effects in Semiconductors* (Academic Press, New York, 1963), p. 12
16. I.M. Tsidil'kovskii, *Thermomagnetic Effects in Semiconductors* (Infosearch, London, 1962)
17. R.T. Delves, Br. J. Appl. Phys. **15**, 105 (1964)
18. H.J. Goldsmid, K.K. Gopinathan, D.N. Matthews, K.N.R. Taylor, C.A. Baird, J. Phys. D **21**, 344 (1988)
19. M.V. Vedernikov, V.L. Kuznetsov, in *CRC Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC Press, Boca Raton, 1994), p. 609

Chapter 5
Minimising the Thermal Conductivity

5.1 Semiconductor Solid Solutions

In the previous chapter, we discussed the selection of materials with a low lattice conductivity under the condition that the phonons are scattered only by other phonons. However, we know that the lattice conductivity can be reduced by the scattering of phonons on various types of defect, including the boundaries of any finite crystal. Such defects may, of course, scatter the charge carriers as well as the phonons. Indeed, since the mean free path is usually greater for electrons or holes than it is for phonons, we might expect there to be a greater effect on the mobility than on the lattice conductivity. In practice, it turns out that, in many cases, the ratio of the mobility to the lattice conductivity can be raised through defect scattering.

A most useful proposal was made in 1956 by Ioffe et al. [1]. They suggested that the formation of a solid solution between two semiconductors that have the same crystal structure should lead to a reduction in the lattice conductivity. It was claimed that the mobility of the charge carriers would not necessarily be reduced by the alloying process. It was argued that the long-range order would be preserved and, since the wavelength associated with the charge carriers is also rather large, they would suffer no additional scattering. On the other hand, the phonons that predominate in the conduction of heat have short wavelengths and are scattered by the disturbances in the short-range order in a solid solution.

In Fig. 5.1, we show the lattice resistivity plotted against the proportion of a second component in certain semiconductors. The effect of alloying is, particularly, large in the Si–Ge system because the lattice conductivity of pure silicon and germanium is so high that it is comparable with the total thermal conductivity of most metals.

Ioffe et al. showed that the lattice conductivity of a solid solution varies approximately according to the rule

$$\frac{1}{\lambda_{\mathrm{L}}} = \frac{1}{(\lambda_{\mathrm{L}})_0} + 4x\,(1-x)\left[\frac{1}{(\lambda_{\mathrm{L}})_{\mathrm{m}}} - \frac{1}{(\lambda_{\mathrm{L}})_0}\right], \tag{5.1}$$

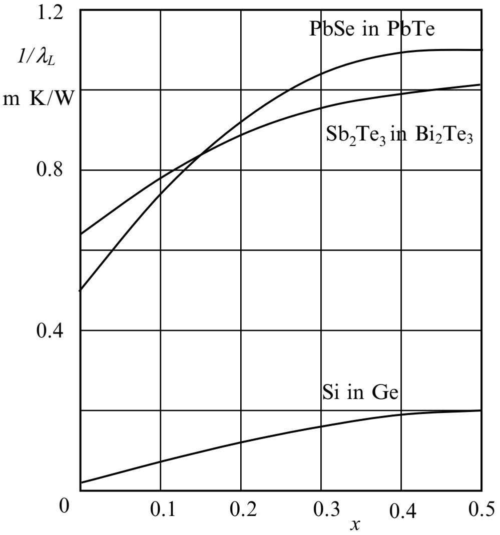


Fig. 5.1 Lattice thermal resistivity plotted against proportion, x, of second component in semiconductor solid solutions

where $(\lambda_L)_0$ is the lattice conductivity when the proportion, x, of the second component is zero and $(\lambda_L)_m$ is the lattice conductivity, when $x = 0.5$.

Airapetyants et al. [2] considered the carrier mobility in solid solutions of compounds. They noticed that the mobility of the electrons is more strongly affected when substitution is made on the electropositive sub-lattice whereas the mobility of holes is more affected by substitution on the electronegative sub-lattice. Figure 5.2 shows the manner in which the ratio of electron to hole mobility varies in solid solutions of PbTe with PbSe and SnSe.

The ideas of Airapetyants et al. can be criticised since there are no strong reasons to associate the motion of electrons or holes with a particular sub-lattice. Moreover, the principle, that the carrier mobility is less strongly affected than the lattice conductivity by the formation of a solid solution, seems to work satisfactorily in solid solutions between elemental semiconductors like silicon and germanium for which electropositive and electronegative sub-lattices cannot be identified. Nevertheless, there is evidence to support the ideas of Airapetants for other systems. For example, it seems preferable to use alloys between Bi_2Te_3 and Sb_2Te_3 for positive thermoelements and alloys between Bi_2Te_3 and Bi_2Se_3 for negative thermoelements.

5.2 Phonon Scattering by Point Defects

The basis of the scattering of phonons in solid solutions is the local changes in density associated with the different atoms. Scattering can also result from local changes in the elastic constants. According to the Rayleigh theory, the scattering cross-section, σ, for point defects is given by

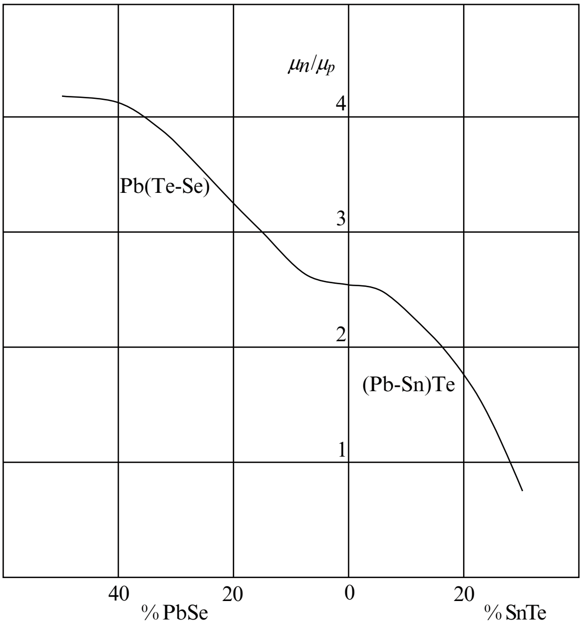


Fig. 5.2 Plot of ratio of electron to hole mobility against concentration of second component in solid solutions of PbTe with PbSe and SnTe

$$\sigma = \frac{4\pi c^6 q_\mathrm{L}^4}{9}\left(\frac{\Delta\chi}{\chi} + \frac{\Delta\rho_\mathrm{d}}{\rho_\mathrm{d}}\right)^2, \tag{5.2}$$

where c is the diameter of the defect, q_L is the magnitude of the phonon wave vector, $\Delta\chi$ is the local change of compressibility, and $\Delta\rho_\mathrm{d}$ is the local change of density. Rayleigh scattering is a classical concept that is applicable only when the defects are much smaller than the wavelength of the phonons. This is not true for the higher frequency phonons but we are justified in using (5.2) because such phonons are so strongly scattered that they make little contribution to the thermal conductivity [3]. For the same reason, we may use the Debye model for the lattice vibrational spectrum since it is a good approximation for the low-frequency phonons.

There is a general problem in dealing with the lattice conductivity. As was shown by Peierls, it is the Umklapp processes that account for the thermal resistance of pure crystals, but we should not neglect the redistribution of phonons due to the normal processes. A powerful technique for handling this problem was developed by Callaway [4]. The normal processes are at first supposed to be just as effective as Umklapp processes in the scattering of phonons. Then, a correction is applied on the basis that any disturbance in the phonon distribution relaxes through the normal processes to a distribution that still carries momentum. Although Callaway's treatment has encountered some criticism, it has the virtue that it can be applied with relative ease. We shall, therefore, give an outline of the Callaway approach.

Suppose that there are processes with a relaxation time τ_{R} that change the momentum or wave vector and other processes with a relaxation time τ_{N} that conserve wave vector; then the distribution function, N, will relax according to the equation

$$\left(\frac{\mathrm{d}N}{\mathrm{d}t}\right)_{\text{scatter}} = \frac{N_0 - N}{\tau_{\mathrm{R}}} + \frac{N_{\mathrm{N}} - N}{\tau_{\mathrm{N}}}, \tag{5.3}$$

where N_0 is the equilibrium distribution function and N_{N} is the distribution function to which the normal processes on their own would lead. The temperature gradient ∇T will change the distribution function according to the equation

$$\left(\frac{\mathrm{d}N}{\mathrm{d}t}\right)_{\text{diffusion}} = -v \cdot \nabla T \frac{\partial N}{\partial T}, \tag{5.4}$$

where v is the sound velocity in the direction of the phonon wave vector, $\boldsymbol{q}_{\mathrm{L}}$. Since the diffusion and scattering processes must balance,

$$\frac{N_0 - N}{\tau_{\mathrm{R}}} + \frac{N_{\mathrm{N}} - N}{\tau_{\mathrm{N}}} - v \cdot \nabla T \frac{\partial N}{\partial T} = 0, \tag{5.5}$$

The distribution function to which the normal processes lead has to carry momentum against the temperature gradient. The Bose–Einstein function that applies to phonons in the absence of a temperature gradient is

$$N_0 = \left[\exp\left(\frac{\hbar\omega}{kT}\right) - 1\right]^{-1}, \tag{5.6}$$

so we suppose that we can use the distribution

$$N_{\mathrm{N}} = \left[\exp\left(\frac{\hbar\omega - \boldsymbol{q}_{\mathrm{L}} \cdot \boldsymbol{l}}{kT}\right) - 1\right]^{-1}. \tag{5.7}$$

In these equations, we have used $\hbar$ to represent $h/2\pi$ and, ω, which is equal to $2\pi f$, to denote the angular frequency. $\boldsymbol{l}$ is a constant vector in the direction of the temperature gradient and is such that $\boldsymbol{q} \cdot \boldsymbol{l} \ll \hbar\omega$. Equation (5.7) differs from (5.6) in that the frequency ω is changed by $-\boldsymbol{q} \cdot \boldsymbol{l}/\hbar$. Thus, the normal processes lead to a change in the distribution function

$$N_{\mathrm{N}} - N_0 = \frac{q_{\mathrm{L}} \cdot \boldsymbol{l}}{kT} \frac{\exp(x)}{(\exp(x) - 1)^2}, \tag{5.8}$$

where $x = \hbar\omega/kT$. Also, from (5.6)

$$\frac{\partial N}{\partial T} = \frac{\hbar\omega}{kT^2} \frac{\exp(x)}{(\exp(x) - 1)^2} \tag{5.9}$$

By combining (5.8) and (5.9) we find that

$$(N_0 - N)\left(\frac{1}{\tau_\mathrm{R}} + \frac{1}{\tau_\mathrm{N}}\right) - \left(v \cdot \nabla \boldsymbol{T} - \frac{q_\mathrm{L} \cdot \boldsymbol{l} T}{\hbar \omega \tau_\mathrm{N}}\right) \frac{\partial N}{\partial T} = 0. \tag{5.10}$$

Since $\boldsymbol{l}$ must be proportional to the temperature gradient, we may express this vector in the form

$$\boldsymbol{l} = -\frac{\hbar}{T} \beta v^2 \nabla T, \tag{5.11}$$

where β is a constant that has the dimensions of time. Then, using the Debye model to replace $\boldsymbol{q}_\mathrm{L}$ by ω / v, (5.10) may be written as

$$(N_0 - N)\left(\frac{1}{\tau_\mathrm{R}} + \frac{1}{\tau_\mathrm{N}}\right)\left(1 + \frac{\beta}{\tau_\mathrm{N}}\right)^{-1} - v \cdot \nabla \boldsymbol{T} \frac{\partial N}{\partial T} = 0. \tag{5.12}$$

This means that there is an effective relaxation time, τ_eff, that is given by

$$\frac{1}{\tau_\mathrm{eff}} = \left(\frac{1}{\tau_\mathrm{R}} + \frac{1}{\tau_\mathrm{N}}\right)\left(1 + \frac{\beta}{\tau_\mathrm{N}}\right)^{-1} = \frac{1}{\tau_\mathrm{c}}\left(1 + \frac{\beta}{\tau_\mathrm{N}}\right)^{-1}. \tag{5.13}$$

In this equation, τ_c is the relaxation time that would be expected if the normal process did not conserve momentum or wave vector. Equation (5.13) shows that the effective relaxation time is obtained by multiplying τ_c by $(1 + \beta / \tau_\mathrm{N})$.

Thus, to allow for the normal processes, we have to evaluate the quantity β. In principle, we can do this by taking account of the fact that the normal processes conserve wave vector. This means that

$$\int \frac{N_\mathrm{N} - N}{\tau_\mathrm{N}} \boldsymbol{q}_\mathrm{L} d^3 \boldsymbol{q}_\mathrm{L} = \int \frac{4\pi\omega^2}{v^3} \frac{N_\mathrm{N} - N}{\tau_\mathrm{N}} \boldsymbol{q}_\mathrm{L} d\omega = 0. \tag{5.14}$$

Now $(N_\mathrm{N} - N)$ may be written as $(N - N_0) + (N_0 - N)$ and is proportional to $(\beta - \tau_\mathrm{eff})\, \partial N / \partial T$. Thus, (5.14) becomes

$$\int_0^{\theta_\mathrm{D}/T} \left(\frac{\beta}{\tau_\mathrm{N}} - \frac{\tau_\mathrm{c}}{\tau_\mathrm{N}} - \frac{\beta \tau_\mathrm{c}}{\tau_\mathrm{N}^2}\right) \omega^4 \frac{\exp(x)}{(\exp(x) - 1)^2} \mathrm{d}x = 0. \tag{5.15}$$

The expression for β is then

$$\beta = \int_0^{\theta_\mathrm{D}/T} \frac{\tau_\mathrm{c}}{\tau_\mathrm{N}} x^4 \frac{\exp(x)}{(\exp(x) - 1^2} \mathrm{d}x \Big/ \int_0^{\theta_\mathrm{D}/T} \frac{1}{\tau_\mathrm{N}} \left(1 - \frac{\tau_\mathrm{c}}{\tau_\mathrm{N}}\right) x^4 \frac{\exp(x)}{(\exp(x) - 1)^2} \mathrm{d}x. \tag{5.16}$$

It is possible to estimate β without difficulty only in certain cases. Thus, when scattering on imperfections is very strong, so that the corresponding relaxation time τ_I

is very much less than τ_N, one may use the approximation $1/\tau_{\text{eff}} \simeq 1/\tau_I + 1/\tau_N$. Another approximation can be used at low temperatures when Umklapp scattering becomes very weak. However, of most interest to us is an approximation that has been given by Parrott [5] for relatively high temperatures, that is $T > \Theta_D$. He supposed that the relaxation times for both Umklapp and normal processes are then proportional to ω^{-2} while the relaxation time for scattering on point defects is proportional to ω^{-4}. We can, therefore, write

$$1/\tau_I = A\omega^4, \quad 1/\tau_U = B\omega^2, \quad 1/\tau_N = C\omega^2,$$

where A, B, and C are constants for a given specimen. Also, in the high-temperature region, $x \ll 1$ for the whole phonon spectrum, whence $x^2 \exp(x)/[\exp(x) - 1]^2 \simeq 1$. It is then found that the lattice conductivity, λ_L, for the material that contains defects is related to the value λ_0 for a pure and perfect sample by the equation

$$\frac{\lambda_L}{\lambda_0} = \left(1 + \frac{5k_0}{9}\right)^{-1} \left[\frac{\tan^{-1} y}{y} + \left(1 - \frac{\tan^{-1} y}{y}\right)^2 \left(\frac{y^4(1 + k_0)}{5k_0} - \frac{y^2}{3} - \frac{\tan^{-1} y}{y}\right)^{-1}\right], \tag{5.17}$$

where k_0 equal to C/B represents the relative strengths of the normal and Umklapp processes. Also, y is defined by

$$y^2 = \left(\frac{\omega_D}{\omega_0}\right)^2 \left(1 + \frac{5k_0}{9}\right)^{-1}. \tag{5.18}$$

and

$$\left(\frac{\omega_D}{\omega_0}\right)^2 = \frac{k}{2\pi^2 v \lambda_0 \omega_D A}. \tag{5.19}$$

The value of k_0 in (5.18) is found experimentally by measuring the lattice conductivity of a pure sample and one that contains imperfections. The same value of k_0 can then be used for samples with other defect concentrations.

We can get a reasonable idea of the effect of scattering in solid solutions at high temperatures, if we assume that Umklapp processes predominate over normal processes. Then, (5.17) reduces to

$$\frac{\lambda_L}{\lambda_0} = \frac{\omega_0}{\omega_D} \tan^{-1}\left(\frac{\omega_D}{\omega_0}\right). \tag{5.20}$$

Here, λ_0 represents the lattice conductivity of a virtual crystal, which is the value that it would have if the solid solution were perfectly ordered with no point-defect scattering. For want of a better procedure, one might determine λ_0 for a binary alloy by linear interpolation between the values for the lattice conductivity of the two components.

In accordance with (5.2), phonons can be scattered by local variations in both the elasticity and the density. Mass-fluctuation scattering is the easier to deal with and, in fact, might be the major contribution in many semiconductor solid solutions.

If the imperfection scattering is associated with density fluctuations, the appropriate value for the parameter A that appears in (5.19) is A_{M}, which is given by

$$A_{\mathrm{M}} = \frac{\pi}{2v^3 N} \sum_{\mathrm{i}} \frac{x_{\mathrm{i}} \left(M_{\mathrm{i}} - \bar{M}\right)^2}{\bar{M}^2}, \tag{5.21}$$

where x_{i} is the concentration of unit cells of mass M_{i}, $\bar{M}$ is the average mass of a unit cell and N is the number of cells per unit volume.

It can be difficult to predict the scattering due to fluctuations in elasticity. Briefly, a foreign atom changes the local value for the compressibility, partly because it has bonds that differ from those of a host atom and partly because it does not fit well into a lattice site, thus straining the crystal. Both these effects will, like mass fluctuations, cause local changes in the speed of sound. If we use the Debye model of an elastic continuum, an impurity atom of diameter δ'_{i} in its own lattice distorts the space that it occupies from the diameter δ of a host atom to a new diameter δ_{i}. These diameters are related to one another by

$$\frac{\delta_{\mathrm{i}} - \delta}{\delta} = \frac{\Delta\delta_{\mathrm{i}}}{\delta} = \frac{\mu}{1+\mu} \frac{\delta'_{\mathrm{i}} - \delta}{\delta}, \tag{5.22}$$

where

$$\mu = \frac{(1+P)\, G_{\mathrm{i}}}{2\,(1+2P)\, G}. \tag{5.23}$$

In this equation, G and G_{i} are the values for the bulk modulus in the host and impurity crystals, respectively, and P is Poisson's ratio for the host crystal. The equation given by Klemens [3] for the parameter A for strain scattering is

$$A_{\mathrm{S}} = \frac{\pi}{v^3 N} \sum_{\mathrm{i}} x_{\mathrm{i}} \left(\frac{\Delta G_{\mathrm{i}}}{G} - 6.4\gamma \frac{\Delta\delta_{\mathrm{i}}}{\delta} \right)^2, \tag{5.24}$$

where $\Delta G_{\mathrm{i}} = G_{\mathrm{i}} - G$ and γ is the Grüneisen parameter.

In Fig. 5.3, we show the ratio of the lattice conductivity of a solid solution to that of a perfect virtual crystal plotted against $\omega_{\mathrm{D}}/\omega_0$ on the basis of (5.20). The curve has been calculated for mass-defect scattering as given by (5.21). The data points show typical observed values for a number of solid solutions [6]. For some of the solid solutions, there is reasonable agreement with the theoretical curve, but in other cases the observed lattice conductivity falls well below the predicted value. We presume that the difference is accounted for by strain scattering. It is noticeable that none of the data points lie above the theoretical curve.

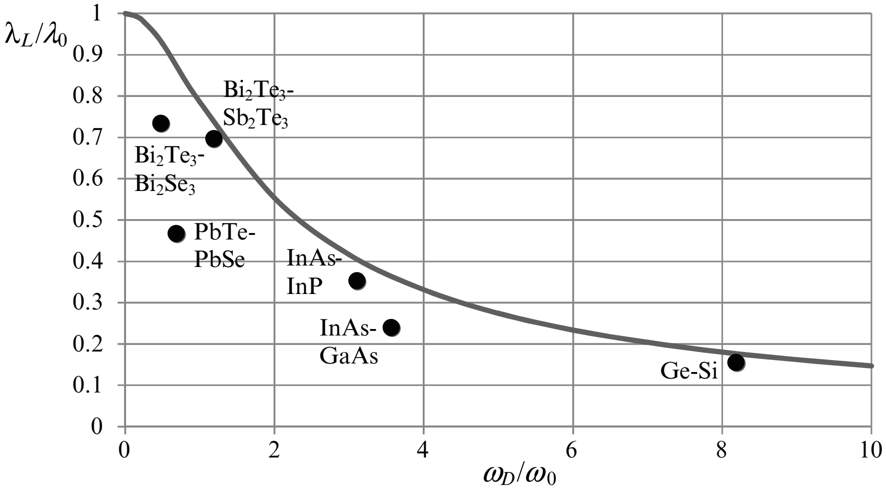


Fig. 5.3 Plot of lattice conductivity against ω_D/ω_0 according to (5.20) assuming mass-defect scattering from (5.21). The data points represent thermal conductivity data for semiconductor solid solutions

5.3 Boundary Scattering

It has long been known that phonons can be scattered on the boundaries of crystals [7] at low temperatures. However, one would only expect the effect to be observed at ordinary temperatures when the grain size is exceedingly small since the mean free path of phonons is usually less than 10^{-9} m. Nevertheless, it was predicted [8] in 1968 that the lattice conductivity might be reduced by boundary scattering of the phonons for grain sizes of the order of 10^{-6} m. The effect was observed [9] for thin sheets of silicon in 1973 and has since been found in other semiconductors. As we shall see, boundary scattering should have a more marked effect for solid solutions than for elements or simple compounds in spite of the reduction in the phonon free path length by alloy scattering.

The key to the understanding of the enhanced boundary scattering effect lies in the fact the free path length for the phonons varies strongly with their vibrational frequency. The low-frequency phonons, though relatively small in number, have a large free path length and, therefore, make a sizeable contribution to the lattice conductivity. In fact, if we assume that the relaxation time for Umklapp scattering is proportional to ω^{-2}, we expect that all frequencies will make comparable contributions to the lattice conductivity, as shown by the upper curve in Fig. 5.4. In this diagram, the relative lattice conductivity $\lambda_L(\omega)$ due to the phonons having an angular frequency ω is plotted against ω. According to the Debye theory, the number of phonons of frequency ω is proportional to ω^2 and we expect the Debye distribution to be valid at low frequencies. Now, point-defect scattering in a solid solution will remove the contribution to the lattice conductivity of most of the high-frequency

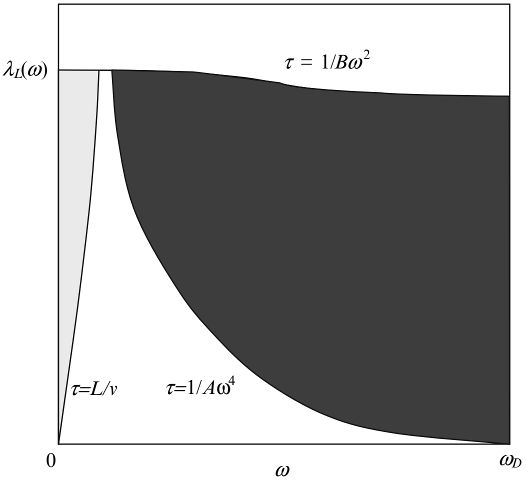


Fig. 5.4 Contribution to the lattice conductivity of phonons of angular frequency ω plotted against ω. The upper curve is applicable when only phonon–phonon scattering occurs. The black region on the right indicates the contribution removed by point-defect scattering. The grey region on the left shows the contribution removed by boundary scattering

phonons. The remaining lattice conductivity is due to low-frequency phonons and, because of their long free path, they are particularly sensitive to boundary scattering.

In Fig. 5.4, the lattice conductivity of a large pure crystal is represented by the area under the upper curve. The black area then represents the loss of lattice conductivity due to point-defect scattering and the grey area is the loss due to boundary scattering. It should be possible to estimate the lattice conductivity in an alloy of small grain size from the ratio of the unshaded area to the total area. Of course, in reality, there are regions in which at least two of the scattering effects are of comparable magnitude and the sharp cut-offs should, therefore, be more gradual. Nevertheless, because the variation of the relaxation time with frequency is quite different for Umklapp, point-defect, and boundary scattering, the error introduced by dividing the plot into three distinct regions is not great. Then, if there is a dominant region for each of these processes, we find that

$$\frac{\lambda_L}{\lambda_S} = 1 - \frac{2}{3}\frac{\lambda_0}{\lambda_S}\sqrt{\frac{l_t}{3L}}, \tag{5.25}$$

where λ_S is the lattice conductivity of a large crystal of the solid solution, λ_0 is that in the absence of alloy scattering, when the mean free path of the phonons is equal to l_t, and L is the effective grain size. L should be close to the actual grain size unless the scattering at the boundaries is substantially specular. It is noted that l_t can be estimated from the thermal conductivity λ_0 using (3.47).

Obviously, (5.25) cannot be applicable when the grain size is very small since it would lead to negative values of λ_L. The reason for this is that there is then no region in which phonon–phonon scattering is dominant unless there is little alloy scattering. There are two special cases of interest.

If alloy scattering is very weak, λ_S becomes equal to λ_0 and only two regions have to be considered. Equation (5.25) is still applicable up to the point at which λ_L

is equal to $\lambda_S/3$; for smaller grain sizes, only boundary scattering is important. On the other hand, if alloy scattering is very strong, (5.25) will apply only until λ_L is equal to $2\lambda_S/3$. In other words, (5.25) can be used until the grain size becomes so small that the lattice conductivity has fallen to between one-third and two-thirds of its value in a large crystal.

Consider, then, what happens when boundary scattering is very strong [10]. For the case of weak alloy scattering, there is no problem. As we enter the region for which (5.25) no longer holds, the lattice conductivity simply becomes proportional to the grain size and is equal to $\lambda_0 L/l_t$. However, the situation is more complicated for strong alloy scattering. In this case, there is an angular frequency ω_B at which alloy scattering takes over from boundary scattering. This frequency is given by $L = v/A\omega_B^4$ and the equation for the lattice conductivity becomes

$$\lambda_L = \left(\frac{cv}{\omega_D^3}\right)\left(\frac{4(v/A)^{3/4}L^{1/4}}{3} - \frac{v}{A\omega_D}\right). \tag{5.26}$$

When ω_B is much less than ω_D, the second term on the right hand side of (5.26) can be neglected and we then expect λ_L to vary as $L^{1/4}$. Thus, in this region, where both alloy scattering and boundary scattering are very strong,

$$\lambda_L = \left(\frac{2\lambda_S}{3}\right)\left[\left(\frac{3L}{l_t}\right)\left(\frac{\lambda_S}{2\lambda_0}\right)^2\right]^{1/4}. \tag{5.27}$$

This equation fails as ω_B approaches ω_D and, when L becomes very small, we obtain the same expression as for weak alloy scattering.

5.4 Scattering of Electrons and Phonons

The considerations of the previous section show that boundary scattering can appear when the mean free path of the phonons is much smaller than the grain size. For this reason alone, it is possible that boundary scattering can improve the ratio of mobility to lattice conductivity even when the mean free path is greater for the charge carriers than for the phonons. It is also possible that boundary scattering of the charge carriers might be more nearly specular than it is for the phonons. However, let us for the moment assume that the boundaries are equally effective in scattering phonons and electrons or holes.

If the charge carriers are scattered by the acoustic-mode lattice vibrations in a large crystal, the mobility is given by

$$\mu_0 = \frac{4}{3\pi^{1/2}}\frac{e\tau_0(kT)^{-1/2}}{m^*}, \tag{5.28}$$

where we have set $r = -1/2$ in (3.30). Thence, the mean free path, l_0, of the charge carriers is given by

$$l_0 = \frac{3\mu_0 \left(2\pi m^* kT\right)^{1/2}}{4e}. \tag{5.29}$$

When boundary scattering is also present, the free path length for the charge carriers is given by

$$l_e = \left(\frac{1}{l_0} - \frac{1}{L}\right)^{-1}, \tag{5.30}$$

where L again represents the effective grain size. Thus, the mobility is expressed as

$$\frac{\mu}{\mu_0} = \frac{L/l_0}{1 + L/l_0}. \tag{5.31}$$

Let us then consider the effect of boundary scattering in a Si–Ge alloy, a material that has found application in thermoelectric generators. Calculations of both the lattice conductivity and the electron mobility have been carried out [10] using (5.25) and (5.31) for a wide range of effective grain sizes. Figure 5.5 shows how μ/μ_0 and λ_L/λ_s vary with L/l_t for Si–Ge with equal proportions of the two elements.

In Fig. 5.6, we show how the ratio of lattice conductivity to electron mobility varies with grain size for the same Si–Ge alloy. The ratio λ_L/μ is, generally, reduced when the grain size becomes smaller, though there is a minimum value for the ratio when the effective grain size is about ten times the mean path length of the phonons.

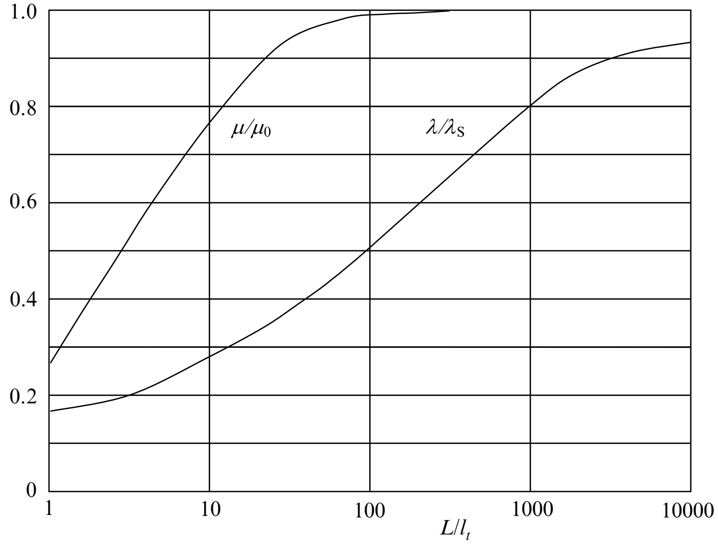


Fig. 5.5 Calculated ratio of lattice conductivity to its value for a large crystal plotted against ratio of effective grain size to phonon mean free path for Si–Ge at 300 K. Also shown is the ratio of electron mobility to its value for a large crystal

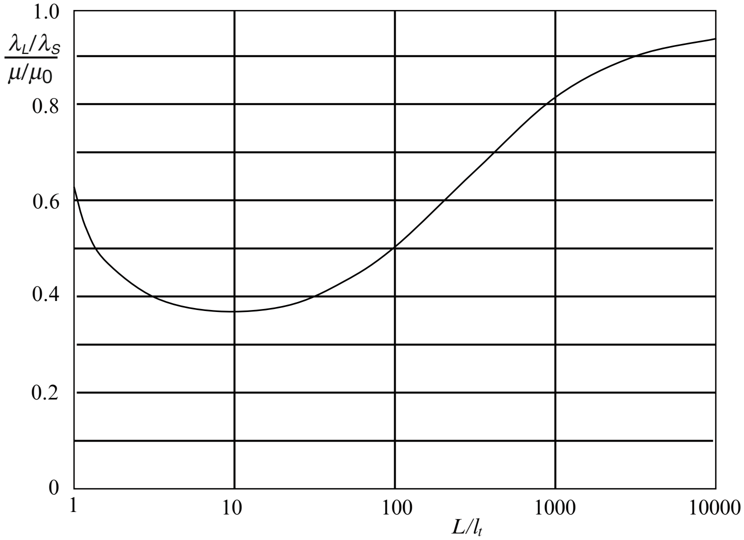


Fig. 5.6 Calculated values of $(\lambda_L \lambda_S) / (\mu \mu_0)$ plotted against ratio of effective grain size to phonon mean free path in a Si–Ge alloy at 300 K

Silicon–germanium alloys are not useful thermoelectric materials at room temperature so the predicted improvement in the ratio of electron mobility to lattice conductivity, shown in Fig. 5.6, is not immediately applicable to practical applications. It is worth noting, however, that there is experimental confirmation that the lattice conductivity can be significantly reduced by boundary scattering. Savvides and Goldsmid [11] found that the lattice conductivity of undoped $Si_{70}Ge_{30}$ at 300 K falls from 8.2 to 4.3 W/m K when the grain size is reduced to 2 μm in polycrystalline-sintered material. At this grain size, one expects very little deterioration of the carrier mobility. Polycrystalline Si–Ge alloys have, in fact, been used in thermoelectric generation.

5.5 Fine-Grained Material with Large Unit Cells

In the above discussion, it has been implicit that the model of a simple continuum is adequate for the phonons that contribute to the heat conduction process. However, most of the materials that are used in thermoelectric devices have several atoms in each unit cell so that there are more optical modes than acoustic modes in the vibrational spectrum. We shall now consider the situation that exists for more complicated crystal structures. We shall still use the Debye theory for the low-frequency acoustic modes, which we understand will make a major contribution to the thermal conductivity. Our object in this section is to estimate the boundary scattering effect when there are optical modes as well as acoustic modes.

We think that most of the heat may still be carried by the acoustic modes of lower frequency. The value of v in (3.47) for the lattice conductivity should be that for the group velocity rather than the phase velocity. The group velocity is determined from the slope of the dispersion curve and, as in Fig. 3.7, this will be relatively low for the high-frequency acoustic modes and all the optical modes. Nevertheless, because of their large number, it may well be that there is a significant contribution from the optical phonons but its magnitude may be difficult to assess. Consequently, we find it best to describe the effect of boundary scattering, in materials that have large unit cells, in terms of the difference between the lattice conductivities in large and small grained material rather than as a ratio, as in (5.25).

The frequency at which Umklapp scattering takes over from boundary scattering is

$$\omega_0 = \sqrt{\frac{v}{BL}}. \tag{5.32}$$

Thus, we shall determine the contribution to the lattice conductivity from the phonons with frequencies up to this limit with and without boundary scattering. Using the Debye model to determine the specific heat of these low-frequency modes, we find that their contribution to the thermal conductivity is

$$\lambda_{\omega<\omega_0} = \frac{1}{9} v L c \omega_0^3, \tag{5.33}$$

whereas the contribution from the same modes in the absence of boundary scattering is

$$\lambda_{\omega<\omega_0} = \frac{1}{3} v^2 \frac{c\omega_0}{B} = \frac{1}{3} v L c \omega_0^3. \tag{5.34}$$

Thence the reduction in the thermal conductivity due to boundary scattering is

$$\Delta\lambda_{\mathrm{L}} = \frac{2}{9} L v c \omega_0^3. \tag{5.35}$$

Our problem, then, is to find ω_0.

The value of the Umklapp scattering parameter B can be estimated from Lawson's relation, (4.10) and this, in turn, leads to the determination of ω_0. It is found that

$$\omega_0 = \sqrt{\frac{a\omega_{\mathrm{D}}^2 C}{9L\alpha_{\mathrm{T}}^2 \rho d\, v^2 T}}, \tag{5.36}$$

where, as before, α_{T} is the thermal expansion coefficient and ρ_{d} is the density. C is the total specific heat per unit volume and a is the lattice constant that we can set equal to the cube root of the volume per unit cell.

This approach [10] was originally used to estimate the boundary scattering effect in the half-Heusler alloy $\mathrm{Zr_{0.5}Hf_{0.5}NiSn}$. It was calculated that a reduction in the lattice conductivity should be noticeable if the grain size were to fall below about 10 μm. Subsequently, thermal conductivity measurements were performed on polycrystalline samples of a similar half-Heusler alloy, $\mathrm{TiNiSn}_{1-x}\mathrm{Sb}_x$, by Bhattacharya et al. [12]. It was found that there was good correlation with the theory that is presented in this section, there being a substantial reduction of the lattice conductivity as the grain size fell from 10 to 1 μm.

5.6 Phonon-Glass Electron-Crystal

The idea of a crystal with a glass-like thermal conductivity was first put forward by Slack [13]. It is well known that a glass or an amorphous substance has the lowest range of thermal conductivity of any type of material. It is, perhaps, strange to refer to a lattice conductivity in a material that does not have a crystal structure but it is convenient to use this term for the non-electronic contribution. If we attempt to use (3.47) to describe this lattice conductivity, inserting measured values for the specific heat and speed of sound, we find that the free path length for the phonons is equal to the diameter of each atom. There will be variations in the thermal conductivity from one glass to another due to variation in v and c_V but the range is very small around an average value of about 0.25 W/m K. This contrasts with the lattice conductivity in crystals, which can vary over several orders of magnitude.

Amorphous semiconductors are unlikely to be good thermoelectric materials because they invariably have very small values for the electrical conductivity, stemming from low carrier mobilities. It is just possible that an amorphous material might prove useful if the low mobility is compensated for by a high density-of-states effective mass. For the thermoelectric materials that are in use today, Chasmar and Stratton's parameter β, as defined in (4.2), has a value of about 0.3. It is unlikely that an amorphous material will have a high figure of merit unless β is already of this order in the crystalline state. The material will, probably, not be improved as we change from a crystalline to a glassy structure unless the mean free path for the phonons in the crystal is greater than the mean free path of the charge carriers.

Nolas and Goldsmid [14] have discussed the criteria that must be met. They supposed that the value of β would actually be equal to 0.3 in the crystalline state and they found that the condition for the charge carriers to have a shorter free path than the phonons is

$$\frac{9}{80\pi}\left(\frac{h^2}{mk}\right)\left(\frac{\Theta_\mathrm{D}}{\alpha^2 T^2 N_\mathrm{v}^{1/3}(m^*/m)}\right) < 1, \qquad (5.37)$$

where key quantities are the density-of-states effective mass, m^*, and the number of valleys, N_v, in the appropriate energy band.

In a thermoelectric material for use at room temperature, the Debye temperature might be of the order of 300 K and for all substances a is about 0.5 nm. Under these conditions, the inequality (5.37) leads to the requirement that $N_\mathrm{v}^{1/3} m^*/m > 17$, an unlikely state of affairs. On the other hand, the requirement becomes less rigorous at high temperatures. If we suppose that Θ_D and T are again similar to one

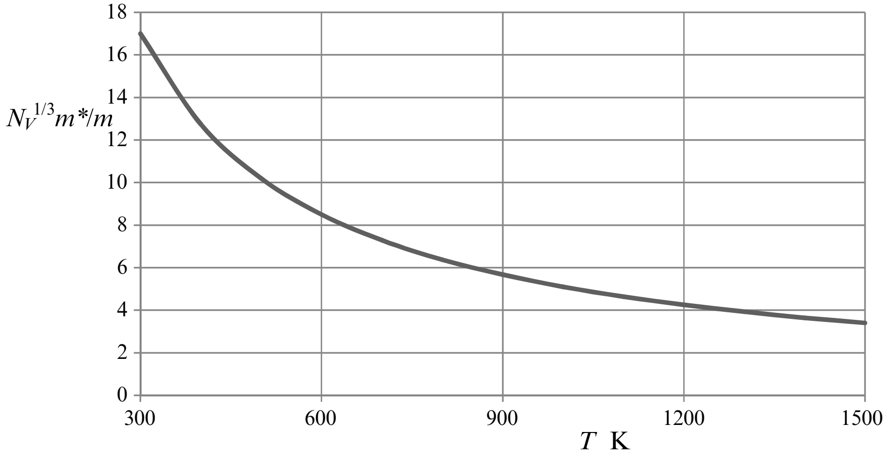


Fig. 5.7 The requirement for improvement in the figure of merit on the transition from the crystalline to the amorphous state. An improvement is likely if the parameter $N_\mathrm{v}^{1/3}m^*/m$ lies above the curve

another, the necessary value for $N_\mathrm{v}^{1/3}m^*/m$ becomes smaller as T becomes larger. The minimum value of $N_\mathrm{v}^{1/3}m^*/m$ is plotted against temperature in Fig. 5.7. At a temperature of, say, 1,200 K, $N_\mathrm{v}^{1/3}m^*/m$ would need to be no more than 4 for the amorphous form of the material to be interesting.

Slack's approach was rather different. He directed attention towards materials that are essentially crystalline rather than amorphous but which have lattice conductivities that are very similar to those in true glasses. Thus, the thermal conductivity of a glass is more-or-less independent of T at high temperatures but varies as T^3 at low temperatures because this is the way that the specific heat behaves. This behaviour is reproduced in the crystalline materials to which Slack directed his attention. Because the electronic properties of these materials are essentially no different from those of other crystals, they have been called phonon glass-electron crystals or PGECs.

The characteristics of crystals that have glass-like thermal conductivities were detailed by Cahill et al. [15]. They contain atoms that are loose in the sense that they do not have unique positions in the lattice. These loose atoms do not have fixed positions relative to each other. They are not like the impurity atoms that one uses to dope semiconductors since their concentrations may be relatively high, at least 3%. The structures that can accommodate these loose atoms exhibit rather large open cages formed by more stable atoms. The so-called rattling motion of the loose atoms is responsible for intense phonon scattering.

The materials that embody the PGEC principle will be discussed in later chapters. Suffice it to say, for the moment, that there do exist certain materials that display glass-like lattice conductivities combined with electronic properties that are not unlike those of other semiconductors.

References

1. A.F. Ioffe, S.V. Airepetyants, A.V. Ioffe, N.V. Kolomoets, L.S. Stil'bans, Dokl. Akad. Nauk. SSSR **106**, 981 (1956)
2. S.V. Airapetyants, B.A. Efimova, T.S. Stavitskaya, L.S. Stil'bans, L.M. Sysoeva, Zh. Tekh. Fiz. **27**, 2167 (1957)
3. P.G. Klemens, Phys. Rev. **119**, 507 (1960)
4. J. Callaway, Phys. Rev. **113**, 1046 (1959)
5. J.E. Parrott, Proc. Phys. Soc. **81**, 726 (1963)
6. J.R. Drabble, H.J. Goldsmid, *Thermal Conduction in Semiconductors* (Pergamon, Oxford, 1961), p. 193
7. H.B.G. Casimir, Physica **5**, 495 (1938)
8. H.J. Goldsmid, A.W. Penn, Phys. Lett. **27**, 523 (1968)
9. N. Savvides, H.J. Goldsmid, J. Phys. C **6**, 1701 (1973)
10. J.W. Sharp, S.J. Poon, H.J. Goldsmid, Phys. Stat. Sol., A **187**, 507 (2001)
11. N. Savvides, H.J. Goldsmid, J. Phys. C **13**, 4657, 4671 (1980)
12. S. Bhattacharya, M.J. Skove, M. Russell, T.M. Tritt, Y. Xia, V. Ponnambalam, S.J. Poon, N. Thadhani, Phys. Rev. B **77**, 184203 (2008)
13. G. Slack, in CRC *Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC Press, Boca Raton, 1994), p. 407
14. G.S. Nolas, H.J. Goldsmid, Phys. Stat. Sol., A **194**, 271 (2002)
15. D.G. Cahill, S.K. Watson, R.O. Pohl, Phys. Rev. B **46**, 6131 (1992)

Chapter 6
The Improvement of a Specific Material – Bismuth Telluride

6.1 Pure Bismuth Telluride

The first report that bismuth telluride, Bi_2Te_3, is an effective thermoelectric material appeared in 1954 [1]. A thermocouple made from a p-type sample of the compound connected to a negative thermoelement made from bismuth was found to yield a cooling effect of 26 K below ambient temperature by means of the Peltier effect.

Bismuth telluride forms single crystals that are markedly anisotropic in their mechanical properties. The crystal structure is such that the bismuth and tellurium atoms are arranged in parallel layers following the sequence:

$$-\mathrm{Te}^{[1]} - \mathrm{Bi} - \mathrm{Te}^{[2]} - \mathrm{Bi} - \mathrm{Te}^{[1]}-,$$

which is continually repeated. Strong covalent-ionic bonds exist between the Bi atoms and the Te atoms on both types of site, but the layers of $\mathrm{Te}^{[1]}$ atoms are bound to neighbouring $\mathrm{Te}^{[1]}$ layers only by weak van der Waals forces [2]. It is found that crystals of bismuth telluride are easily cleaved along the direction of the layers, normal to the trigonal or c-direction. Perpendicular to the c-axis, there are two a-axes inclined at 60° to each other.

It is not only the mechanical properties that are different in the plane of the a-axes and the c-direction. For example, the electrical and thermal conductivities are higher parallel to the cleavage planes than perpendicular to them.

Bismuth telluride was selected as a material to be studied on account of its high mean atomic weight. It also has the relatively low melting temperature of 585°C and satisfies the criteria set by both Ioffe and Ioffe [3] and by Keyes [4] for a low lattice conductivity. The material in the first experiments was produced by the zone melting of a mixture of the elements in the correct proportions. No attempt was made to optimise the doping level but it so happens that melt-grown bismuth telluride does not have the stoichiometric formula Bi_2Te_3. Instead, there is an excess of Bi atoms, with corresponding vacancies on some of the Te sites. The excess Bi atoms appear to act as acceptor impurities leading to p-type conduction. The number of acceptors is such that the Fermi energy lies close to the optimum value, with the Seebeck coefficient equal to $220\,\mu\mathrm{V/K}$. The original sample was aligned with current flow perpendicular to the c-direction and the electrical and thermal conductivities were found to be $4.0\times 10^{4}\,\Omega^{-1}\,\mathrm{m}^{-1}$ and 2.1 W/m K, respectively. The figure of merit z_p was no more than about $0.9\times 10^{-3}\,\mathrm{K}^{-1}$ and, at that time, there was no n-type bismuth telluride to complete the couple. In fact, a properly aligned crystal of bismuth would have provided a negative branch with at least the figure of merit of the positive bismuth telluride branch but only a polycrystalline sample was available. Poor as was the performance of the Bi_2Te_3–Bi couple, it established the fact that semiconductor thermoelements are superior to metals.

There is no reason why the negative and positive branches of a thermocouple should consist of the same element or compound. Nevertheless, it so happens that the figure of merit of optimised n-type bismuth telluride is almost the same as that of optimised p-type material. Thus, by 1955, after n-type bismuth telluride had been obtained through the addition of the donor impurity, iodine, it was found that a couple made from both types of the compound gave a Peltier cooling effect of 40 K below room temperature [5]. The figure of merit Z for the couple could have been no more than about $1.2\times 10^{-3}\,\mathrm{K}^{-1}$ with ZT equal to about 0.35. During the next couple of years, the techniques for producing uniformly doped bismuth telluride of both conductivity types were improved so that the properties could be truly optimised. It was found [6] that the optimum electrical conductivity is close to $1.0\times 10^{5}\,\Omega^{-1}\,\mathrm{m}^{-1}$ and the dimensionless figure of merit for the best couple made from Bi_2Te_3 is then about 0.6.

These early results for bismuth telluride validated the theoretical work on the selection and optimisation of materials. It also brought into prominence the deleterious effect of minority carriers. Thus, the plot of Seebeck coefficient against electrical conductivity in Fig. 6.1 shows not only the increase of Seebeck coefficient as the electrical conductivity is reduced in the extrinsic region but also the decrease of $|\alpha|$ in the mixed and intrinsic regions. Undoped melt-grown Bi_2Te_3 is p-type with the electrical conductivity slightly higher than that for which the Seebeck coefficient has its maximum value. Higher electrical conductivities are obtained by doping with an acceptor impurity such as lead. If, instead, a donor impurity, such as iodine, is added the material becomes intrinsic and then n-type.

In Fig. 6.2, we show how the thermal conductivity varies with electrical conductivity for the same samples of bismuth telluride as those used to obtain Fig. 6.1. As one might have expected, the thermal conductivity generally rises with increasing electrical conductivity due to the increase in the electronic component. However, at low values of the electrical conductivity, the thermal conductivity again becomes larger. Clearly, the Lorenz number is becoming much greater than expected for a single type of charge carrier. This is a convincing demonstration of the importance of the bipolar heat conduction effect. For intrinsic bismuth telluride, the Lorenz number is about $25\,(k/e)^2$ compared with about $2\,(k/e)^2$ for extrinsic material. One can extrapolate the almost linear plot in the extrinsic region towards the vertical axis to obtain the lattice conductivity which has the value of about 1.6 W/m K.

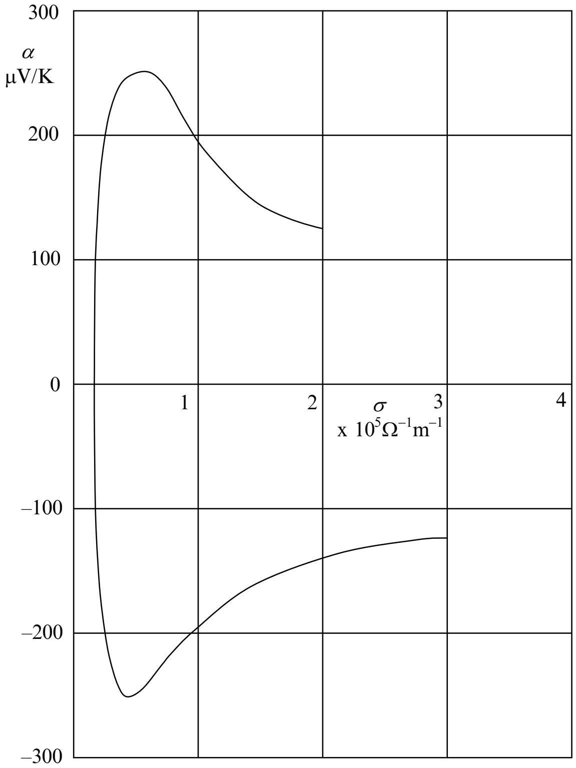


Fig. 6.1 Seebeck coefficient plotted against electrical conductivity for p-type and n-type bismuth telluride at 300 K

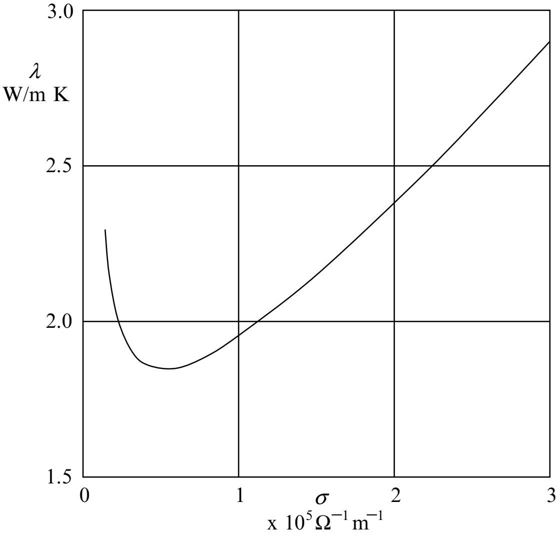


Fig. 6.2 Plot of thermal conductivity against electrical conductivity for bismuth telluride at 300 K

6.2 Band Structure of Bismuth Telluride

One of the consequences of the detailed study of the thermoelectric properties of bismuth telluride was the availability of high-quality single crystals of the compound. This allowed a large number of physical measurements to be made, thereby enhancing our knowledge of electronic transport in the material.

As is only to be expected for a substance with such marked anisotropic mechanical properties, the electrical and thermal conductivities parallel to the cleavage planes are different from those in the perpendicular direction. The lattice conductivity in the c-direction is less than that in the plane of the a-axes by a factor of 2.1. This would make the c-direction preferable for thermoelectric applications were it not for the fact that the electrical conductivity is even more strongly anisotropic. The anisotropy of the electrical conductivity is different for n-type and p-type material and, at least for the former, it varies with the doping level as shown in Fig. 6.3. The anisotropy of about 2.7 for hole conduction is not much greater than the anisotropy of the lattice conductivity so, although the p-type figure of merit is somewhat the smaller for current flow in the c-direction, the difference is not great. On the other hand, the anisotropy factor for electron conduction is equal to at least 4, so it is most unfavourable for the current flow to be in the c-direction for n-type thermoelements made from bismuth telluride. This is an important factor if one wishes to make use of randomly oriented polycrystalline specimens. Although such material is preferable to single crystals from the mechanical viewpoint, it leads to a substantial degradation of the figure of merit for the negative branches. In actual fact, single crystals are usually too fragile for practical applications but one can make use of melt-grown material in which the cleavage planes all lie parallel to the growth direction but not parallel to one another.

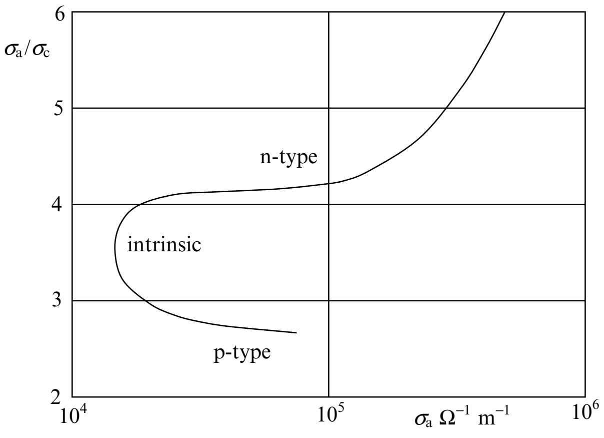


Fig. 6.3 Ratio of electrical conductivity in the a-direction to that in the c-direction plotted against σ_a for single crystal bismuth telluride at 300 K

One of the quantities that one wishes to know in any new semiconductor is the mobility of the electrons and holes. This should be independent of the position of the Fermi level, if the conductor is non-degenerate and extrinsic, provided that the scattering is entirely due to the lattice vibrations. In isotropic semiconductors, the mobility is easily found by measuring the Hall coefficient, R_H, and the electrical conductivity. In an extrinsic conductor with only one type of carrier the Hall coefficient is given by [7]

$$R_H = \pm \frac{a_H}{ne}, \tag{6.1}$$

where a_H is a parameter close to unity that depends on the scattering law. For acoustic mode lattice scattering a_H is equal to $3\pi/8$, whereas for a fully degenerate conductor it is equal to 1. Since the electrical conductivity is equal to $ne\mu$, the mobility is given by $|R_H\sigma/a_H|$. It is often convenient to make use of the so-called Hall mobility given by $|R_H\sigma|$.

For a single crystal with the structure of bismuth telluride, there will be two different Hall coefficients depending on the orientation of the current and the magnetic field. One might hope to obtain the carrier concentration, n, by taking some average of the Hall coefficients but it turns out that this leads to an error by a factor of more than 2. The true carrier concentration can be determined from the Hall coefficient only if the ratios between the inertial effective masses in each energy band are known.

It is possible to find the ratios between the effective masses as well as the number of valleys and the tilt of the constant energy surfaces in wave vector space by making comprehensive galvanomagnetic measurements. Such measurements were first carried out for single crystals of p-type and n-type Bi_2Te_3 by Drabble et al. [8, 9]. There are two components of the electrical resistivity, two Hall coefficients and eight magnetoresistance coefficients. Less than half these coefficients are needed once the number of valleys in the band is given but, if all are measured, the additional information can be used to confirm the validity of any hypothetical model. Reference is made to the basic cell in wave vector space known as the first Brillouin zone, its shape for bismuth telluride being shown in Fig. 6.4.

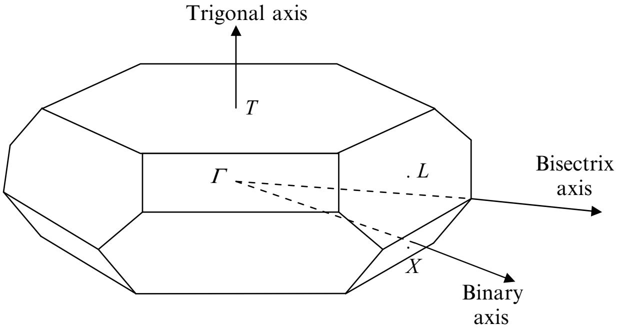


Fig. 6.4 First Brillouin zone for Bi_2Te_3 and other crystals with the same structure

The surfaces of constant energy in wave vector space should be ellipsoidal provided that the electrons or holes have energies that are close to those at the band edges. The surfaces may be tilted with respect to the axes provided that crystal symmetry is maintained. Various possibilities exist for the number of valleys but restrictions are placed so as to satisfy the symmetry conditions. Drabble et al. were able to show that only a six-valley model could be made to fit the experimental data. They described the bands in terms of reciprocal effective mass tensors of the form α_{ij}/m which are such that the energy of a charge carrier is given by

$$E = E_0 - \frac{h^2}{2m}\left(\alpha_{11}k_2^1 + \alpha_{22}k_2^2 + \alpha_{33}k_3^2 + 2\alpha_{23}k_2k_3\right), \tag{6.2}$$

where $\boldsymbol{k}$ is the wave vector for the carriers. It should be noted that the galvanomagnetic measurements establish the shape of the energy surfaces but not the absolute values of the effective masses. The latter require the additional knowledge of the position of the Fermi level. The extra data were provided by Bowley et al. [10], who observed the Seebeck coefficient and its change in a magnetic field. The Seebeck coefficient on its own gives a reasonable indication of the Fermi energy but a more accurate value requires knowledge of the scattering law.

The scattering law can be determined by measurement of either the magneto-Seebeck coefficient or the Nernst coefficient. If a sufficiently large transverse magnetic field could be applied so that the condition $(\mu B)^2 \gg 1$ were satisfied, the Seebeck coefficient would eventually reach the value that it would have in zero field if r were equal to zero. One could, therefore, find the value of r by subtracting the Seebeck coefficient, α_0, in zero field from that, α_∞, in a very large field. Thus, from (3.32)

$$|(\alpha_\infty - \alpha_0)| = -kr/e. \tag{6.3}$$

However, even at liquid nitrogen temperature, the mobility of the carriers is too small for the high field condition to be met with any available magnet. Instead, one has to be satisfied with measurements under the low field condition $(\mu B)^2 \ll 1$. It can then be shown [10] that

$$\frac{\sigma\Delta\alpha}{\Delta\sigma} = \frac{k}{e}(2 - \frac{1}{\beta^*})r, \tag{6.4}$$

where β^* is equal to $\left(1 - \rho\Delta\rho/R_{\mathrm{H}}^2B^2\right)$. $\Delta\sigma$ and $\Delta\rho$ are the changes in electrical conductivity and resistivity, respectively, in the magnetic field B. $\Delta\alpha$ has the same sign as α if r is negative and the opposite sign if r is positive. Bowley et al. found that their measurements were consistent with r being equal to $-1/2$. It would also have been possible to determine r from the measurement of the Nernst coefficient as was done by Mansfield and Williams [11].

Table 6.1 shows the band parameters that were determined for p-type material by Drabble et al. after the Fermi energy of their samples had been established. For n-type material, the improved data of Caywood and Miller [12] are used. The band extrema lie on the reflection planes that contain the trigonal and bisectrix directions.

Table 6.1 Band parameters for p-type and n-type Bi_2Te_3

Parameter	Valence band	Conduction band
Number of valleys	6	6
Location in ***k***-space	On reflection planes	On reflection planes
α_{11}	19.8	26.8
α_{22}	3.26	4.12
α_{33}	4.12	3.72
α_{23}	1.0	2.4

The reciprocal effective mass tensor α_{11}/m is referred to axes x and z in the reflection plane and y in the perpendicular direction.

In actual fact, although the band model has been described in terms of six valleys, the galvanomagnetic data would fit equally well for three valleys with the extrema at the boundaries of the Brillouin zone. This uncertainty was removed by work on the de Haas-van Alphen effect [13] and by studies of the reflectance minima associated with the plasma edges [14]. Both experiments showed that there are six valleys with the extrema inside the zone. Thus, the good thermoelectric properties of bismuth telluride lend credibility to the idea that the parameter β should be large for semiconductors with both a high mean atomic weight and a multi-valley band structure.

There is empirical support for the principle that a good material will have as small an energy gap as is consistent with the requirement that there should be only one type of carrier. This means that the energy gap will probably be a few times kT in width. That, in turn, suggests that materials with different energy gaps will be needed in each region of temperature. This suggestion is borne out by the experimental data.

The energy gap of a semiconductor can be found either from the variation of the carrier concentration with temperature in the intrinsic region [15] or from optical transmission studies. It is usually satisfactory to determine the rate of change of the electrical conductivity with temperature, since the intrinsic carrier concentration has an exponential temperature dependence. It is easy enough to take account of the much slower temperature variation of the mobility.

The observation of the wavelength of the optical absorption edge is probably the most reliable method of finding the energy gap if due account is taken of the fact that the Fermi level may lie inside the valence or conduction band. Austin's value [16] of 0.13 eV for the gap was later confirmed by Greenaway and Harbeke [17]. The optical studies reveal what cannot be determined from conductivity measurements, namely the dependence of the gap width on temperature. It is found that the gap becomes larger as the temperature falls, the temperature coefficient being -9.5×10^{-5} eV K^{-1} down to -155°C.

It is noted that the energy gap has a value of no more than about $5.2kT$ at room temperature so it does not quite satisfy our requirement of at least $6kT$ to avoid minority carrier conduction. This should not matter too much for refrigeration below room temperature but should be significant in applications involving higher temperatures. To some extent, the problem can be overcome by doping the compound more heavily.

Recently, consideration has been given to the possibility of using ionised-impurity scattering to improve the figure of merit of a narrow-gap material. It was pointed out by Ioffe [18] that the increase of the Seebeck coefficient brought about by the change in the scattering law due to impurity scattering might more than compensate for the fall in the carrier mobility. The effect has never been exploited in materials with a single type of carrier and the improvement would only be marginal in that case. However, when there are both electrons and holes in a narrow-gap semiconductor or semimetal, the situation is different [19]. In a bismuth telluride-like model with zero energy gap, the figure of merit becomes almost twice as great if lattice scattering is augmented by ionised-impurity scattering. The improvement would be less for an energy gap as wide as it actually is in bismuth telluride but there could still be a significant advantage in having some ionised-impurity scattering, particularly, for applications above room temperature. It would, of course, be necessary to optimise the carrier concentration, so counter-doping with donors and acceptors would have to be carried out.

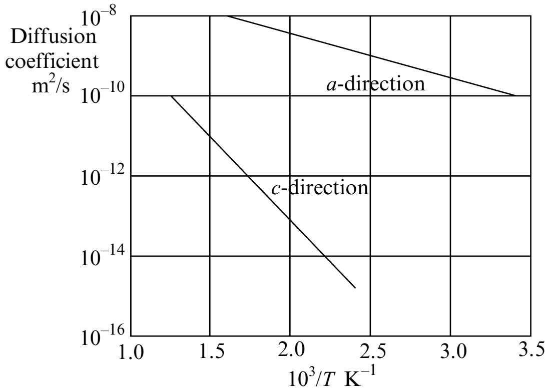
6.3 Diffusion in Bismuth Telluride

Early attempts to manufacture thermoelectric coolers based on bismuth telluride received a setback when it was discovered that the observed performance fell far short of that expected from laboratory measurements of the parameters involved in the figure of merit. It is not easy to solder directly to bismuth telluride though some success has been achieved using bismuth for this purpose. More reliable contacts of low electrical resistance are obtained after first electroplating the ends of the thermoelements with metallic layers. The poor performance of the thermocouples was observed when copper was employed as the plating material. It seemed that copper was diffusing rapidly through the thermoelements, acting as a donor impurity and changing the Seebeck coefficient and the electrical conductivity. The problem appeared to be solved when nickel was used to replace copper as the contact material.

However, even when nickel-plated thermoelements were used, further problems were experienced. In one particular assembly procedure, the solder in contact with the nickel-plated ends remained molten for several minutes. Thermoelectric modules that had been made using this procedure were found to have an extremely low cooling power. It was subsequently established that copper, which had been dissolved in the solder, was able to diffuse through the plating and then travel into the bismuth telluride. It was clear that the diffusion coefficient of copper in bismuth telluride must be very high, indeed.

The phenomenon of fast copper diffusion can be explained qualitatively by assuming that a Cu^+ ion is very small and is able to fit easily between the adjacent layers of $Te^{[1]}$ atoms that are so weakly bound together. This idea was confirmed by

Fig. 6.5 Diffusion coefficient of copper in bismuth telluride in the *c*-direction and along the plane of the *a*-axes



Carlson [20] who measured the diffusion coefficient for copper in bismuth telluride, as a function of temperature, in both the *a*- and *c*-directions. His results are shown in Fig. 6.5. The diffusion coefficient in the plane of the *a*-axes is higher at 300 K than it is in the *c*-direction at 800 K.

It is noteworthy that copper will diffuse out from bismuth telluride just as readily as it diffuses in. The copper that contaminates a sample can be removed by immersing it in an aqueous solution; dilute hydrochloric acid has been found suitable for this purpose.

These remarks about copper diffusion apply not only to single crystals but also to the material that is produced by zone melting in which there is a clear path for the copper ions through the cleavage planes from one end of a sample to the other. However, copper does not seem to be able to diffuse readily through sintered bismuth telluride with its randomly orientated grains. Thus, copper cannot be ruled out as a useful doping agent in this type of material.

In view of the reasons for the rapid diffusion of copper in bismuth telluride, it is not surprising that other elements from the same group of the periodic table, namely silver and gold, behave in more-or-less the same way.

6.4 Solid Solutions Based on Bismuth Telluride

It was not long after the establishment of bismuth telluride as a thermoelectric material that the ideas of Ioffe et al. [21] on semiconductor solid solutions became known. There seemed to be no reason why the beneficial effect on the thermal conductivity of PbTe, on alloying it with PbSe or SnTe, should not be duplicated on alloying Bi_2Te_3 with isomorphous compounds such as Sb_2Te_3 and Bi_2Se_3.

If the further conclusions of Airapetyants et al. [22], on the importance of substituting on the appropriate sub-lattice, are valid, then one would expect solid solutions between Bi_2Te_3 and Sb_2Te_3 to be the best p-type materials. Similarly, alloys of

Bi_2Te_3 with Bi_2Se_3 would make better n-type thermoelements. In fact, the solid solutions based on bismuth telluride do conform to this pattern, though not necessarily for the reasons that Airapetyants et al. put forward. As has already been mentioned, bismuth telluride itself, when grown from the melt, is non-stoichiometric and at ordinary temperatures is a p-type extrinsic semiconductor. As the compound is alloyed with Sb_2Te_3, the non-stoichiometry becomes even more pronounced and it would be rather difficult to add sufficient donor impurities, in a controlled manner, to make these solid solutions n-type. On the other hand, when Bi_2Se_3 is added to Bi_2Te_3, the material becomes less strongly p-type and is more easily doped so as to have the optimum electron concentration for a negative thermoelement.

There is no doubt that the solid solutions have lattice conductivities that are lower than that of pure bismuth telluride. This is borne out by the measurements of all groups of workers [23–25]. The actual values for λ_L differ from one set of results to another but this is possibly due to the different methods for estimating the electronic thermal conductivity. In Figs. 6.6 and 6.7, we show the lattice conductivity in the $(Bi–Sb)_2\,Te_3$ and $Bi_2\,(Te–Se)_3$ systems, respectively, based on the results of Rosi et al. [23]. The peculiar maximum in the $Bi_2\,(Te–Se)_3$ system when the concentration of Bi_2Se_3 is about 60% was not noticed by Birkholz [24] although a similar peak was observed at about 25% Bi_2Se_3 by Champness et al. [26]. It so happens that the best n-type materials have relatively low amounts of Bi_2Se_3 and the observations for higher concentrations do not have any practical significance.

Of course, the important question that must be answered is whether or not the mobility of the charge carriers is affected by the formation of the solid solutions. In the $(Bi–Sb)_2\,Te_3$ system, there actually seems to be an increase in the value of $\mu\,(m^*/m)^{3/2}$ as the proportion of Sb_2Te_3 rises, as is apparent from the slight increase in the electrical conductivity at a given value of the Seebeck coefficient. There is, thus, no doubt that $(Bi–Sb)_2\,Te_3$ is superior to Bi_2Te_3 for p-type thermoelements. All observers seem to agree that the minimum lattice conductivity is

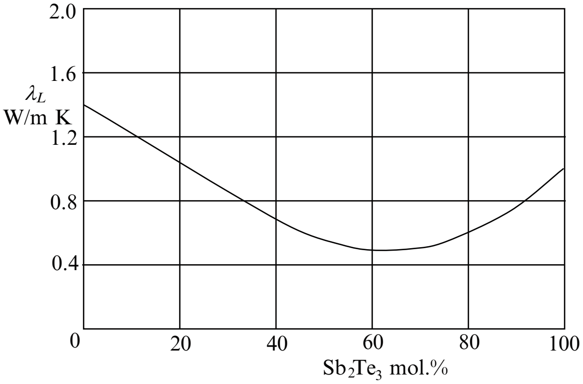


Fig. 6.6 Lattice thermal conductivity at 300 K plotted against concentration of Sb_2Te_3 in $(Bi - Sb)_2\,Te_3$ solid solutions (based on the data of Rosi et al. [23])

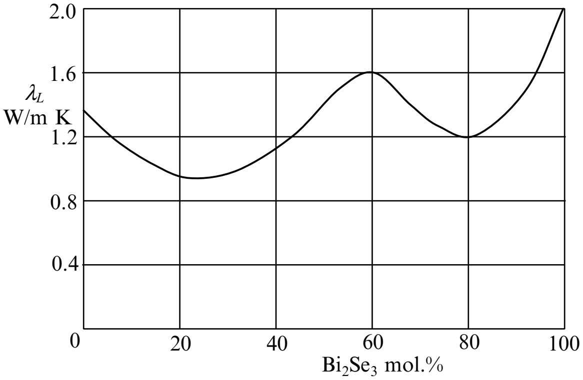


Fig. 6.7 Lattice thermal conductivity at 300 K plotted against concentration of Bi_2Se_3 in $(Bi - Sb)_2\,Te_3$ solid solutions (based on the data of Rosi et al. [23])

to be found when the concentration of Sb_2Te_3 is about 75%. There is some speculation that even more Sb_2Te_3 might be advantageous but it is difficult to achieve the optimum Seebeck coefficient when there is more than 75% of this compound in the alloy. Sb_2Te_3 itself always displays near-metallic properties with a Seebeck coefficient of less than $100\,\mu V\,K^{-1}$ but it is uncertain whether this is due to overlapping valence and conduction bands or to the inability to add enough donor impurities to compensate for gross non-stoichiometry. It is interesting that materials with the composition $Bi_{0.5}Sb_{1.5}Te_3$ are still referred to as bismuth telluride alloys or even, loosely, as bismuth telluride, in spite of the fact that they are composed mostly of Sb_2Te_3.

The optimum n-type composition has not been determined with the same confidence. All that seems to be established is that the proportion of Bi_2Se_3 in the $Bi_2\,(Te–Se)_3$ alloy should be quite small. The quantity $\mu\,(m^*/m)^{3/2}$ for the electrons certainly becomes less as Bi_2Se_3 is added, but the fall in the lattice conductivity more than compensates for this up to a certain concentration. Beyond that concentration, the deterioration in $\mu\,(m^*/m)^{3/2}$ is such that the small benefit from a further fall in λ_L is not worthwhile. In any case, although other workers do not agree with Rosi et al. [23] on the magnitude of λ_L, they all set the minimum lattice conductivity at a Bi_2Se_3 concentration of no more than 20%. The composition $Bi_2Se_{0.3}Te_{2.7}$ is often selected for the n-type material that is used in modules and this cannot be far from the optimum composition.

Goldsmid and Delves [27] made a direct comparison between the thermoelectric properties of selected bismuth telluride alloys and the best Bi_2Te_3 that was available. This comparison has the advantage that the materials were prepared using the same zone-melting apparatus and the measurements were all carried out using the same equipment. The reliability of the results was confirmed by assessing the performance of thermocouples made from the compound and alloys. The figure of merit, of course, varies with the carrier concentration, but the optimum electrical conductivity, not unexpectedly, has more-or-less the same value for both p-type and

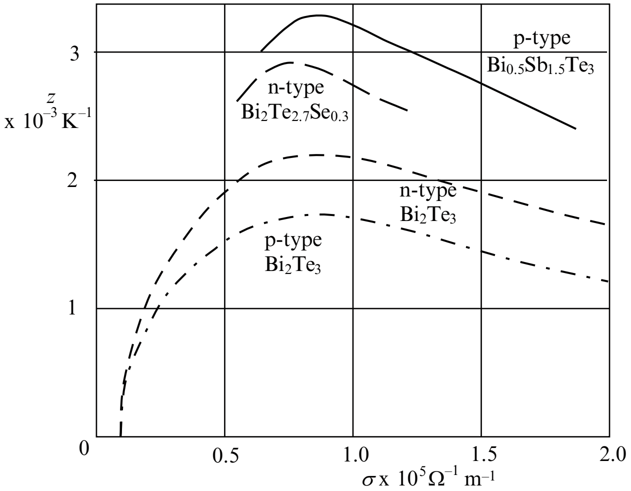


Fig. 6.8 Plot of the thermoelectric figure of merit against electrical conductivity for p-type and n-type Bi_2Te_3 and specified solid solutions

n-type material and for both the solid solutions and the pure compound. The figure of merit at 20°C is plotted against electrical conductivity in Fig. 6.8.

The compositions that were developed in the late 1950s and the early 1960s have been used in the manufacture of modules since that time. It is possible that marginal improvements might result from the simultaneous addition of Sb_2Te_3 and Bi_2Se_3 to Bi_2Te_3. Yim and Rosi claimed that $Bi_{0.5}Sb_{1.5}Te_{2.91}Se_{0.09}$ with excess Te is the best p-type material and that $Bi_{1.8}Sb_{0.2}Te_{2.85}Se_{0.15}$ doped with SbI_3 gives the highest n-type figure of merit. They observed values for z equal to $3.3 \times 10^{-3}K^{-1}$ at 300 K for both materials and it was widely accepted until recently that a dimensionless figure of merit, zT, equal to unity is the best that can be achieved using bismuth telluride alloys.

6.5 Practical Developments

One of the problems that was encountered by the manufacturers of thermoelectric modules was the fact that the thermoelements rarely displayed the best properties that had been observed in the laboratory. We shall discuss the zone melting technique in a later chapter but we shall mention here that material that is grown too rapidly has a non-uniform composition. There can be variations in the proportions of the major components in an alloy and, what is probably more important, variations in the concentration of the dopants that are used to control the carrier concentration. Non-uniform material obviously cannot have the optimum composition throughout its volume but one might hope that small variations would not have too much effect. However, the tolerable departures in a material like bismuth telluride with a rather small energy gap are not very large since the figure of merit falls off rapidly as soon as mixed conduction appears. Furthermore, non-uniformity in a thermoelectric material can give rise to internal circulating currents associated with the Seebeck effect, producing unwanted heat flow. There is also evidence for additional phases, such as the deposition of tellurium in the grain boundaries, appearing under some circumstances, and this may adversely affect the transport properties. One of the major achievements since the introduction of the bismuth telluride alloys has been the understanding of the growth factors leading to better control of materials production.

Bismuth telluride and its alloys are brittle materials. Single crystals are particularly fragile, and in preparing the samples that were used by Drabble et al. [8, 9] in their galvanomagnetic measurements, it was necessary to etch away several millimetres on all sides before it could be certain that there were no cracks. It has been stated that zone melting and similar processes produce materials in which the cleavage planes all lie parallel to the direction of motion of the liquid–solid interface. This is not quite true. There is often some tilt of the cleavage planes, particularly, if the interface is not planar. This means that the cleavage planes in a thermoelement that is cut from the grown material may be inclined to the length direction. Subsequent fracture can have a catastrophic effect on a module in which such a thermoelement is incorporated. One of the achievements has been better alignment of the cleavage planes so that any cracks have a minimal effect on the flow of current.

Although bismuth telluride has such weak bonding between the $\mathrm{Te}^{[1]}$ layers, it is otherwise strongly bonded. Thus, randomly oriented polycrystalline material has superior mechanical properties to single crystals or aligned polycrystals. Randomly oriented material can be produced by various sintering processes but the resulting thermoelements usually have inferior properties to those of aligned material.

The mechanical strength of polycrystals combined with the high figure of merit of single crystals can be achieved in aligned polycrystalline material of small grain size. We shall discuss the processes for obtaining such alignment later but it is convenient to consider the properties of partially aligned thermoelements at this point. In a sintered sample, the grains are symmetrically arranged around the direction of pressing. Following a suggestion by Penn [29], we suppose that, in such a situation, the amount of material having its c-axis inclined at a polar angle θ to the direction of symmetry is proportional to $\cos^n \theta \mathrm{d}\theta$. The quantity n is then a measure of the degree of alignment. It is found that the thermoelectric properties perpendicular to the axis of symmetry are

$$\alpha_{\perp} = \frac{(2+n)\alpha_{\mathrm{a}}\sigma_{\mathrm{a}} + \alpha_{\mathrm{c}}\sigma_{\mathrm{c}}}{(2+n)\sigma_{\mathrm{a}} + \sigma_{\mathrm{c}}}, \tag{6.5}$$

$$\sigma_{\perp} = \frac{(2+n)\sigma_{\mathrm{a}} + \sigma_{\mathrm{c}}}{n+3}, \tag{6.6}$$

$$\lambda_{\perp} = \frac{(2+n)\lambda_{\mathrm{a}} + \lambda_{\mathrm{c}}}{n+3}. \tag{6.7}$$

while in the direction of symmetry

$$\alpha_{\parallel} = \frac{(1+n)\alpha_{\mathrm{c}}\sigma_{\mathrm{c}} + 2\alpha_{\mathrm{a}}\sigma_{\mathrm{a}}}{(1+n)\sigma_{\mathrm{c}} + 2\alpha_{\mathrm{a}}}, \tag{6.8}$$

$$\sigma_{\parallel} = \frac{(1+n)\sigma_{\mathrm{c}} + 2\sigma_{\mathrm{a}}}{n+3}, \tag{6.9}$$

$$\lambda_{\parallel} = \frac{(1+n)\lambda_{\mathrm{c}} + 2\lambda_{\mathrm{a}}}{n+3}. \tag{6.10}$$

The above equations allow for anisotropy of the Seebeck coefficient but are probably in error if such anisotropy exists. This is because non-uniformity of the Seebeck coefficient will give rise to the circulating currents that have already been mentioned and which may significantly increase the thermal conductivity. Fortunately, there is no anisotropy of the Seebeck coefficient in the extrinsic bismuth telluride alloys that are used for making devices.

Equations (6.6) to (6.10) can be used to assess the degree of alignment in samples of bismuth telluride but must be employed with caution for sintered material since pressing faults can distort the behaviour. For this reason, an X-ray diffraction technique for assessing the alignment was thought to be more reliable [30].

Cold or hot pressing followed by sintering and annealing does not produce substantial alignment but Kim et al. [31] found that an extrusion process produces material in which there is significant orientation of the grains. The extruded $(\mathrm{Bi{-}Sb})_2\,\mathrm{Te}_3$ produced by these authors appeared to be just as good as monocrystalline material but that is not too surprising since the figure of merit is not strongly anisotropic in p-type crystals. However, the strong alignment achieved by extrusion should be particularly valuable for n-type material where the anisotropy is much stronger.

Experts in sintering technology usually regard the density of the final product as a measure of their success. Density is probably not a good yardstick for sintered thermoelements since sometimes highly dense samples have a poor figure of merit. In fact, there could be some advantage in using less dense thermoelements as long as the figure of merit remains high. Provided that the ratio of electrical to thermal conductivity stays unchanged, it does not matter if the material is porous. This is demonstrated by the low-density samples with a good figure of merit that were produced by Durst et al. [32]. It has recently been shown [33] that porous bismuth telluride should be useful in the production of synthetic transverse thermoelements.

Another area in which advances are being made is in the production of thick or thin films of thermoelectric material. There is a trend towards the reduction of the amount of material in a thermoelectric module. In theory, by reducing both the length and the cross-sectional area, one should be able to economise on the amount of material that is needed for a given cooling power. In practice, as we shall see later, there are problems associated with electrical contact resistance and thermal losses as the size of the thermoelements is reduced but that does not seem to deter the manufacturers from their attempts to miniaturise their devices. Recent work [34] using the flash evaporation technique shows that thin films of $Bi_{0.4}Sb_{1.6}Te_3$ have a power factor of 3.49 $W/m\,K^2$ which means that they are comparable in quality with bulk material for which the power factor is typically about 4 $mW\,m^{-1}\,K^{-2}$.

6.6 Extension of the Temperature Range

Most applications of bismuth telluride and its alloys are restricted to cooling from a heat sink at room temperature using a single stage. There is sometimes the need, however, to use the materials outside the range of, say, 250 to 300 K. For example, the low temperature stages of a thermoelectric cascade operate below this temperature region and generators of electricity from low-grade heat sources need modules that are designed for higher temperatures.

The selection of materials for use outside the usual region of temperature has been considered by Kutasov et al. [35]. The properties below 200 K have been discussed by Vedernikov et al. [36]. As a general principle, thermoelements operated at low temperatures should be less heavily doped. They may, in fact, display substantial transport by minority carriers at room temperature in order that the Seebeck coefficient is optimised at some lower temperature. Figure 6.9 shows how the Seebeck coefficient and electrical resistivity vary with temperature for two different n-type bismuth telluride alloys. Sample 1 would be suitable for use at ordinary temperatures whereas sample 2 would be better at lower temperatures.

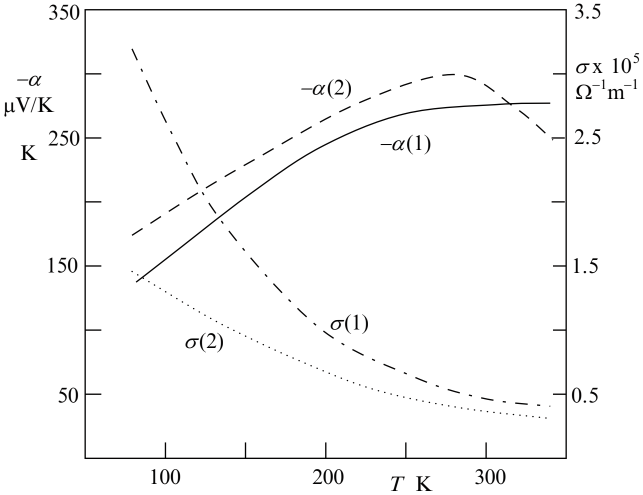


Fig. 6.9 Seebeck coefficient and electrical conductivity plotted against temperature for n-type alloys of bismuth telluride. A schematic plot based on the data of Kutasov et al. [35]. Sample 1 is suitable for use at ordinary temperatures while sample 2 is a low-temperature material

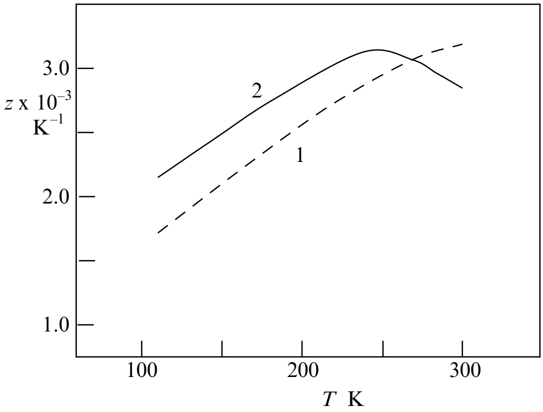


Fig. 6.10 Figure of merit plotted against temperature for two bismuth telluride alloys. Samples 1 and 2 are similar to those in Fig. 6.9. A schematic plot based on the data of Vedernikov et al. [36]

The difference between the figures of merit of bismuth telluride alloys with different compositions can be substantial. Thus, Fig. 6.10 shows plots of z against temperature for two different compositions similar to those of the samples in Fig. 6.9.

A feature that becomes apparent when bismuth telluride or any of its alloys are studied at low temperatures is that the effective mass is not really a constant. If we assume that classical statistics are valid, the Seebeck coefficient varies with temperature according to the relation

$$|\alpha| = \frac{k}{e}\left[\text{constant} + \ln\left(m^*\right) + \frac{3}{2}\ln(T)\right], \tag{6.11}$$

which may be derived from (3.29) and (3.31).

Even the earliest measurements [37] on bismuth telluride indicated that, although the plots of the Seebeck coefficient against $\ln(T)$ were linear, the slope was not equal to $3k/2e$. The temperature dependence of the effective mass, as derived from Seebeck coefficient and resistivity measurements, has been highlighted by Kutasov et al. [35]. Their observations are somewhat confused, with differences between the temperature variation of the effective mass and its magnitude from one sample to another. Nevertheless, it must be agreed that their data are consistent with the fact that the energy band structure of bismuth telluride is more complicated than was at first supposed. There is good evidence for the presence of a second conduction band with its edge 0.03 eV above that of the main band [12, 38].

Turning now to the use of bismuth telluride alloys above room temperature, the most important change to make is the addition of dopants to maintain, as far as possible, extrinsic conduction over the full operating region. It would, of course, be helpful if the energy gap could be increased and it appears that this does occur if the alloys contain substantial concentrations of bismuth selenide. This is indicated by the plot of energy gap against composition due to Greenaway and Harbeke [17] and shown in Fig. 6.11.

Fig. 6.11 Variation of energy gap with concentration of Bi_2Se_3 in $Bi_2(Te\text{-}Se)_3$ alloys based on the observations of Greenaway and Harbeke [17]

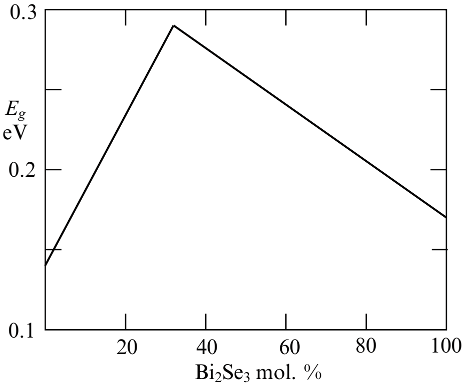
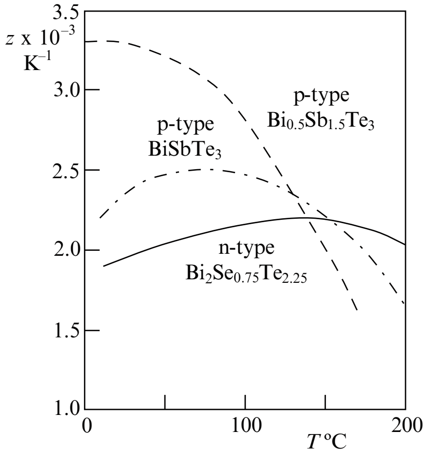


Fig. 6.12 Figure of merit for p-type and n-type bismuth–telluride alloys above ambient temperature. Data selected by Matsuura and Rowe [39]



The discontinuity in Fig. 6.11 at a composition containing about 35% Bi_2Se_3 is probably due to a shift in the position of the band extrema. However, Caywood and Miller [12] reported that a six-valley model for the conduction band in Bi_2Se_3 is similar to that in Bi_2Te_3, though the surfaces of constant energy are more nearly spherical.

In view of the interest in generating electricity from low-grade heat sources, it is surprising that relatively little work seems to have been carried out on the improvement of bismuth telluride alloys above ambient temperature. Matsuura and Rowe [39] included a review of the properties of bismuth telluride alloys up to about 150°C in their work on low-temperature heat conversion and their data are included in Fig. 6.12. It does not appear that these materials have been optimized, but it is noteworthy that in their best n-type material the proportion of Bi_2Se_3 was 25%.

6.7 Recent Advances

During the past few years, there have been several reports of values of zT substantially greater than unity. Most of these advances have been associated with the use of nanostructures in one form or another, although Yamashita and Tomiyoshi [40] have claimed that zT can be as high as 1.41 for p-type $Bi_{0.5}Sb_{1.5}Te_3$ and 1.19 for n-type $Bi_2Te_{2.82}Se_{0.18}$. These authors used a large excess of tellurium in their p-type specimens and CuBr as a dopant for the n-type material. However, these high values remain to be confirmed by other workers.

We shall discuss nanostructures in detail in a later chapter. Briefly, they can lead to improvements in the figure of merit due to both an enhancement of the electronic properties and a decrease in the lattice conductivity. Perhaps, the best validated improvement in the bismuth telluride system is that obtained [41] from a superlattice based on Bi_2Te_3 and Sb_2Te_3. Successive layers of these two compounds were of the order of 100 nm in thickness. In p-type material zT at 300 K was found to be as high as 2.4 while an n-type superlattice consisting of Bi_2Te_3 and $Bi_2Te_{2.83}Se_{0.17}$ yielded a value for zT of 1.4. The flow of current was in the cross-plane direction; that is, in the direction of the c-axis. The lattice conductivity is low in this orientation even for a bulk crystal but, in the superlattice, it was found to be as small as 0.22 W/m K. In fact, the high values for the figure of merit can be attributed entirely to the decrease in the lattice conductivity. It seems possible that nanostructures on an even smaller scale are needed before significant changes in the electronic properties can take place. It seems remarkable the carrier mobility remains of the same order as in bulk material in spite of the strong scattering of phonons in the superlattice structure.

A substantial improvement of the figure of merit of bismuth telluride alloys has also been reported for a completely different arrangement. Ghoshal et al. [42] applied multiple-point metal contacts between flat surfaces of p-type and n-type material. The same materials in a commercial module displayed a figure of merit ZT of 0.84, but in Ghoshal's device, this rose to 1.4. As in the superlattices described by Venkatasubramanian et al. [41], the improvement of zT is presumably due entirely to a reduction in the lattice conductivity. This work will be discussed in more detail later.

Since the lattice conductivity has already been reduced to what must be very close to the value for amorphous material, it seems unlikely that there can be further advances using the bismuth telluride alloys unless band structure changes can be produced in new nanostructures. However, if the reduction of lattice conductivity can be incorporated in commercial thermoelectric modules, this will have a major effect on their performance. Perhaps, the best prospects lie in what are called bulk nanostructures. These materials would have embedded nanoscale inclusions to produce strong phonon scattering and would probably be much easier to manufacture than superlattices and the like.

References

1. H.J. Goldsmid, R.W. Douglas, Br. J. Appl. Phys. **5**, 386 (1954)
2. J.R. Drabble, C.H.L. Goodman, J. Phys. Chem. Solids **5**, 142 (1958)
3. A.V. Ioffe, A.F. Ioffe, Dokl. Akad. Nauk. SSSR **97**, 821 (1954)
4. R.W. Keyes, Phys. Rev. **115**, 564 (1959)
5. H.J. Goldsmid, J. Electronics **1**, 218 (1955)
6. H.J. Goldsmid, A.R. Sheard, D.A. Wright, Br. J. Appl. Phys. **9**, 365 (1958)
7. J.S. Blakemore, *Solid State Physics (second edition)* (Cambridge University Press, Cambridge, 1985), p. 167
8. J.R. Drabble, R. Wolfe, Proc. Phys. Soc. B **69**, 1101 (1956)
9. J.R. Drabble, R.D. Groves, R. Wolfe, Proc. Phys. Soc. **71**, 430 (1958)
10. A.E. Bowley, R.T. Delves, H.J. Goldsmid, Proc. Phys. Soc. **72**, 401 (1958)
11. R. Mansfield, W. Williams, Proc. Phys. Soc. **72**, 733 (1958)
12. L.P. Caywood, G.R. Miller, Phys. Rev. B **2**, 3209 (1970)
13. L.R. Testardi, P.J. Stiles, E.H. Burshstein, Bull. Amer. Phys. Soc. **7**, 548 (1962)
14. R. Sehr, L.R. Testardi, J. Appl. Phys. **34**, 2754 (1963)
15. J. Black, E.M. Conwell, L. Seigle, C.W. Spencer, J. Phys. Chem. Solids **2**, 240 (1957)
16. I.G. Austin, Proc. Phys. Soc. **72**, 545 (1958)
17. D.L. Greenaway, G. Harbeke, J. Phys. Chem. Solids **26**, 1585 (1965)
18. A.F. Ioffe, *Semiconductor Thermoelements and Thermoelectric Cooling* (Infosearch, London, 1957), p. 146
19. H.J. Goldsmid, J. Thermoelectricity, **4**, 31 (2006)
20. R.O. Carlson, J. Phys. Chem. Solids **13**, 65 (1960)
21. A.F. Ioffe, S.V. Airepetyants, A.V. Ioffe, N.V. Kolomoets, L.S. Stil'bans, Dokl. Akad. Nauk. SSSR **106**, 981 (1956)
22. S.V. Airapetyants, B.A. Efimova, T.S. Stavitskaya, L.S. Stil'bans, L.M. Sysoeva, Zh. Tekh. Fiz. **27**, 2167 (1957)
23. F.D. Rosi, B. Abeles, R.V. Jensen, J. Phys. Chem. Solids **10**, 191 (1959)
24. U. Birkholz, Z. Naturforsch. A **13**, 780 (1958)
25. H.J. Goldsmid, J. Appl. Phys. **32**, 2198 (1961)
26. C.H. Champness, W.B. Muir, P.T. Chiang, Can. J. Phys. **45**, 3611 (1967)
27. H.J. Goldsmid, R.T. Delves, GEC Jnl. **28**, 102 (1961)
28. W.M. Yim, F.D. Rosi, Solid State Electron **15**, 1121 (1972)
29. A.W. Penn, Private communication
30. H.J. Goldsmid, F.A. Underwood, Adv. Energy Conversion **7**, 297 (1968)
31. S.S. Kim, S. Yamamoto, T. Azaiwa, J. Alloys Compounds **375**, 107 (2004)
32. T. Durst, H.J. Goldsmid, L.B. Harris, J. Mater. Science Lett. **16**, 2632 (1981)
33. H.J. Goldsmid, Phys. Stat. Solidi A **12**, 2966 (2008)
34. M. Takashiri, K. Miyazaki, H. Tsukamoto, Thin Solid Films **516**, 6336 (2008)
35. V.A. Kutasov, L.N. Lukyanova, M.V. Vedernikov, in *Thermoelectrics Handbook: Macro to Nano*, ed. by D.M. Rowe (CRC Taylor and Francis, Boca Raton, 2006), p. 37–1
36. M.V. Vedernikov, V.A. Kutasov, L.N. Luk'yanova, P.P. Konstantinov, *Proceedings of the 16th International Conference on Thermoelectrics* (IEEE, Dresden, 1997), p. 56
37. H.J. Goldsmid, Ph.D. thesis, University of London (1957)
38. R.B. Mallinson, J.R. Rayne, R.W. Ure, Phys. Rev. **175**, 1049 (1968)
39. K. Matsuura, D.M. Rowe, in *CRC Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC Press, Boca Raton, 1994), p. 573
40. O. Yamashita, S. Tomiyoshi, J. Appl. Phys. **95**, 161 (2004)
41. R. Venkatasubramanian, E. Silvola, T. Colpitts, B. O'Quinn, Nature **413**, 597 (2001)
42. U. Ghoshal, S. Ghoshal, C. McDowell, L. Shi, S. Cordes, M. Farinelli, Appl. Phys. Lett. **80**, 3006 (2002)

Chapter 7
Methods for the Production of Materials

7.1 General Principles

The development of thermoelectric refrigerators and generators after the 1950s stemmed from the work on semiconductors for use in transistors and other electronic devices. These applications of semiconductors were made possible through the availability of highly pure and perfect single crystals that could be doped with minute quantities of donor and acceptor impurities. The effects that were exploited were for the most part of an interfacial nature generally associated with internal barriers in silicon and germanium and, later, the III–V compounds. The problems in producing good thermoelectric materials were different but often equally demanding. Thus, the carrier concentrations in thermoelectric materials are usually orders of magnitude greater than those in conventional semiconductor devices but the proportional variation has to be more strictly controlled. Most of the conventional semiconductors have cubic structures whereas thermoelectric materials are often composed of anisotropic crystals. Although single crystals are not often required, orientation of grains is sometimes of great significance.

The range of elements and compounds that can be classed as thermoelectric materials is very wide. At the low temperature end, we have semiconductors and semimetals with relatively low melting points, like bismuth and bismuth telluride. At the high temperature end, we encounter refractory oxides and silicides. In between there are materials with simple crystal structures, such as the Si–Ge alloys, and complex materials like some of the clathrates and skutterudites. Thus, although thermoelectric materials may all have similar transport properties within their particular range of operation, they may need quite different methods of production.

We have to discuss growth from the melt and sintering of bulk materials as well as thin and thick film deposition by physical and chemical techniques. In this chapter, we shall attempt to discuss the various production techniques as generally as possible but reference will be made to specific materials to illustrate the different methods.

7.2 Growth From the Melt

It is often possible to produce a thermoelectric compound by melting together the constituent elements. Such a melt and freeze technique can be useful in producing the starting materials for, say, a powder metallurgy process but will not generally lead to a controlled distribution of impurities nor will it lead to any preferred orientation, if that is needed. Usually, then, some kind of directional freezing technique is adopted.

A simple example of direction freezing is that in which a sealed quartz capsule containing the thermoelectric material is slowly lowered out of a furnace that is maintained above the melting temperature of the charge. Solidification starts at the bottom of the charge and proceeds upwards. This type of procedure is used in the growth of single crystals by the Bridgman method. The end of the capsule is conical in shape and both the temperature gradient and the rate of growth are accurately controlled.

Bismuth telluride and its alloys are sometimes produced by such a vertical technique but horizontal methods are also encountered. The charge is placed in a quartz boat and, very often, no more than a short length is molten at any time. The molten zone is gradually moved along the boat. Zone melting was introduced by Pfann [1] as a method of purifying silicon and germanium but is useful in controlling the distribution of impurities in thermoelectric materials.

We must consider the composition of a solid that is deposited at the freezing interface with a molten material. In Fig. 7.1, we show the equilibrium phase diagram for two cases, (a) in which the impurities raise the melting point and (b) in which they lower it. The distribution coefficient, k, is the ratio of the concentration in the solid to that in the liquid at equilibrium.

We suppose that the melt is lowered in temperature until the point A is reached. Then, in case (a), the material that solidifies will have an impurity concentration that is higher than that in the liquid, as indicated by the point B. This will deplete the

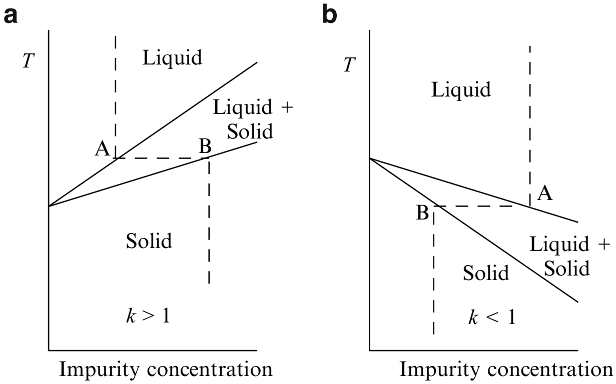


Fig. 7.1 Equilibrium phase diagrams showing the liquidus and solidus lines for different values of the segregation coefficient k. (**a**) $k > 1$ and (**b**) $k < 1$

impurities in the liquid as the freezing process continues so that the points A and B will gradually move to the left. In case (b), the liquid that freezes at point A will be in equilibrium with a solid at point B that has a lower concentration of impurity. As the interface moves on, the impurities in the liquid will become more concentrated. This is the basis of the zone refining technique, in which repeated passes of molten zones sweep the impurities from one end of an ingot to the other. Even before the advent of zone refining, directional freezing was used to purify metals, but the repeated traverse of narrow molten zones is much more effective.

When zone melting is used in the production of thermoelectric materials like bismuth telluride, it serves a dual purpose. It aligns the grains so that, if there is a preferred direction for the current flow, thermoelements can be cut appropriately from the resulting ingot. Furthermore, the impurities can be distributed uniformly if their distribution in the starting material is suitably arranged.

Suppose that we wish to have an impurity concentration c in the solid; then, the liquid that deposits solid with this composition at the interface must have a composition c/k. We may use as our starting material an ingot with the impurity concentration c that has been melted and rapidly frozen so that it is uniform on a macroscopic scale. We remove a section of the cast ingot equal to the zone length and replace it with the same length of casting that has the impurity concentration c/k. Then, as shown in Fig. 7.2, an ingot that is uniform, except at the last end to freeze, results from the passage of a single zone. This is the loaded zone technique.

Another technique, known as reverse-pass zone melting, also produces reasonably uniform material, at least over the central part of an ingot. A molten zone is formed at one end of a casting and passed to the opposite end, where, if $k < 1$, it will have acquired an excess concentration of impurities. The zone is then made to traverse the ingot in the opposite direction.

We have assumed that the liquid and solid are in equilibrium at all times. Of course, this is not so unless the zone moves at an infinitesimal speed. Let us, therefore consider what happens when the zone moves forward at some arbitrary speed. The composition of the liquid is no longer uniform and there is a transition region between the liquid and the solid that, because of segregation, has a reduced melting temperature. Referring to Fig. 7.1, it will be seen that, irrespective of whether k is greater than or less than unity, the effect of segregation is to reduce the liquidus temperature. Figure 7.3 shows how the liquidus temperature is affected close to the interface.

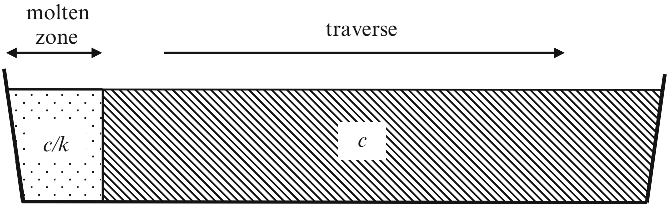


Fig. 7.2 The loaded zone technique. The molten zone with an impurity concentration c/k traverses the casting that has an impurity concentration c

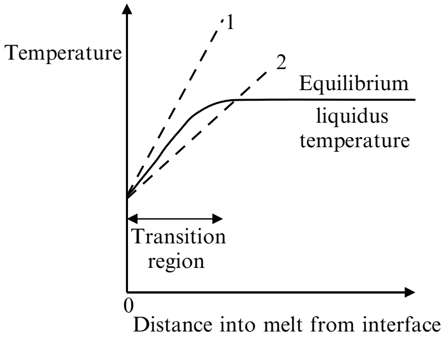


Fig. 7.3 Temperature in the molten zone near the solid–liquid interface. The solid curve represents the equilibrium liquidus temperature and the broken lines give the temperature gradient in the melt for two different situations as discussed in the text

The width of the transition region is controlled by the speed with which the molten zone moves away from the frozen solid and on the rate of diffusion of the impurities in the liquid. If the temperature gradient is high enough, as indicated by line 1, no particular problems arise. However, if the temperature gradient is low, as for line 2, a phenomenon known as constitutional supercooling occurs. Freezing tends to occur in advance of the interface. In practice, solid cells form in advance of the main front and liquid with a greater impurity concentration and a lower melting temperature is trapped between these cells. This causes the ingot produced by the zone melting process to be inhomogeneous on a microscopic scale. This usually means that the thermoelectric properties will be non-uniform and the figure of merit will be reduced.

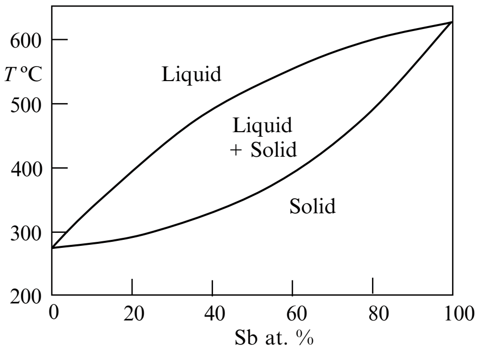
The problem was recognised by Tiller et al. [2] who specified the conditions under which constitutional supercooling can be avoided. These authors assumed that the melt is unstirred and it seems clear that stirring would reduce the width of the transient region. However, it was shown by Hurle [3] that the condition for constitutional supercooling is independent of whether or not the liquid is stirred since a boundary layer with a gradient of composition will always remain. Hurle's condition for the avoidance of constitutional supercooling is

$$\frac{\mathrm{d}T}{\mathrm{d}x} > v\left(\frac{L\rho_\mathrm{d}}{\lambda_\mathrm{S}} - \frac{mC_\mathrm{S}\lambda_\mathrm{L}(1-k)}{D\lambda_\mathrm{S}k}\right), \tag{7.1}$$

where v is the speed of the zone, L is the latent heat, ρ_d is the density, m is the slope of the liquidus line, C_S is the concentration of solute in the solid, and D is the diffusion coefficient in the liquid. The thermal conductivities in the solid and liquid are λ_S and λ_L, respectively. It may often be assumed that the first term in the brackets on the right-hand side of the inequality is small compared with the second term. Also, near the melting temperature, there will probably be little difference between the thermal conductivities in the liquid and solid states. We can then use a simpler expression for avoiding constitutional supercooling, namely

$$\frac{\mathrm{d}T}{\mathrm{d}x} > \frac{\mathcal{V}\Delta T}{D}, \tag{7.2}$$

Fig. 7.4 Schematic phase diagram for the Bi–Sb system [5]



where ΔT is the difference between the liquidus and solidus compositions at the temperature in which we are interested.

We may illustrate the problem by referring to the Bi–Sb alloy system. Within a certain range at the bismuth-rich end, the alloys are narrow-gap semiconductors with promising thermoelectric properties at low temperatures. The phase diagram for the Bi–Sb system is shown in Fig. 7.4. It is apparent that there is a very great difference between the liquidus and solidus temperatures so the quantity ΔT in (7.2) is very large as soon as the Sb concentration reaches more than 1 or 2%. The concentration of antimony in the liquid is only about one-tenth of that in the crystallising solid. Also, because of the low melting temperature, it is difficult to establish a very large temperature gradient near the growth front. Thus, the growth of single crystals of Bi–Sb from the melt requires the use of an extremely low rate of movement of the liquid–solid interface, if the material is to be uniform in its properties. This has been realised by all recent workers on Bi–Sb. Thus, Uher [5], who produced single crystals containing 10% Sb by a horizontal zone melting process, allowed the interface to move at no more than $1\,\mathrm{mm\,h^{-1}}$. Uher attempted to increase the temperature gradient by using a water-cooled jacket near the end of the molten zone but he found that too steep a gradient led to lineage in his crystals.

The use of a slow growth rate requires the ambient temperature to be kept constant, the heater power to be very precisely controlled, and the drive for the movement of the zone to be very steady. As a result of his studies, Uher was able to present a diagram that specifies the conditions for preventing constitutional supercooling in Bi–Sb alloys over a wide range of compositions. His diagram is shown in Fig. 7.5.

Brown and Heumann [7] grew single crystals of Bi–Sb containing 13% Sb. They used a steeper gradient than Uher but their results agreed with the conditions that he specified. Thus, they found that when $(\mathrm{d}T/\mathrm{d}x)/v$ was $5.2 \times 10^{10}\,\mathrm{K\,s^{-1}\,m^{-2}}$, their crystal was homogeneous but when this ratio was equal to $1.3 \times 10^{10}\,\mathrm{K\,s^{-1}\,m^{-2}}$, it was not. The observations by Yim and Amith [8] of the conditions for the appearance of cellular growth in $\mathrm{Bi_{0.95}Sb_{0.5}}$ also agree with Fig. 7.5. There can, therefore, be no doubt that the principles for the avoidance of supercooling given in this section are soundly based on experimental evidence.

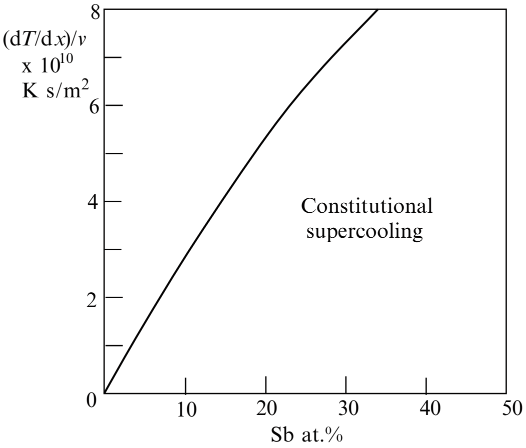


Fig. 7.5 The conditions for the avoidance of constitutional supercooling in Bi–Sb [6]. The ratio of the temperature gradient to the speed of the molten zone should lie to the left of the curve

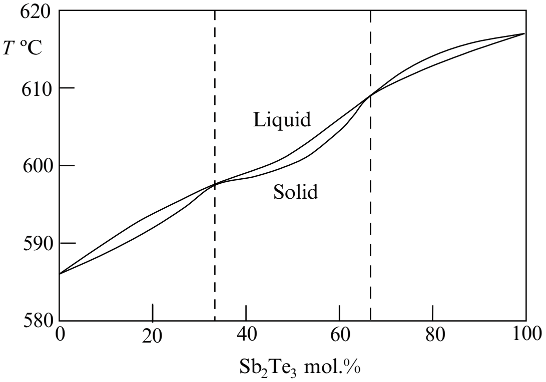


Fig. 7.6 Schematic phase diagram for the Bi_2Te_3–Sb_2Te_3 alloy system. The meeting of the liquidus and solidus curves at Sb_2Te_3 concentrations of one-third and two-thirds suggests that there may be some measure of ordering at these compositions

Inhomogeneity in Bi–Sb is important because the transport properties depend very strongly on the relative concentrations of the two constituents. However, the problem in the bismuth telluride alloys is different. The liquidus and solidus curves lie very close together in the $(Bi\text{–}Sb)_2Te_3$ alloys system [9] as shown in Fig. 7.6 and, in any case, the band parameters change only slowly as the ratio of Sb_2Te_3 to Bi_2Te_3 varies. What is much more important is any inhomogeneity in the impurity concentration.

The separation of the liquidus and solidus curves [10] is larger for the $Bi_2(Se\text{–}Te)_3$ system than for $(Bi\text{–}Sb)_2Te_3$, as shown in Fig. 7.7, but is still not very great for the rather small proportions of Bi_2Se_3 in which we are usually interested. Once again, it is variations of the impurity concentration rather than of the alloy composition that are the more significant.

It is probable that the improvement over the years in the figure of merit of the bismuth telluride alloys that are used in the production of modules stems largely from better understanding of the necessary growth conditions. This is suggested by the experiments carried out by Cosgrove et al. [11] on $BiSbTe_3$. Cosgrove et al.

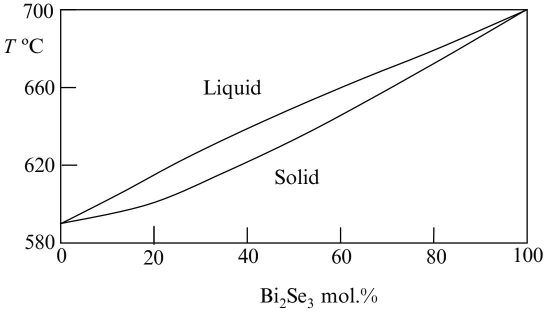


Fig. 7.7 Schematic phase diagram for the Bi_2Te_3–Bi_2Se_3 alloy system

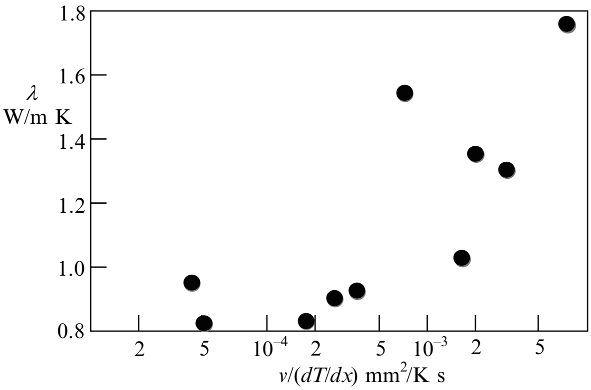


Fig. 7.8 Thermal conductivity of $BiSbTe_3$ grown at different speeds under different temperature gradients. Diagram based on the data of Cosgrove et al.[11]

varied the conditions for the growth of their material in a Bridgman furnace. The temperature gradient was 2.5, 5.0, or $25\,\mathrm{K\,mm^{-1}}$ and the growth rate varied between 2.2×10^{-4} and $4.2 \times 10^{-2}\,\mathrm{mm\,s^{-1}}$. One of the parameters that is most sensitive to non-uniformity of the impurity concentration is the thermal conductivity as it becomes larger when circulating thermoelectric currents are present. In Fig. 7.8, we show the variation of the thermal conductivity with the ratio $v/(\mathrm{d}T/\mathrm{d}x)$. It will be seen that there is a substantial rise in λ when this ratio rises above $2 \times 10^{-4}\,\mathrm{mm^2\,K^{-1}\,s^{-1}}$. These results emphasise the need for a slow growth speed and a steep temperature gradient when bismuth telluride alloys are grown by Bridgman or zone melting techniques.

7.3 Sintering

There is no doubt that thermoelectric materials that are produced by a sintering process have better mechanical properties than those that are usually prepared by growth from the melt. For those materials that have a cubic structure, there seems to be no reason why sintered samples should not be as good as large crystals provided that the grain size is not so small that the mobility of the charge carriers is reduced. There may, in fact, be an improvement in the figure of merit if the lattice conductivity is lowered by boundary scattering of the phonons.

The advantages of the sintering process have long been recognised for the production of bismuth telluride alloys. For these materials, that do not possess a cubic structure, it has also been realised that randomly orientated polycrystals may display some deterioration in the figure of merit if, as is usual, the properties differ substantially in the different crystallographic directions. In such materials, it is an advantage if some degree of alignment can be achieved during the sintering process.

Thermoelements have been successfully produced by sintering for many systems. For example, Si–Ge alloys have been prepared by a hot-pressing technique [12]. Sintering is one of the most common methods for producing PbTe and its alloys [13]. However, we shall discuss sintering with specific reference to the bismuth telluride alloys since we then have to deal with not only the quality of the thermoelectric properties in each grain but also the problem of grain alignment.

The basic steps in the sintering process are common for all materials. The starting components are prepared in the form of a powder which is then pressed in some type of mould. The material in this so-called green state is then heated so that the powder particles become bonded together. Further heating, or annealing, may then be employed to homogenise the product through solid-state diffusion and to remove imperfections in the structure.

There are, of course, differences in detail between the processes that are used for one material as compared with another. The temperature at which sintering occurs depends on the melting temperature. The temperature at which one might sinter bismuth telluride would be enough to melt bismuth or Bi–Sb. There are wide variations in the pressure that is used in the compaction of the powders. The powders themselves may consist of elements that combine chemically during sintering or they may consist of already reacted compounds. Heat may be applied while the pressing is carried out or cold-pressing may be used. Let us, then, look at some of the ways in which bismuth telluride and its alloys have been produced by sintering.

A typical procedure has been described by Cope and Penn [14]. The powder was prepared by milling material that had already been reacted in a furnace. They used a wide range of powder sizes, with particle diameters between 150 μm and 1 mm. A disadvantage of using fine powders is that there is then greater risk of atmospheric contamination, oxidation being known to affect the thermoelectric properties [15]. Normally, one selects particles within a certain range of sizes by using a pair of sieves with appropriate meshes. It seems important that the particles should be significantly smaller than the smallest dimension of a thermoelement since it is undesirable that the whole width should be spanned by a single grain. A wide range of sintering temperatures is possible. However, if the temperature lies below 300°C the process is very slow, while above 450°C distortion occurs. The sintered material produced by Cope and Penn was an n-type bismuth seleno–telluride alloy and was clearly inferior to zone-melted material. This is not surprising since the n-type figure of merit is so much less in the c-direction than it is along the cleavage planes.

The problem of aligning the grains in cold-pressed and sintered n-type bismuth telluride has been studied by Situmorang et al. [16]. In the first place, the powder was prepared in the form of plate-like particles as the shavings from ingots were turned on a lathe. These powders were then suspended in a dielectric liquid under a high electric field and allowed to settle in a steel die as the field was reduced. The samples were pressed at between 0.4 and 1 MPa and sintered under argon for 60 h at 420°C. X-ray diffraction was used to determine whether or not alignment had taken place. The possibility of alignment was also investigated by observation of the Seebeck effect parallel and perpendicular to the pressing direction. The Seebeck coefficient is isotropic in an extrinsic single crystal but becomes anisotropic as the intrinsic region is approached. Thus, α was measured up to a temperature of 260°C and it did indeed differ by over $20\,\mu\mathrm{V\,K^{-1}}$ in the two directions, at this temperature. This alone indicated substantial alignment though the difference between the Seebeck coefficients in a single crystal in the principal directions would be about $50\,\mu\mathrm{V\,K^{-1}}$.

In Sect. 6.5, we introduced a quantity n as a measure of the degree of alignment in polycrystalline bismuth telluride alloys. X-ray studies yielded a value for n of 4.0 and this was consistent with the Seebeck measurements. It was calculated that such alignment would allow sintered n-type material to attain a figure of merit within 10% of the single crystal value in the preferred direction.

The technique described by Situmorang et al. is encouraging in that it leads to an acceptable degree of alignment in a sintering process that involves cold pressing. However, it is unlikely to be practical for large-scale production so we should look at the approaches made by other workers.

The electrical conductivity often appears to be greatest in a direction perpendicular to that of pressing but this is commonly due to the existence of pressing faults rather than alignment. That is why the X-ray technique for determining preferred orientation is so valuable.

One obvious way to improve the process of sintering is by using hot pressing as was done in the early years of research on bismuth telluride by Airapetyants and Efimova [17]. Their samples were pressed at a temperature of 400°C at a pressure of 7×10^5 kPa. However, much of the recent work on the sintering of bismuth telluride has used rather different procedures. Thus, very good thermoelements have been made using spark plasma sintering [18].

In spark plasma sintering, a pulsed direct current flows between a graphite die and the powder charge within the die. This current promotes the heating of the powders during the pressing process. The advantages that are claimed for the spark plasma process are homogeneity, low operating temperatures, and a shortened sintering time. Jiang et al. [18] reported a high mechanical strength and exceptionally good thermoelectric properties for samples of $(\mathrm{Bi{-}Sb})_2\mathrm{Te}_3$ prepared in this way. A bending test showed a strength of 80 MPa compared with 10 MPa for the same alloy when produced by zone melting. Zone-melted material yielded a value of 1.0 for zT at 350 K, whereas the material produced by spark plasma sintering gave a value of 1.15. It is possible that zT has been over-estimated for both types of material since the thermal conductivity was measured indirectly using a laser-flash method to find the thermal diffusivity. This method has some advantages over the direct measurement of the thermal conductivity but can lead to greater errors. In spite of this reservation, it is obvious that the spark plasma sintered material has excellent properties and is probably a better material for use in devices than melt-grown $(Bi–Sb)_2Te_3$. Whether or not the same can be said for n-type material is another matter, but X-ray data on the samples produced by Jiang and his co-workers indicated significant alignment of the grains.

It was stated that the spark plasma sintered material was produced in the form of ingots that could then be sliced and diced to yield thermoelements for use in modules. If a simple sintering process is used, it is probably more convenient to press the individual thermoelements. A possible danger associated with the cutting process is the introduction of surface damage. It is known that surface damage can be minimised by using spark-erosion cutting [19] or an acid saw but it is much more practical to use rotating diamond-impregnated cutting wheels that are well lubricated. The care that must be taken to avoid extensive damage to the surfaces of thermoelements was highlighted by Alieva et al. [20].

One of the advantages of sintering is that the size of the grains in the final product can be controlled. Jaklovsky et al. [21] have investigated the effect on the thermal conductivity of reducing the grain size in sintered bismuth telluride. They found a substantial reduction in the lattice conductivity for grain diameters of less than about 200 μm and reported a peak in the figure of merit with a grain size of the order of 80 μm. However, in neither n-type nor p-type material was the figure of merit greater than can be obtained with zone-melted material and it cannot be taken as established that z can be improved by reducing the grain size in bismuth telluride-based thermoelectric alloys.

In recent years, the starting point for the sintering process has often been a mechanical alloying technique [22]. Elemental powders can be mechanically alloyed during a milling process and this procedure seems to be widely used in the metallurgical industry. One development of the method has been the bulk mechanical alloying of bismuth telluride alloys [23]. In this process, powder mixtures of the starting materials are compacted using cyclical loading, avoiding the possibility of contamination in a conventional milling method. Bulk mechanical alloying followed by hot pressing has led to n-type $Bi_2Te_{2.85}Se_{0.15}$ with the relatively high value for z of $2.3 \times 10^{-3}\ K^{-1}$.

Another development based on mechanical alloying has been the pulverised and intermixed elements sintering (PIES) technique [24]. The powdered elements are pulverised and intermixed before cold pressing and sintering. Good thermoelectric properties can be achieved by the PIES method provided that oxidation can be prevented.

A most important development in the production of polycrystalline bismuth telluride alloys has been the use of extrusion to assist in the alignment of the grains. The process has been described by Seo et al. [25]. Billets of $Bi_2Te_{2.85}Se_{0.15}$ were first produced by a hot pressing and sintering method. These billets were then hot-extruded at temperatures between 300°C and 510°C with an extrusion ratio of 20:1. The extruded material was strong, free from defects, and had a density equal to 99.5% of the single crystal value. The best value for the figure of merit at room temperature was $2.62 \times 10^{-3}\,\mathrm{K}^{-1}$, this being achieved at an extrusion temperature of 440°C. Although this value of z is smaller than for the best n-type zone-melted material, it is sufficiently close to indicate substantial grain alignment.

The evolution of mechanical stress during the extrusion process has been investigated by Pelletier et al. [26]. Their extruded n-type and p-type bismuth telluride alloys gave figures of merit equal to $2.8 \times 10^{-3}\,\mathrm{K}^{-1}$ and $3.3 \times 10^{-3}\,\mathrm{K}^{-1}$, respectively, so it is clear that the process yields material that is virtually the equal of zone-melted samples in its thermoelectic properties with much better mechanical strength.

A more extensive study of extruded p-type materials has been carried out by Ivanova et al. [27]. They were able to optimise the composition of $(\mathrm{Bi{-}Sb})_2\mathrm{Te}_3$ over the temperature range 100–300 K, showing that the addition of Se improved the figure of merit at 200 K. Once again, the properties were found to be as good as those of single crystals.

The alignment that can be produced by extrusion is much more important for n-type $\mathrm{Bi}_2(\mathrm{Te{-}Se})_3$ than for p-type $(\mathrm{Bi{-}Sb})_2\mathrm{Te}_3$. Fan et al. [28] have shown that one can use a hot extrusion technique in which the cross-sectional area remains constant. In applying this method to $\mathrm{Bi}_2\mathrm{Te}_{2.85}\mathrm{Se}_{0.15}$, they produced material with zT equal to 0.66. This is not particularly impressive in itself but the Seebeck coefficient had not been optimised. zT would, probably, have been much closer to unity if the Seebeck coefficient had been increased from about -150 to $-200\,\mu\mathrm{V\,K}^{-1}$.

7.4 Thick and Thin films

There are an increasing number of applications for which the thermoelectric material takes the form of thick or thin films. There would be a great saving of material if the elements in a module could be made very short, retaining the same length to cross-section as in conventional products. There are severe heat transfer problems when the length falls below about 100 μm and there are also difficulties associated with electrical contact resistance but this has not dissuaded people from the aim of using micro-modules.

For most purposes, it would be expected that the thermal and electrical flows would be perpendicular to the surface of the film but there are devices in which the flows would be parallel to the surface. In that case, heat losses through the substrates have to be taken into account.

Films of compound semiconductors can be produced simply by thermal evaporation of the elements from multiple sources in a vacuum. Unless the substrate is heated, chemical reaction may not have occurred and annealing has to be carried out. Annealing is also necessary to achieve uniformity and relative freedom from structural defects. It is sometimes difficult to control the rates at which the different components are deposited, particularly, if they have widely different melting points. Thus, it is more usual to evaporate from a single source that has a composition close to that required for the final product. Flash evaporation is a technique which ensures that the deposited film and the source have almost the same chemical composition. It has recently been shown [29] that p-type $Bi_{0.4}Sb_{1.6}Te_3$ films produced by flash evaporation can display a power factor of 3.5 mW/m K, which compares quite favourably with a value of about 4 mW/m K for bulk material in the preferred direction.

Electrodeposition is, generally, used to produce rather thicker films. Thus, films of about 30 μm thickness of $Bi_2(Te–Se)_3$ have been prepared [30] by deposition on gold and stainless steel substrates from solutions of the constituents in dilute nitric acid. The properties of such films were found to depend on the type of substrate but, whatever the substrate, the observed Seebeck coefficient was much too low for practical applications. Thicker films of bismuth telluride have also been produced. One report describes electrodeposition of 200 μm films on gold-sputtered aluminium substrates [31]. Such thick films overlap the range that can be covered by bulk materials. The rather small magnitude of the Seebeck coefficient of less than 80 $\mu V\,K^{-1}$, even after annealing, and a power factor of only 0.23 mW/m K show that there is a long way to go before such material is useful. However, if thick films improve with time, as thin films have done, they will eventually have properties that are comparable with those of bulk materials. They will then, undoubtedly, find their place in the manufacture of modules.

References

1. W.G. Pfann, *Zone Melting* (Wiley, New York, 1958)
2. W.A. Tiller, K.A. Jackson, J.W. Rutter, B. Chalmers, Acta Metall. **1**, 428 (1953)
3. D.T.J. Hurle, Solid State Electron. **3**, 37 (1961)
4. G.E. Smith, R. Wolfe, J. Appl. Phys. **33**, 841 (1962)
5. M. Hansen, *Der Aufbau der Zweistofflegierungen* (Springer, Berlin, 1936)
6. C. Uher, PhD. thesis, University of New South Wales (1975)
7. D.M. Brown, F.K. Heumann, J. Appl. Phys. **35**, 1947 (1964)
8. W.M. Yim, A. Amith, Solid State Electron. **15**, 1641 (1972)
9. M.J. Smith, R.J. Knight, C.W. Spencer, J. Appl. Phys. **33**, 2186 (1962)
10. J.P. McHugh, W.A. Tiller, Trans. Metall. Soc. AIME **215**, 651 (1959)
11. G.J. Cosgrove, J.P. McHugh, W.A. Tiller, J. Appl. Phys. **32**, 621 (1961)
12. C.B. Vining, W. Laskow, J.O. Hanson, R.R. Van der Beck, P.D. Gorsuch, J. Appl. Phys. **69**, 4333 (1991)
13. V. Fano, in CRC *Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC Press, Boca Raton, 1994), p. 257
14. R.G. Cope, A.W. Penn, J. Mater. Sci. **3**, 103 (1968)
15. R.A. Horne, J. Appl. Phys. **30**, 393 (1959)
16. M. Situmorang, H.J. Goldsmid, D.H. Morton, J. Mater. Sci.: Materials in Electronics **1**, 148 (1990)
17. S.V. Airapetyants, B.A. Efimova, Sov. Phys. Tech. Phys. **3**, 1632 (1958)
18. J. Jiang, L. Chen, S. Bai, Q. Yao, Q. Wang, Scripta Materialia **52**, 347 (2005)
19. M. Cole, I.A. Bucklow, C.W.B. Grigson, Brit. J. Appl. Phys. **12**, 296 (1961)
20. T.D. Alieva, S. Sh. Abdinov, E.Y.U. Salaev, Inorg. Mater. **17**, 311 (1971)
21. J. Javlovsky, R. Ionescu, N. Nistor, A. Chiculita, Phys. Stat. Solidi A **27**, 329 (1975)
22. J.Y. Yang, J. Wu, W. Hua, Physica B **279**, 241 (2000)
23. J.Y. Yang, T. Aizawa, A. Yamamoto, T. Ohta, J. Alloys and Compounds **312**, 326 (2000)

24. T. Ohta, T. Uesugi, T. Tokiai, N. Nosaka, T. Kajikawa, in *Proceedings of the 8th International Conference on Thermoelectrics* (Nancy, INPL, France, 1989), p. 7
25. J. Seo, K. Park, C. Lee, Materials Res. Bull. **33**, 553 (1998)
26. R. Pelletier, S. Turenne, A. Moreau, D. Vasilevskiy, R.A. Masut, in *Proceedings of the 26th International Conference on Thermoelectrics* (IEEE, Jeju, Korea, 2007), p. 49
27. L.D. Ivanova, L.I. Petrova, Y.U.I. Granatkina, V.S. Zemskov, O.B. Sokoov, S.Y.A. Skipidarov, N.I. Duvankov, Inorganic Materials **44**, 687 (2008)
28. X.A. Fan, J.Y. Yang, W. Zhu, S.Q. Bao, X.K. Duan, C.J. Xiao, K. Li, J. Phys. D: Appl. Phys. **40**, 5727 (2007)
29. M. Takashiri, K. Miyazaki, H. Tsukamoto, Thin Solid Films **516**, 6336 (2008)
30. L. Bu, W. Wang, H. Wang, Mater. Res. Bull. **43**, 1808 (2008)
31. S. Li, H.M.A. Soliman, J. Zhou, M.S. Toprak, M. Muhammed, D. Platzek, P. Ziolkowski, E. Müller, Chem Mater. **20**, 4403 (2008)

Chapter 8
Measurement Techniques

8.1 General Considerations

The measurement of the thermoelectric properties of a material presents some special problems. In the early stages of the development of a new material, a high accuracy may not be needed, but, once it is to be used in practical applications, these quantities must be known precisely. The performance is strongly dependent on the figure of merit, z, which itself involves three other parameters, the square of the Seebeck coefficient, the electrical conductivity, and the thermal conductivity. It is important that these three properties be measured independently but, as we shall see, it is possible to determine zT directly. In fact, it can be claimed that, when the Seebeck coefficient and the electrical conductivity are known, the direct measurement of zT provides the most accurate way of establishing the thermal conductivity, at least at ordinary temperatures for good thermoelectric materials.

It is, generally, preferable for all the measurements to be made on a single sample. There are examples in the literature where falsely optimistic predictions have been made about potential new thermoelectric materials on the basis of data obtained from different specimens. It is understandable how such a problem arises. The electrical conductivity is best found for a sample that is long and of small cross-sectional area since this minimises possible errors due to non-linear current flow near the contacts. On the other hand, a short sample with a large cross-sectional area is preferred for thermal conductivity measurements since the relative effect of heat losses, due, for example, to radiation, is then smaller. When the use of different samples from a given ingot is unavoidable, it is recommended that several pieces be cut and that the determination of, say, the thermal conductivity of one piece be accompanied by electrical conductivity measurements for neighbouring pieces on either side.

Even a simple one-stage Peltier cooler operates over a range of temperature and some devices, such as thermoelectric generators and multi-stage coolers, may involve wide ranges. Thus, the thermoelectric properties need to be determined as a function of temperature. Measurements at different temperatures also assist us in the understanding of the transport processes. However, it must not be forgotten that one often needs rapid but nevertheless accurate measurement of the Seebeck coefficient or the electrical resistance for the control of quality in the production of thermoelements, and it is usually satisfactory for this purpose to make observations at a single temperature.

It is helpful in the basic study of a material to extend the measurements to cover the Hall effect and useful information can be gained from the changes with magnetic field of the thermoelectric properties. Similarly, it is worthwhile determining such quantities as the Nernst coefficient, in particular, for the information that it gives about the scattering law. Of course, the transverse thermomagnetic effects can themselves be used in energy conversion and this may be an extra incentive to determine the Nernst and Ettingshausen coefficients.

In this chapter, we shall deal mainly with principles rather than experimental details.

8.2 Electrical Conductivity

Electrical resistance is a property that has been measured for such a long time that it might be thought unnecessary to devote much attention to it here. However, it turns out that there are special problems associated with electrical resistance measurements on thermoelectric materials.

The electrical resistance, R, of a piece of metal wire is found by passing a known current through it and observing the potential difference between the ends. The electrical conductivity, σ, is equal to A/Rl. It is a little more difficult to find the electrical conductivity of a semiconductor because there is often a resistance associated with the contacts. The conventional way to overcome this problem involves the use of inset probes, as shown in Fig. 8.1. The two inset probes should be far enough from the end contacts for any departures from planar equipotential surfaces to be eliminated.

The current is introduced by large area contacts at the ends of the sample which must have a uniform cross-sectional area. The determination of the electrical conductivity requires an accurate knowledge of the spacing between the

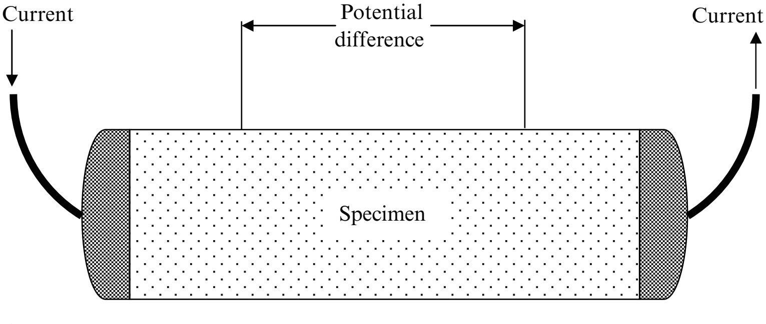


Fig. 8.1 Arrangement of sample with inset probes for measurement of the electrical conductivity

potential probes. These probes may be held in contact by pressure or by welding. Sometimes, they may be inserted in small holes that are drilled in the specimen, though this distorts the current flow.

Most techniques involve some uncertainty in determining the precise spacing. For this reason, the use of inset probes is often avoided when it is known that the electrical resistance at the current contacts is small. In principle, this resistance can be found by making measurements on samples of different length. In practice, for semiconductors like bismuth telluride and lead telluride, the electrical resistance at plated contacts is so small as to be negligible provided that the sample is at least a few millimetres in length.

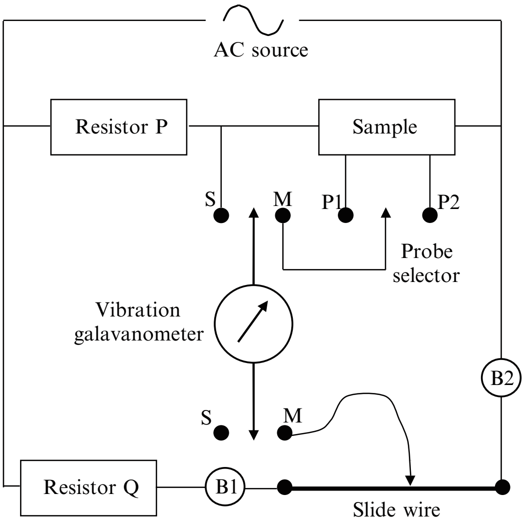
The main problem, in measuring the electrical resistance for the materials in which we are interested, is associated with the thermoelectric effects that are always present unless isothermal conditions are maintained. If there is a temperature difference between the two ends, this will give rise to a Seebeck voltage that augments or opposes the resistive voltage. When that temperature difference arises from some asymmetry in the experimental arrangement, the problem may be overcome simply by reversing the current flow. However, there is a more subtle effect that can be very large for good thermoelectric materials [1]. The flow of current causes a temperature difference to be established by means of the Peltier effect. When the current is reversed, the Peltier heating and cooling are also reversed and there is an increase in the potential difference above the resistive value whichever way the current flows. This effect may actually be used in determining the thermoelectric figure of merit but, when only the electrical conductivity is of interest, it is a source of error that has to be eliminated.

One of the ways of avoiding thermoelectric effects in the measurement of electrical resistance is to use alternating current instead of direct current [2]. An alternating current bridge that was developed for this purpose is shown in Fig. 8.2. The frequency must be high enough for there to be no appreciable build up of thermoelectric voltage during each cycle. The thermal mass of soldered contacts is sufficient to allow frequencies of the order of 50 Hz to be satisfactory.

In the apparatus shown in Fig. 8.2, the sample is shown with inset probes. The resistance between the two probes is determined by obtaining balance on the vibration galvanometer at two positions of the moving contact on the calibrated slide wire. The balancing resistors are adjusted so that both balance points are within the length of the slide wire. Balance is first obtained with the selectors in the positions S; measurements are then made with the selectors in the positions M. The resistance of the length of the sample between the probes is given by PR_S/Q, where P and Q are the resistances of the standard resistors P and Q, and R_S is the resistance of the slide wire between the balance points with the probe selector in the positions P1 and P2.

There is no doubt that the direct-current potentiometer is a powerful tool for the precise measurement of small voltages. In spite of the problems that we have described, this instrument can still be used for thermoelectric materials if the mechanical chopping device that was designed by Dauphinee and Woods [3] is incorporated in the apparatus. The potential differences that are encountered in measuring the galvanomagnetic coefficients are sometimes very small and masked

Fig. 8.2 Alternating current bridge for measurement of electrical conductivity of thermoelectric materials. B1 and B2 are the balancing resistors



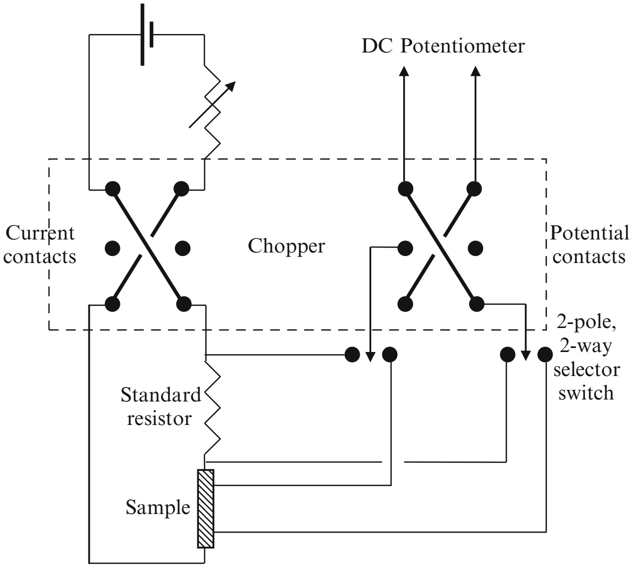


Fig. 8.3 Measurement of electrical conductivity of a sample of thermoelectric material using a DC potentiometer and a mechanical chopper

by noise if an alternating current is used. Thus, the Dauphinee and Woods chopper, although useful in the measurement of electrical conductivity, is particularly valuable when other parameters are also being determined. The principle of the device is illustrated in Fig. 8.3, which refers specifically to the measurement of electrical resistance. Figure. 8.4 shows the variation with time of the current input and the voltage output.

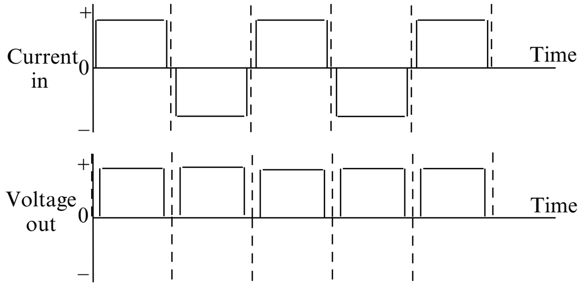


Fig. 8.4 Input current and output voltage during operation of a synchronous chopper

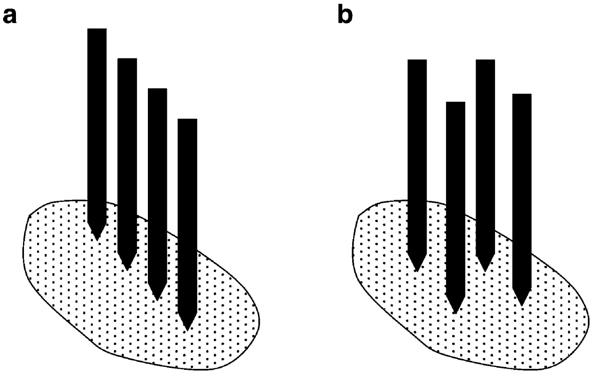


Fig. 8.5 Four-point probe configurations; **(a)** probes regularly spaced in line and **(b)** probes at the corners of a square. The probes are usually spring-loaded and pressed against the plane surface of a semiconductor

The principle of the measurement is simple enough but there are some precautions that have to be taken. The periodic reversal of the current through the specimens prevents the buildup of temperature gradients from the Peltier effect. The reversal of the voltage output ensures that the polarity of the signal to the potentiometer is always the same. However, during the brief period when the switching of the current is taking place, there is the danger of a large induced spurious signal. Thus, as shown in Fig. 8.4, the voltage pick up does not occur until after the current switching is complete. Likewise, the potential contacts are broken before the reversal of the current takes place. It is probably not a good idea for the chopper to be operated at the mains supply frequency, and in the experiments of Drabble et al. [4] on the galvanomagnetic effects in bismuth telluride, the chopping frequency was 40 Hz. It was found that potential measurements that were accurate to within less than 1 μV could be achieved.

Some comments may be made about the 4-point probe methods that are often used in measuring the electrical conductivity or resistivity of semiconductors like silicon and germanium. Typical arrangements for the probes are shown in Fig. 8.5.

When the linear arrangement (a) is used, an electric current is passed between the outer probes and the potential difference is measured between the inner probes. If the probe spacing is equal to a and the sample is much larger than the distance between the outer probes, the electrical conductivity is given by [5]

$$\sigma = \frac{I}{2\pi a V}, \tag{8.1}$$

where I is the current and V is the observed potential difference. The square arrangement (b) has the advantage that it occupies less space and it can, therefore, be used for smaller samples. In this configuration, the current is passed between adjacent probes and the potential difference between the other pair of probes is measured. Another advantage of the square configuration is that it allows the Hall coefficient to be found. In the measurement of this quantity, the current is passed between an opposite pair of probes and the potential difference observed between the other pair, the magnetic field being applied perpendicular to the surface of the semiconductor.

Neither of the 4-point probe measurements is considered as accurate as a measurement in which the current is passed along a rectilinear specimen with the potential gradient determined using probes that are spaced as widely as possible. There is a further disadvantage if the electrical conductivity is anisotropic. In this case, 4-probe measurements are almost meaningless for polycrystalline samples. Even when the probes are applied to a surface that is perpendicular to one of the axes of a single crystal, the observations cannot be interpreted unless one knows the ratios between the conductivities in the different directions.

8.3 Seebeck Coefficient

In many ways, the Seebeck coefficient is one of the easiest quantities to measure and for this reason it is often used in preference to, say, the electrical conductivity in checking whether or not production material falls within its specification. One merely has to apply a known temperature difference between the ends of a sample and observe the potential difference that results.

Suppose that the temperature difference is measured using, say, copper–constantan thermocouples, the copper wires can also be used to determine the potential difference but due account must be taken of the small but not negligible absolute Seebeck coefficient of the metal. The question arises as to whether it is better to attach the thermocouples to the sample as in Fig. 8.6(a) or to the metal blocks that constitute the heat source and the sink as in Fig. 8.6(b).

In the arrangement (a), the heat conducted along the thermocouple wires may influence the temperature at the point of contact. Also, there may be a temperature gradient within the thermocouple junction so that the observed temperature is different from that on the sample at the point of contact. Both these problems may

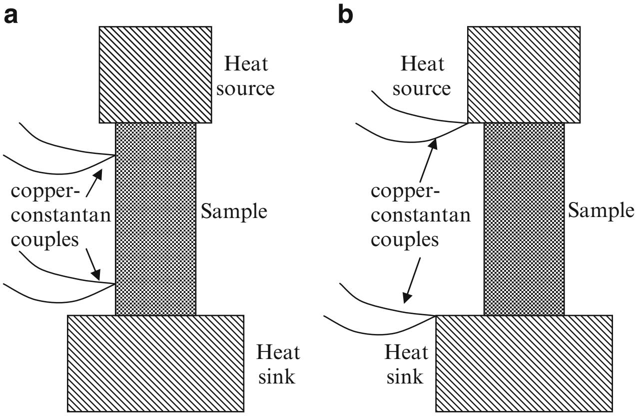


Fig. 8.6 Two different arrangements for the thermocouples in the measurement of the Seebeck coefficient; **(a)** thermocouple attached to the sample and **(b)** thermocouples attached to the source and sink

be minimised if the wires are of very small diameter, perhaps less than 60 μm. Alternatively, the wires may be inserted in small holes that are drilled in the sample, as in some measurements of the electrical conductivity. However, the holes may not be isothermal enclosures and this can lead to errors if the electric potential and the temperature are not observed at the same point.

It is not immediately obvious that arrangement (b) is any better unless the source and sink are soldered to the sample. However, the arrangement seems to work satisfactorily even if the source and sink are just pressed against the sample. To be sure, there may then be thermal resistance between the metal blocks and the semiconductor but the temperatures on either side of the actual points of contact must be the same. Virtually, all the drop in temperature between the source and the sink will exist in the semiconductor if it is, in fact, a thermoelectric material with a low thermal conductivity. The metal blocks should preferably be made from a highly conducting metal like copper. If the material is non-uniform, there may be a difference between the observed Seebeck coefficients for the blocks soldered to the specimen or merely pressed against it. For pressed contacts, most of the temperature drop will occur near the interfaces and these regions will have a predominant influence on the measurement. However, there are other problems in all the transport measurements if the samples are non-uniform. It is not good enough to take average values since any non-uniformity may affect the different parameters in different ways.

A small probe for determining the Seebeck coefficient at a localised region of a semiconductor may actually be the easiest way of detecting non-uniformity. The apparatus for this purpose uses the arrangement in Fig. 8.6(b) with the large heat source block being replaced by a much smaller one that is tapered so that the region of contact is small. The precise thermal distribution does not have to be known as long as one can be sure that virtually all the temperature difference between the source and sink is experienced over a small region of the sample.

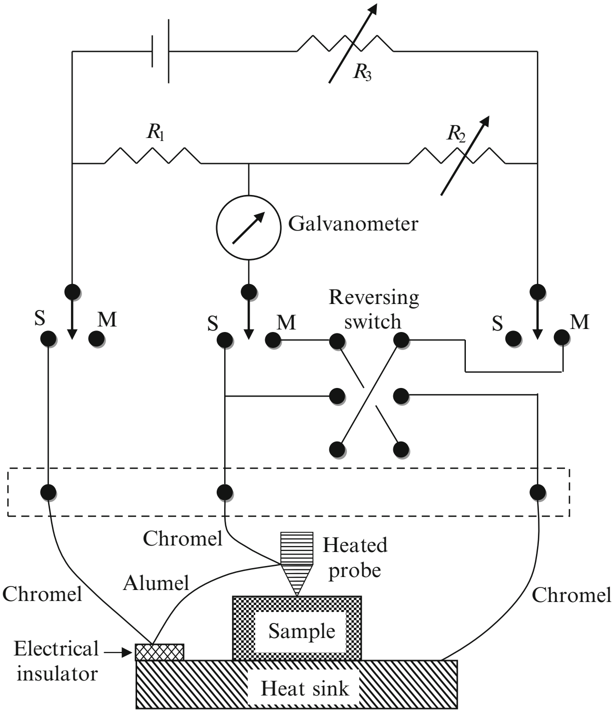


Fig. 8.7 Measurement of the Seebeck coefficient using the technique devised by Cowles and Dauncey [6]. The chromel wires (left to right, a, c, b) are connected to copper wires in an enclosure at a constant temperature

One does not actually have to measure the temperature difference and the thermoelectric voltage. A single observation can determine the ratio of the Seebeck coefficient to that of the thermocouple. A simple and elegant technique that uses this principle was devised by Cowles and Dauncey [6]. The principle of their method is clear from Fig. 8.7.

Cowles and Dauncey used chromel–alumel thermocouples in their apparatus and they determined the Seebeck coefficient of the test sample with respect to chromel. When the heated probe is placed in contact with the test specimen, a thermal EMF is generated between the chromel wires a and b. There is also the EMF across the chromel–alumel couple that appears between the chromel wires a and c. The aim, then, is to determine the ratio between these EMFs at the same time. In practice, the temperature difference soon reaches a steady state so that a short time interval between the observations of the two EMFs is permissible.

The ratio between the EMFs is determined as the ratio between a fixed standard resistor R_1 and a calibrated variable resistor R_2. With the switches in the "set" position S, the uncalibrated variable resistor R_3 is adjusted until the galvanometer indicates that balance has been achieved. The switches are then moved to the "measure" position M and balance is again obtained, this time by adjustment of R_2.

This, of course, upsets the balance in the position S so R_3 has to be adjusted again. However, if R_3 is very much greater than R_1 and R_2, simultaneous balance in both the S and M positions can be reached very quickly.

The condition for balance in position S is

$$IR_1 = \alpha_0 \Delta T, \tag{8.2}$$

where α_0 is the differential Seebeck coefficient of the chromel–alumel couple. The balance condition for position M is

$$IR_2 = \alpha \Delta T, \tag{8.3}$$

where α is the differential Seebeck coefficient between the test specimen and chromel.

When simultaneous balance is achieved

$$\alpha = \alpha_0 \frac{R_2}{R_1}. \tag{8.4}$$

The reversing switch shown in Fig. 8.7 allows measurements to be made on both p-type and n-type materials.

When the Seebeck coefficient is to be measured over a wide range of temperature, another approach may be used. One end of the sample is held at a fixed temperature, say, 300 K, while the other end is slowly heated. The thermoelectric EMF between the ends is continuously monitored at the same time as the temperature of the hot end. The Seebeck coefficient at any temperature is then given by the slope of the plot of thermoelectric voltage against temperature. The method is not as accurate as one in which the whole sample is heated and the voltage is measured for a small temperature difference between the ends. However, it allows data to be collected rapidly and this may be important if the properties of the sample are likely to change with time at the higher temperatures.

8.4 Thermal Conductivity

It has always been recognised that it is much more difficult to measure the thermal conductivity than the electrical conductivity. There are a number of reasons for this. Thus, while extremely good electrical insulators can be found, it is virtually impossible to separate a body from its surroundings from the viewpoint of heat transfer. Admittedly, high vacuum prevents heat transfer by conduction and convection but radiation is always possible. In principle, either of the arrangements shown in Fig. 8.6 for the determination of the Seebeck coefficient can be adapted for measuring the thermal conductivity. If one assumes that all the heat lost by the source is conducted through the specimen, one can determine the thermal conductivity as the ratio of the input power per unit area to the temperature gradient. However, if the configuration of Fig. 8.6(b) is used, care must be taken to ensure that the opposite faces of the sample are in intimate contact with the heat source and the heat sink.

Even when the sample is soldered to the source and sink, there will always be some thermal contact resistance that has to be taken into account, especially, since alloys used in soldering have a lower thermal conductivity than most pure metals. The relative effect of this resistance can be reduced by making the sample very long but then transfer of heat from the source to the sink via the surrounding medium becomes more significant. Remembering that good thermoelectric materials are invariably poor heat conductors, the sample should probably be short rather than long. Traditionally, thermal conductivity has been measured on samples of large volume as this helps us to reduce the relative corrections due to contact resistance and heat losses. However, most thermoelectric materials are only available in the form of small bars having a volume considerably less than a cubic centimetre. Some attempt may be made to calculate the correction factors from first principles but generally it is necessary for them to be determined by measurements on samples of different length and cross-section.

In selecting a technique for the measurement of thermal conductivity, one usually has a choice between an absolute method and a comparative method. A comparative method has the advantage that heat lost from the source, other than through the test specimen, can be ignored. On the other hand, successful comparative techniques require the availability of a standard material with a thermal conductivity similar to that of the material being tested. Surprisingly, few reliable standard materials are at hand. In fact, we are usually aiming for greater accuracy in the measurement of the thermal conductivity of thermoelectric materials than has been achieved in possible standards of comparison. At ordinary temperatures, it is more usual to adopt an absolute method, but at high temperatures, when the losses from the heater are unavoidably greater, a comparative method is often used.

One can also choose between static and dynamic methods. It usually takes a long time for thermal equilibrium to be established after power is introduced to the heat source so a great deal of time can be saved by using a dynamic technique but there is no doubt that a static method is more precise. In establishing a relation between lattice conductivity and mean atomic weight, Ioffe and Ioffe [7] made use of a dynamic method [8, 9] so that they could obtain many measurements in a short time.

Ioffe and Ioffe's apparatus is illustrated in Fig. 8.8. No doubt different materials would be used in its construction if the measurements were to be repeated now but the principles remain sound after about 50 years. The test sample is sandwiched between two copper blocks and held in place by pressure applied using a screw. In the original measurements, grease or glycerine was used to improve the thermal contact between the surfaces. A thin sheet of mica in one of the contacts ensured electrical insulation between the thermocouples. The acrylic glass enclosure is fitted fairly closely to the specimen so as to minimise convection losses through the air, the space around the sample not being evacuated as is usual with more precise measurements. Corrections were made for thermal contact resistance and for heat losses.

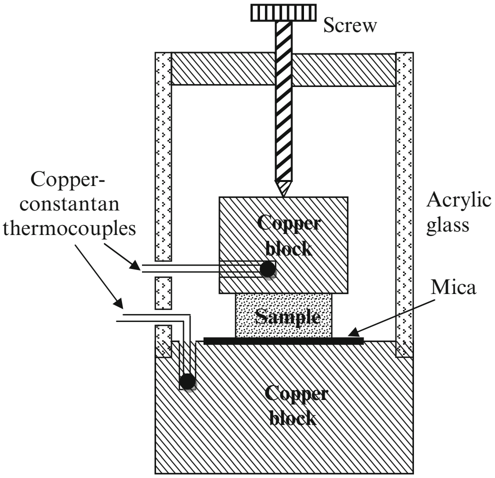


Fig. 8.8 Ioffe and Ioffe's apparatus for the dynamic measurement of thermal conductivity

At the start of a measurement, the whole apparatus was in equilibrium with the surrounding air. The lower block was then inserted into a bath of refrigerant. Observations were made of the temperature of one block and of the temperature difference as a function of time.

Suppose that the temperatures of the lower and upper block are respectively T_1 and T_2 and that C_2 is the known thermal capacity of the upper block, then as a first approximation, the rate at which the upper block loses heat is equal to the rate at which heat passes through the sample. Thence

$$-C_2 \frac{\mathrm{d}T_2}{\mathrm{d}t} = \lambda \left(T_2 - T_1\right) \frac{A}{L}, \tag{8.5}$$

where A/L is the ratio of cross-sectional area to length of the sample. However, allowance must be made for the fact that part of the heat reaching the lower block comes from the sample. It was found appropriate to add one-third of the heat capacity, C, of the sample to C_2 giving

$$\lambda = \frac{C_2 + C/3}{T_2 - T_1} \frac{\mathrm{d}T_2}{\mathrm{d}t}. \tag{8.6}$$

Since it is usual for C_2 to be very much greater than C, the latter does not have to be known very accurately.

The correction for heat loss from the upper block to the surrounding walls was found from the equilibrium temperature difference between the blocks. By placing a sample of known thermal conductivity between the blocks, it was possible to determine the heat loss to the lower block through the air. Finally, the correction for contact resistance across the mica and greased layers was found from the use of samples of different length. The whole measurement has been put on a firmer foundation by the theoretical treatments of Kaganov [10]and Swann [11]. An important conclusion from the more refined theory is that the measurements should neither be

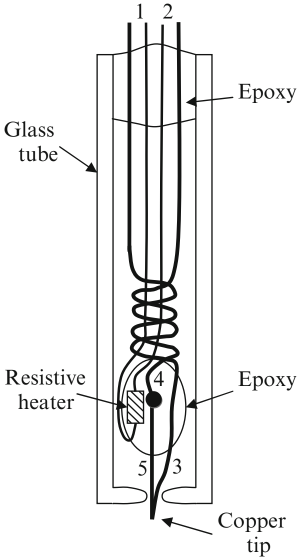


Fig. 8.9 Miniature thermal comparator. 1 and 2 are the leads to the resistive heater. 3 and 4 are the constantan thermocouple wires and 5 is the copper branch

attempted immediately after the lower block is placed in the refrigerating bath nor when the temperatures are approaching their equilibrium values.

The thermal comparator is another device for the rapid evaluation of the thermal conductivity. It is based on a principle first developed by Powell [12]. A miniaturised version of the comparator [13] is shown in Fig. 8.9.

The comparator is made up of a copper–constantan thermocouple, in which one of the junctions is close to a small resistive heater. The other junction lies close to the copper tip, which protrudes from the glass tube. In the rest position, a small temperature difference will exist between the two thermocouple junctions. When the tip is pressed against the sample under test, a much larger temperature difference appears and this will give rise to a large increase in the output voltage. Clearly, the voltage will be larger if more heat is conducted away from the tip by a good thermal conductor. The output voltage will be much smaller, if the tip is pressed against a poor conductor of heat. The temperature difference reaches a steady state in a few seconds so the measurement is very rapid. The validity of the test depends on the contact area remaining the same from one sample to another. One might have expected this area to be more or less the same only for samples that are harder than copper, since the force is generated by the coiled constantan wires and should have a constant value when the tip is retracted to the base of the holder. In practice, the effective area of the contact seems to remain the same even for quite soft materials. Thus, if the heater power is kept steady, the steady-state output voltage should be a measure of the thermal conductivity of the test sample. Although, a theoretical estimate of the output voltage can be made, it is much better to calibrate the instrument using materials of known thermal conductivity. A thermal comparator of this design is used frequently in tests of gemstones, diamond, in particular, having a much

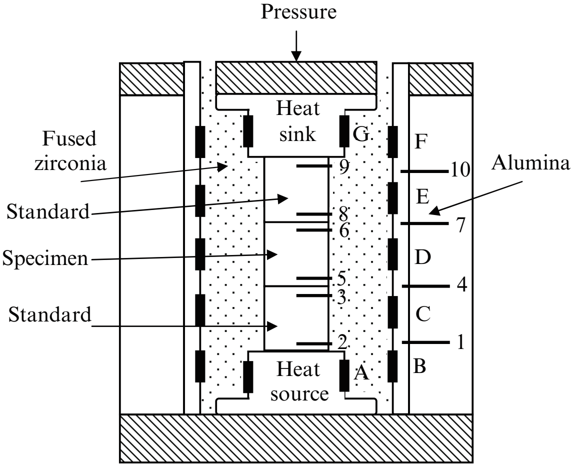


Fig. 8.10 A comparative method for the measurement of the thermal conductivity of thermoelectric materials at elevated temperatures. Electric heaters are represented by the letters A–G and thermocouples by the numerals 1–10

higher thermal conductivity than any of its simulants. However, the comparator has also found use in preliminary tests on potential thermoelectric materials.

Precise comparison methods have been used for measuring the thermal conductivity of generator materials at elevated temperatures. The apparatus shown in Fig. 8.10 has been used up to a temperature of 800°C in measurements on III–V compounds [14].

The two standard samples are selected so as to have a similar conductance to that of the test specimen that is sandwiched between them. Heat transfer from the sides of the three blocks is prevented by a heat shield with multiple heaters. In any case, lateral heat flow is minimised by the filling material, fused zirconia powder, which has a very low thermal conductivity. The length of the stack means that it takes a long time to reach equilibrium, but the method is an accurate one.

We shall discuss the determination of the thermal conductivity by measuring the figure of merit later and we shall also devote a section to thermal diffusivity measurements. However, we shall now discuss a method that has come into prominence in recent years and that is useful in work on thin specimens of poor heat conductors. This is the $3\,\omega$ technique [15], so called because the heat input varies with time at an angular frequency ω and the analysis is based on the observation of an electrical signal at the frequency $3\,\omega$. Radial flow methods are useful in reducing losses by thermal radiation but they often need large samples to be available. The $3\,\omega$ system is suitable for use on small specimens. The essential features of the experimental sample arrangement are shown in Fig. 8.11.

A metal strip is laid down on specimen so that it is in good thermal contact but electrically isolated from it. The outer side arms are used for the introduction of an electric current and the inner side arms are for picking up the voltage across the

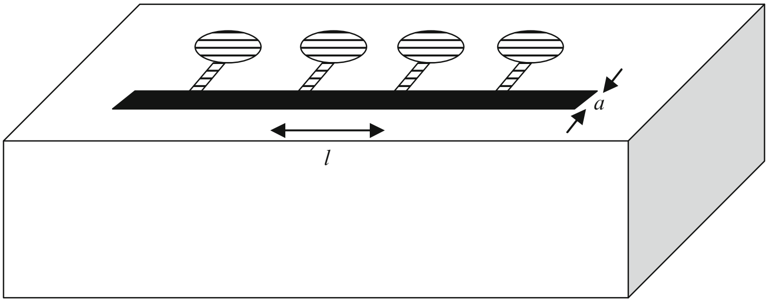


Fig. 8.11 Arrangement of the metal strip and side arms on the sample in the 3ω method

central section of length l. It is noted that the resistance of the metal strip depends on the temperature so it can be used as a thermometer. The strip is heated by the passage of an alternating current of angular frequency ω and a temperature wave of frequency $2\,\omega$ enters the sample since, of course, heat is generated in each half-cycle. The temperature wave travels radially outwards from the metal strip, suffering exponential damping as it proceeds. The sample must be thick enough to prevent any significant interference from the reflected wave at the time that it reaches the strip; it is satisfactory for the thickness to be not less than $5a$, where a is the width of the strip. Voltage measurements at a specific frequency are observed using a lock-in amplifier. The voltage oscillations in the strip will have a component at a frequency ω associated with the current flow and a component at a frequency $2\,\omega$ associated with the change of resistance due to the temperature wave. When these two contributions are multiplied together, the resultant will include a term that oscillates at the frequency $3\,\omega$. The thermal conductivity of the sample is obtained after voltage measurements, $V_{3,1}$ and $V_{3,2}$, at two frequencies $3\,\omega_1$ and $3\,\omega_2$, respectively, where ω_1 and ω_2 are the frequencies of the applied current. It may be shown [16] that the thermal conductivity can be calculated from the relation

$$\lambda = \frac{V^3 \ln\left(\omega_1/\omega_2\right)}{4\pi l R^2 [V_{3,1} - V_{3,2}]} \frac{\mathrm{d}R}{\mathrm{d}T}. \tag{8.7}$$

Here, V is the voltage at some frequency ω and R is the resistance of the metal strip between the inner side arms. It is noted that, although the technique is based on thermal diffusion, it is essentially the thermal conductivity that is determined.

8.5 Thermal Diffusivity

Steady-state thermal conductivity measurements become increasingly difficult as the temperature becomes higher because of the rapidly rising radiation factor. For this reason, the thermal diffusivity is often determined instead of the thermal conductivity.

The thermal diffusivity, κ, is defined by the relation

$$\kappa = \frac{\lambda}{c_V}. \tag{8.8}$$

Variations on the original thermal diffusivity measurements of Ångström [17] are still encountered. A brief discussion of his technique shows that, in common with other thermal diffusivity measurements, it allows the thermal losses to be eliminated.

In Ångström's method, a sinusoidal temperature variation is applied to one end of a long sample. The amplitude of the temperature wave, as it travels along the sample, is monitored at two points that are separated by a distance l. Then, in the absence of lateral heat loss, the thermal diffusivity is given by

$$\kappa = \frac{\omega l^2}{2\ln^2(\alpha)} = \frac{\omega l^2}{2\beta^2}, \tag{8.9}$$

where α is the ratio between the amplitudes of the temperature wave at the two points and β is the phase difference. If there are losses equation (8.9) has to be replaced by

$$k = \frac{\omega l^2}{2\beta\ln(\alpha)}. \tag{8.10}$$

It will be seen that the use of this equation allows the losses to be ignored.

Adaptations of Ångström's method have been used for thermoelectric materials by Nii [18] and Abeles et al. [19] but one of the more interesting developments was the method of Green and Cowles [20] in which a Peltier heat source was employed. This allowed the temperature wave to be initiated without any overall heating of the sample. These authors were also able to use the sample as its own thermometer, thus exploiting its high Seebeck coefficient. It may be noted that, even when the input is not strictly sinusoidal, the higher harmonics are rapidly attenuated as the wave moves along the sample but, in any case, Fourier analysis of the temperature fluctuations at any point is not difficult.

A disadvantage of the Ångström principle is that it requires a longer sample than is commonly available. However, thermal diffusivity measurements are easily made on short samples. Thus, Goldsmid et al. [21], adopted a thermal diffusivity measurement for thin crystals of Cd_3As_2, comparing the transverse electric fields, generated by the Nernst effect, for steady and intermittent thermal radiation.

Most thermal diffusivity measurements, nowadays, make use of thin samples and laser heat sources [22]. The essential components for a typical laser flash measurement are shown in Fig. 8.12. The sample is selected so that its cross-section is large compared with its thickness. One of its faces is irradiated by a pulsed laser and the fluctuations in temperature at the opposite face are observed using, say, an infrared sensor. In a typical measurement, the temperature rise at the back surface is compared with that which is reached in the steady state when the sample is continuously

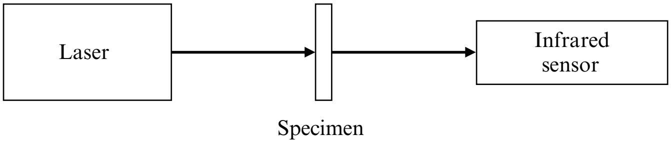


Fig. 8.12 Essential features of the laser flash method for determining the thermal diffusivity

heated. The time, $t_{1/2}$, taken to reach one-half of the steady-state temperature rise allows the thermal diffusivity to be found from the relation

$$\kappa = \frac{1.37d^2}{\pi^2 t_{1/2}}. \tag{8.11}$$

It must, of course, be remembered that a thermal diffusivity measurement needs to be accompanied by knowledge of the specific heat if one is to determine the thermal conductivity, the quantity that is actually needed for the figure of merit. Nevertheless, the advantages in the avoidance of heat loss corrections are sufficient to make the laser-flash technique the preferred approach by many workers.

8.6 The Figure of Merit

The thermoelectric figure of merit can be found from independent measurements of the Seebeck coefficient and the electrical and thermal conductivity. It is, however, possible to determine the figure of merit directly and, having made this determination, one of the three quantities, α, σ or λ, can then be found if the other two are known. The usual procedure is to measure the Seebeck coefficient and the electrical conductivity, and to deduce the thermal conductivity. In fact, this is often the most accurate method for determining λ but, as we shall see, there are some dangers in this approach.

The direct measurement of zT was first proposed by Harman [23] and the procedure is usually named after him. In principle, all that needs to be done is to observe the ratio of the electrical conductivities, σ_a and σ_i, under adiabatic and isothermal conditions respectively. Then,

$$zT = \frac{\sigma_i}{\sigma_a} - 1. \tag{8.12}$$

Two possible experimental arrangements are shown in Fig. 8.13. In (a) the sample is suspended by its current and thermocouple leads in a vacuum enclosure while in (b) one end of the sample is attached to a heat sink. The advantage of arrangement (a) is that the heat loss corrections are smaller, but arrangement (b) is more practical as it enables equilibrium at different temperatures to be established more quickly. In both arrangements, the electric current is introduced by copper wires and the temperatures at the ends of the sample are measured using copper–constantan thermocouples. The copper branches of the thermocouples are also used to measure

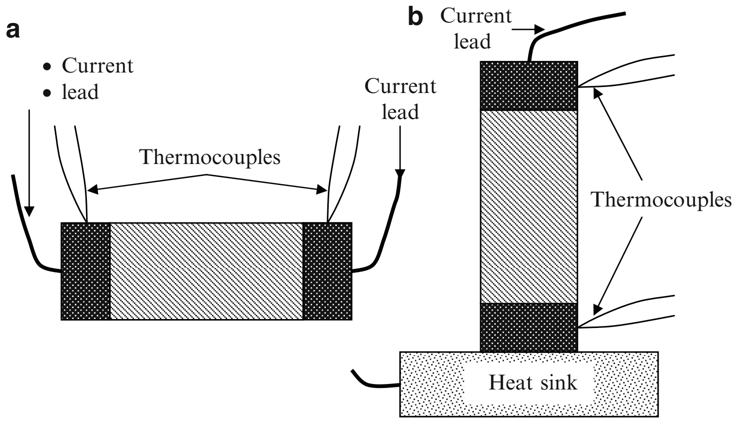


Fig. 8.13 Experimental arrangements for the direct measurement of the figure of merit. In **(a)** the sample is suspended in vacuum and in **(b)** one end of the sample is attached to a heat sink

the potential difference. The electrical resistance between the sample and the copper end plates is small and often neglected but it can be included if necessary.

Suppose that a current I is passed along the sample; then, if we ignore Joule heating for the time being, the temperature difference, ΔT between the ends is given by

$$\frac{\lambda A \Delta T}{L} = |\alpha| I T, \tag{8.13}$$

where α is the Seebeck coefficient of the sample with respect to copper, A is the cross-sectional area, and L is the length. The temperature difference leads to the development of a thermoelectric voltage given by

$$|\alpha| \Delta T = \frac{\alpha^2 I L T}{\lambda A}. \tag{8.14}$$

Under isothermal conditions, the potential difference between the contacts is

$$V_\mathrm{i} = \frac{I L \rho}{A}. \tag{8.15}$$

However, under adiabatic conditions, the total potential difference becomes

$$V_\mathrm{a} = V_\mathrm{i} + \frac{\alpha^2 I L T}{\lambda A} = V_\mathrm{i}\left(1 + \frac{\alpha^2 T}{\rho \lambda}\right). \tag{8.16}$$

Rearranging (8.16), we find that

$$zT = \frac{V_\mathrm{a}}{V_\mathrm{i}} - 1. \tag{8.17}$$

One of the advantages of the Harman procedure is that the heat losses are smaller than in a conventional static thermal conductivity measurement. In particular, there is no separate heater that might lose heat to the surroundings. Also, in arrangement (a) of Fig. 8.13, the losses from the faces of the sample are substantially reduced. We shall use arrangement (a) as the basis for the following calculations.

In calculating the heat losses, we assume that the isothermal surfaces are nearly planar so the problem becomes one-dimensional. We must determine the heat loss by radiation from the end contacts and from the exposed faces of the sample (assuming that the vacuum in the enclosure is good enough to eliminate both gaseous conduction and convection). We also have to find the losses along the electrical wires, which we suppose to be anchored at the temperature of the enclosure.

The radiation from each of the end contacts is equal to $\beta_c A_c \Delta T/2$ when the current is small enough for $\Delta T \ll T$. One end of the sample is supposed to rise to $T_0 + \Delta T/2$ and the other end to fall to $T_0 - \Delta T/2$, where T_0 is the ambient temperature. β_c is the rate of radiation per unit area per unit temperature difference and A_c is the area of each contact. β_c is proportional to T_0^3 so radiation becomes a severe problem at higher temperatures.

The radiation loss per unit length from the part of the sample at temperature T is equal to $\beta P\,(T - T_0)$, where β is the rate of radiation per unit area per unit temperature difference (which may be significantly different from β_c) and P is the perimeter of the sample.

The heat transferred by conduction along the wires at each end is equal to $K_1 \Delta T/2$, being a loss at one end and a gain at the other. K_1 is the conductance in parallel of each set of leads.

At any part of the sample, the rate of heat flow is

$$q = -\lambda A \frac{\mathrm{d}T}{\mathrm{d}x}, \tag{8.18}$$

and

$$\frac{\mathrm{d}q}{\mathrm{d}x} = -\beta P\,(T - T_0) = -\lambda A \frac{\mathrm{d}^2 T}{\mathrm{d}x^2}. \tag{8.19}$$

At the ends of the sample, when $x = \pm L/2$, q is equal to $\pm q_0$ which is given by

$$\pm q_0 = \alpha I T - \frac{\beta_c A_c \Delta T}{2} - \frac{K_1 \Delta T}{2}. \tag{8.20}$$

The solution of the differential equation is then

$$T - T_0 = \pm \frac{q_0}{(\lambda A \beta P)^{1/2}} \frac{\exp[(\beta P/\lambda A)^{1/2}\,x] - \exp[-\,(\beta P/\lambda A)^{1/2}\,x]}{\exp[\frac{1}{2}\,(\beta P/\lambda A)^{1/2}\,L] + \exp[-\frac{1}{2}\,(\beta P/\lambda A)^{1/2}\,L]}. \tag{8.21}$$

When the temperature gradient is close to being uniform, we find

$$\pm q_0 = \frac{\lambda A}{L}\Delta T + \frac{\beta P L}{12}\Delta T \tag{8.22}$$

On combining (8.20) and (8.22), we obtain

$$\frac{\alpha IT}{\Delta T} = \lambda \frac{A}{L} + \frac{\beta PL}{12} + \frac{\beta_{\mathrm{c}} A_{\mathrm{c}}}{2} + \frac{K_{\mathrm{l}}}{2}. \tag{8.23}$$

The first term on the right-hand side represents the heat conducted through the sample while the remaining terms represent the losses. The second term has only one-quarter of the value that it would have in a conventional steady-state thermal conductivity apparatus and the third and fourth terms have been reduced by a factor of 2. Also, of course, there is a reduction in A_{c} since the end contacts are so much smaller than a heating block.

Up to this point, it has been assumed that the current is so small that Joule heating can be ignored. It can be shown [24] that the effect of Joule heating is to shift the position on the sample for which $T = T_0$ to $x = a$, where

$$a\Delta T = \frac{I^2 \rho L / A}{\beta P + 2\left(\beta_{\mathrm{c}} A_{\mathrm{c}} + K_l\right)/L}. \tag{8.24}$$

Asymmetry is introduced into the temperature distribution and (8.23) becomes

$$\frac{\alpha IT}{\Delta T} = \lambda \frac{A}{L} + \frac{\beta P\,(L \pm 2a)}{4} + \frac{(\beta_{\mathrm{c}} A_{\mathrm{c}} + K_{\mathrm{l}})}{2}\left(1 \pm \frac{2a}{L}\right) \mp \frac{I^2 \rho L}{A \Delta T}. \tag{8.25}$$

After combining (8.24) and (8.25) to eliminate a, we again obtain (8.23). Thus, whether or not there is any significant Joule heating, the form of (8.17) that takes account of the heat losses is

$$zT = \left(\frac{V_{\mathrm{a}}}{V_{\mathrm{i}}} - 1\right)\left(1 + \frac{\beta PL^2}{12\lambda A} + \frac{\beta_{\mathrm{c}} A_{\mathrm{c}} L}{2\lambda A} + \frac{K_{\mathrm{l}}}{2\lambda A}\right). \tag{8.26}$$

If the apparatus has the form shown in Fig. 8.13(b), the heat loss terms become larger and the modified form of (8.23) is

$$\frac{\alpha IT}{\Delta T} = \lambda \frac{A}{L} + \frac{\beta PL}{3} + \beta_{\mathrm{c}} A_{\mathrm{c}} + K_{\mathrm{l}}. \tag{8.27}$$

The losses are then virtually the same as they would be in a conventional thermal conductivity apparatus, apart from the reduction due to the absence of a heater.

The magnitudes of the parameters involved in the loss terms are best obtained experimentally. Measurements are made using samples of different length and shape. It is also advisable to check that the electrical contact resistance is negligible by conducting experiments on very short samples. In fact, the Harman technique offers what is probably the best way to determine the contact resistance for thermoelectric materials. The usual method for semiconductors involves the use of a probe that is scanned across the contact region. However, the contact resistances experienced using present-day methods for the attachment of electrodes are too small to be detected by such a method. Any deterioration in the observed figure of merit using the Harman procedure, when the sample becomes very short, may be attributed to the contacts.

Various authors have used different methods to collect their data from Harman-type measurements. It is usual to determine not only the figure of merit but the electrical conductivity and the Seebeck coefficient as well. The Seebeck coefficient is found by observing the temperature difference and the thermoelectric voltage after the current is interrupted but before ΔT has had time to change. If the value of α is to be accurate, the material must have a reasonably high value of zT, otherwise the temperature difference will be rather small.

One way of obtaining the adiabatic and isothermal electrical conductivities is to use direct and alternating currents. However, it is more usual to make use of a rapid data collection system to observe the potential difference and temperature difference as a function of time. Typical profiles are shown schematically in Fig. 8.14. It is assumed that the current has a steady value during the times that it is switched on. It is possible to measure the thermoelectric and resistive voltages and the temperature difference after the current is switched on or switched off.

The Harman method was used by Sharp et al. [25] to measure the thermal conductivity of polycrystalline Bi–Sb. However, there were anomalous features in the experimental results. Since the transport parameters in a transverse magnetic field were also observed, it was possible to determine the Lorenz number both theoretically and experimentally. The experimental Lorenz number was found to be much smaller than the theoretical value and no real explanation could be found. What was particularly puzzling was the fact that Jandl and Birkholz [26] had made similar measurements on single crystal Bi–Sb and had found no anomalies. It was only later that a reason for the unexpected observations of Sharp et al. was uncovered.

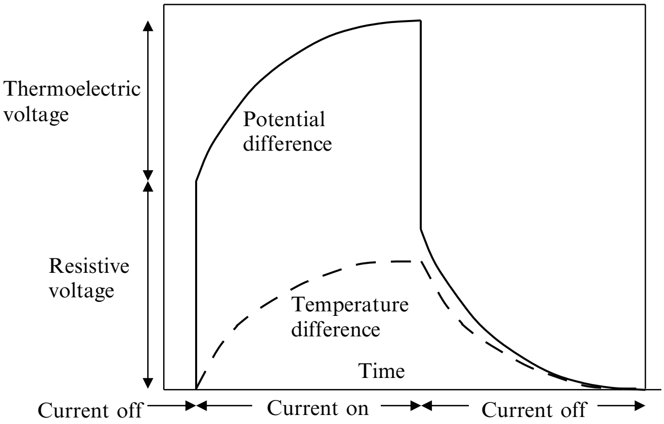


Fig. 8.14 Schematic variation of potential and temperature differences with time when the current is switched on and off in a Harman-type measurement

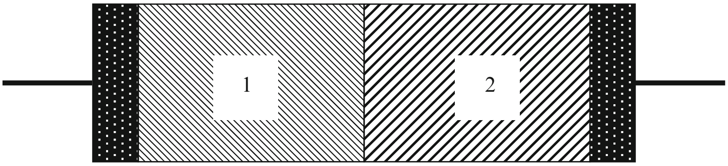


Fig. 8.15 Simple heterogeneous model consisting of two elements in series

It was realised that the Seebeck coefficient might vary within a polycrystalline sample since this quantity is quite strongly anisotropic in single crystals. However, it had been hoped that the experimental parameters for polycrystalline material would be some kind of average of the single crystal properties in the different directions. That such averaging is invalid becomes clear from the following considerations [27].

We suppose that a certain sample is made up of two components connected in series, as shown in Fig. 8.15. The two components are supposed to have electrical resistances R_1 and R_2 and thermal resistances W_1 and W_2, respectively. Then, a conventional thermal conductivity measurement will yield a value W equal to $W_1 + W_2$. Let us suppose that the Harman procedure leads to a different thermal resistance W^*.

In the Harman measurement, the isothermal voltage is given by

$$\frac{V}{I} = R_1 + R_2. \tag{8.28}$$

Under adiabatic conditions, the usual Harman theory can be applied to each component in turn to give

$$V_1 = IR_1\left(1 + z_1 T\right), \tag{8.29}$$

and

$$V_2 = IR_2\left(1 + z_2 T\right). \tag{8.30}$$

The overall adiabatic voltage is then related to the current by

$$\frac{V}{I} = R_1\left(1 + z_1 T\right) + R_2\left(1 + z_2 T\right). \tag{8.31}$$

The apparent figure of merit, z^*, is found from

$$\left(R_1 + R_2\right)\left(1 + z^* T\right) = R_1\left(1 + z_1 T\right) + R_2\left(1 + z_2 T\right). \tag{8.32}$$

In order to determine the thermal conductivity, we need to know the Seebeck coefficient for the system. Let us assume that it is found by applying a temperature difference ΔT between the end contacts. This temperature difference will distribute itself according to the relative thermal resistances of the components. Then, the overall Seebeck coefficient is given by

$$\alpha = \frac{\alpha_1 W_1 + \alpha_2 W_2}{W}, \tag{8.33}$$

and the total electrical resistance is

$$R = R_1 + R_2. \tag{8.34}$$

The apparent thermal resistance W^* can then be calculated from z^*, α and R.

By way of example, let us assume that both components have the same figure of merit but different Seebeck coefficients. We shall, then, let the ratio W_1/W_2 be equal to n, which will also be equal to R_1/R_2. In this example, α_1/α_2 also has the value n. In this case, the true value for the thermal resistance is

$$W = (1+n)\,W_1, \tag{8.35}$$

and the apparent thermal resistance is

$$W^* = \frac{(1+n)^2}{4n} W_1. \tag{8.36}$$

In general, W^* and W will be different. For example, if $n = 3$, $(1+n)^2/4n$ [2] becomes equal to 4/3 and W^* differs from the true thermal resistance by over 30%. A striking case is that for which α_1 and α_2 are equal and opposite, with $W_1 = W_2$ and $R_1 = R_2$. The apparent figure of merit z^* is then equal to the figure of merit of each component but the overall Seebeck coefficient is zero. The apparent thermal resistance W^* is, then, infinite. It is clear that under some circumstances, the Harman technique can lead to substantial errors in the thermal conductivity if the material is non uniform.

Calculations for $R_1 \neq R_2$ and for $W_1 \neq W_2$ have shown that there is no difference between W^* and W provided that $\alpha_1 = \alpha_2$. It seems clear, then, that the Harman method must be used with caution for heterogeneous materials. However, it should be perfectly satisfactory for polycrystalline extrinsic samples of bismuth telluride or any other anisotropic thermoelectric material in which the Seebeck coefficient is independent of crystal orientation.

There is a closely related technique that has sometimes been used for determining the figure of merit of modules [28, 29] and couples [30]. The thermal resistance is determined under the open-circuit and short-circuit conditions. The open-circuit thermal conductance, K, has the value that is calculated from the thermal conductivity and the dimensions of the thermoelements. When the open ends are short-circuited, the thermal conductance rises to a value K^* because of the extra heat transport associated with the thermoelectric current. This current is equal to $(\alpha_p – \alpha_n)\,\Delta T/R$ and it generates a Peltier heat flow equal to $(\alpha_p – \alpha_n)\,T\Delta T/R$ [2]. Thus, we see that

$$K^* = K\,(1 + ZT)\,. \tag{8.37}$$

8.7 Thermomagnetic Measurements

We shall discuss some of the principles that are involved in the measurement of the thermogalvanomagnetic coefficients, partly because of their immediate relevance to energy conversion, and partly because they are often needed for the understanding of thermoelectric materials.

Figur. 8.16 shows a typical arrangement of sample, heater, heat sink, and the various potential probes and thermocouples for the measurements. The sample will normally be located inside a cryostat so that measurements can be made under vacuum down to low temperatures. The thermomagnetic coefficients are often too small to be measured accurately at room temperature but they become larger at low temperatures because of an increase in the carrier mobility. The external dimensions of the cryostat should be kept as small as possible if an electromagnet is used to provide the field though nowadays it is common practice to make use of a superconducting magnet. A complete set of data requires the application of both transverse and longitudinal magnetic fields but very often a transverse field will suffice. This is the case, for example, in the measurement of the Hall and Nernst coefficients.

One must usually take care to avoid distortion of the equipotentials and isothermals from the end contacts unless, of course, it is the end effects that are being studied. A rule of thumb states that probes used in the determination of transverse coefficients should be inset from the ends by at least twice the sample width, while probes for observing longitudinal coefficients should be inset by one sample width. This implies that the sample should be at least four times as long as it is wide.

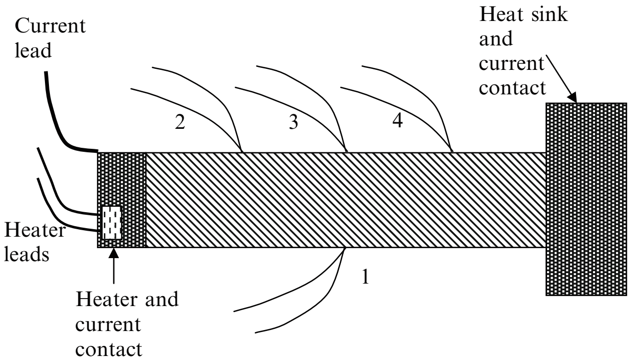


Fig. 8.16 Arrangement of contacts for thermogalvanomagnetic measurements. 1, 2, 3, and 4 are thermococouples, one lead of each also serving as a potential probe. The magnetic field is directed perpendicular to the diagram

One of the branches of each thermocouple can act as a potential probe. The wires, generally, should be of as fine a gauge as possible so that the disturbance of the flow of heat or electricity in the sample is minimal. They are often spark-welded in place to ensure good thermal contact. The heater block should be small since its thermal mass controls the rate at which equilibrium is reached. The measurements on a sample, mounted as in Fig. 8.16, normally yield the adiabatic coefficients. However, it is the isothermal Nernst coefficient that is implicit in the theory of thermomagnetic energy conversion presented in Sect. 2.5. Account must, therefore, be taken of the transverse temperature gradient and its influence on the transverse potential difference through the Seebeck effect. One must also include contributions from the Hall and Nernst effects in transforming the observed adiabatic electrical resistivity into the isothermal value.

Guthrie and Palmer [31] have shown that the thermomagnetic figure of merit Z_{NE} can be determined directly rather than through the measurement of individual transport parameters. Their method is based on the fact that the transverse temperature gradient has a much shorter time constant than the longitudinal temperature gradient. This is because samples are generally much longer than they are wide.

The measurement requires that the sample be provided with inset potential probes. The voltage between these probes is observed as a function of time. Immediately after the current is switched on, a resistive voltage is observed. Thereafter, the voltage changes as the longitudinal and transverse temperature gradients become established.

The longitudinal temperature gradient, arising from the Peltier effect, produces a voltage contribution from the Seebeck effect as in the Harman experiment. The transverse temperature gradient due to the Ettingshausen effect produces a longitudinal temperature gradient through the Righi–Leduc effect and this also gives rise to a Seebeck voltage. The transverse temperature gradient also leads to a longitudinal potential gradient directly through the Nernst effect. It is this last contribution that is related to the thermomagnetic figure of merit. It is also the only one of the three contributions to the changing longitudinal potential difference that does not involve a longitudinal temperature gradient. Since the transverse temperature gradient is established much more rapidly than the longitudinal temperature gradient, it is easily separated out as the effect that is established soon after the current is switched on.

If the sample is, say, 10 times as long as it is wide, the longitudinal thermal time constant is 100 times the transverse value. This means that, in a period of four transverse time constants, the transverse temperature gradient will be almost fully established but the longitudinal temperature gradient will still be negligible. In any case, the Seebeck coefficient is likely to be small in a good thermomagnetic material since it will probably be an intrinsic conductor. Figure 8.17 shows schematically a typical plot of the longitudinal voltage against time. The time over which the current is switched on, t_1–t_0, is supposed to be considerably less than the longitudinal thermal time constant. The depth, V_{N}, of the plateau below the resistive voltage, V_0, is due to the Nernst effect acting on the transverse temperature gradient. It is also equal in magnitude to the voltage that is present immediately after the current is switched off.

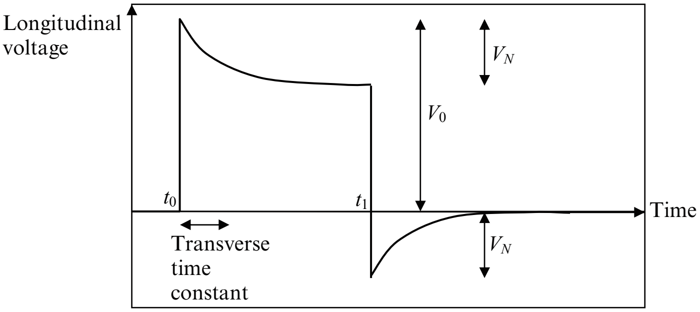


Fig. 8.17 Schematic plot of longitudinal voltage against time in the measurement of the thermomagnetic figure of merit using the technique of Guthrie and Palmer [28]

It is readily shown that the thermomagnetic figure of merit can be obtained from the values of V_0 and V_N. The transverse temperature gradient due to the Ettingshausen effect is

$$\frac{\mathrm{d}T}{\mathrm{d}y} = P i_\mathrm{x} B_\mathrm{z} = \frac{N T i_\mathrm{x} B_\mathrm{z}}{\lambda}. \tag{8.38}$$

The longitudinal field due to the Nernst effect is

$$E_\mathrm{N} = \frac{N^2 T i_\mathrm{x} B_\mathrm{z}^2}{\lambda}. \tag{8.39}$$

This is superimposed on the electric field, E_0, associated with the resistivity, ρ, which is given by

$$E_0 = i_\mathrm{x}\rho. \tag{8.40}$$

Combining (8.38) and (8.39), we find that

$$\frac{V_\mathrm{N}}{V_0} = \frac{E_\mathrm{N}}{E_0} = Z_\mathrm{NE}^\mathrm{i}. \tag{8.41}$$

In this equation, Z_NE^i is the isothermal thermomagnetic figure of merit, whereas it is the adiabatic thermomagnetic figure of merit, Z_NE, which is appropriate in transferring the thermoelectric energy conversion relations to the Nernst and Ettingshausen effects. Then,

$$Z_\mathrm{NE} T = \frac{V_\mathrm{N}}{V_0 - V_\mathrm{N}}. \tag{8.42}$$

The same adiabatic thermomagnetic figure of merit can also be found by measuring the maximum temperature depression, $(T_2 – T_1)_\mathrm{max}$ that can be achieved using an Ettingshausen cooler. Thus,

$$(T_2 - T_1)_\mathrm{max} = \frac{1}{2} Z_\mathrm{NE} T_1^2. \tag{8.43}$$

References

1. E.H. Putley, Proc. Phys. Soc. B **68**, 35 (1955)
2. H.J. Goldsmid, Proc. Phys. Soc. **71**, 633 (1958)
3. T.M. Dauphinee, S.B. Woods, Rev. Sci. Instrum. **26**, 693 (1955)
4. J.R. Drabble, R.D. Groves, R. Wolfe, Proc. Phys. Soc. **71**, 430 (1958)
5. J.P. McKelvey, *Solid State and Semiconductor Physics* (Harper and Row, New York, 1966), p. 380
6. L.E.J. Cowles, L.A. Dauncey, J. Sci. Instrum. **39**, 16 (1962)
7. A.V. Ioffe, A.F. Ioffe, Dokl. Akad. Nauk SSSR **97**, 821 (1954)
8. A.V. Ioffe, A.F. Ioffe, Zh. Tekh. Fiz. **22**, 2005 (1952)
9. A.V. Ioffe, A.F. Ioffe, Sov. Phys. Tech. Phys. **3**, 2163 (1958)
10. M.A. Kaganov, Sov. Phys. Tech. Phys. **3**, 2169 (1958)
11. W.F.G. Swann, J. Franklin Inst. **267**, 363 (1959)
12. R.W. Powell, J. Sci. Instrum. **34**, 485 (1957)
13. H.J. Goldsmid, J. Phys. E **12**, 1129 (1979)
14. R. Bowers, R.W. Ure, J.E. Bauerle, A.J. Cornish, J. Appl. Phys. **30**, 930 (1959)
15. D.G. Cahill, H.E. Fischer, T. Klitsner, E.T. Swartz, R.O. Pohl, J. Vac. Sci. Technol. A **7**, 1259 (1989)
16. D.G. Cahill, Rev. Sci. Instrum. **61**, 802 (1990)
17. A.J. Ångström, Ann. Phys. **114**, 513 (1861)
18. R. Nii, J. Phys. Soc. Japan **13**, 769 (1958)
19. B. Abeles, G.D. Cody, D.S. Beers, J. Appl. Phys. **31**, 1585 (1960)
20. A. Green, L.E.J. Cowles, J. Sci. Instrum. **37**, 349 (1960)
21. H.J. Goldsmid, M.E. Ertl, A.J. Dixon, J. Phys. D **7**, 131 (1974)
22. T.M. Tritt, D. Weston, in *Thermal Conductivity: Theory, Properties, and Applications*, ed. by T.M. Tritt (Kluwer/Plenum, New York, 2004), p. 197
23. T.C. Harman, J. Appl. Phys. **29**, 1373 (1958)
24. G.S. Nolas, J. Sharp, H.J. Goldsmid, *Thermoelectrics: Basic Principles and New Materials Developments* (Springer, Berlin, 2001), p. 102
25. J.W. Sharp, E.H. Volckmann, H.J. Goldsmid, Phys. Stat. Solidi A **185**, 257 (2001)
26. P. Jandl, U. Birkholz, J. Appl. Phys. **76**, 7351 (1994)
27. H.J. Goldsmid, J. Thermoelecticity, **1**, 5 (2006)
28. I.S. Lisker, Sov. Phys. Solid State **8**, 1042 (1966)
29. T.A.A. Al-Obaidi, H.J. Goldsmid, Energy Convers. **9**, 131 (1969)
30. W.B. Berry, Energy Convers. **9**, 133 (1969)
31. G.L. Guthrie, R.L. Palmer, J. Appl. Phys. **37**, 90 (1966)

Chapter 9
Review of Thermoelectric Materials

9.1 Bismuth and Bismuth–Antimony

Bismuth was one of the first thermoelectric materials to be studied, and for many years it was used in radiation thermopiles. It has a negative Seebeck coefficient when pure, and it was common practice to complete the thermocouple with another Group V element, antimony, which has a positive Seebeck coefficient.

Bismuth and antimony have the same crystal symmetry as bismuth telluride and cleave easily along the basal planes. The Brillouin zone of bismuth is similar to that of bismuth telluride shown in Fig. 6.4, but it is more extensive in the c-direction and shows only a slight distortion from cubic symmetry.

Both bismuth and antimony are semimetals; i.e., they have overlapping valence and conduction bands. The overlap is more pronounced in antimony than in bismuth for which the overlap is only about 0.02 eV [1].

Galvanomagnetic measurements have been performed on both bismuth and antimony. For bismuth, the results can be interpreted in terms of a three-valley conduction band with the extrema at the L-points in the Brillouin zone [2]. The ellipsoids are only slightly tilted away from the principal axes and it is a reasonable approximation to ignore the tilt [3]. There is a high effective mass in each valley along the bisectrix direction with rather small effective masses in the binary and trigonal directions. The valence band consists of a single valley with the surfaces of constant energy centred at the T-points in the zone. As required by the crystal symmetry, the surfaces are spheroidal about the trigonal axis.

The conduction band of antimony is also of the three-valley type [4] but the tilt angle is far larger than for bismuth. Antimony also has a three-valley valence band with almost spheroidal surfaces of constant energy tilted by about 60° from the trigonal direction.

It is of particular interest that, although both bismuth and antimony have overlapping bands, there is a range of Bi–Sb alloys that is semiconducting. This behaviour was first noticed by Jain [1], who suggested that the compositions between 4 and 40 molar% of Sb have a positive energy gap. The maximum gap of about 0.014 eV was thought to occur for the composition $Bi_{0.88}Sb_{0.12}$.

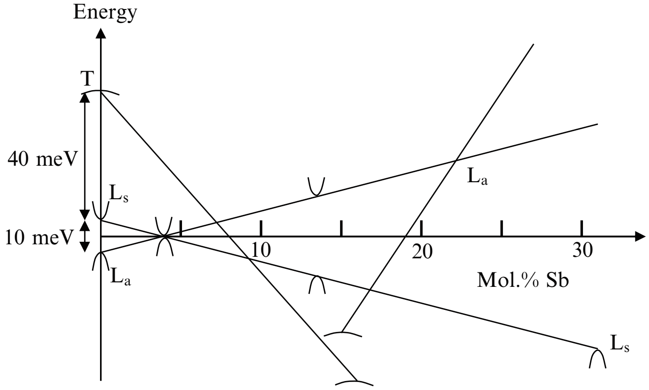


Fig. 9.1 Schematic band diagram for Bi–Sb alloys based on the data of Lenoir et al. [5]

Since the observations by Jain were performed, it has been realised that homogeneous alloys of bismuth and antimony are exceedingly difficult to produce because of the problems of constitutional supercooling. Not surprisingly, Jain's description of the bands has had to be modified. A more precise band scheme has been presented by Lenoir et al. [5], and is shown schematically in Fig. 9.1

There is a positive direct gap of 10 meV for pure bismuth but the heavy electron band at the T-point overlaps the conduction band by 40 meV. As antimony is added, the hole band and the heavy electron band move down and the light electron band moves up. The light bands cross at a concentration of 4% antimony and the heavy electron band moves below the light hole band at 7% antimony. At this point, the alloy becomes a semiconductor and remains so until the hole band crosses a heavy electron band at 22% antimony. Thus, the semiconductor region extends from 7 to 22% antimony with a maximum positive gap of about 30 meV at an antimony concentration between 15% and 17%.

Bismuth itself is close to being a good thermoelectric material at room temperature. It, certainly, has a very high value for the product $\mu\,(m^*/m)^{3/2}$ in the trigonal direction, probably higher than any other known material, but the presence of minority carriers causes the Seebeck coefficient to remain rather low, however the element is doped. Comprehensive measurements of the thermoelectric properties were carried out by Gallo et al. [6] and their results are summarised in the plots of Figs. 9.2 and 9.3 for the directions normal to and parallel to the trigonal axis, respectively.

It is immediately apparent from Figs. 9.2 and ,9.3 that the thermoelectric figure of merit is highest for bismuth with the current in the trigonal direction, that is the direction for which z is lowest in bismuth telluride. This means that thermoelements made from single crystal bismuth, aligned in the preferred direction, tend to cleave across the line of current flow. Although the figure of merit, z, is no more than $1.3 \times 10^{-3}\,\mathrm{K}^{-1}$ at 300 K, it rises to $1.7 \times 10^{-3}\,\mathrm{K}^{-1}$ at 100 K.

The high mobility of electrons in bismuth means that the thermogalvanomagnetic effects are large and easily observed. Thus, it has been possible to use the magneto-thermal resistance effect to separate the lattice and electronic components

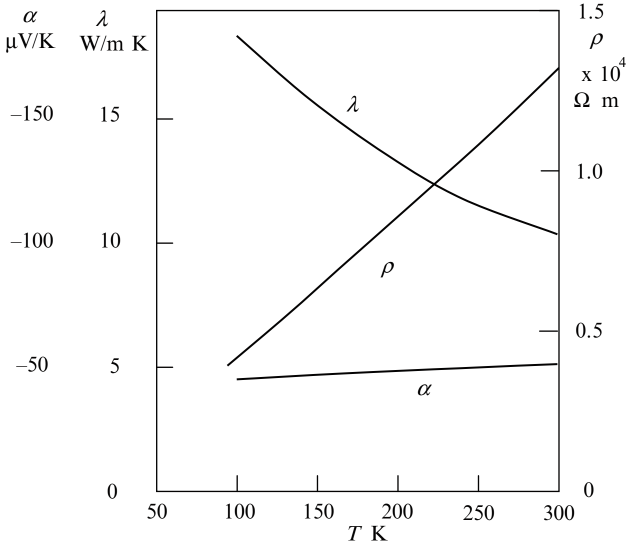


Fig. 9.2 Thermoelectric properties of bismuth normal to the trigonal axis

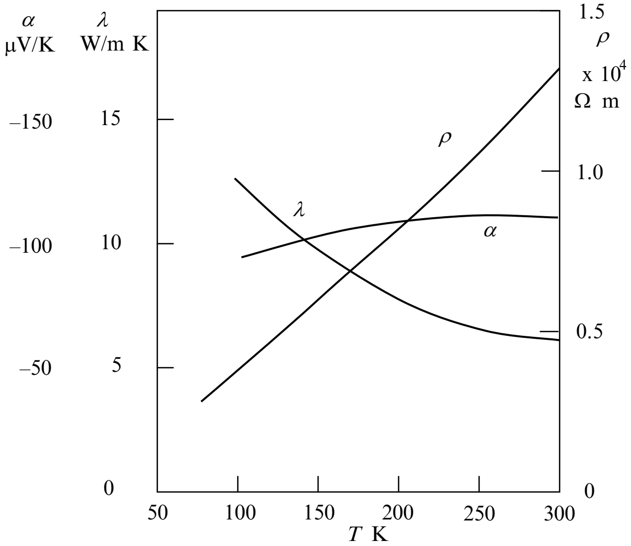


Fig. 9.3 Thermoelectric properties of bismuth parallel to the trigonal axis

of the thermal conductivity. For most conductors, this separation can be performed by extrapolating the observed thermal conductivity to the value for infinite magnetic field. However, there is a problem that arises for bismuth because it happens to have a high thermomagnetic figure of merit. The electronic thermal conductivity in a very large magnetic field does not then tend towards zero [7]. Uher and

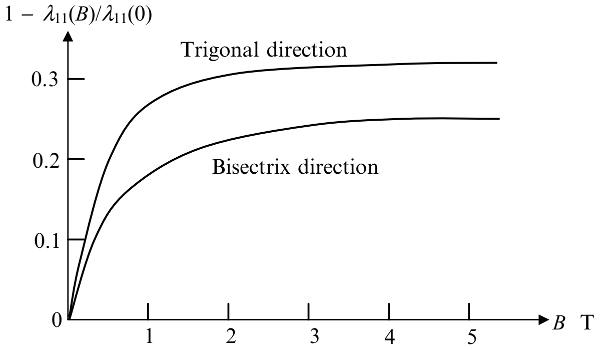
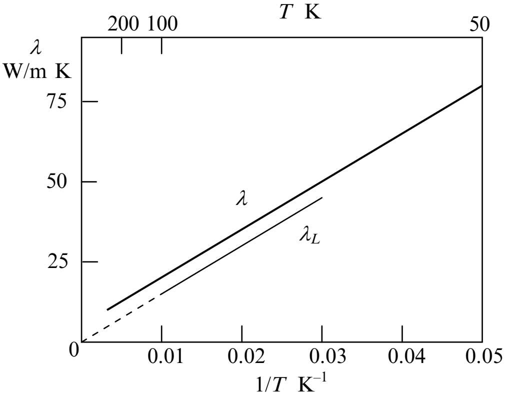


Fig. 9.4 Plots of the relative change in the total thermal conductivity of bismuth in the binary direction against the strength of the magnetic field in the bisectrix and trigonal directions. The temperature is about 115 K

Fig. 9.5 Total and lattice thermal conductivity of bismuth plotted against the reciprocal of the absolute temperature



Goldsmid [8] were able to overcome this difficulty by measuring the thermal conductivity with the magnetic field aligned successively in the bisectrix and trigonal directions. It is easy to predict the ratio of the electronic thermal conductivities for infinite magnetic fields in the two directions.

Figure 9.4 shows the relative change in the thermal conductivity, at about 115 K, of bismuth in a binary direction as a function of the magnetic field in the bisectrix and trigonal directions. The substantial difference between the two sets of readings confirms that the electronic thermal conductivity does not become zero whatever is the magnetic field strength. Figure 9.5 shows the total thermal conductivity and lattice conductivity of bismuth plotted against $1/T$. Over the range covered by the observations, λ_L is inversely proportional to the absolute temperature, in accordance with Eucken's law.

One of the interesting phenomena that can be observed in bismuth is the Umkehr effect. As mentioned in Chap. 1, the value of the Seebeck coefficient in a magnetic field may not remain the same when the direction of the field is reversed. Smith et al. [9] found that the Seebeck coefficient, of a particular crystal of bismuth in the bi-

Table 9.1 Basic parameters for bismuth at 80 and 300 K. The 1, 2, and 3 directions lie along the binary, bisectrix, and trigonal axes respectively. The lattice conductivities at 300 K have been derived from those at 80 K assuming λ_L proportional to 1/T

Parameter	80 K	300 K
Ne $\left(C/m^3\right)$	7.37×10^4	3.52×10^3
Electron mobility		
$\mu_n(1)\ \left(m^2\,V^{-1}\,s^{-1}\right)$	55.7	3.18
$\mu_n(2)\ \left(m^2\,V^{-1}\,s^{-1}\right)$	1.40	0.08
$\mu_n(3)\ \left(m^2\,V^{-1}\,s^{-1}\right)$	33.3	1.90
Hole mobility		
$\mu_p(1) = \mu_p(2)\ \left(m^2\,V^{-1}\,s^{-1}\right)$	12.4	0.77
$\mu_p(3)\ \left(m^2\,V^{-1}\,s\right)^{-1}$	3.33	0.21
Partial Seebeck coefficient $(\mu B)^2 \gg 1)$		
α_n (μV/K)	−100	−125
α_p (μV/K)	105	107
Lattice conductivity		
λ_L(normal to trigonal axis)$\left(W\,m^{-1}\,K^{-1}\right)$	11.0	2.9
λ_L(parallel to trigonal axis) $\left(W\,m^{-1}\,K^{-1}\right)$	7.5	2.0

sectrix direction at 80 K, was equal to $-150\,\mu V\,K^{-1}$ in a magnetic field of 1T at an angle of 60° to the binary direction. Upon reversal of the field, the Seebeck coefficient changed sign to a value of $170\,\mu V\,K^{-1}$. It is a general rule [10] that the Umkehr effect will be present for any semiconductor with non-spherical surfaces of constant energy, when the magnetic field does not lie along a reflection plane. The effect is, particularly, large in bismuth as it contains electrons and holes, both carriers being highly mobile.

The basic parameters at 80 and 300 K that are relevant to the thermoelectric, galvanomagnetic and thermomagnetic effects in bismuth are given in Table 9.1. These parameters were obtained from the data of Abeles and Meiboom [3], ignoring the tilt of the ellipsoids for the conduction band. The thermoelectric parameters obtained by Gallo et al. [6] were also used. The high field Seebeck coefficients were calculated from the expressions given by Tsidil'kovskii [11].

Bismuth is no longer used as a thermoelectric material at room temperature. In the absence of a magnetic field, it is inferior to bismuth telluride and the mobility near 300 K is not high enough to allow the thermomagnetic effects to be exploited. However, it is possible that it will find application at ordinary temperatures as a component in synthetic transverse thermoelements [12].

At low temperatures, it is a different matter. The negative energy gap is still a major disadvantage for thermoelectric applications in zero magnetic field, but near liquid nitrogen temperature, the mobility is large enough for μB to become of the order of unity or greater, as is needed for thermomagnetic applications. For this purpose, the semiconducting Bi–Sb alloys may be superior as they have lower values of the lattice conductivity but the mobility for a given carrier concentration is also less. The parameters given in Table 9.1 have been used to predict the thermomagnetic

Table 9.2 Calculated thermomagnetic figure of merit for single crystal bismuth in different orientations

Direction of temperature gradient	Direction of electric current	Z_{NE} at 80 K $((\mu B)^2 \gg 1)$	Z_{NE} at 300 K $((\mu B)^2 \gg 1)$
Bisectrix	Binary	$0.63 \times 10^{-3}\,K^{-1}$	$0.82 \times 10^{-3}\,K^{-1}$
Binary	Bisectrix	$0.63 \times 10^{-3}\,K^{-1}$	$0.82 \times 10^{-3}\,K^{-1}$
Binary	Trigonal	$2.1 \times 10^{-3}\,K^{-1}$	$2.9 \times 10^{-3}\,K^{-1}$
Bisectrix	Trigonal	$0.84 \times 10^{-3}\,K^{-1}$	$1.1 \times 10^{-3}\,K^{-1}$
Trigonal	Binary	$1.2 \times 10^{-3}\,K^{-1}$	$1.8 \times 10^{-3}\,K^{-1}$
Trigonal	Bisectrix	$1.2 \times 10^{-3}\,K^{-1}$	$1.8 \times 10^{-3}\,K^{-1}$

figure of merit for various orientations in Table 9.2. The high values for 300 K are probably of no practical interest in view of the very high magnetic field that would be needed to approach the condition $(\mu B)^2 \gg 1$.

As shown in Table 9.2, the thermomagnetic figure of merit might reach a value in excess of $2 \times 10^{-3}\,K^{-1}$ at 80 K and the required magnetic field could probably be attained with a permanent magnet [13]. However, $Z_{NE}T$ would still be no more than about 0.17 so the cooling that would be achieved with an Ettingshausen refrigerator based on bismuth would be rather small.

The thermomagnetic figure of merit has been measured by Yim and Amith [14] over the temperature range 70–300 K in a magnetic field of 0.75T. They confirmed that the highest value is obtained for the predicted orientation. Because the high magnetic field condition is far from being reached at 300 K, Z_{NE} at this temperature was found to be no more than about 0.025, but at 80 K a value of 0.24 was observed, which may be compared with the calculated value of 0.17. The agreement is remarkably good in view of the approximations that have been made.

Bi–Sb alloys have attracted interest for both thermoelectric and thermomagnetic applications at low temperatures because of the possibility of obtaining both a positive energy gap and a reduced lattice conductivity. It is important, therefore, to determine the effect of alloying on the effective mass and the mobility.

Observations by Smith [15] of cyclotron resonance in $Bi_{0.95}Sb_{0.05}$ have shown that, although the shape of the ellipsoids does not change appreciably on adding antimony to bismuth, the values of the effective masses are reduced. This is not unexpected if the bands are non-parabolic since the effective mass should then become smaller as the Fermi level moves closer to the band edge. This idea is supported by the work of Brandt et al. [16] on alloys containing 1.7–4 molar% antimony. Unfortunately, the decreased effective mass is not accompanied by an increase of the mobility. Thus, Jain [1] found no significant differences between the electron and hole mobilities in $Bi_{0.95}Sb_{0.05}$ and bismuth, and he noted a decrease by a factor of 2 for the average mobility in $Bi_{0.93}Sb_{0.07}$. However, the decrease in the lattice conductivity may be sufficient compensation for the reduction in $\mu\,(m^*/m)^{3/2}$. Figure 9.6 shows the observations of Cuff et al. of the total thermal conductivity of different Bi–Sb alloys at 80 K, as a function of the transverse magnetic field. For the reasons that we have already discussed, the thermal conductivity

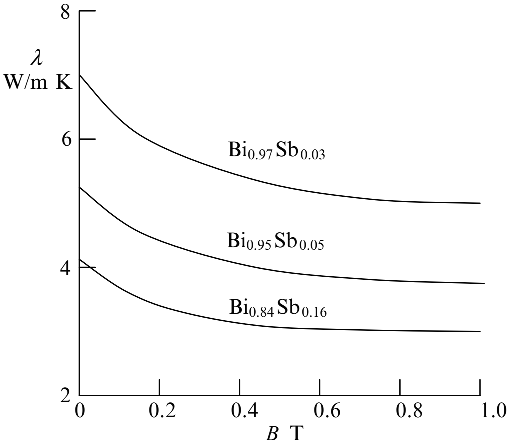


Fig. 9.6 Thermal conductivity against magnetic field for Bi–Sb alloys at 80 K. The temperature gradient and magnetic field are in the binary and bisectrix directions

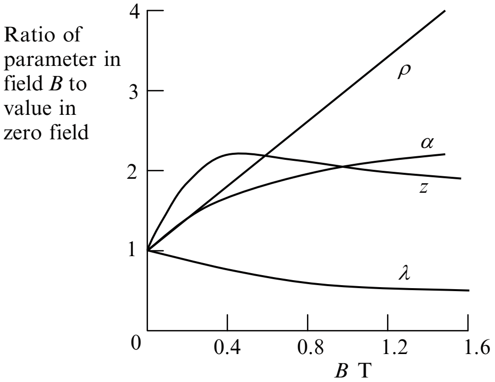
may not be approaching the lattice value at the higher fields, particularly, since the results were obtained for the orientation that yields the highest value of Z_{NE}. However, it is clear enough that the lattice conductivity of all three alloys is considerably less than the value of 11 W m^{-1} K^{-1} for bismuth in the same direction. The results obtained by Horst and Williams were obtained for other orientations and probably give a more reliable indication of the lattice conductivity. Their values for λ_L are 3.7 W m^{-1} K^{-1} for $Bi_{0.95}Sb_{0.05}$ and 3.1 W m^{-1} K^{-1} for $Bi_{0.88}Sb_{0.12}$.

Bismuth and Bi–Sb alloys can be doped with donor impurities such as tellurium and acceptor impurities such as tin. However, it was thought for many years that the undoped material has close to the optimum properties for the negative branch of a couple. Positive Seebeck coefficients are the result of doping with acceptors, but the figure of merit is less than for the negative material.

The thermoelectric properties of both undoped and doped Bi–Sb have been given by Wolfe and Smith [17]. The value of z for undoped $Bi_{0.88}Sb_{0.12}$ is 5.2×10^{-3} K^{-1} at 80 K, but this falls to 2.0×10^{-3} K^{-1} at 200 K and to 1.0×10^{-3} K^{-1} at 300 K. A more dilute alloy, $Bi_{0.95}Sb_{0.05}$ yields a z value of only 4.8×10^{-3} K^{-1} at 80 K, but at 300 K, the value of 1.8×10^{-3} K^{-1} is superior to that of $Bi_{0.88}Sb_{0.12}$.

Rather better values were obtained by Jandl and Birkholz [18] for tin-doped $Bi_{0.95}Sb_{0.05}$. A sample with a tin content of 145 ppm gave a figure of merit in the trigonal direction of about 3×10^{-3} K^{-1} over a wide range of temperature from about 120 K to 280 K. This material yielded a superior value for the figure of merit to that of n-type bismuth–telluride alloys over most of this range. However, there are practical reasons for preferring the use of bismuth–telluride. Aligned polycrystals suffice for bismuth–telluride alloys whereas Bi–Sb needs to be in the form of a single crystal. Moreover, Bi–Sb crystals aligned in the preferred direction are liable to cleavage fracture across the flow lines. An important feature of the results obtained by Jandl and Birkholz is that they dispel the idea that undoped material has the best thermoelectric properties.

Fig. 9.7 Thermoelectric properties of $Bi_{0.88}Sb_{12}$ at 160 K. Electric and thermal flows in the trigonal direction and magnetic field along a bisectrix axis [17]



The thermoelectric figure of merit can be improved significantly, particularly, at low temperatures, by the application of a magnetic field. For example, Wolfe and Smith [17] found that z for $Bi_{0.88}Sb_{0.12}$ at 160 K is more than doubled in a magnetic field of 0.6T, as shown in Fig. 9.7. Jandl and Birkholz [18] obtained a figure of merit as high as $5 \times 10^{-3}\,K^{-1}$ at 293 K but only after applying a magnetic field of 0.9T.

Bi–Sb alloys doped with tin have been seriously considered as p-type thermoelements. Yim and Amith [14] reported a figure of merit of $0.3 \times 10^{-3}\,K^{-1}$ at 90 K for p-type $Bi_{0.88}Sb_{0.12}$ containing 300 ppm tin, but at 85 K, z became equal to $2.3 \times 10^{-3}\,K^{-1}$ in a magnetic field of 0.75T. This value was not exceeded in any sample studied by Jandl and Birkholz [18] and must be regarded as close to the best that can be achieved in such a magnetic field.

The magneto-Seebeck effect in bismuth and Bi–Sb is clearly influenced by the transverse thermomagnetic phenomena. This was demonstrated Ertl et al. [19] in their measurement of the Seebeck coefficient at 80 K on a crystal of $Bi_{0.93}Sb_{0.07}$. The temperature gradient lay in the trigonal direction and a magnetic field was applied in a bisectrix direction. The length of the sample was varied so that the ratio of length to width changed from 0.71 to 2.55. The Seebeck coefficient is plotted against magnetic field in Fig. 9.8.

Although the shape dependence of the magneto-Seebeck coefficient is evidence for the influence of the transverse effects, this may not be the whole explanation of the observations. Thomas and Goldsmid [20] found that the Seebeck coefficient of $Bi_{0.95}Sb_{0.05}$ certainly showed a greater change in a magnetic field when there were two types of carrier, as one would expect if the transverse effects were contributing. However, the change from −60 to −123 μV K^{-1} resulting from the application of a field of 1.6T was thought to be evidence of a non-parabolic dispersion law.

Bi–Sb alloys may, in theory, be superior to pure bismuth as thermomagnetic materials but, in practice, the reduction in the mobility may be too great a disadvantage since it means that a higher magnetic field is needed to satisfy the condition $(\mu B)^2 \gg 1$. Figure 9.9 shows the dimensionless thermomagnetic figure of merit of bismuth and $Bi_{0.99}Sb_{0.01}$ plotted against temperature for a magnetic field of 0.75T. The alloy is superior below 130 K but the pure bismuth has the higher figure of

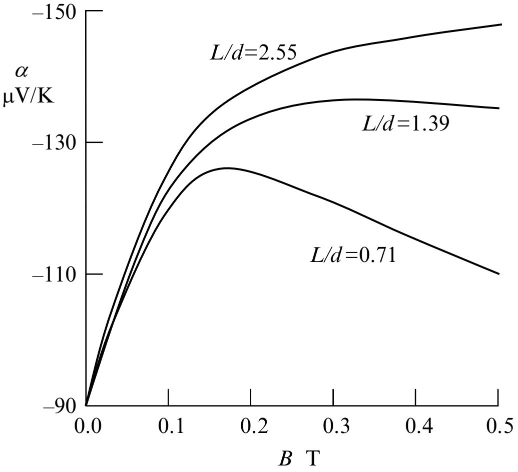


Fig. 9.8 Seebeck coefficient plotted against magnetic field for $Bi_{0.93}Sb_{0.07}$ at 80 K. The length L is in the trigonal direction and the magnetic field is along a bisectrix direction. The width d is perpendicular to the magnetic field

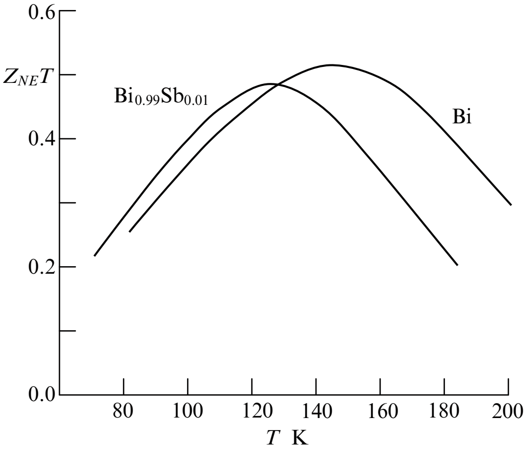
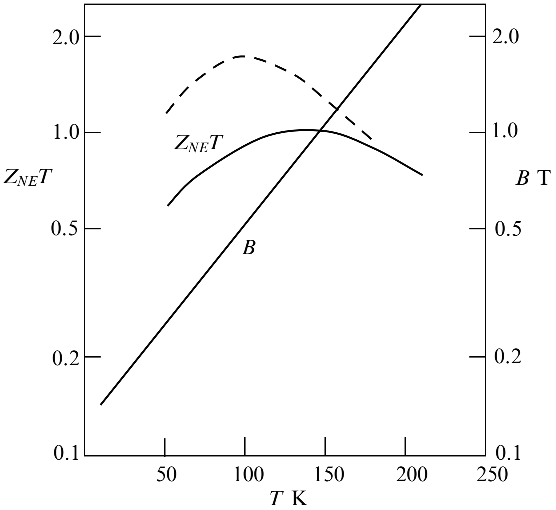


Fig. 9.9 Dimensionless thermomagnetic figure of merit plotted against temperature for Bi and $Bi_{0.99}Sb_{0.01}$. The magnetic field is 0.75 T. Observations of Yim and Amith [14]

merit above this temperature. No doubt the Bi–Sb alloy would be the better material at higher temperatures if a larger magnetic field were available.

The thermomagnetic figures of merit shown in Fig. 9.9 fall short of the values achieved by later workers. Horst and Williams [21] showed that the purity of the material is of the utmost importance in attaining high values of Z_{NE}. They were able to reach a value for $Z_{NE}T$ of about unity for $Bi_{0.97}Sb_{0.03}$ at 150 K in a magnetic field of 1T. Their material had no more than 10^{20} excess carriers per cubic metre. The curve in Fig. 9.10 shows the values of $Z_{NE}T$ observed by Horst and Williams plotted against temperature and the straight line on the same diagram shows the magnetic field strength that was required to reach these values. The broken curve is a tentative representation of the value of Z_{NE} to be expected, if the number of excess carriers could be reduced to 10^{16} per cubic metre. It is possible that the previous results, which indicated a preference for pure bismuth, are an indication of the difficulty of achieving purity and homogeneity in Bi–Sb.

Fig. 9.10 Schematic representation of the variation of $Z_{NE}T$ with temperature for $Bi_{0.97}Sb_{0.03}$ according to Horst and Williams [21] (solid curve). The line shows the requisite strength of magnetic field. The broken curve is an estimate of the value of $Z_{NE}T$ that might be reached if the excess carrier concentration could be reduced to $10^{16}/m^3$



9.2 Lead Telluride and Related Compounds

When Ioffe et al. realized, in the early 1950s, that semiconductors of high mean atomic weight were likely to be good thermoelectric materials, they became interested in lead telluride and isomorphous compounds. For example, those semiconductors were used to exemplify the principle [22] that solid solutions have a reduced lattice conductivity without there necessarily being a change in the carrier mobility.

When lead telluride is compared with bismuth telluride, higher melting temperature, 923°C compared with 585°C, and an energy gap of 0.32 eV compared with 0.13 eV, is noted. This means that, although the value of the figure of merit of PbTe is lower than for Bi_2Te_3, it can be used up to considerably higher temperatures without chemical stability problems or unwanted contributions from the minority charge carriers. PbTe, then, has been considered more as a material for thermoelectric generation at moderately high temperatures rather than for refrigeration at room temperature and below.

Both p-type and n-type materials can be produced either by departures from stoichiometry or by doping with donor or acceptor impurities [23]. There is a wide choice of dopants; Na, Au, Ti, and O behave as acceptors and Zn, Cd, In, Bi, and Cl are donors. PbTe has the cubic rock salt structure so the thermoelectric properties are isotropic. Hall effect measurements show that the mobilities of both types of carrier are rather high [24]. Thus, at a temperature of 295 K, μ_n is equal to $0.16\,m^2\,V^{-1}\,s^{-1}$ and μ_p is $0.075\,m^2\,V^{-1}\,s^{-1}$. However, the density-of-states effective masses of both electrons and holes are only about 0.03 m. Taking these factors into account, together with a lattice conductivity of $2.0\,W\,m^{-1}\,K^{-1}$ at room temperature, it is found [25] that the maximum value of z is little more than $1 \times 10^{-3}\,K^{-1}$.

In any practical application, one would expect to use a solid solution of the form $Pb_xSn_{1-x}Te_ySe_{1-y}$ or a similar alloy rather than PbTe or one of the other pure

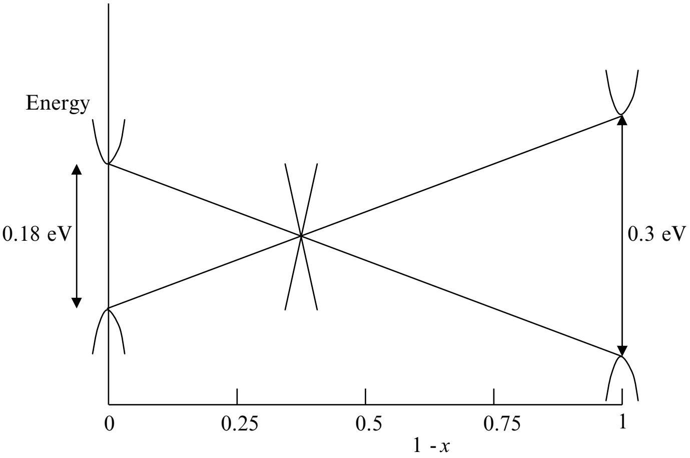


Fig. 9.11 Schematic plot against x of energy at the band edges in $Pb_xSn_{1-x}Te$. Based on the data of Fano [23] for a temperature of 12 K

compounds. Early work [26] on these alloys suggested that $x = 1$ is the appropriate choice for p-type thermoelements and $y = 1$ is preferable for n-type material. This seems to be true if we take account only of the mobilities of the electrons and holes but the lattice conductivity is lower when $y = 1$ than when $x = 1$ and this outweighs the relative effects on the mobilities [25]. Thus, $Pb_xSn_{1-x}Te$ is preferred for both types of thermoelement, if one is restricted to this range of alloys.

There are, however, problems with the use of $Pb_xSn_{1-x}Te$. The energy gap falls rapidly as x becomes smaller and becomes zero when $x \simeq 0.6$. On either side of the zero point, the energy gap is positive though, beyond this point, the positions of the band extrema in wave–vector space are interchanged as shown in Fig. 9.11. The energy gap becomes larger on the PbTe-rich side and the cross-over point changes as the temperature rises. Rosi et al. [27] suggested the use of the composition $Pb_{0.75}Sn_{0.25}Te$ for practical applications, but one really needs to consider carefully the temperatures at which the thermoelectric material will be used.

The figure of merit is higher for n-type $Pb_xSn_{1-x}Te$ than for p-type material and reaches a value of about unity at temperatures of the order of 500 K. However, over the temperature range for which PbTe and its alloys are suitable, the positive material is commonly one of the so-called TAGS formulations. TAGS is an acronym for alloys that contain the elements Te, Ag, Ge, and Sb. These alloys are solid solutions between $AgSbTe_2$ and GeTe, the latter being closely related to PbTe, since Pb and Ge are in the same group of the periodic table. GeTe has the rock salt structure but $AgSbTe_2$ is rhombohedral and there is a phase transition at a composition that contains about 80% GeTe. Although it might be thought that one should avoid compositions close to that of the phase transition, it turns out that the alloys with 80 and 85% GeTe have exceptional thermoelectric properties [25]. It seems that the lattice conductivity is, particularly, small in this region, presumably due to strain scattering.

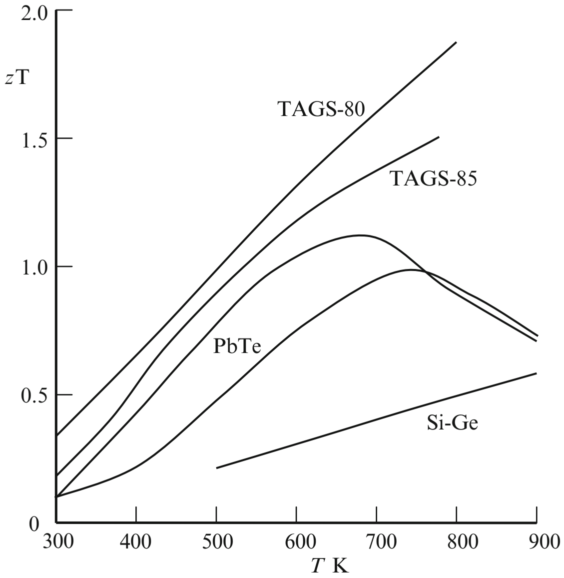


Fig. 9.12 Plots of zT against temperature for TAGS formulations and other positive materials. TAGS-80 and TAGS-85 are alloys of $AgSbTe_2$ with 80% and 85% GeTe respectively. The two curves for PbTe represent different doping constituents. Based on the data of Skrabek and Trimmer [28]

One might expect some mechanical problems for alloys that lie close to the phase transition and, indeed, the alloy containing 85% GeTe is less prone to cracking. Figure 9.12 shows the dimensionless figure of merit plotted against temperature for the TAGS materials together with data for p-type PbTe and Si–Ge.

Recent observations by Heremans et al. [29] suggest an interesting route to the improvement of PbTe and, indeed, other thermoelectric materials. They studied the properties of PbTe doped with thallium, pointing out that this element creates resonant energy levels, as do gallium and indium. They showed that this can lead to an enhanced density of states in the valence band, and hence an improvement in the figure of merit. The improvement can be quite substantial as shown in Fig. 9.13. In this figure, zT is plotted against temperature for PbTe doped with a normal acceptor impurity, Na, and two different levels of Tl. The Na-doped sample has a zT value of about 0.7 at 700 K whereas the value for PbTe doped with 2% Tl is about 1.5. There is some support for the suggestion that the density of states has been modified from the observation that the electrical resistivity has an anomalously high value below 200 K. It is possible that enhanced densities of states may be obtained in other materials if suitable doping agents can be found. Perhaps, the key to finding these dopants will be the observation of similar resistivity anomalies.

Although it does not have the same structure as PbTe, $AgSbTe_2$ (i.e., TAGS without GeTe) is a promising material in its own right. $AgSbTe_2$ and the closely related $AgBiTe_2$ have recently been studied by Morelli et al. [30]. They found that, while $AgInTe_2$ has a lattice conductivity that is not much smaller than that of PbTe at

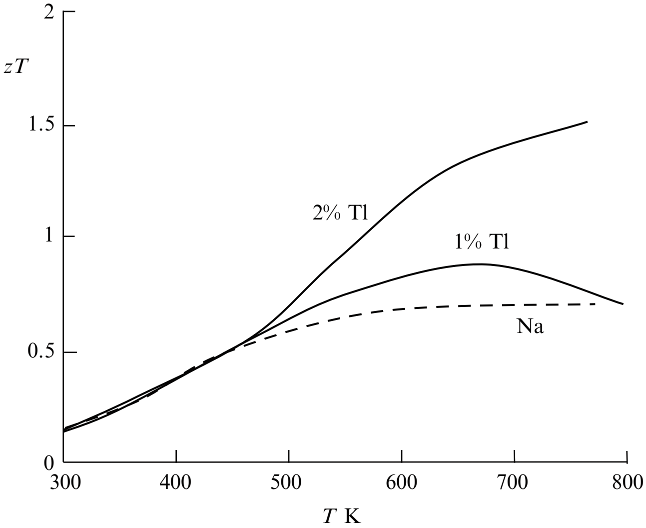


Fig. 9.13 Dimensionless figure of merit plotted against temperature for PbTe doped with thallium. Values for sodium-doped PbTe are also shown. Schematic plot based on the data of Heremans et al. [29]

ordinary temperatures, $AgSbTe_2$ and $AgBiTe_2$ both have lattice conductivities that are close to the calculated minimum value, the value that would be expected in the amorphous state. This small lattice conductivity is independent of impurity or defect concentration and is regarded as an intrinsic property of the material. Morelli and his colleagues did not determine the electronic properties of their material but they referred to earlier work on $AgSbTe_2$ in which a value of zT equal to 1.3 at 720 K was reported [31, 32].

Closely related to $AgSbTe_2$ is $AgPb_mSbTe_{2+m}$, which has a "rock salt"-like structure. This material is remarkable in that zT rises continuously from room temperature upwards, reaching a value of 2.1 at 800 K [33]. The material is n-type and the authors speculate on the possibility that the high figure of merit may be due to a quantum dot effect. Electron microscopy studies revealed the presence of quantum dot-sized regions that are rich in Ag and Sb.

9.3 Silicon–Germanium Alloys

Both silicon and germanium have rather high lattice conductivities, though they can yield reasonably large values for the power factor since both elements have high carrier mobilities. Thus, if the lattice conductivity can be reduced, the figure of merit will rise to a worthwhile value. We shall see in Chap. 12 that a large value of zT can be obtained at room temperature using silicon nanowires. Here, we shall discuss the use of silicon, or rather Si–Ge alloys, in bulk thermoelements.

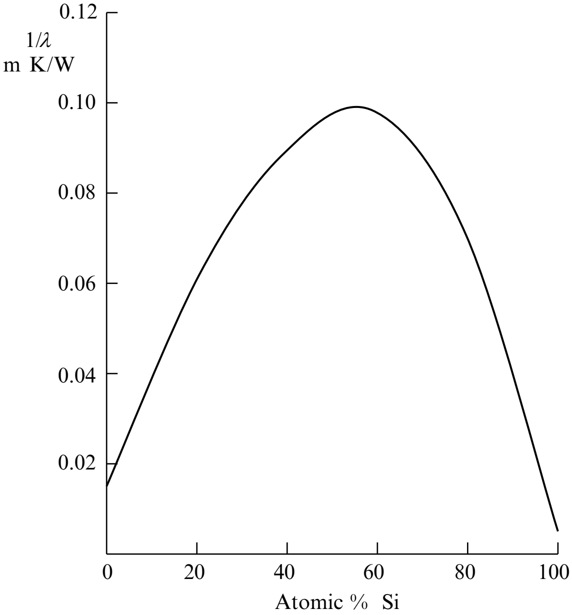


Fig. 9.14 Thermal resistivity of silicon-germanium alloys at 300 K. Schematic plot based on the data of Steele and Rosi [35]

The lattice conductivities of silicon and germanium at 300 K are 145 and 64 W m^{-1} K^{-1}, respectively [34]. The value of λ_L falls rapidly on adding germanium to silicon, as is apparent from Fig. 9.14 in which the thermal resistivity is plotted against the concentration of silicon in germanium for the whole range of Si–Ge alloys [35]. In fact, even larger increases in thermal resistivity have been observed by later workers. For example, Vining [36] gives a value for the thermal resistivity in the range 0.16–0.20 m K/W for $Si_{0.7}Ge_{0.3}$ at room temperature. It is possible that the relatively low thermal resistivity observed by Steele and Rosi for the Si–Ge alloys could have been due to inhomogeneity. The liquidus and solidus curves are widely separated and constitutional supercooling is a significant problem for melt-grown material.

In spite of the dramatic decrease in the lattice conductivity when germanium is alloyed with silicon, the Si–Ge alloys cannot compete with other thermoelectric materials at ordinary temperatures. However, they come into their own at, say, 600 K. Above this temperature, zT for both n-type and p-type Si–Ge reaches a value of about 0.5 and remains at or above this level up to temperatures in excess of 1,000 K [25]. $Si_{0.7}Ge_{0.3}$ has a solidus temperature of about 1,500 K and remains stable over long periods at 1,300 K [31]. The energy gaps for silicon and germanium are 1.15 and 0.65 eV, respectively so the silicon-rich alloys, when heavily doped, remain effectively free of minority carriers up to high temperatures.

A powder metallurgical technique is the preferred method for producing Si–Ge thermoelements [37]. Since these materials have the cubic diamond structure, there is no question of anisotropy of the thermoelectric properties. The parameters that are chosen for a hot-pressing and sintering process do not seem to be critical but there remains the choice of the size of the starting powders.

Because of the major contribution of alloy scattering to the thermal resistance, most of the heat is carried by low-frequency phonons. As shown in Sect. 5.3, this means that boundary scattering can have a large effect on the lattice conductivity. However, because of the high carrier mobility, we might expect this quantity also to be sensitive to boundary scattering. Although Slack and Hussain [38] have doubted the possibility of improving the figure of merit by reducing the grain size in sintered Si–Ge, an improvement for fine-grained material has been claimed by Rowe and Bhandari [39].

Slack and Hussain [38] have carried out a complete review of the properties of Si–Ge alloys with the aim of specifying the maximum efficiency that might be achieved for a generator operating between 300 and 1,300 K. Their calculations focussed on the $Si_{0.7}Ge_{0.3}$ alloy since this seems to have the most favourable combination of properties. Due account was taken of the complexities of the band structure. For example, the valence band maximum at the centre of the Brillouin zone actually has three components. There is a heavy mass band, a light mass band, and a split-off light mass band. Turning to the conduction band, there are minima at the X and L points that have to be considered. Such complications have an effect on the variation of Seebeck coefficient with carrier concentration and electrical conductivity. The carrier mobility, too, varies in an irregular manner with the carrier concentration. When it came to optimising the carrier concentration, it was found by Slack and Hussain that there were two maxima for the power factor in n-type material. The carrier concentrations for the two peaks did not depend strongly on temperature. One was at just over 10^{26} electrons per cubic metre and the other at around 10^{27} electrons per cubic metre. This means, of course, that there are two maxima when the figure of merit is plotted against carrier concentration.

Slack and Hussain supposed that some means is found to reduce the lattice conductivity below its value when only Umklapp and alloy scattering exist. As has already been mentioned, they doubted that boundary scattering would be useful but micro-inclusions might have a beneficial effect. Whatever mechanism is introduced to lower the lattice conductivity, it will never fall below a value $\lambda_{\min}$ that represents the value for amorphous material. A quantity f signifies the reduction of the lattice conductivity due to these unspecified processes. f is equal to zero without this additional scattering, and to unity when the lattice conductivity is equal to $\lambda_{\min}$. Slack and Hussain presented data for the maximum figure of merit as a function of temperature for different values of f but Fig. 9.15 is restricted to the results for f equal to zero. It shows the figure of merit zT plotted against the absolute temperature for both n-type and p-type $Si_{0.7}Ge_{0.3}$. Practical figures of merit have fallen some way short of the projected values. Nevertheless, there seems to be the possibility of exceeding these projections, if forms of phonon scattering can be introduced that increase f above zero.

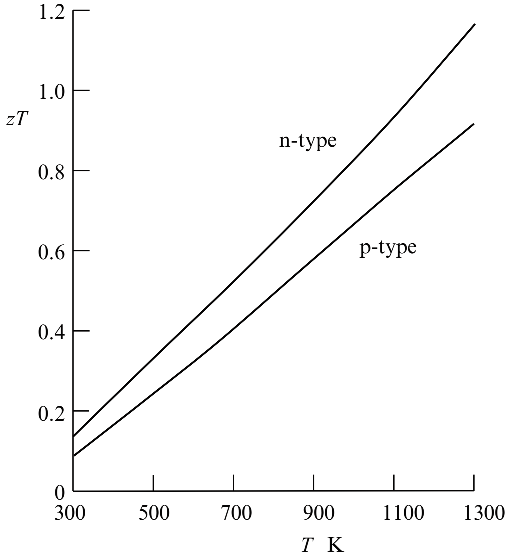


Fig. 9.15 Projected values of zT for $Si_{0.7}Ge_{0.3}$ plotted against temperature. A schematic plot based on the calculations of Slack and Hussain [38]

9.4 Skutterudites and Clathrates

9.4.1 Skutterudites

We now turn to the materials that have been studied as the likely embodiments of the phonon-glass electron-crystal (PGEC) concept that was discussed in Sect. 5.6. The types of material that might have a low lattice conductivity and at the same time have a large power factor have been considered by Nolas et al. [40]. They drew attention to the criteria that were enumerated by Cahill et al. [41] for a glass-like lattice conductivity.

The materials should possess loose atoms or molecules that have more than one metastable position in the lattice and do not have well-defined positions or motion. These atoms or molecules have no long range correlation with one another. They are an integral part of the compound rather than impurities, contributing at least 3% to the total mass. It is important to realise that the role of these loose atoms is quite different from that of the impurities that usually lead to point-defect scattering.

One of the classes of material that have been studied is that of the skutterudites. Skutterudite is the name first given to the mineral $CoAs_3$ and since extended to other compounds in the same family, such as $CoSb_3$. Such materials satisfy the requirements of Cahill et al. for a small lattice conductivity and possess reasonably mobile charge carriers.

A key feature of the unit cell of a skutterudite is that it contains empty spaces. In $CoSb_3$, the cobalt atoms form an almost cubic framework with square arrangements of each set of four antimony atoms, there being six such squares for every eight pseudo cubes. The voids that exist in such an arrangement can be occupied by loosely bound atoms that are known as “rattlers”. It is these rattlers that reduce the lattice conductivity to an extremely low level.

The general formula for the basic skutterudites can be written as MX_3 where M is Co, Rh, or Ir and X is P, As, or Sb. An alternative representation is $[\,]_2\,M_8X_{24}$, which takes account of the two voids for every 8 M atoms and 24 X atoms. We may then introduce guest atoms into the voids to form what can be called filled skutterudites. An indication of the ability of the guest atoms to rattle, and thereby reduce the lattice conductivity, is the atomic displacement parameter. One partly filled mixed skutterudite that has been studied is $La_{0.75}Fe_3CoSb_{12}$. Here, the atomic displacement parameter [42] at 300 K of La is about 0.02 Å^2 compared with about 0.007 Å^2 for Sb and 0.005 Å^2 for Fe and Co. It is expected, then, that the lattice conductivity of $La_{0.75}Fe_3CoSb_{12}$ should be small. This is, indeed, the case as is illustrated in Fig. 9.16 in which the lattice conductivity of the closely related compound $La_{0.75}Th_{0.2}Fe_3CoSb_{12}$ is plotted against temperature. Also shown are the values for the unfilled skutterudite $CoSb_3$ and the amorphous material, vitreous silica. The minimum thermal conductivity of the filled skutterudites could, in principle, reach even lower values.

Sales and his colleagues measured all the thermoelectric parameters on their material and were, thus, able to assess the figure of merit. Two of the compositions that gave, particularly, favourable results were $Ce_{0.9}Fe_3CoSb_{12}$ and $La_{0.9}Fe_3CoSb_{12}$. It was observed that zT became equal to about unity for both these compounds at 700 K, as shown in Fig. 9.17, and it was predicted that a value of about 1.4 might be reached at 1,000 K. It is noted that both materials displayed a Seebeck coefficient equal to about 200 $\mu V\,K^{-1}$ at 700 K, this probably being close to the optimum level. At lower temperatures, the Seebeck coefficient became smaller and it is expected that the figure of merit in this region could be improved.

The skutterudites are unlike many semiconductors in that the density-of-states effective mass is rather large so that the optimum Seebeck coefficient is reached at

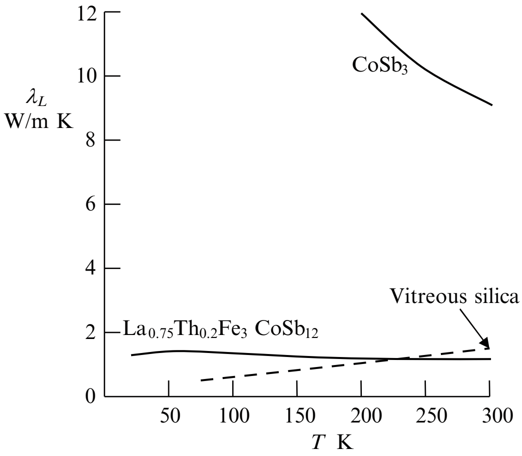


Fig. 9.16 Lattice conductivity of a filled skutterudite plotted against temperature. Also shown are values for $CoSb_3$ and vitreous silica. Based on the data of Sales et al. [42]

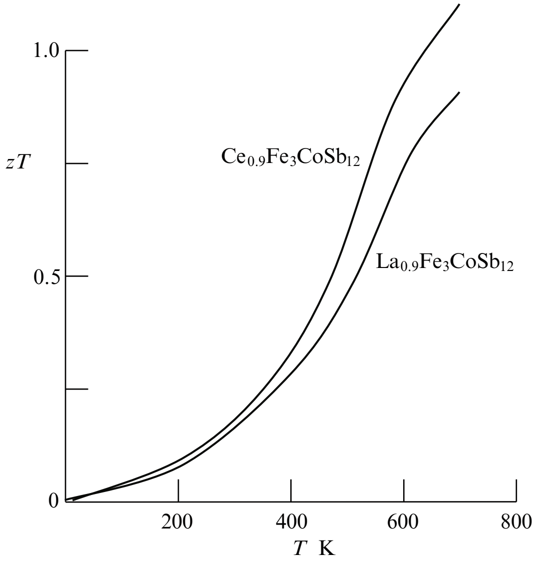


Fig. 9.17 Schematic plots of zT against temperature for $Ce_{0.9}Fe_3CoSb_{12}$ and $La_{0.9}Fe_3CoSb_{12}$ based on the data of Sales et al. [42]

carrier concentrations that would lead to metallic conduction in other materials. The effective mass for p-type skutterudites is of the same order as the free electron mass, and for n-type compositions is an order of magnitude greater [40]. The guest atoms that are introduced into the voids act as dopants and the high effective masses allow their concentration to be high without the Seebeck coefficient becoming too small.

There are a large number of both host compositions and guest atoms from which to choose. Many seem to have promising thermoelectric properties [43] and provide dimensionless figures of merit of the order of unity or greater in the temperature range 500–700 K.

A curious effect has been reported by He et al. [44]. He and his co-workers prepared the skutterudite $Co_{1-x}Ni_xSb_3$ by adding $Co_{1-x}Ni_xSb_{3.05}$ and found that, after annealing, the material was porous. The presence of the pores, of more than 1 μm in diameter, was found to change the thermoelectric properties. The electrical conductivity became greater at the expense of the Seebeck coefficient and the thermal conductivity was substantially reduced. It was found that the porous material had a value of zT equal to 0.6 at 400 K, about twice the value for the original material. Perhaps, when this effect is properly understood, a general improvement of the thermoelectric properties of the skutterudites will be achieved.

9.4.2 Clathrates

The clathrates are another group of compounds that have open structures into which loosely bound guest atoms can be incorporated. The original clathrates were crystalline complexes of H_2O with trapped atoms or molecules. Ice clathrates were known to have very small thermal conductivities. They typically have a very large number of host atoms in the unit cell. For example, Type I ice clathrates have 46 H_2O molecules in the unit cell and Type II ice clathrates have 136 such molecules. These clathrates are, of course, electrical insulators but there also clathrates with semiconducting properties and it is these in which we are interested.

Type I conducting clathrates have the general formula $X_2Y_6E_{46}$ where X and Y are guest atoms on two different sites and E is Si, Ge, or Sn. The corresponding formula for the Type II clathrates is $X_8Y_{16}E_{136}$. The elements in Group IV of the periodic table are usually found to have the diamond structure, each atom being covalently bonded to four other atoms. These bonds are retained in the clathrates but there are no longer groups of four atoms forming tetrahedra. Instead, the Group IV atoms form dodecahedra and either tetrakaidecahedra or hexakaidecahedra [45]. In Type I clathrates, the unit cell is made up of six tetrakaidecahedra and two dodecahedra, whereas in Type II clathrates, the cell comprises 16 dodecahedra and eight hexakaidecahedra. It is, therefore, possible for each unit cell in a Type I clathrate to accommodate eight guest atoms and in a Type II clathrate it can accommodate 24 guest atoms.

Most of the experimental work has been performed on Type I clathrates. Figure 9.18 shows how the lattice conductivity of a number of these materials varies with temperature. Also shown is the lattice conductivity of amorphous germanium. It is clear that the lattice conductivity of the clathrates approaches the value for amorphous germanium at the higher temperatures and in some cases reaches it.

The lattice conductivity of all the clathrates at room temperature is much lower than for silicon or germanium in the usual crystalline state. At lower temperatures, it becomes even lower for most samples suggesting that the guest atoms are particularly effective in scattering the low-frequency phonons. The exception is the Sn

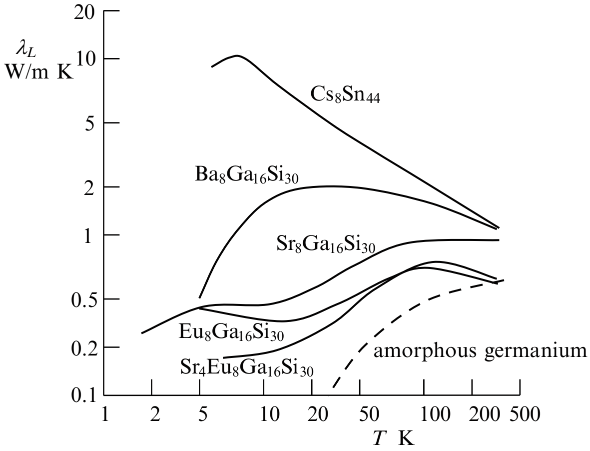


Fig. 9.18 Lattice conductivity plotted against temperature for selected Type I clathrates. A schematic plot based on the data of Cohn et al. [46]

clathrate and it is thought by Cohn et al. [46] that this is because the Cs ion has a large radius and only just fits into the Sn cage. Nevertheless, at higher temperatures, this clathrate, too, has a low thermal conductivity. Even when the phonon free path length is not reduced, the very large size of the unit cell is instrumental in reducing the heat conduction by the lattice.

We have already discussed atomic displacement parameters for the skutterudites. The concept has the same significance for the clathrates. Thus, in $Sr_8Ga_{16}Ge_{30}$, the Sr ions in one type of cage site have a very large atomic displacement parameter [47] and this can be associated with the small lattice conductivity at low temperatures.

Of course, the lattice conductivity is only one factor when a potential thermoelectric material is being selected. The electronic properties are equally important.

The thermoelectric properties of various clathrates have been studied by several workers. Kuznetsov et al. [48] have made observations in the temperature range 100–870 K on $Ba_8Ga_{16}Si_{30}$, $Ba_8Ga_{16}Ge_{30}$, $Ba_8Ga_{16}Sn_{30}$, and $Sr_8Ga_{16}Ge_{30}$. These compounds were all found to have negative Seebeck coefficients as shown in Fig. 9.19. The electrical resistivity of the same samples is shown in Fig. 9.20.

Kuznetsov et al. did not measure the thermal conductivity but they estimated its value from the data for $Ba_8Ga_{16}Ge_{30}$ and $Sr_8Ga_{16}Ge_{30}$ that were available for the temperature range 10–300 K [46, 49]. They assumed that the lattice conductivity would remain constant above 300 K, which seems to be reasonable from the data shown in Fig. 9.18. Using calculated values for the electronic component of the thermal conductivity, they estimated that zT would be equal to 0.7 for $Ba_8Ga_{16}Ge_{30}$ at 700 K and 0.87 for $Ba_8Ga_{16}Si_{30}$ at 870 K. Inspection of Fig. 9.19 suggests that neither of these compositions has the optimum carrier concentration. It appears that the sample of $Ba_8Ga_{16}Ge_{30}$ is entering the mixed conduction region at 700 K and would be improved at this temperature by the addition of donor impurities. On the other hand, the Seebeck coefficient of the specimen of $Ba_8Ga_{16}Si_{30}$ seems to lie below its optimum value at 870 K and a reduction in the donor concentration should improve

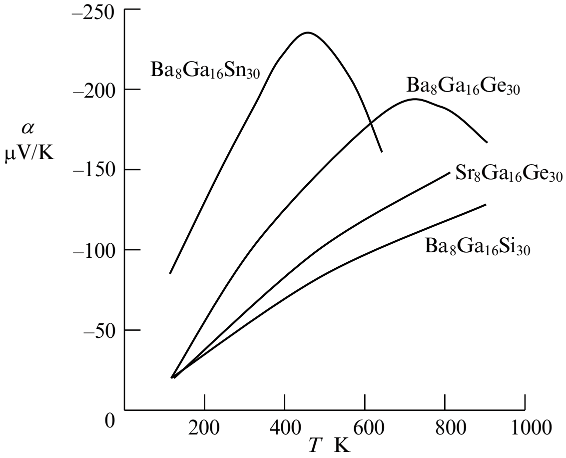


Fig. 9.19 Seebeck coefficient plotted against temperature for some Type I clathrates. Schematic representation of the data of Kuznetsov et al. [48]

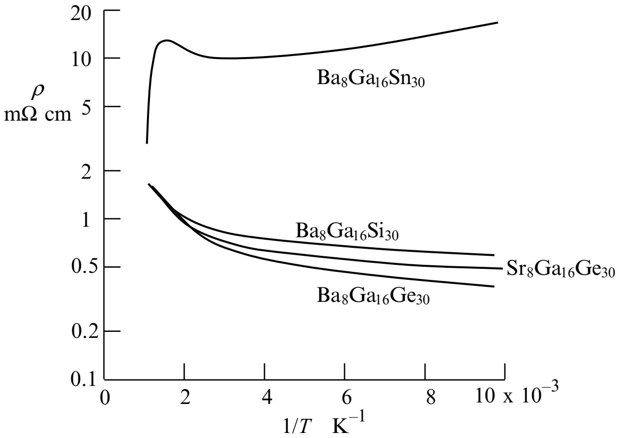


Fig. 9.20 Electrical resistivity plotted against inverse temperature for some Type I clathrates. Schematic representation of the data of Kuznetsov et al. [48]

its properties. A figure of merit zT as high as 1.35 at 900 K has been reported [50] for single crystal $Ba_8Ga_{16}Ge_{30}$. Martin et al. [51] have found that the substitution of a certain proportion of 20% Si for Ge in $Ba_8Ga_{16}Ge_{30}$ can enhance the performance. There is no doubt then that the clathrates can provide worthwhile n-type thermoelements.

It is possible to produce p-type material using Al as a doping agent. Thus, Deng et al. [52] have prepared p-type samples with the formula $Ba_8Ga_{16}Al_xGe_{30-x}$, where $x = 1, 2, 3, 4$ or 5. The atomic displacement parameter is not affected by the aluminium substitution so the thermal conductivity should remain low at all temperatures. In fact, mass fluctuation scattering lowers the lattice conductivity still further. The reasonably large value of 0.61 for zT at 760 K has been observed for $Ba_8Ga_{16}Al_3Ge_{27}$.

9.5 Oxides

Oxides are attractive for high-temperature applications as they are potentially stable and chemically inert. However, they must still have high values of the figure of merit if they are to be of any use. Scientists were unaware of the possibility of using them as thermoelements until the observation of a reasonably large figure of merit for $NaCo_2O_4$. Yakabe et al. [53] prepared material by sintering, using both hot and cold pressing techniques. The figure of merit of the hot-pressed material remained above 0.5×10^{-3} K^{-1} over the temperature range 100–400°C. The Seebeck coefficient varied between 100 and 140 μV K^{-1} and was clearly less than the optimum value. It was shown that higher Seebeck coefficients could be reached using doping agents including Ba, Cu, and Mn. The largest Seebeck coefficient reported in this early work was 180 μV K^{-1} at 400°C for Na $(Co_{0.95}Mn_{0.05})_2\,O_4$ though the highest figure of merit of more than 0.8×10^{-3} K^{-1} was observed for Na $(Co_{0.95}Cu_{0.05})_2\,O_4$ because this formulation had the exceedingly small thermal conductivity of about 1.0 W m^{-1} K^{-1} over the same temperature range. In these measurements, the highest value of zT was about 0.54 at 673 K, certainly a most promising starting point for further investigation of oxide systems.

$NaCo_2O_4$ with improved thermoelectric properties was reported by Ohtaki et al. [54]. An essential feature of the preparation of this improved material seems to be a double-sintering procedure. This led to an increase in both the Seebeck coefficient and the electrical conductivity with a doubling of the power factor, admittedly from a rather low starting level. The value of zT for this p-type conductor reached 0.78 at 1,053 K. Similar values are found for another p-type oxide [55], $Ca_3Co_4O_9$ so it is clear that $NaCo_2O_4$ is not unique among the oxides in possessing good thermoelectric properties.

Perhaps, the best of the n-type oxides [56] is $SrTiO_3$. Although the mobility is low, the effective mass is high and the power factor at room temperature compares well with that of material like Bi_2Te_3. However, the thermal conductivity is rather high, at about 8 W m^{-1} K, and this means that zT at room temperature is only about 0.08. Because of the high effective mass and the low mobility, there are good prospects for reducing the lattice conductivity without affecting the electronic properties. Muta et al. [57] reduced the thermal conductivity to 3.4 W m^{-1} K^{-1} at 300 K by partially substituting Dy for Sr. A value for zT of 0.37 at 1,000 K has been achieved for $SrTiO_3$ with a high concentration of Nb [58]. One can certainly do better than this with other thermoelectric materials so further progress needs to be made before it can be claimed that there is a satisfactory n-type oxide at 1,000 K. However, one should remember that $SrTiO_2$ has the high melting temperature of 2, 080°C so it may be possible to use this material under conditions that are unsuitable for other thermoelements.

9.6 Other Thermoelectric Materials

9.6.1 Zinc Antimonide

Zinc antimonide, Zn_4Sb_3, exists in three crystalline forms. One of these, the β-phase, which is stable between 263 and 765 K, has very good thermoelectric properties in a particular temperature range. The key to its high figure of merit is a small lattice conductivity associated with a disordered crystal structure [59].

It is a p-type conductor with a figure of merit zT that rises from about 0.6 at 200°C to about 1.3 at 400°C. Over this range, it is probably as good as any other material that is available. Caillat et al. [60], who reported the thermoelectric properties in 1997, mentioned that it is difficult to grow large crack-free crystals because of the phase transitions. However, they were able to produce good material using a sintering process. Their typical samples displayed a Seebeck coefficient of nearly $200\,\mu\mathrm{V\,K}^{-1}$ at the upper end of the temperature range, but at 200°C, it was no more than $150\,\mu\mathrm{V\,K}^{-1}$, which is almost certainly below its optimum value. Pedersen et al. [59] studied the effect of Mg as a dopant but were unable to report any improvement in the figure of merit. It is possible that there will be no substantial improvement on the results obtained by Caillat and his colleagues for pure Zn_4Sb_3 though Pedersen and co-authors mentioned that later work by Caillat et al. showed some benefit from the substantial Cd content in the composition $Cd_{0.8}Zn_{3.2}Sb_3$. The lattice conductivity already lies below $1\,\mathrm{W\,m^{-1}\,K^{-1}}$ so the reduction in this quantity by alloy scattering of the phonons is never likely to be very great.

9.6.2 Half-Heusler Compounds

The basic Heusler alloy Cu_2MnAl is a ferromagnetic material with a structure in which the copper atoms form a primitive cubic lattice with alternative cells containing Mn and Al atoms. The half-Heusler structure is the same except that half the atoms on the copper sites are missing. In the prototype half-Heusler compound, AgAsMg, the Mg and Ag atoms form a rock salt structure, and the As and either the Mg or Ag atoms form a zinc blende structure. A group of half-Heusler compounds with the formula MNiSn, where $M = H$, Zr, or Ti, is known to have good n-type thermoelectric properties even though the lattice conductivity is rather high.

A typical half-Heusler compound, ZrNiSn, has a lattice conductivity that is equal to about $10\,\mathrm{W\,m^{-1}\,K^{-1}}$, but this can be reduced by forming a solid solution such as $Zr_{0.5}Hf_{0.5}NiSn$ [60]. The prospects for improvement are good since the effective mass is of the order of $5m$ and the mobility is correspondingly small and not likely to be reduced much when attempts are made to reduce the lattice conductivity still further.

If it is assumed that most of the heat conduction is due to the acoustic-mode phonons because of their large group velocity, it turns out that the mean free path of the phonons exceeds that of the electrons. Thus, it is likely that ZrNiSn in the amorphous state would be superior to crystalline material [62].

Many of the half-Heusler compounds, such as ZrNiSn, HfNiSn, and TiNiSn, normally display n-type conduction but others including HfPtSn and ZrPtSn are p-type conductors [63]. However, the latter have rather high values for the electrical resistivity with zT less than 0.03 at all temperatures. The effective mass is smaller for holes than for electrons and the hole mobility is not particularly high.

Muta et al. [64] have found that the thermal conductivity at high temperatures can be reduced by the addition of impurities but they showed that the effect was due to the elimination of the bipolar electronic conduction rather than a decrease in the lattice conductivity.

Ono et al. [65] were able to improve the power factor of spark-plasma-sintered and arc-melted half-Heusler compounds by substitution on both the Nb and Sn sites in NbCoSn. The highest power factor of $2.2\,\mathrm{mW\,m^{-1}\,K^{-2}}$ was observed for $Nb_{0.99}Ti_{0.01}CoSn_{0.9}Sb_{0.1}$ at 700 K. However, the lattice conductivity was not significantly reduced and zT did not rise above 0.3 at 850 K. This would certainly be a useful material if the lattice conductivity could be made smaller. A similar result was obtained by Sekimoto et al. [66] for $TiCoSn_{0.1}Sb_{0.9}$ at 959 K while zT rose to 0.45 at 958 K for $ZrCoSn_{0.1}Sb_{0.9}$. It does not seem that the formation of such solid solutions is going to reduce the lattice conductivity sufficiently for zT to approach unity. The exceptionally high power factor of $4.1\,\mathrm{mW\,m^{-1}\,K^{-2}}$ at 673 K observed by Matsubara et al. [67] for $(Ti_{0.5}Zr_{0.25}Hf_{0.25})_{0.99}\,Y_{0.01}NiSn_{0.99}Sb_{0.01}$ suggests that this composition may have a high figure of merit but its thermal conductivity is not known. Matsubara et al. observed a chain-like nanostructural feature in their material and it could be that this type of structure is responsible for the promising properties. The half-Heusler compounds, particularly, those with n-type conduction, remain as contenders for supremacy in thermoelectric applications at certain temperatures.

9.6.3 Metal Silicides

Iron disilicide, $FeSi_2$, has long been considered as a suitable thermoelectric generator material even though it does not have the highest figure of merit at any temperature. Its great virtues are its stability and strength, while it is made from two elements that are cheap and plentiful. The compound exists in two phases, the α-phase being produced on solidification of the melt and the β-phase being the more stable form below 955°C. $FeSi_2$ can be produced by a powder metallurgy technique [68] with an optimum annealing temperature for the production of the β-phase of about 800°C. The compound can be obtained in either n-type or p-type form by doping with either cobalt or aluminium, respectively. Rather high dopant concentrations are needed as the effective masses of the carriers are large. The dimensionless figure of merit for a thermocouple in which both legs are made from $FeSi_2$ rises from about 0.03 at 100°C to 0.2 at 700°C. Stable contacts at the hot junction can be made by direct sintering. There is no doubt that $FeSi_2$ has practical advantages over other materials and, in some circumstances, these may outweigh the poor figure of merit.

There are a number of transition metal silicides that share many of the advantages of $FeSi_2$. Their properties have been reviewed by Fedorov and Zaitsev [69]. The highest value of zT seems to be around 0.4, this being exhibited by Rh-doped Ru_2Si_3 at about 900°K and by $Fe_{0.95}Co_{0.05}Si_2$ at 800°K.

Rather better thermoelectric properties are found for alloys based on magnesium silicide, Mg_2Si. The properties of Mg_2Si and the isomorphous compounds Mg_2Ge and Mg_2Sn, together with their solid solutions, have been reviewed by Zaitsev et al. [70]. The solid solutions are, particularly, interesting because of the substantial increase in thermal resistivity compared with that of any of the pure compounds [71].

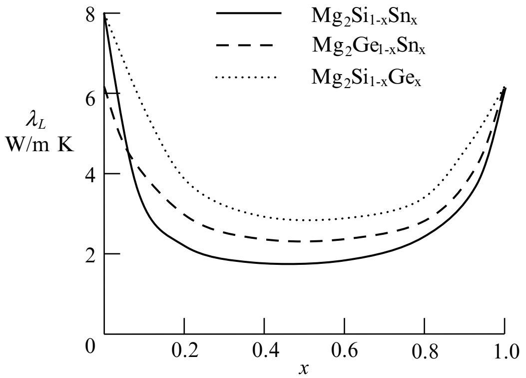


Fig. 9.21 Schematic plot of lattice conductivity at room temperature against concentration of second component in Mg_2 (Si,Ge,Sn) alloys. Based on data of Zaitsev et al. [70]

The variation of lattice conductivity with composition for the three solid solution systems is shown in Fig. 9.21.

All the solid solutions display favourable n-type thermoelectric properties at temperatures above, say, 500 K. The best results were reported for $Mg_2Si_{0.6}Sn_{0.4}$ with zT rising from just over 0.4 at 500 K to about 0.9 at 800 K. Zaitsev et al. believe that further work should be done on this group of compounds and alloys.

9.6.4 Boron Carbide

In this chapter, an attempt has been made to portray the properties of a wide variety of thermoelectric materials. It is remarkable how many different systems yield values of zT that are equal to about unity or greater over some temperature range. No mention has been made, for example, of organic semiconductors and there are other omissions of systems that might eventually give good thermoelectric properties, However, it is hoped that the readers of this book will be able to evaluate new materials as they appear, basing their judgement on the principles that have been outlined.

We shall, finally, mention the thermoelectric properties of boron carbide, a refractory material of exceptional hardness [72]. The thermoelectric properties of this material have been measured by Bouchacourt and Thevenot. The value of zT observed by these authors was 1.06 at 1,250 K so boron carbide is clearly a potential thermoelectric material for use at high temperatures.

Boron carbide is normally a p-type conductor, and in this context, Mori and Nishimura [74] have drawn attention to rare earth boron cluster compounds that have similar refractory characteristics but are n-type. At present, these materials do not have high figures of merit but this is not surprising in view of the fact that the Seebeck coefficient lies far below its optimum value. Mori and Nishimura consider that these boron cluster compounds are worth further study.

References

1. A.L. Jain, Phys. Rev. **114**, 1518 (1959)
2. A.L. Jain, S.H. Koenig, Phys. Rev. **127**, 442 (1962)
3. B. Abeles, S. Meiboom, Phys. Rev. **101**, 544 (1956)
4. S.J. Freedman, H.J. Juretschke, Phys. Rev. **124**, 1379 (1961)
5. B. Lenoir, A. Dauscher, M. Cassart, Y.I. Ravich, H. Scherrer, J. Phys. Chem. Solids **59**, 129 (1998)
6. C.G. Gallo, B.S. Chandrasekhar, P.H. Sutter, J. Appl. Phys. **34**, 144 (1963)
7. Y.A. Korenblit, M.E. Kuznetsov, V.M. Muzhdaba, S.S. Shalyt, Sov. Phys. J. Exp. Theor. Phys. **30**, 1009 (1969)
8. C. Uher, H.J. Goldsmid, Phys. Stat. Solidi A **65**, 765 (1974)
9. G.E. Smith, R. Wolfe, S.E. Haszko, in *Proceeedings of the International Conference on Physics of Semiconductors, Paris* (Dunod, Paris, 1964), p. 399
10. G.E. Smith, R. Wolfe, *Proceeedings of the International Conference on Physics of Semiconductors, Kyoto*, J. Phys. Soc. Japan **21**, 651 (1966)
11. I.M. Tsidil'kovskii, *Thermomagnetic Effects in Semiconductors* (Infosearch, London, 1962)
12. T.S. Gudkin, E.K. Iodanishvili, E.E. Fiskind, Sov. Phys. Phys. Lett. **4**, 844 (1978)
13. J.J. Becker, J. Appl. Phys. **41**, 1055 (1970)
14. W.M. Yim, A. Amith, Solid State Electron. **15**, 1141 (1972)
15. G.E. Smith, Phys. Rev. Lett. **9**, 487 (1962)
16. N.B. Brandt, L.G. Lyubutina, N.A. Kryukova, Sov. Phys.-J. Exp. Theor. Phys. **26**, 93 (1968)
17. R. Wolfe, G.E. Smith, Appl. Phys. Lett. **1**, 5 (1962)
18. P. Jandl, U. Birkholz, J. Appl. Phys. **76**, 7351 (1994)
19. M.E. Ertl, G.R. Pfister, H.J. Goldsmid, Br. J. Appl. Phys. **14**, 161 (1963)
20. C.B. Thomas, H.J. Goldsmid, Phys. Lett. A **27**, 369 (1968)
21. R.B. Horst, L.R. Williams, in *Proceedings of the Third International Conference on Thermoelectrics, Arlington, Texas* (IEEE, New York, 1980), p. 183
22. A.F. Ioffe, A.V. Airapetyants, A.V. Ioffe, N.V. Kolomoets, L.S. Stil'bans, Dokl. Akad. Nauk SSSR **106**, 981 (1956)
23. V. Fano, in *CRC Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC Press, Boca Raton, 1994), p. 257
24. R.S. Allgaier, W.W. Scanlon, Phys. Rev. **111**, 1029 (1958)
25. D.M. Rowe, C.M. Bhandari, *Modern Thermoelectrics* (Holt, Rinehart and Winston, London, 1983), p. 69
26. S.V. Airapetyants, B.A. Efimova, T.S. Stavitskaya, L.S. Stil'bans, L.M. Sysoeva, Zh. Tekh, Fiz **27**, 2167 (1957)
27. F.D. Rosi, E.F. Hocking, N.E. Lindenblad, RCA Rev. **22**, 82 (1961)
28. E.A. Skrabek, D.S. Trimmer, in *CRC Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC Press, Boca Raton, 1994), p. 267
29. J.P. Heremans, V. Jovovic, E.S. Toberer, A. Saramat, K. Kurosaki, A. Charoenphakdee, S. Yamanaka, G.J. Snyder, Science **321**, 554 (2008)
30. D.T. Morelli, V. Jovovic, J.P. Heremans, Phys. Rev. Lett. **101**, 035901 (2008)
31. C. Wood, Rep. Prog. Phys. **51**, 459 (1988)
32. F.D. Rosi, E.F. Hockings, N.E. Lindenblad, RCA Rev. **22**, 82 (1961)
33. K.F. Hsu, S. Loo, F. Guo, W. Chen, J.S. Dyck, C. Uher, T. Hogan, E.K. Polychroniadis, M.G. Kanatzidis, Science **303**, 818 (2004)
34. J.A. Carruthers, T.H. Geballe, H.M. Rosenberg, J.M. Ziman, Proc. Roy. Soc. A **238**, 502 (1957)
35. M.C. Steele, F.D. Rosi, J. Appl. Phys. **29**, 1517 (1958)
36. C.B. Vining, in *CRC Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC Press, Boca Raton, 1994), p. 329
37. C.B. Vining, W. Laskow, J.O. Hanson, R.R. Van der Beck, P.D. Gorsuch, J. Appl. Phys. **69**, 4333 (1991)
38. G.A. Slack, M.A. Hussain, J. Appl. Phys. **70**, 2694 (1991)

39. D.M. Rowe, C.M. Bhandari, App. Energy **6**, 347 (1980)
40. G.S. Nolas, J. Sharp, H.J. *Goldsmid, Thermoelectrics: Basic Principles and New Materials Developments* (Springer, Berlin, 2001), p. 177
41. D.G. Cahill, S.K. Watson, R.O. Pohl, Phys. Rev. B **46**, 6131 (1992)
42. B.C. Sales, D. Mandrus, B.C. Chakoumakos, V. Keppens, J.R. Thompson, Phys. Rev. B **56**, 15081 (1997)
43. C. Uher, in *Recent Trends in Thermoelectric Materials Research I, Semiconductors and Semimetals*, ed. by T.M. Tritt (Academic Press, San Diego, 2001), p. 139
44. Q. He, S. Hu, X. Tang, Y. Lan, J. Yang, X. Wang, Z. Ren, Q. Hao, G. Chen, Appl. Phys. Lett. **93**, 042108 (2008)
45. G.S. Nolas, G.A. Slack, S.B. Schujman, in *Recent Trends in Thermoelectric Materials Research I, Semiconductors and Semimetals*, ed. by T.M. Tritt (Academic Press, San Diego, 2001), p. 255
46. J.L. Cohn, G.S. Nolas, V. Fessatidis, T.H. Metcalf, G.A. Slack, Phys. Rev. Lett. **82**, 779 (1999)
47. B.C. Chakoumakos, B.C. Sales, D.G. Mandrus, G.S. Nolas, J. Alloys Comp. **296**, 801 (1999)
48. V.L. Kuznetsov, L.A. Kuznetsova, A.E. Kaliazin, D.M. Rowe, J. Appl. Phys. **87**, 7871 (2000)
49. G.S. Nolas, J.L. Cohn, G.A. Slack, S.B. Schujman, Appl. Phys. Lett. **73**, 178 (1998)
50. A. Saramat, G. Svensson, A.E.C. Palmqvist, C. Stiewe, E. Mueller, D. Platzek, S.G.K. Williams, D.M. Rowe, J.D. Bryan, G.D. Stucky, J. Appl. Phys. **99**, 023708 (2006)
51. J. Martin, G.S. Nolas, H. Wang, J. Yang, J. Appl. Phys. **102**, 103719 (2007)
52. S.K. Deng, X.F. Tang, P. Li, Q.J. Zhang, J. Appl. Phys. **103**, 73503 (2008)
53. H. Yakabe, K. Kikuchi, I. Terasaki, Y. Sasago, K. Uchinokura, in *Proceedings of the Sixteenth International Conference on Thermoelectrics, Dresden* (IEEE, New York, 1997), p. 523
54. M. Ohtaki, Y. Nojiri, E. Maeda, in P*roceedings of the Nineteenth International Conference on Thermoelectrics, Cardiff* (Babrow Press, Wales, 1997), p. 190
55. H. Ohta, K. Sugiura, K. Koumoto, Inorganic Chem. **47**, 8429 (2008)
56. K.K. Koumoto, I. Terasaki, T. Kajitani, M. Ohtaki, R. Funahashi, in *Thermoelectrics Handbook: Macro to Nano*, ed. by D.M. Rowe (CRC Taylor and Francis, Boca Raton, 2006), p. 35–1
57. H. Muta, K. Kurosaki, S. Yamanaka, J. Alloys Comp. **350**, 292 (2003)
58. S. Ohta, T. Nomura, H. Ohta, M. Hirano, H. Hosono, K. Koumoto, Appl. Phys. Lett. **87**, 092108 (2005)
59. B.L. Pedersen, H. Birkendal, M. Nygren, P.T. Frederiksen, B.B. Iverse, in *Proceedings of the Twenty Sixth International Conference on Thermoelectrics, Jeju, Korea* (IEEE, New York, 2007), p. 417
60. T. Caillat, J.P. Fleurial, A. Borshchevsky, J. Phys. Chem. Solids **58**, 1119 (1997)
61. C. Uher, J. Yang, S. Hu, D.T. Morelli, G.P. Meisner, Phys. Rev. B **59**, 8615 (1999)
62. G.S. Nolas, H.J. Goldsmid, Phys. Stat. Solidi A **194**, 271 (2002)
63. Y. Kimura, A. Zama, Y. Mishima, in *Proceedings of the Twenty Fifth International Conference on Thermoelectrics, Vienna* (IEEE, New York, 2006), p. 115
64. H. Muta, T. Kanemitsu, K. Kurosaki, S. Yamanaka, in *Proceedings of the Twenty Fifth International Conference on Thermoelectrics, Vienna* (IEEE, New York, 2006), p. 120
65. Y. Ono, S. Inayama, H. Adachi, T. Kajitani, in *Proceedings of the Twenty Fifth International Conference on Thermoelectrics, Vienna* (IEEE, New York, 2006), p. 124
66. T. Sekimoto, K. Kurosaki, H. Muta, S. Yamanaka, in *Proceedings of the Twenty Fifth International Conference on Thermoelectrics, Vienna* (IEEE, New York, 2006), p. 128
67. M. Matsubara, R. Asahi, T. Nakagaki, D. Isheim, D.N. Seidman, in *Proceedings of the Twenty Sixth International Conference on Thermoelectrics, Jeju, Korea* (IEEE, New York, 2007), p. 280
68. U. Birkholz, E. Gross, U. Stöhrer, in *CRC Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC Press, Boca Raton, 1994), p. 287
69. M.I. Fedorov, V.K. Zaitsev, in *Thermoelectrics Handbook: Macro to Nano*, ed. by D.M. Rowe (CRC Taylor and Francis, Boca Raton, 2006), p. 31–1
70. V.K. Zaitsev, M.I. Fedorov, I.S. Eremin, E.A. Gurieva, in *Thermoelectrics Handbook: Macro to Nano*, ed. by D.M. Rowe (CRC Taylor and Francis, Boca Raton, 2006), p. 29–1

71. V.K. Zaitsev, E.N. Tkalenko, E.N. Nikitin, Sov. Phys. Solid State **11**, 221 (1969)
72. T.L. Aselage, D. Emin, in *CRC Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC Press, Boca Raton, 1994), p. 373
73. M. Bouchacourt, F. Thevenot, J. Mater. Sci. **20**, 1237 (1985)
74. T. Mori, T. Nishimura, in *Proceedings of the Twenty Fifth International Conference on Thermoelectrics, Vienna* (IEEE, New York, 2006), p. 164

Chapter 10
Thermoelectric Modules and Their Application

10.1 The Modular Concept

In principle, a single thermocouple can be adapted to provide any required cooling capacity as a refrigerator by altering the ratio of length to cross-sectional area. However, such a couple would operate from a very small voltage and the current would be very large except for minimal cooling power. For most practical purposes, a number of thermocouples are connected in series electrically, while operating thermally, in parallel. Such an arrangement of thermocouples is known as a thermoelectric module. The essential features of a typical module are shown in Fig. 10.1.

The thermoelements in a module are linked together by strips of a good electrical conductor, such as copper. The connecting links are usually held in good thermal contact with electrically insulating plates that should be made from material that conducts heat well. Very often, alumina plates are used but these are not particularly good thermal conductors and beryllia, for example, is a better material for this purpose. Sometimes, the ceramic plates are metallised to assist in the connection to a heat source and a sink.

The number, N, of thermocouples in a module is determined primarily by the required cooling capacity and the maximum electric current. Suppose, for example, that 10 W of cooling is required for a particular application, such a level of cooling power might be adequate for a small portable refrigerator. One could design the module for operation at maximum cooling power or optimum coefficient of performance or, probably, some intermediate condition. It might seem attractive to attempt to obtain the maximum possible cooling from a module, but the lower the coefficient of performance, the greater the amount of heat that has to be removed by the sink. One could very well economise on the cost of the module only to spend more on increasing the capacity of the heat sink. On the other hand, it would be absurd to attempt to operate at maximum coefficient of performance if the temperature difference across the module were very small. This is because the cooling power has then to be very small if (2.8) is to be satisfied.

The coefficient of performance depends on the temperature difference between the thermocouple junctions. This temperature difference will be larger than that between the source and sink since there will always be some thermal resistance

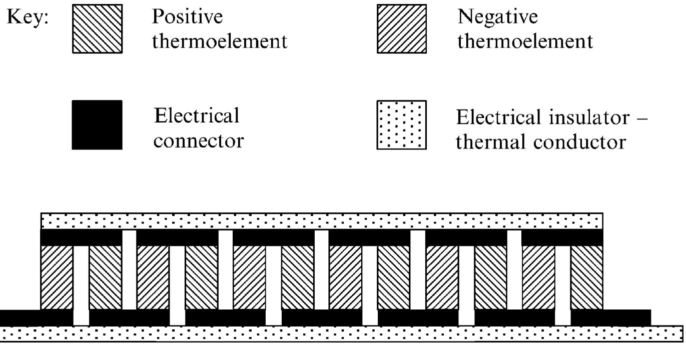


Fig. 10.1 Schematic diagram of a thermoelectric module

other than that of the module. This should be borne in mind when the coefficient of performance is calculated using (2.7) or (2.9).

To proceed further, we must be more specific about the operating temperatures and the properties of the materials in the module. Suppose that the hot side of the module is maintained at 300 K and that a temperature difference between the junctions of 40° must be maintained, we shall also assume that the value of ZT is equal to 0.8 and that the differential Seebeck coefficient is $400\,\mu\mathrm{V\,K^{-1}}$. The value of ZT is not the best that can be achieved but it is typical of a good production module. The optimum COP is then found, from (2.9), to be equal to 0.40. From (2.8), we find that the current through each thermoelement is $46.8 \times 10^{-3} / \left(R_\mathrm{p} + R_\mathrm{n}\right)$ A, where the resistance is expressed in ohms. If we assume that the current is not to exceed, say, 5 A, the resistance $\left(R_\mathrm{p} + R_\mathrm{n}\right)$ must not be less than about $9.4\,\mathrm{m}\Omega$. The electrical power per couple is found using (2.4) and is not greater than 0.31 W. The coefficient of performance of 0.4 then indicates that the cooling power per couple is 0.125 W. Thus, to obtain 10 W of cooling, we need about 80 couples. This would be a convenient number as we could arrange the 160 thermoelements in a 16 × 10 matrix.

We have not as yet determined the dimensions of each thermoelement though we know that its resistance must be about $4.7\,\mathrm{m}\Omega$. The electrical resistivity of a typical thermoelectric material is some $10^{-5}\,\Omega$ m, so we require L/A to be $0.47\,\mathrm{mm}^{-1}$. For reasons that we discuss later in this chapter, it would be difficult to reduce the length of a thermoelement below 1 mm. Then, each would have a cross-sectional area of about $2.1\,\mathrm{mm}^2$. In actual fact, the thermoelements in most commercial modules have rather larger values of L/A than $0.47\,\mathrm{mm}^{-1}$ suggesting that a lower current than 5 A is preferable in typical applications.

The user of thermoelectric refrigeration is faced with a somewhat different problem. It is usually a matter of selecting one or more modules from the range offered by a manufacturer to meet a specific requirement. The procedure in this case has been outlined by Buist [1]. It is supposed that the user is supplied with a set of design curves and a range of module sizes. The range of modules that are available is now much greater than it was when Buist described his selection process, but

Table 10.1 Range of thermoelectric modules used to illustrate Buist's selection process [1]

Model number	N	$L/A\,\mathrm{mm}^{-1}$	NA/L mm
MI 1020	7	2.52	2.78
MI 1021	17	2.52	6.75
MI 1022	31	2.52	12.30
MI 1023	71	2.52	28.17
MI 1050	35	0.921	38.00
MI 1060	7	0.727	9.63
MI 1061	17	0.727	23.38
MI 1062	31	0.727	42.64
MI 1063	71	0.727	97.66
MI 1092	31	0.523	59.27
MI 1120	31	0.414	74.88
MI 1142	31	0.327	94.80

Data on current thermoelectric modules are available from www.marlow.com

to illustrate the steps that must be taken, we shall restrict ourselves to the standard modules listed by Marlow Industries in the 1970s. The data for these modules are listed in Table 10.1.

One needs a set of performance curves that relate to the series of modules under consideration. Such curves are based on the assumption that all modules make use of the same thermoelectric materials. Figure 10.2 shows how four parameters that are independent of the size or number of thermoelements vary with the hot junction temperature. These curves can be updated as new thermoelectric materials are introduced. The independent parameters are $I_q L/A$, $q_{\max} L/NA$, V_q/N, and $\Delta T_{\max}$ and relate to the maximum temperature difference and cooling power and the corresponding current and voltage.

Figure 10.3, perhaps, needs some explanation. It shows plots of $\Delta T/\Delta T_{\max}$ against the ratio $q_1/q_{\max}$ of the ratio of the cooling power to its maximum value for various ratios of the current to the current for maximum cooling power. The intersection of each of the lines with the broken curve corresponds to operation at the optimum coefficient of performance. It is obvious that the cooling power will be zero when $\Delta T = \Delta T_{\max}$ and, as mentioned previously, the optimum COP also requires a very small cooling power when $\Delta T \ll \Delta T_{\max}$.

Figures 10.2 and 10.3 are used in the following way. The hot junction temperature T_2 is first selected and this determines the four parameters in Fig. 10.2. One then turns to Fig. 10.3 to select a suitable value for I/I_q so that the module operates in the region to the right of the broken curve; that is, between the condition for optimum COP and that for maximum cooling power. From the specified cooling power, one can then determine an acceptable value for NA/L.

Buist used a realistic example to illustrate the procedure. He supposed that heat has to be removed at the rate of 10 W from a heat source at 290 K and transferred into a heat sink at 350 K. Strictly speaking, these are the temperatures of the cold and hot junctions, respectively.

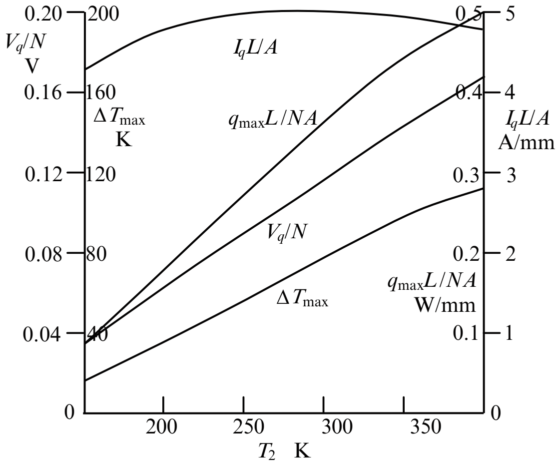


Fig. 10.2 Maximum design parameters plotted against temperature of hot junctions for modules [1] in Table 10.1

Fig. 10.3 Plots of $\Delta T/\Delta T_{max}$ against ratio of cooling power to maximum cooling power for modules in Table 10.1. Each *solid line* represents a different ratio of the current to the current for maximum cooling power. *The broken curve* corresponds to the conditions for optimum coefficient of performance

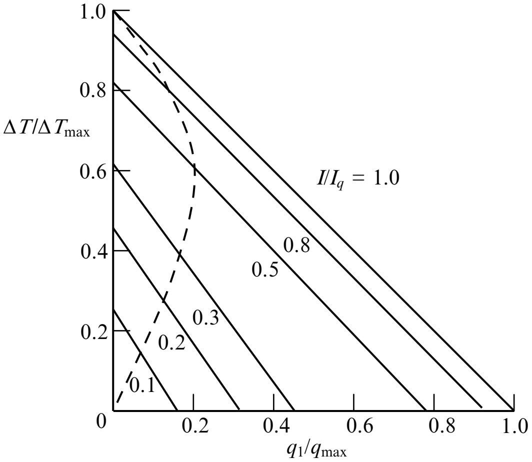


Figure 10.2 shows that $\Delta T/\Delta T_{max}$ is equal to 0.63 and that $q_{max}L/NA$ is 0.45 W/mm^2.

Then, from Fig. 10.3, one finds that q_1/q_{max} should lie between 0.18 for optimum COP and 0.36 for maximum cooling power. Since q_1 is to be equal to 10 W, q_{max} should be between 55.6 and 27.8 W. This means that NA/L must be in the range 123 mm and 61.7 mm. By referring to Table 10.1, we see that a number of modules might meet this requirement. Thus, MI 1063 and MI 1142 both have values of NA/L that are reasonably close to the value of 123 mm for optimum COP and MI 1120 has a value that is closer to that for maximum cooling power. Buist's conclusion was that MI 1063 would be the best choice since it is advantageous to operate closer to the optimum COP condition than that for maximum cooling power and it is usually preferable to use the smallest possible current. Having selected the appropriate module, one then finds from Fig. 10.2 that the operating current and voltage are 4.18 A and 6.53 V, respectively.

Marlow et al. [2] have pointed out that it is often necessary to work as close as possible to the condition of optimum COP. This is the case, for example, if the supply source is a storage battery rather than the electrical mains. Thus, in the example that we have considered, the use of two modules of the type MI 1092 might have been a better option. This would have brought the overall value of NA/L to 118.5 mm, which is very close to the value of 123 mm that we obtained when the COP was optimised.

10.2 Heat Transfer Problems

One of the aims of the manufacturers of thermoelectric modules is the reduction in size of the thermoelements. The cost of the raw materials is a significant factor and there are other advantages that would accrue if modules could be made smaller and lighter. However, when we reduce the size of a module, we encounter problems associated with the transfer of heat. The smaller the cross-sectional area of the end plates, the more difficult it is to transfer heat from the source and to the sink without excessive temperature differences. It is certainly possible to alleviate this problem by increasing the space between the thermoelements but this makes greater the heat losses by convection, conduction, and radiation. Thus, there is an optimum spacing giving the best balance between excessive thermal resistance at the end plates and unwanted heat losses around the thermoelements. We analyse this problem [3] using the model shown in Fig. 10.4.

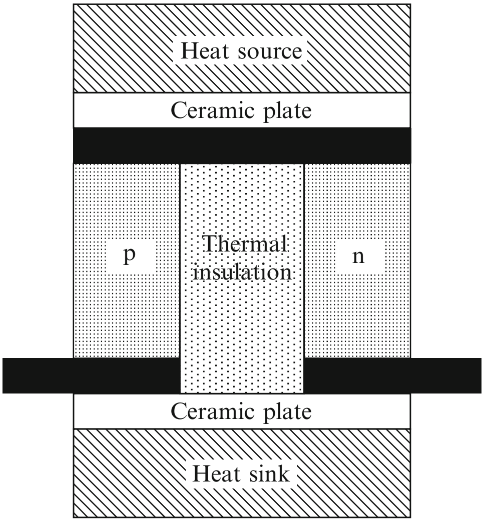


Fig. 10.4 Model for the calculation of heat losses and thermal resistance

It is supposed that there is thermal insulation between the thermoelements that occupies a fraction g of the area occupied by the thermoelements. This insulation has a thermal conductivity λ_I and it is assumed that there are no radiation or convection losses. The conduction of heat through the insulating material increases the thermal conductance by a factor $(1 + \lambda_I g/\lambda)$, where λ is the average thermal conductivity of the thermoelectric material. The effective figure of merit then becomes $Z/(1 + \lambda_I g/\lambda)$.

By inserting the insulation between the thermocouples the total cross-sectional area is increased from A to $A(1 + g)$. Thus, if K_c is the thermal conductance per unit area of the end plates, the thermal conductance of each plate is $K_c A(1 + g)$. When heat is being extracted from the source at a rate q_1, it is delivered to the sink at the rate $q_1(1 + 1/\phi)$, where ϕ is the coefficient of performance. The unwanted temperature differences across the end plates at the source and sink are equal to $q_1/[K_c A(1 + g)]$ and $q_1(1 + 1/\phi)/[K_c A(1 + g)]$, respectively. Thus if ΔT^* is the required temperature difference between the sink and the source, the temperature difference across the thermocouples has to be

$$\Delta T = \Delta T^* + \frac{q_1(2 + 1/\phi)}{K_c A(1 + g)} \tag{10.1}$$

The coefficient of performance is reduced since $\Delta T > \Delta T^*$.

In order to make further progress, we need to specify the condition of operation. For our present purposes, it will be supposed that we want to obtain the maximum temperature difference when the heat drawn from the source is zero. We know that this temperature difference is equal to $ZT_1^2/2$ when there are no heat losses and no thermal resistance at the end plates. According to (2.6), the current that gives the maximum temperature difference is equal to $\left(\alpha_p - \alpha_n\right) T_1/R$ and we find that the power input is $\left(\alpha_{p-}\alpha_n\right)^2 T_1 T_2/R$. If we set $\left(\alpha_{p-}\alpha_n\right)^2/R$ equal to $Z\lambda A/L$, we obtain

$$\Delta T_{\max} = \Delta T^*_{\max} + \frac{\lambda Z T_1 T_2}{K_c L(1 + g)}, \tag{10.2}$$

which shows that the maximum temperature difference $\Delta T^*_{\max}$ between the sink and the source is less than that across the thermocouples.

If we combine (10.1) and (10.2) to include both the heat losses through the insulation and the thermal resistance across the end plates, we obtain

$$\Delta T^*_{\max} = \frac{1}{2} Z T_1^2 \left[\frac{1}{1 + \lambda_I g/\lambda} - \frac{2\lambda T_2}{K_c L(1 + g) T_1}\right]. \tag{10.3}$$

The aim is to make the quantity in the square brackets as close to unity as possible.

Let us apply (10.3) to a specific example. We shall suppose that the module is made of excellent thermoelectric materials with ZT close to unity and T_2/T_1 having a maximum value of about 1.4. For good thermoelectric materials, the thermal

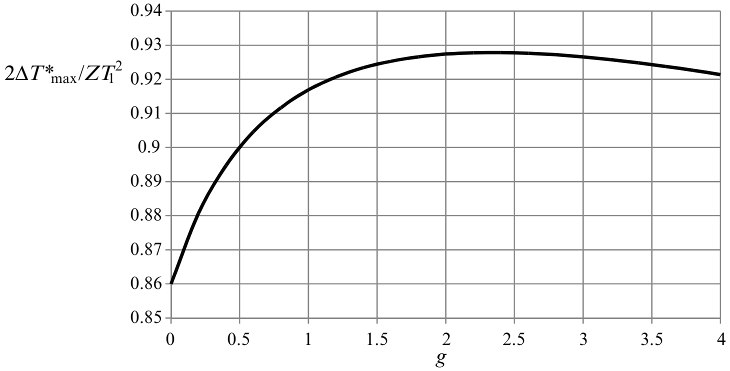


Fig. 10.5 Ratio of maximum temperature difference between sink and source to $ZT_1^2/2$ plotted against the spacing factor g

conductivity λ would be equal to about $1.5\,\mathrm{W\,m^{-1}\,K^{-1}}$ and the spaces between the thermoelements might be filled with expanded polyurethane that has a thermal conductivity of $0.02\,\mathrm{W\,m^{-1}\,K^{-1}}$. The thermal conductivity [4] of alumina is about $30\,\mathrm{W\,m^{-1}\,K^{-1}}$ and for the purpose of this calculation we shall assume a thickness of 1 mm and a thermal conductance of $3 \times 10^4\,\mathrm{W/m^2\,K}$.

Figure 10.5 shows a plot of the term in the square brackets of (10.3) against the spacing factor g. It will be seen that the maximum temperature difference has its highest value when g is about 2. This value is equal to about 93% of $ZT_1^2/2$. It is noted that the thermal resistance of the end plates can be reduced by making them from aluminium nitride [5] instead of alumina while an even better thermal conductor is diamond, though the cost of the latter would usually rule it out.

In fact, in many practical applications, a major factor in the degradation of the performance of a thermoelectric refrigerator lies in the thermal resistance between the surfaces of the end plates and the real source and sink. Even if solid copper conductors are used to spread the heat flux, there can be a substantial temperature difference over their length.

One solution of this problem lies in the use of fluid heat exchangers. Attey [6] has shown that in a typical application using solid-state heat exchangers, the coefficient of performance of a thermoelectric cooling system was found to be 0.4. In this application, the temperature difference between the source and sink was only 20° and, for the particular thermoelectric materials that were used in the module, the theoretical COP was as high as 1.3. The value of *ZT* for the thermocouples was 0.65 but, because of the ineffectiveness of the solid-state heat transfer system, the module was behaving as if *ZT* were no more than 0.28. Using Attey's liquid heat transfer system, the COP rose to 0.8 corresponding to an effective value of 0.43 for *ZT*. Obviously, there is some way to go before the full potential of the thermoelectric cooling modules is realized, but Attey's work highlights the need to improve heat transfer as well as the figure of merit.

10.3 Electrical Contact Resistance

Heat transfer is not the only problem that arises when we attempt to reduce the size of a thermoelectric module. If there is any significant electrical resistance at the contacts, this will become more apparent as the length of the thermoelements is made smaller.

It is actually extremely difficult to measure the electrical contact resistance when present-day methods for attaching metal connectors to thermoelements are used. The only real test is the performance of the thermocouples as coolers or generators when the length is made as small as possible. Semenyuk [5] estimates the electrical resistance of his contacts to be $0.84 \times 10^{-10}\,\Omega\,\mathrm{m}^2$.

The effect of electrical contact resistance on the performance of thermoelectric refrigerators was analysed by Parrott and Penn [7] and we shall give an outline of their treatment here. Although the methods for making electrical contact to thermoelectric materials have improved since this analysis was performed, the thermoelements themselves have become shorter, so the problem is still a real one.

The aim of Parrott and Penn was not to eliminate the effect of contact resistance but rather to take account of it in minimising the amount of thermoelectric material needed for a module. It is supposed that the contact resistance is equal to r_c for unit cross-sectional area. Then, (2.3) becomes

$$q_1 = \left(\alpha_p - \alpha_n\right) I T_1 - (T_2 - T_1)\left(K_p + K_n\right) - I^2\left(R_p + R_n + \frac{r_c}{A}\right)/2. \tag{10.4}$$

Equation (2.4) for the rate of consumption of electrical energy becomes

$$w = \left(\alpha_p - \alpha_n\right) I\,(T_2 - T_1) + I^2\left(R_p + R_n + \frac{r_c}{A}\right). \tag{10.5}$$

The effect is the same as that of increasing the resistance of the two branches from $\left(R_p + R_n\right)$ to $\left(R_p + R_n + r_c/A\right)$. Alternatively, the average electrical resistivity can be regarded as increasing from ρ to $(\rho + r_c/L)$.

Parrott and Penn stated that one can proceed further by assuming either maximum cooling power or optimum coefficient of performance. It is easier to use the first of these assumptions using (10.4) with the current given by (2.6) modified by the inclusion of the contact resistance. Then the maximum cooling power per unit volume is given by

$$\frac{q_{\max}}{AL} = \frac{2\lambda T}{r_c^2\rho^2}\Gamma_q, \tag{10.6}$$

where Γ_q is a function that was defined by Parrott and Penn and is given by

$$\Gamma_q = \frac{r_c^2}{\rho^2 L^2}\left[\frac{ZT}{4}\left(1 + \frac{r_c}{\rho L}\right)^{-1} - \frac{\Delta T}{2T}\right]. \tag{10.7}$$

Parrott and Penn defined a similar but more complex function Γ_ϕ to be used when the coefficient of performance rather than the cooling power is maximised.

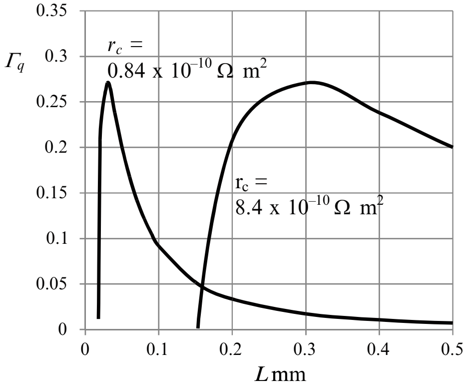


Fig. 10.6 Plot of Parrott and Penn's function Γ_q against length of thermoelements for a contact resistance r_c equal to 0.84×10^{-10} and $8.4 \times 10^{-10} \Omega\,\mathrm{m}^2$. $ZT = 1$ and $\Delta T/T = 1/15$

Figure 10.6 shows the function Γ_q plotted against the length of the thermoelements for contact resistance values of 0.84×10^{-10} and $8.4 \times 10^{-10} \Omega\mathrm{m}^2$, the latter figure being close to the lowest value observed by Plekhotkin et al. [8]. Typical values of ZT and $\Delta T/T$ have been selected. In both cases, there is a maximum at a particular length. Below this maximum, there are two values of L that give a specific ratio of cooling power to volume of thermoelectric material and it is always preferable to choose the greater length since this corresponds to the higher coefficient of performance. It will be seen that Γ_q becomes negative below a certain value of L indicating that the specified temperature difference can no longer be met. In other words, it would not be possible to obtain the rather modest value for ΔT of about 20° if $L = 0.1$ mm and r_c were as high as $8.4 \times 10^{-10} \Omega\mathrm{m}^2$. On the other hand, using Semenyuk's value for r_c a temperature difference of 20° could still be reached with a thermoelement length of about 20 μm.

The expression for the coefficient of performance at maximum cooling power is

$$\phi_q = \frac{ZT - 2\left(1 + \frac{r_c}{\rho L}\right)\frac{\Delta T}{T}}{2ZT\left(1 + \frac{\Delta T}{T}\right)}. \tag{10.8}$$

The coefficient of performance will become appreciably reduced by the effect of contact resistance when $r_c/\rho L$ becomes a significant fraction of unity. For a typical thermoelectric material with ρ of the order of $10^{-5} \Omega\mathrm{m}$ and, assuming r_c equal to Semenyuk's value, $r_c/\rho L$ is of the order of $10^{-2}/L\ \mathrm{mm}^{-1}$. Contact resistance will start to make itself felt as the length of the thermoelements falls below about 0.1 mm.

The miniaturisation of thermoelectric modules has recently been discussed in detail by Semenyuk [9]. He pointed out that there is a need for coolers with a high power density for use with semiconductor lasers and other electronic devices. Thermoelements of up to 50 μm in length can be produced by thin film techniquesm, but the figure of merit is less than can be achieved with bulk materials. Very short samples with excellent thermoelectric properties have been cut from extruded bismuth telluride alloys. Such materials have a figure of merit of

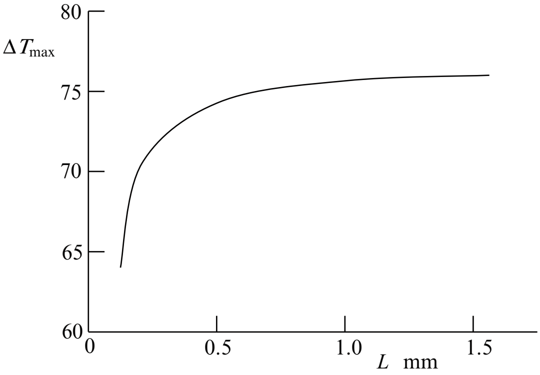


Fig. 10.7 Dependence of maximum temperature depression on length of thermoelements in a module made from extruded bismuth–telluride alloys, according to Semenyuk [9]. The heat sink was at 30°C

over $3 \times 10^{-3}\,\mathrm{K}^{-1}$ with no difference between thermoelements of 200 and 130 μm length. However, because of contact resistance, the maximum temperature difference that can be reached is less for modules made from the shorter samples. Thus, with the heat sink at 30°C, the maximum temperature depression was observed to be 70.6° for thermoelements of 200 μm length and 64.2° when the length was reduced to 130 μm. The variation of $\Delta T_{\max}$ with thermoelement length is shown in Fig. 10.7. It was noted that the ceramic end plates were made from AlN rather than alumina so that the fall in $\Delta T_{\max}$ as the length is reduced is associated with electrical contact resistance. Semenyuk's work confirms that effective thermoelectric coolers can be made with thermoelements of the order of 100 μm in length but some degradation in performance due to contact resistance must be accepted.

10.4 Applications of the Peltier Effect

At the present time, the performance of thermoelectric refrigerators is inferior to that of conventional compressor-type machines. It is likely that the difference between the two types of cooler will become narrower with the passage of time but, until this takes place, Peltier cooling will be restricted to those areas in which it has obvious advantages.

One of the characteristics of a Peltier device is that its performance is almost independent of its capacity. It, therefore, has a definite advantage for the cooling of small enclosures. Manufacturers have considered it to be advantageous to use thermoelectric refrigeration for small portable cold boxes, particularly, when the available power source is a 12 V automobile battery. In this field, there is competition from absorption refrigerators that have the advantage that they can be gas-operated, but when they make use of an electric power source, they are no more efficient than thermoelectric coolers. The thermoelectric devices are also insensitive to movement or inclination and, because of this characteristic, they are attractive for use on board ship.

A feature of the Peltier effect is that it can be used for heating as well as cooling. Thus, in principle, it is possible to maintain foodstuffs in an enclosure that is refrigerated until their time of use whereupon they can be heated to the required temperature for consumption.

Domestic applications of thermoelectric cooling have been reviewed by Banney et al. [10]. These authors state that thermoelectric refrigerators outperform absorption units in mobile applications. They are already effective in dispensers of chilled water. They also claim that even now household refrigerators using the Peltier effect can perform almost as well as compressor units if an efficient heat transfer system, probably based on liquid flow, is incorporated.

One of the largest potential fields of application is in air conditioning and heat pumping. Very often, the required temperature difference between the source and sink is small compared with the maximum that can be reached. Compressor units deal with this situation by being switched on for short periods of time but Peltier devices can be operated continuously at the optimum current. This allows them to adapt readily to changes of the sink temperature whether cooling or heating is needed.

As long ago as 1958, Lindenblad [11] demonstrated the air conditioning of a room by incorporating a thermoelectric cooling unit in one of its walls. Basu [12] claimed that his thermoelectric air conditioner that was powered by a solar battery compared well with a compressor unit from the viewpoints of maintenance, life, cost, and power consumption. Kulagin and Makov [13] used a solar thermoelectric generator to power their Peltier air conditioner claiming the system to be self-regulating, since the cooling power rises with increase of the incoming thermal radiation. It may be noted that Vella et al. [14] used a thermoelectric generator to power a thermoelectric refrigerator claiming that this combination enabled the number of thermocouples to be minimised.

One of the most successful projects involving the use of thermoelectric air conditioning was reported by Stockholm et al. [15]. The objective was the control of the temperature in a train carriage on the French national railways. At different times of the year, both heating and cooling are required though the temperature difference from that of the surrounding air is usually not large. The flexibility of the Peltier effect would seem to make this application particularly suitable.

Stockholm [16] claimed that it is advantageous in large-scale applications to integrate the heat exchangers with the thermoelements. Such integrated heat exchangers minimise the thermal resistance at the source and sink, and were used in the railway application. However, they can present structural problems and the modular approach is preferred for small- to medium-scale applications [17].

The structural problems were certainly overcome by Stockholm et al. The 20 kW air conditioning unit was operated successfully on a particular route for over 10 years without a single failure of the thermoelectric components. However, in spite of this success, thermoelectric air conditioners have not been installed elsewhere in the French railways since there is always some reluctance to replace an existing system with a new one unless the latter has an overwhelming advantage.

Perhaps, the greatest advantage of thermoelectric cooling over other systems lies in the ease with which it can be controlled. This was recognised in the early days with applications such as dew point hygrometers [18] and thermocouple reference enclosures. The maintenance of an enclosure at a constant temperature can present problems, if conventional systems are used. For example, suppose that the required temperature is only a few degrees above that of the ambient air; a simple heater can then raise the temperature of the enclosure to the specified level, but once it passes this level, the loss of heat can be a slow process. However, if the Peltier effect is employed, the ability to heat or cool by small or large amounts by adjusting the current is a considerable asset.

It is in the cooling of small electronic and electro-optic devices that thermoelectric cooling has undoubtedly come into its own [19]. Multi-stage units are used in this type of application when the need is a large temperature depression with a small cooling power. On the other hand, in other situations, high thermal flux densities with more modest temperature differences are often needed with expense being a secondary consideration. Thus, if necessary, the device to be cooled might be mounted on a diamond substrate, which would provide electrical insulation with the minimum of thermal resistance. Semenyuk has presented data that show what can be achieved with different substrate materials. His results are presented in graphical form in Fig. 10.8.

In Fig. 10.8, the maximum temperature difference and maximum cooling power are plotted against the length of the thermoelements. The region for $L < 10\,\mu\text{m}$ is covered by thin film thermoelements and requires diamond substrates. For $10\,\mu\text{m} < L < 60\,\mu\text{m}$, one uses thick-film thermoelements and either diamond or aluminium nitride substrates. Techniques for making thermoelements with L between 60 and $200\,\mu\text{m}$ are in the course of development and, again, diamond or aluminium nitride substrates are needed. The region $200\,\mu\text{m} < L < 300\,\mu\text{m}$ is covered by existing modules with aluminium nitride end plates, whereas for $L > 300\,\mu\text{m}$, alumina end plates can be employed, though aluminium nitride will give a superior performance.

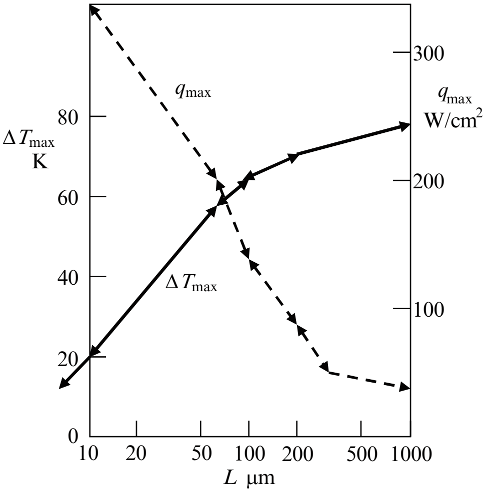


Fig. 10.8 Representation of Semenyuk's data on thermoelectric modules for cooling optoelectronic components

10.5 Transient Cooling

We are interested in the transient behaviour of thermoelectric coolers from two points of view. In some applications, it is useful to know how rapidly the system responds to changes in the load conditions. Also, there always exists the possibility of increasing the temperature difference above its usual maximum value for a short time by applying a large current pulse.

The theory of transient cooling used by Babin and Iordanishvili [20] is based on the assumption that the thermocouple consists of two infinitely long thermoelements that are joined at the position $x = 0$. The thermal load is assumed to be negligible. The distribution of temperature in either branch has to satisfy the equation

$$\frac{\partial^2 T}{\partial x^2} + \frac{i^2 \rho}{\lambda} = \frac{1}{\kappa}\frac{\partial T}{\partial t}, \tag{10.9}$$

where, as before, ρ, λ, and κ are the electrical resistivity, the thermal conductivity, and the thermal diffusivity, respectively. When the time, t, is zero, the temperature is T_2 at all points. When $x = \infty$, $\partial T/\partial x = 0$, while for $x = 0$, $\lambda \partial T/\partial x = \alpha I T$, where it has been assumed that $\pm\alpha$ is the Seebeck coefficient of each branch of the couple. Applying these conditions, (10.9) becomes

$$T_2 - T_1 = T_2\left[\left(1 - \exp\left(A^2\right)\operatorname{erfc}\left(A\right)\right)\left(\frac{zT_2 + 1}{zT_2}\right) - \frac{2}{\pi^{1/2}}\frac{A}{zT_2}\right]. \tag{10.10}$$

In this equation, erfc represents the complementary error function and $A = \alpha\kappa^{1/2} i t^{1/2} \lambda$. The depression of temperature at the junction rises, reaches a maximum, and then falls. The maximum is reached when A satisfies the equation

$$zT_2 = \frac{\pi^{1/2} A \exp\left(A^2\right)\operatorname{erfc}(A)}{1 - \pi^{1/2} A \exp\left(A^2\right)\operatorname{erfc}(A)}. \tag{10.11}$$

In Fig. 10.9, we show a plot of $(T_2 - T_1)/T_2$ against time in arbitrary units, according to (10.10), for $zT_2 = 1$. If the couple were to be operated under steady-state conditions, the maximum value of $(T_2 - T_1)/T_2$ would be equal to 0.265. The reason that the maximum temperature difference in the transient mode is smaller is that there is no heat sink. It should be noted that the maximum temperature difference does not depend on the current; a change of current merely alters the time scale so that $it^{1/2}$ remains constant for a particular point on the curve.

Of course, the assumption of infinite length is unrealistic but the theory holds for samples of finite length up to a certain time. As an empirical rule, the infinite-length approximation is valid up to the maximum temperature difference if the current density is at least twice that which produces maximum cooling in the steady state.

Babin and Iordanishvili showed that one can obtain transient temperature depressions that are significantly greater than the steady-state maximum value. They used a technique in which a steady current giving maximum cooling power is first applied

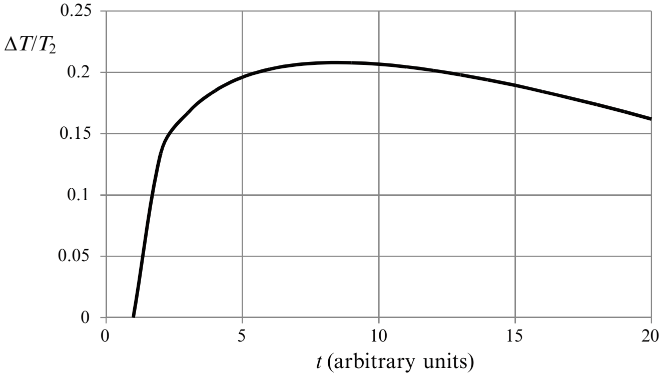


Fig. 10.9 Temperature depression in the transient mode. $\Delta T/T_2$ is plotted against time in arbitrary units. zT_2 is equal to unity

to a thermocouple and then, after equilibrium has been established, the current is increased to some higher value. One can still use (10.9) but the temperature at zero time now has the value

$$T(x) = T_2 - \Delta T_{\max}\left(1 - \frac{x}{L}\right)^2, \tag{10.12}$$

where $\Delta T_{\max}$ is the maximum temperature depression in the steady state. The current density is then increased by the superposition of i^* on i_q, whereupon

$$\begin{aligned}\Delta T = (T_2 - \Delta T_{\max})\Bigg\{&\left(1 - \exp\left(A^2\right)\operatorname{erfc}(A)\right)\left[\gamma + \frac{1}{z(T_2 - \Delta T_{\max})}\right] \\ &- \frac{2}{\pi^{1/2}}\frac{A}{z(T_2 - \Delta T_{\max})}\Bigg\},\end{aligned} \tag{10.13}$$

where

$$\gamma = 1 - \frac{2\Delta T_{\max}}{(T_2 - \Delta T_{\max})\left[(1 + 2zT_2)^{1/2}\right]\left(1 + i^*/i_q\right)}. \tag{10.14}$$

When $i^* >> i_q, \gamma \to 1$ and (10.13) becomes identical with (10.10) with the exception that T_2 is replaced by $(T_2 – \Delta T_{\max})$. This means that the temperature depression portrayed in Fig. 10.9 is superimposed on the maximum temperature difference in the steady state. Babin and Iordanishvili stated that one can increase the maximum temperature difference for a thermocouple with $Z = 2.5 \times 10^{-3}\,\mathrm{K}^{-1}$ from 70 to 105 K by this method.

A more refined calculation has been carried out by Hoysos et al. [21, 22] performed experiments in which short current pulses of large magnitude were superimposed on a steady current using a thermocouple made from bismuth–telluride alloys. The transient behaviour was improved by tapering the branches of the thermocouple so that the cross-sectional area was much smaller near the junction. With the heat sink maintained at 290 K, the minimum temperature of the cold junction in the steady state was found to be 220 K. On the application of current pulses with $i^* \simeq 8i_q$, a cold junction temperature as low as 175 K was obtained. The duration of each pulse was 50 ms and recovery took place in less than 2 s because of the tapering of the thermoelements. Similar experiments on pulsed currents have been performed by Field and Blum [23].

Most of the work on enhanced cooling using pulsed currents has been of an exploratory nature but Yamamoto [24] applied the effect to improve the performance of GaAs electroluminescent diodes and lasers. The same current was passed through the thermocouple and the GaAs diode, which was interposed between the positive and negative thermoelements. The doubling of the emitted radiation from the diode indicated a reduction of the temperature by an additional 50°.

Woodbridge and Ertl [25] have shown that the transient cooling can be enhanced by using shaped pulses. The optimisation of the pulse shape has been discussed by Landecker and Findlay [26]. In principle, one can reduce the temperature to any desired level by continuously increasing the current within a pulse. It is necessary to keep on increasing the Peltier effect to compensate for the Joule heat arriving at the cold junction.

In the experiments performed by Landecker and Findlay, the temperature at the junction was measured after the application of the current pulse. Let us suppose that the pulse is supplied up to time t_1 and the temperature is measured at time t_2. Also, the current within the pulse is proportional to $(t_2 - t)^{-1/2}$. Then, the temperature at the time of observation is given by

$$T = \frac{ZT_2^2}{\pi} \ln\left(\frac{t_2 - t_1}{t_2}\right). \tag{10.15}$$

This equation shows that the temperature can become vanishingly small as the time of observation approaches the time that the pulse is terminated, but this also requires that the current approaches infinity. Nevertheless, it does appear that very low temperatures can be reached using shaped pulses for the current. The experimental results that were obtained by Landecker and Findlay for one particular bismuth–telluride couple are shown in Fig. 10.10.

The results obtained by Landecker and Finlay are supported by the observations on ramp-shaped pulses by Woodbridge and Ertl [25] and by the studies of Idnurm and Landecker [27], who were able to generate shaped pulses in which the current was proportional to $(t_2 - t)^{-1/2}$.

Woodbridge and Ertl [28] also carried out experiments on transient cooling using the transverse thermomagnetic effects. They pointed out that the advantage of this mode of operation is the virtual elimination of any thermal mass at the cold surface, an advantage that should also be found for synthetic transverse thermoelements. They observed a temperature depression of 4 K below 80 K for bismuth in a magnetic field of 0.8 T when they used a pulsed current whereas the maximum temperature difference was only 1.2 K for a steady current.

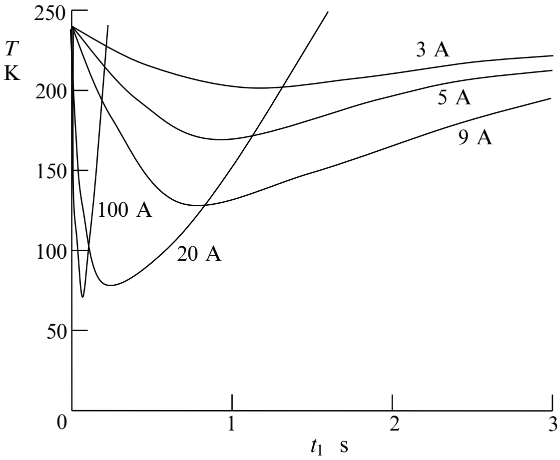


Fig. 10.10 Plot of cold junction temperature against pulse length for different values of the pulse current. The pulses are superimposed on a steady current of 5 A. Data obtained for a bismuth–telluride couple by Landecker and Finlay [26]

From the practical viewpoint, one really needs to discuss transient cooling with substantial thermal loads. The behaviour of a two-stage module during the cooling-down period has been investigated by Hendricks and Buist [29]. A problem arises when one wishes to obtain a rapid response from a conventional module. From this point of view, one would like to reduce the thermal capacity of the copper links but any decrease in thickness can lead to a significant electrical resistance. One means of decreasing the time taken to reach a given temperature is to increase the current above the steady-state optimum value during the cool-down period. Also, in a cascade cooler, it is an advantage to make the different stages closer in cooling capacity than is the case if only the equilibrium condition is being considered. These features were accurately modelled by the calculations of Hendricks and Buist and the work was later extended to 3-stage refrigerators [30]. The transient response of single-stage modules is, of course, a much simpler proposition.

10.6 Seebeck Devices

Although most of the activity in the field of thermoelectric energy conversion over the past 50 years has been concentrated on refrigeration, it is likely that thermoelectric generation will become the more important aspect if the figure of merit can be substantially improved. Unconventional methods for the generation of electricity will be sought as new sources of heat become available. For example, it may be necessary to make use of low-grade heat, with unusually small differences of temperature between the source and sink. Of course, there will always remain those fields of application of the Seebeck effect that are already attractive in which reliability is of more importance than efficiency.

Radio-isotope-powered thermoelectric generators have been used for many years in space vehicles and, more recently, fission power has been utilised. The advantages of thermoelectric generators over other systems include not only reliability but also robustness, long life, capacity for uninterrupted operation, and insensitivity to radiation and degradation due to the environment [31]. Unlike many other types of generator, thermoelectric devices are almost independent of scale and, having no moving parts, are free from noise and vibration. Their linear current–voltage characteristic makes them easy to control.

The first radio-isotope-powered generators used lead telluride thermoelements. This compound and its alloys were deemed to be the best materials for operation with heat sources at temperatures that were moderate but still too high for bismuth telluride and its alloys; an efficiency of just over 5% could be achieved [32]. To improve the efficiency, one can either increase the figure of merit of the thermocouple materials or raise the temperature difference. Over the course of time, both these approaches have been used. Thus, p-type lead telluride has been replaced by TAGS (Te–Ag–Ge–Sb) alloys [33] and, for higher temperatures, Si–Ge alloys have been used. These measures have led to efficiencies in the range 6–7%.

Efficiency is an important factor in space vehicles, because as the efficiency increases, the weight of both the generator and the power source can be reduced. The power and mass of space generators covers a wide range [31]. A generator for a U.S. earth orbit satellite launched in1961 gave an electrical output of 2.7 W whereas a Russian generator powered from a nuclear reactor had an output of 5.5 kW and a mass of 1,000 kg.

Turning to terrestrial applications, there is an interest in generating electricity from waste heat. An analysis of the sources of waste heat, with special reference to Japan, has been carried out by Kajikawa [34]. It is evident from his observations that there is a great variation in the temperatures of the different sources, ranging from about 100 to $1,000$°C. Since it is best to make use of the highest available temperature, it is clear that a range of generators is needed. In a specific design [35], a Si–Ge alloy thermocouple is used with hot and cold junction temperatures equal to 1,123 K and 323 K, respectively. The heat flux is $80.4\,\mathrm{kW/m^2}$ and the thermoelectric efficiency is 10.1%. However, only a fraction of the available heat can be used by the generator so the system efficiency is less than this value. It is clearly important to ensure that as much as possible of the heat of combustion passes to the heat sink via the generator.

There are a number of possible low-temperature heat sources that could be exploited using thermocouples [36]. Geothermal energy is available at temperatures of up to about 200°C. This energy may be found in the form of hot water or steam. At the higher temperatures, it can be used to drive steam turbines but when the temperature is no more than 100°C, some other means of generation is preferred. Thermoelectricity seems to be an obvious possibility. It must be remembered that even the Carnot cycle efficiency is rather low when the temperature difference between the source and sink is less than 100°. A low efficiency is not necessarily too bad a factor in itself since the heat source is free but the smaller the efficiency the larger and more costly is the generator and heat transfer system.

With such a low-temperature source, bismuth telluride alloys are the preferred thermoelectric materials but they should have higher carrier concentrations than they would have in refrigeration applications. The connections between the thermoelements should, of course, be capable of withstanding the operating temperatures, but, otherwise, it is probably most economical to make use of modules of the same kind as are used in cooling.

In some parts of the world, there is a substantial difference between the temperature at the surface of the ocean and that at a depth of, say, 500 m. This temperature difference is of the order of 20° so the efficiency is bound to be low and the size of the plant large, if this source of energy is converted to electricity. Nevertheless, the resource is large and already there are pilot plants for ocean thermal energy conversion (OTEC) schemes. Nihous [37] has pointed out that large-scale use of OTEC systems could, eventually, deplete the resource and this is something that might have to be borne in mind for the future. Nevertheless, there is the potential for large-scale use of thermal gradients in the ocean [38], and thermoelectric generation with an improved efficiency may form part of an integrated system.

Closely related to the use of ocean thermal gradients is the exploitation of solar ponds. Under normal circumstances, water in the sea, which is heated by the sun, rises to the surface because of its lower density. It is, however, possible to invert the usual temperature gradient using a saline gradient to stabilise the system. Straatman and van Sark [39] have discussed the use of the Rankine cycle to capture the energy of this resource and propose that solar ponds should be employed to augment OTEC systems. A thermoelectric generator could be used as an alternative to the Rankine engine.

The efficiency of a thermoelectric generator is always going to be greater if the whole temperature range between that of the source and a sink at ambient temperature can be utilised. Thus, although the motivation may be different, thermoelectric cascades may be used in generation as well as refrigeration. Two- and three-stage cascades for generation have been described by Zhang et al. [40]. These authors used bismuth–telluride alloys for the low-temperature stage and oxides for the higher temperatures. In their two-stage cascade, titanium oxide and strontium titanate were used as the p- and n-type high-temperature materials. The three-stage cascade yielded an efficiency of 13.5% with the hot junctions at 1,223 K. The usual practice when a wide temperature range is to be covered is to use segmented legs rather than a cascade. The question then arises as to whether or not different thermoelectric materials are compatible with one another. This problem has been discussed by Ursell and Snyder [41].

At first sight, it might appear that it is best to make a segmented thermoelement from different materials, each having the highest figure of merit for its range of operation. For example, one might propose a segmented element consisting of bismuth telluride for the low temperature end and Si–Ge for the high temperature end. However, this would not be a good choice as the so-called compatibility factor is substantially different for these two materials. If the compatibility factor differs greatly between the materials, one cannot match the electrical and thermal flux densities without a considerable reduction in efficiency.

Ursell and Snyder considered the reduced efficiency, η_r, of a given segment at the limit of very small $\Delta T/T$, η_r being defined as the ratio of the efficiency to that of a Carnot cycle. Their expression for the reduced efficiency is

$$\eta_r = \frac{\frac{u\rho\lambda}{\alpha}\left(1 - \frac{u\rho\lambda}{\alpha}\right)}{\frac{u\rho\lambda}{\alpha} + \frac{1}{zT}}, \tag{10.16}$$

where u is the ratio of electric current density to heat flux density. The value of u that yields the maximum efficiency is called the compatibility factor s. If the values of s are similar for different materials in a segmented leg, a compromise value for u can be found for which the efficiency of each component is not far from its optimum value. Figure 10.11 shows the variation of reduced efficiency with u for p-type generator materials. It is clear that a compromise value of u can be found for bismuth telluride, zinc antimonide, and $CeFe_4Sb_{12}$ but none of these materials are compatible with Si–Ge.

A basic requirement for two components is that the compatibility factor of one should correspond to a positive value of u for the other. A multi-staged cascade operating between room temperature and 1,000°C might make use of $(Bi\text{–}Sb)_2Te_3$, Zn_4Sb_3, TAGS, $CeFe_4Sb_{12}$, and Si–Ge as the positive branches of the thermocouples. The negative branches for the different stages could be made from $Bi_2(Te\text{–}Se)_3$, PbTe alloys, $CoSb_2$, and La_2Te_3. However, segmented legs in simple thermocouples could not utilise all these materials because of differences in compatibility factor.

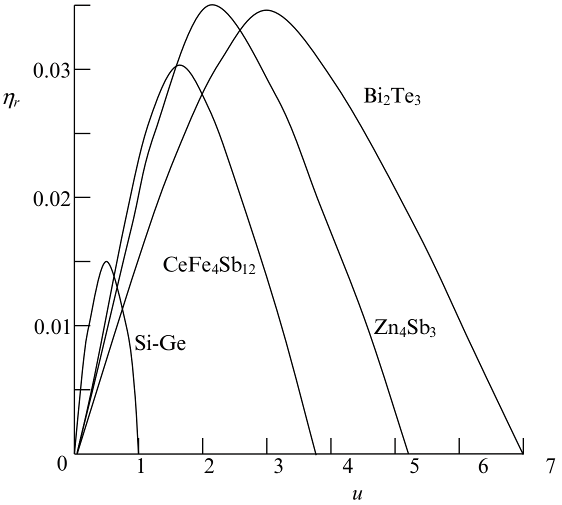


Fig. 10.11 Reduced efficiency as a function of the ratio of electric current density to heat flux density. Schematic plot based on the data of Ursell and Snyder [41]

Snyder [42] has given values for zT and s for a number of p-type and n-type generator materials for use at temperatures of up to 1,000°C. His data for p-type materials are given in Figs. 10.12 and 10.13, whereas those for n-type materials are shown in Figs. 10.14 and 10.15.

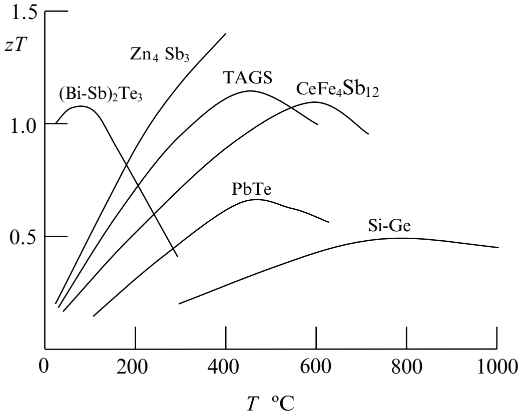


Fig. 10.12 Dimensionless figure of merit of p-type generator materials. Schematic plot based on data of Snyder [42]

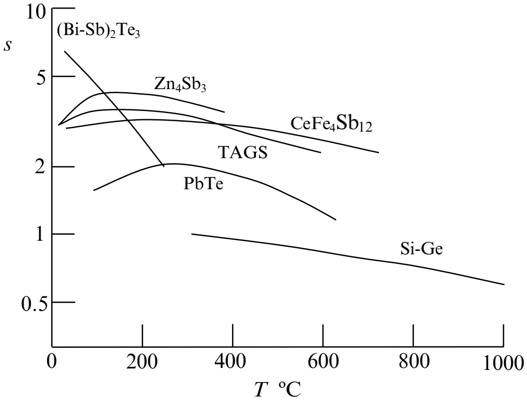


Fig. 10.13 Compatibility factor of p-type generator materials. Schematic plot based on data of Snyder [42]

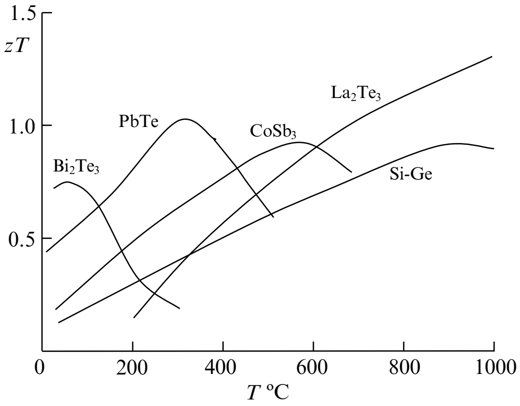
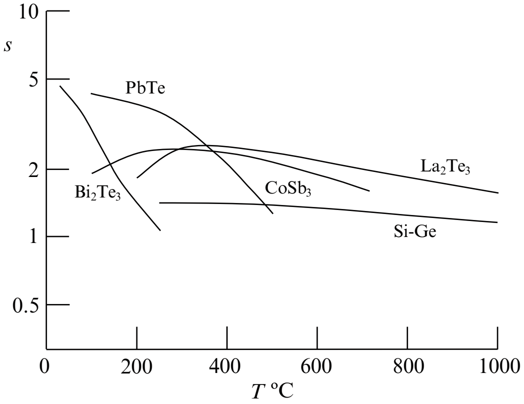


Fig. 10.14 Dimensionless figure of merit of n-type generator materials. Schematic plot based on data of Snyder [42]

Fig. 10.15 Compatibility factor of n-type generator materials. Schematic plot based on data of Snyder [42]



Snyder has proposed as a rule of thumb that segmented materials should have values of s that do not differ by more than a factor of 2. He has been able to explain why the addition of a TAGS segment to a PbTe or SnTe segment produced little extra power since there was a mismatch of compatibility factors. Si–Ge is a material that is particularly difficult to match in segmented generators. Snyder has concluded that, important though the figure of merit is, it is essential to take account of the compatibility factor in selecting generator materials.

The radiation thermopile existed long before it was realised that the use of semiconductors would improve the efficiency of thermoelectric generators. There are other factors besides the figure of merit Z that are important when we consider thermoelectric radiation detectors. In fact, there are two other figures of merit that one might use to characterise such devices. The *responsivity, R,* is the ratio of output voltage to the incident radiation power. This quantity gives no indication of the rate of response of the device and is not as relevant to high-sensitivity detection as the *specific detectivity, D**. The detectivity is the reciprocal of the noise equivalent power (NEP). By dividing the square root of the product of surface area and frequency band width by the NEP, one obtains a quantity that allows direct comparison between different sensors. The NEP is equal to the smallest detectable radiation input, so D^* allows us to compare different sensing technologies.

Graf et al. [43] have carried out an extensive review of radiation thermopiles and compared the different types of device that are available. The highest responsivity of $500\,\mu\mathrm{V\,W}^{-1}$ was reported [44] for a detector based on a couple between $(\mathrm{Bi{-}Sb})_2\,\mathrm{Te}_3$ and Bi–Sb but a Bi/Sb couple, although having only about half the responsivity, had a slightly higher specific detectivity equal to $88 \times 10^5 \mathrm{m/Hz}^{1/2}\,\mathrm{W}$. There is no doubt that the thermoelectric figure of merit is most useful in selecting materials for radiation thermopiles but R and D^* also depend on the techniques that are used in the construction of the devices. One of the most important considerations is the rate of response to the incoming signal. The thermal capacity of the receiver is, therefore, of great significance and, in this respect, transverse devices are most attractive. The transverse thermomagnetic effects have already been mentioned and transverse thermoelectric devices are discussed in the next chapter.

References

1. R.J. Buist, in *Proceedings of the Third International Conference on Thermoelectrics, Arlington, Texas* (IEEE, New York, 1980), p. 130
2. R. Marlow, R.J. Buist, J. Nelson, in *Proceedings of the Fourth International Conference on Thermoelectrics, Arlington, Texas* (IEEE, New York, 1982), p. 125
3. H.J. Goldsmid, *Electronic Refrigeration* (Pion, London, 1986), p. 173
4. R. Sun, M.A. White, in *Thermal Conductivity: Theory, Properties and Applications*, ed. by T.M. Tritt (Kluwer/Plenum, New York, 2004), p. 239
5. V. Semenyuk, in *Proceedings of the Twentieth International Conference on Thermoelectrics, Beijing* (IEEE, New York, 2001), p. 391
6. G.S. Attey, in *Proceedings of the Seventeenth International Conference on Thermoelectrics, Nagoya* (IEEE, New York, 1998), p. 519
7. J.E. Parrott, A.W. Penn, Solid State Electron. **3**, 91 (1961)
8. V.P. Plekhotkin, M.F. Tkachenko, T.M. Melikhova, L.A. Bondarevskaya, Appl. Sol. Energy **17**, 32 (1981)
9. V. Semenyuk, in *Proceedings of the Twenty Fifth International Conference on Thermoelectrics, Vienna* (IEEE, New York, 2006), p. 322
10. B.P. Banney, P.T. Clarke, B.R. Manners, R.M. Weymouth, in *Thermoelectrics Handbook: Macro to Nano, ed. by* D.M. Rowe (CRC Taylor and Francis, Boca Raton, 2006), p. 59–1
11. N.E. Lindenblad, Electr. Eng. **7**, 802 (1958)
12. R.N. Basu, in *Proceedings of the Third International Conference on Thermoelectrics, Arlington, Texas* (IEEE, New York, 1980), p. 117
13. A.I. Kulagin, N.V. Makov, Appl. Sol. Energy **3**, 54 (1967)
14. G.J. Vella, L.B. Harris, H.J. Goldsmid, Sol. Energy **18**, 355 (1976)
15. J.G. Stockholm, L. Pujol-Soulet, P. Sternat, in *Proceedings of the Fourth International Conference on Thermoelectrics, Arlington, Texas* (IEEE, New York, 1982), p. 136
16. J.G. Stockholm, in *CRC Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC Press, Boca Raton, 1994), p. 657
17. J.G. Stockholm, in *CRC Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC Press, Boca Raton, 1994), p. 667
18. J.E. Bean, Ind. Electr. **1**, 132 (1962)
19. V.A. Semenyuk, in *Thermoelectrics Handbook: Macro to Nano*, ed. by D.M. Rowe (CRC Taylor and Francis, Boca Raton, 2006), p. 58–1
20. V.P. Babin, E.K. Iordanishvili, Sov. Phys. Tech. Phys. **14**, 293 (1969)
21. G.E. Hoysos, K.R. Rao, D. Jerger, Energy Convers. **17**, 23 (1977)
22. G.E. Hoysos, K.R. Rao, D. Jerger, Energy Convers. **17**, 45 (1977)
23. R.L. Field, H.A. Blum, Energy Convers. **19**, 159 (1979)
24. T. Yamamoto, Proc. IEEE **56**, 230 (1968)
25. K. Woodbridge, M.E. Ertl, Phys. Stat. Solidi A **44**, K123 (1977)
26. K. Landecker, A.W. Findlay, Solid State Electron. **3**, 239 (1961)
27. M. Idnurm, K. Landecker, J. Appl. Phys. **34**, 1806 (1963)
28. K. Woodbridge, M.E. Ertl, J. Phys. F **8**, 1941 (1978)
29. T.J. Hendricks, R.J. Buist, in *Proceedings of the Third International Conference on Thermoelectrics, Arlington, Texas* (IEEE, New York, 1980), p. 135
30. T.J. Hendricks, in *Proceedings of the Fifth International Conference on Thermoelectrics, Arlington, Texas* (IEEE, New York, 1984), p. 107
31. D.D. Abelson, in *Thermoelectrics Handbook: Macro to Nano*, ed. by D.M. Rowe (CRC Taylor and Francis, Boca Raton, 2006), p. 56–1
32. G.L. Bennett, in *CRC Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC Press, Boca Raton, 1994), p. 515
33. E.A. Skrabek, D.S. Trimmer, in *CRC Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC Press, Boca Raton, 1994), p. 267

34. T. Kajikawa, in *Thermoelectrics Handbook: Macro to Nano*, ed. by D.M. Rowe (CRC Taylor and Francis, Boca Raton, 2006), p. 50–1
35. T. Kajikawa, in *Proceedings of the Fifteenth International Conference on Thermoelectrics, Pasadena* (IEEE, New York, 1996), p. 343
36. K. Matsuura, D.M. Rowe, in *CRC Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC Press, Boca Raton, 1994), p. 573
37. G.C. Nihous, Ocean Engineering **34**, 2210 (2007)
38. P.K. Takahashi, S.M. Masutani, Marine Technology Society Journal **31**, 63 (1997)
39. P.J.T. Straatman, W.G.J.H.M. van Sark, Solar Energy **82**, 520 (2008)
40. L. Zhang, T. Tosho, N. Okinaka, T. Akiyama, Materials Transactions **49**, 1675 (2008)
41. T.S. Ursell, G.J. Snyder, in *Proceedings of the Twenty First International Conference on Thermoelectrics, Long Beach, California* (IEEE, New York, 2002), p. 412
42. G.J. Snyder, in *Thermoelectrics Handbook: Macro to Nano*, ed. by D.M. Rowe (CRC Taylor and Francis, Boca Raton, 2006), p. 9–1
43. A. Graf, M. Arndt, M. Sauer, G. Gerlach, Measurement Sci. Technol. **18**, R59 (2007)
44. F. Völklein, A. Wiegand, V. Baier, Sensors Actuators A **29**, 87 (1991)

Chapter 11
Transverse Devices

11.1 Features of Transverse Coolers and Generators

There are two methods by which one can obtain a transverse temperature gradient from a longitudinal electric current or a transverse electromotive force from a longitudinal heat flow. We have already discussed the transverse thermomagnetic effects and have shown that they can be exploited in low-temperature energy conversion. Transverse thermoelectric effects will also occur for arbitrary orientations in conductors that possess anisotropic Seebeck coefficients.

We first consider the inherent characteristics of transverse devices. Apart from the fact that there are disturbances in the thermal and electrical flow lines near the end contacts, the principal difference lies in the fact that the electrical and thermal resistances can be adjusted independently. As was shown in Chap. 2, the electrical resistance in the direction of current flow is equal to $\rho_x L_x / \left(L_z L_y\right)$ and the thermal conductance in the direction of heat flow is equal to $\lambda_y L_x L_z / L_y$. Thus, while in an ordinary thermoelement the product of the electrical resistance and the thermal conductance cannot be altered by changing the dimensions, this is not the case when a transverse effect is being used. The product of these two quantities in the transverse device is equal to $\rho_x \lambda_y L_{x^2} / L_{y^2}$. In a transverse refrigerator, we can make the sample long in the direction of current flow and short in the direction of heat flow so that the applied voltage matches the output of any available supply, while at the same time providing the required cooling power. A transverse generator will have a short response time, if the length in the direction of heat flow is small, but the output voltage can be large, if the length in the perpendicular direction is increased.

Another feature of a transverse device that was discussed in Chap. 2, in the context of thermomagnetic refrigeration, is the capability of making an infinite-staged cascade by suitable shaping of the sample. This feature is shared by all transverse devices. However, the transverse thermoelements to be discussed in the next section possess an advantage over thermomagnetic devices in that the length can be increased indefinitely by adopting a serpentine configuration, as shown in Fig. 11.1. The need for the magnetic field to be applied in a certain direction prevents that configuration being used in a Nernst–Ettingshausen device, though, in principle, a spiral arrangement could be devised with the magnetic field directed along the axis and the heat flow in a radial direction.

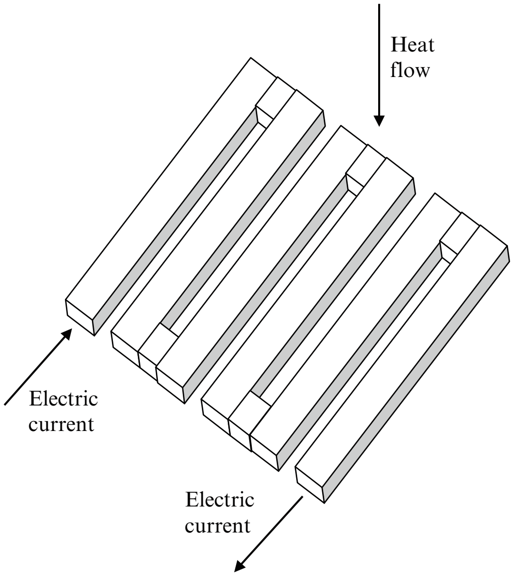


Fig. 11.1 Labyrinth form of transverse thermoelement giving increase in length in a compact arrangement

11.2 Synthetic Transverse Thermoelements

An essential requirement in a material that is to be used for a transverse thermoelement is a substantial anisotropy of the Seebeck coefficient. It also needs a high electrical conductivity and a low thermal conductivity just like any other thermoelectric device. In most extrinsic semiconductors, the Seebeck coefficient displays little if any anisotropy even if the crystal structure allows it and the electrical and thermal conductivities are orientation dependent. It is certainly possible to find a large anisotropy of the Seebeck coefficient in an intrinsic conductor but the other properties are then unfavourable. The problem has been solved by using a synthetic transverse thermoelement made from two conductors that have different Seebeck coefficients. A requirement is that one of the conductors should have much higher electrical and thermal conductivities than the other. The principle of the system may be understood with reference to Fig. 11.2, which shows a layered structure composed of two materials, A and B. This model was used by Babin et al. [1, 2] as the basis for their theory of synthetic transverse thermoelements. It is supposed that $\sigma_A\lambda_A \gg \sigma_B\lambda_B$. We can, therefore, select the relative thicknesses of the layers so that the electrical resistances satisfy the inequality $R_B \gg R_A$ and the thermal conductances obey the inequality $K_A \gg K_B$.

Let the Seebeck coefficients of the two conductors be α_A and α_B. In the x_0 direction, each component will contribute to the overall Seebeck coefficient in proportion to the temperature difference between its surfaces. The temperature difference across each layer will be inversely proportional to its thermal conductance. Thus,

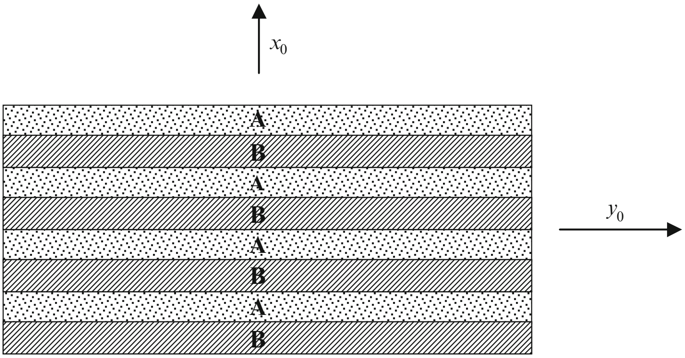


Fig. 11.2 Diagram showing the principle of a synthetic transverse thermoelement

$$\alpha_{x_0} = \frac{\alpha_\mathrm{A}/K_\mathrm{A} + \alpha_\mathrm{B}/K_\mathrm{B}}{1/K_\mathrm{A} + 1/K_\mathrm{B}}. \tag{11.1}$$

Applying the condition $K_\mathrm{A} \gg K_\mathrm{B}$, $\alpha_{x0} \simeq \alpha_\mathrm{B}$

The Seebeck coefficient of the composite in the y_0 direction is given by

$$\alpha_{y_0} = \frac{\alpha_\mathrm{A}/R_\mathrm{A} + \alpha_\mathrm{B}/R_\mathrm{B}}{1/R_\mathrm{A} + 1/R_\mathrm{B}} \tag{11.2}$$

Since $R_\mathrm{B} \gg R_\mathrm{A}$, it follows that $\alpha_{y0} \simeq \alpha_\mathrm{A}$.

A layered composite of the form that we have described does not display a transverse Seebeck or Peltier effect, if the current passes parallel to or normal to the layers. To obtain such an effect, a sample must be cut at some angle. Let us suppose that this sample is oriented at some angle ϕ to the perpendicular to the layers. We also assume that the thicknesses of the layers, d_A and d_B, are such that $d_\mathrm{B}/d_\mathrm{A}$ is equal to n. When the angle ϕ is equal to zero, (11.1) and (11.2) become

$$\alpha_{x_0} = \frac{\alpha_\mathrm{A}/\lambda_\mathrm{A} + n\alpha_\mathrm{B}/\lambda_\mathrm{B}}{1/\lambda_\mathrm{A} + n/\lambda_\mathrm{B}}. \tag{11.3}$$

and

$$\alpha_{y_0} = \frac{\alpha_A \rho_B + n\alpha_B \rho_A}{n\rho_A + \rho_B} \tag{11.4}$$

We may also obtain expressions for the electrical resistance and the thermal conductance of the layered structure in the x_0 and y_0 directions. In the x_0 direction, the effective electrical resistivity of the layers in series is

$$\rho_{x_0} = \frac{\rho_\mathrm{A} + n\rho_\mathrm{B}}{n + 1}. \tag{11.5}$$

and the effective thermal conductivity is

$$\lambda_{x_0} = \frac{n+1}{1/\lambda_A + n/\lambda_B}. \tag{11.6}$$

In the y_0 direction the electrical resistivity of the layers in parallel becomes

$$\rho_{y_0} = \frac{n+1}{1/\rho_A + n/\rho_B}. \tag{11.7}$$

and the thermal conductance is

$$\lambda_{y_0} = \frac{\lambda_A + n\lambda_B}{n+1}\left(1 + Z_{AB}T_m\right). \tag{11.8}$$

It is noted that the thermal conductivity in the y_0 direction is augmented by a factor that includes the figure of merit Z_{AB} of a longitudinal thermocouple made up of the materials A and B. This is necessary because there will be circulating thermoelectric currents in the layers that will produce a Peltier effect. The value of Z_{AB} is given by

$$Z_{AB} = \frac{(\alpha_A - \alpha_B)^2}{(\lambda_A + n\lambda_B)(\rho_A + \rho_B/n)}. \tag{11.9}$$

We now consider the situation, when $\phi \neq 0$. Then, the transverse Seebeck voltage in the y direction due to a temperature gradient in the x direction is given by

$$\alpha_{y\phi x\phi} = \left(\alpha_{x_0} - \alpha_{y_0}\right)\sin\phi\cos\phi. \tag{11.10}$$

It is apparent that the transverse Seebeck coefficient can be quite large if there is a large difference between the Seebeck coefficients of the two components. The effective thermal conductivity and electrical resistivity of the composite, in the y_ϕ and x_ϕ directions respectively, are

$$\lambda_{y\phi\, y\phi} = \lambda_{x_0}\sin^2\phi + \lambda_{y_0}\cos^2\phi, \tag{11.11}$$

and

$$\rho_{x\phi\, x\phi} = \rho_{x_0}\cos^2\phi + \rho_{y_0}\sin^2\phi. \tag{11.12}$$

The transverse figure of merit is

$$Z_\phi = \frac{\alpha^2_{y\phi x\phi}}{\lambda_{y\phi y\phi}\,\rho_{x\phi x\phi}}. \tag{11.13}$$

Babin et al. showed that there is a simple expression for the optimum transverse figure of merit when $\sigma_A\lambda_A \gg \sigma_B\lambda_B$. It is then found that

$$Z_{\phi}^{\max} = z_{\mathrm{A}} \frac{(1 - \alpha_{\mathrm{B}}/\alpha_{\mathrm{A}})^2}{\left[1 + \sqrt{(\lambda_{\mathrm{B}}\rho_{\mathrm{B}}/\lambda_{\mathrm{A}}\rho_{\mathrm{A}})\,(1 + Z_{\mathrm{AB}}T_{\mathrm{m}})}\right]^2}, \tag{11.14}$$

where z_{A} is the longitudinal figure of merit for component A. This expression is identical to that of a longitudinal couple made from A and B apart from the term $(1 + Z_{\mathrm{AB}}T_{\mathrm{m}})$. The figure of merit is always less than that of a longitudinal couple but the loss of performance may be quite small.

The optimum angle ϕ is may be found from

$$\tan\phi_{\mathrm{opt}} = \frac{\sqrt{n}}{n+1}\left[\frac{\rho_{\mathrm{A}}\lambda_{\mathrm{B}}}{\rho_{\mathrm{B}}\lambda_{\mathrm{A}}}\,(1 + Z_{\mathrm{AB}}T_{\mathrm{m}})\right]^{1/4}. \tag{11.15}$$

Babin et al. pointed out that (11.14) does not contain the quantity n that defines the relative thicknesses of the two layers. They stated that $Z_{\phi}^{\max}$ is not critically sensitive to the value of n.

If the inequality $\sigma_{\mathrm{A}}\lambda_{\mathrm{A}} \gg \sigma_{\mathrm{B}}\lambda_{\mathrm{B}}$ is not satisfied, the equations for the maximum figure of merit and the optimum angle become more complicated. This situation was also considered by Babin et al. who obtained analytical expressions in the more general case. However, it does not take long to find $Z_{\phi}^{\max}$ from (11.10) to (11.13) by computation, and the optimum values of ϕ and n can then be determined by inspection. An approximate value for the optimum ratio of the layer thicknesses is

$$n_{\mathrm{opt}} \simeq \left(\frac{2\frac{\lambda_{\mathrm{B}}\rho_{\mathrm{B}}}{\lambda_{\mathrm{A}}\rho_{\mathrm{A}}}}{1 + \frac{\lambda_{\mathrm{B}}\rho_{\mathrm{B}}}{\lambda_{\mathrm{A}}\rho_{\mathrm{A}}}}\right)^{1/2}. \tag{11.16}$$

If both components satisfied the Wiedemann–Franz law $\lambda_{\mathrm{B}}\rho_{\mathrm{B}}$ would be equal to $\lambda_{\mathrm{A}}\rho_{\mathrm{A}}$ and n_{opt} would be equal to 1. Babin et al. discussed the case where material A is a semiconductor and B is a metal. Then, material B would satisfy the Wiedemann–Franz law, but the thermal conductivity would be larger than this law would indicate for material A. However, if A is a good thermoelectric material, the lattice conductivity would be small and the total thermal conductivity would be greater than the Wiedemann–Franz value by a relatively small factor. For example, in a typical bismuth–telluride alloy, this factor might be no more than 4. In this case, n_{opt} would be about 0.6. Generally speaking, it seems that it is likely to be satisfactory to use layers of approximately equal thickness.

11.3 Materials for Transverse Thermoelements

The first demonstration of transverse thermoelectric energy conversion was reported by Korolyuk et al. [3] who made use of the anisotropy of the Seebeck coefficient in a single crystal of cadmium antimonide. However, the performance was poor, as it would be for any other known single-phase material. Much better results were obtained by Gudkin et al. [4] who made a multi-layer composite from bismuth–antimony telluride and bismuth. These authors achieved a figure of merit Z_ϕ of $0.85 \times 10^{-3}\,\mathrm{K}^{-1}$ and actually observed cooling of 23° below room temperature using the transverse Peltier effect in a rectangular bar in which the angle ϕ was equal to 60°. A trapezoidal sample with a ratio of 10:1 between the widths of the hot and cold faces, acting as a cascade, yielded a maximum temperature difference of 35°. More recently, Kyarad and Lengfellner [5] have obtained cooling through 22° using a multi-layer structure consisting of bismuth telluride and lead.

Kyarad and Lengfellner produced their composites by stacking plates of lead and n-type Bi_2Te_3 of $10 \times 20\,\mathrm{mm}^2$ cross-section. The stacks were heated in argon at 320°C under pressure. The bismuth telluride layers were 1 mm thick while experiments were made with different thicknesses of the lead component. The observed anisotropy of the Seebeck coefficient is shown in Fig. 11.3, where the curve represents the variation with the ratio n expected from (11.3) and (11.4).

Kyarad and Lengfellner opted for a thickness ratio n equal to 1 and adopted a tilt angle of 25° in cutting samples from their stack. Typically, the length of a sample was 20 mm, its thickness 10 mm and its width 2 mm. Figure 11.4 shows a plot of temperature difference against current for such a sample.

Ideally, the two materials that are used in a synthetic transverse thermoelement should have a high figure of merit when used together in a conventional thermocouple. They should also have very different values for the electrical and thermal conductivity so that the condition $\sigma_A\lambda_A \gg \sigma_B\lambda_B$ can, at least, be approached. Unfortunately, in most thermocouples with high figures of merit, the positive and negative branches have similar electrical and thermal conductivities. That is the reason, of course, why Gudkin et al. [4] and Kyarad and Lengfellner [5] chose a semimetal or metal rather than a semiconductor as the second component of their composites.

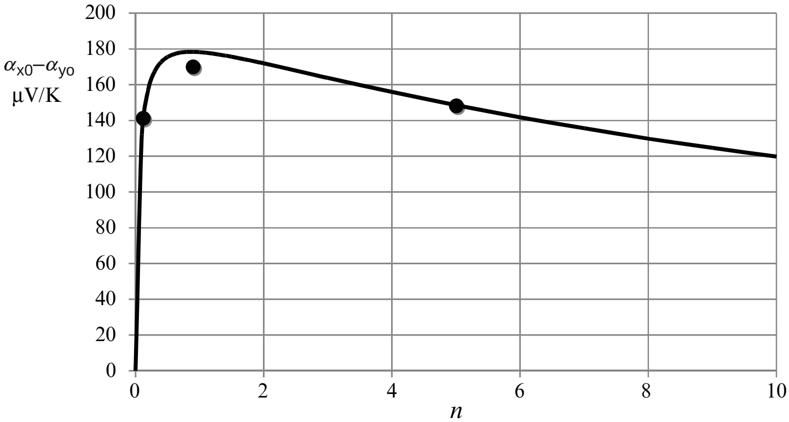


Fig. 11.3 Anisotropy of Seebeck coefficient in a Bi_2Te_3–Pb composite plotted against ratio of layer thicknesses. The curve is the theoretical variation from (11.3) and (11.4) and the points are the experimental data of Kyarad and Lengfellner [5]

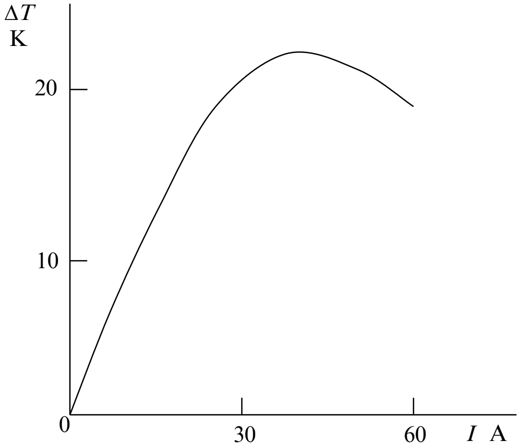


Fig. 11.4 Temperature difference produced by the transverse Peltier effect in a Bi_2Te_3–Pb composite. The heat sink was kept at 295 K. Observations of Kyarad and Lengfellner [5]

There are some general steps that can be taken to make the best of the situation. Thus, if one opts for a semiconductor–metal combination, at least a metal that has a reasonably large Seebeck coefficient of opposite sign to that of the semiconductor can be selected. As we shall see later, narrow-gap semiconductors that have Seebeck coefficients of about $\pm 100\,\mu V\,K^{-1}$ with a high power factor might be used in combination with a conventional thermoelectric material to form composites with a reasonably high transverse figure of merit. One could probably improve on the results of Gudkin et al. by using single crystal bismuth oriented with the trigonal axis in the plane of the layers. In this direction, the Seebeck coefficient of bismuth at 300 K is $-105\,\mu V\,K^{-1}$ and the figure of merit of a longitudinal couple with a p-type bismuth–telluride alloy could be almost as great as that of a couple made entirely from bismuth–telluride alloys. The condition $\sigma_A \lambda_A \gg \sigma_B \lambda_B$ would not really be satisfied, but the ratio $(\sigma_A \lambda_A) / (\sigma_B \lambda_B)$ could be improved by using a sintered bismuth–telluride alloy rather than aligned crystalline material.

It has recently been suggested that one might be able to obtain $\sigma_A \lambda_A \gg \sigma_B \lambda_B$ for a pair of good thermoelectric semiconductors by making one of the materials porous or, perhaps, discontinuous in some other fashion [6]. In principle, the electrical and thermal conductivities should become less in a porous material but their ratio should remain the same.

The n-type semiconductor $Bi_2\,(Te–Se)_3$ has its highest figure of merit when oriented with current flow perpendicular to the trigonal axis but orientation is not a major factor for p-type $(Bi–Sb)_2\,Te_3$. Thus, a thermocouple made from sintered p-type material and aligned n-type material has a value of ZT equal to about 0.85. A synthetic transverse thermoelement made from fully dense samples of these materials would have a very small figure of merit Z_ϕ, but a value in excess of 0.7 would be obtained if the porosity of the sintered component were such as to reduce its conductivities by a factor of 20. Figure 11.5 shows how the transverse figure of merit of a composite of aligned $Bi_2\,(Te–Se)_3$ and sintered $(Bi–Sb)_2\,Te_3$ varies with the porosity factor p of the latter component. The porosity factor is defined as the

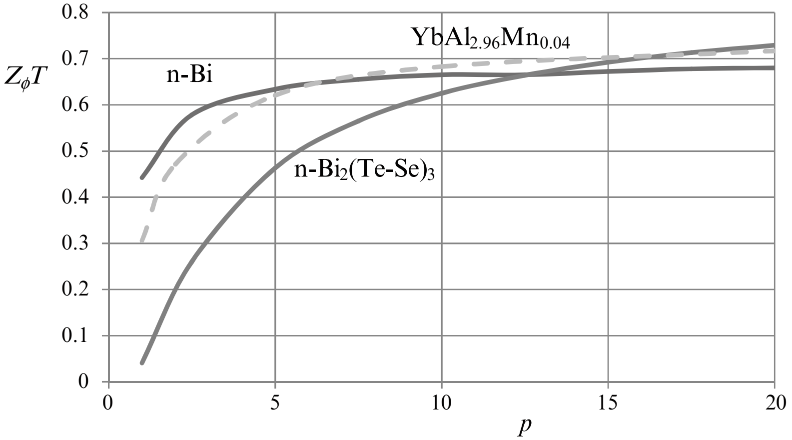


Fig. 11.5 Plot of optimum dimensionless transverse figure of merit against the porosity factor of the positive component $(Bi–Sb)_2\,Te_3$. For the solid curves, the negative component is either $Bi_2\,(Te–Se)_3$ or Bi. For the broken curve, the negative component is $YbAl_{2.96}Mn_{0.04}$

Table 11.1 Properties of materials for use in synthetic transverse thermoelements

Material	Seebeck coefficient μV K^{-1}	Electrical resistivity μΩ m	Thermal conductivity W/m K
Oriented n-type $Bi_2\,(Te–Se)_3$	−180	7.2	1.51
Sintered p-type $(Bi–Sb)_2\,Te_3$*	245	17.5	1.25
n-type bismuth parallel to layers	−105	1.1	13.0
n-type bismuth perpendicular to layers	−50	1.3	6.0
$YbAl_{2.96}Mn_{0.04}$	−90	1.3	3.1

* The values in this row are for fully dense material i.e., $p = 1$

ratio of the conductivity of fully dense material to that of the porous material. At moderate porosities, one should be able to relate p to the density using the formula given by Kingery [7]. As an alternative to making use of porosity to lower the effective conductivities, one might make use of one of the components in the form of widely spaced thin strips.

In the calculations on which the plot in Fig. 11.5 is based, the p-type material has been selected with a higher Seebeck coefficient than its optimum in a conventional couple whereas the n-type material has a lower than optimum Seebeck coefficient. The values for the thermoelectric parameters of the two materials are given in Table 11.1.

Figure 11.5 shows that a transverse dimensionless figure of merit of over 0.7 might be achieved with a porosity factor for the p-type component in excess of 15.

Such a high porosity might be difficult to achieve in practice, so a similar calculation [8] has been carried out for a composite in which the negative component is single crystal bismuth, aligned with the *c*-axis parallel to the layers. The properties of bismuth in the two directions are also included in Table 11.1.

Bismuth has an advantage over Bi_2 $(Te–Se)_3$ as the n-type component in a transverse thermoelement since its ratio of thermal conductivity to electrical resistivity is somewhat greater. Thus, as shown in Fig. 11.5, the value of Z_ϕ is higher for the $(Bi–Sb)_2$ Te_3–Bi composite than for $(Bi–Sb)_2$ Te_3–Bi_2 $(Te–Se)_3$ with porosity factors of less than about 12 for the p-type component.

Some indication of the effect of changing the angle ϕ and the ratio n of the layer thicknesses for the $(Bi–Sb)_2$ Te_3–Bi_2 $(Te–Se)_3$ composite may be assessed from Figs. 11.6 and 11.7. Figure 11.6 shows how $Z_\phi T$ varies with ϕ when $n = 1.1$ and $p = 20$, whereas in Fig. 11.7, $Z_\phi T$ is plotted against n when $\phi = 18.4°$. For the $(Bi–Sb)_2$ Te_3–Bi composite, the preferred value of n remains at about 1.3 for all

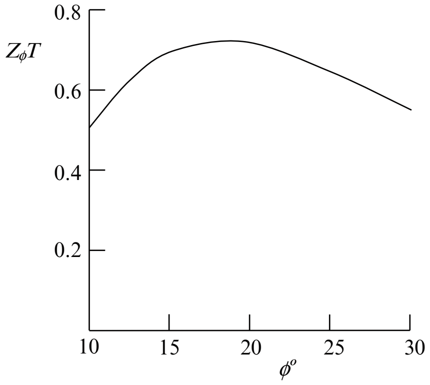


Fig. 11.6 Plot of dimensionless transverse figure of merit against angle of orientation of layers in a $(Bi–Sb)_2$ Te_3–Bi_2 $(Te–Se)_3$ composite. The thickness ratio $n = 1.1$ and the porosity factor for the p-type component is 20

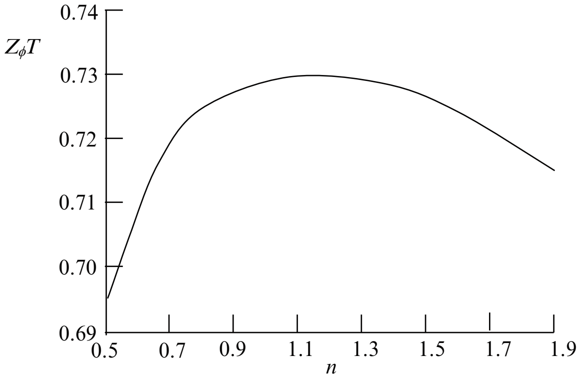


Fig. 11.7 Dimensionless transverse figure of merit plotted against ratio of layer thicknesses for a $(Bi–Sb)_2$ Te_3–Bi_2 $(Te–Se)_3$ composite. The porosity factor for the p-type component is 20 and the angle of orientation of the layers is 18.4°

porosities but the optimum value for ϕ varies from 40° when $p = 1$, through 30° when $p = 5$, to less than 20° when $p = 20$.

It has recently been reported [9] that $YbAl_{2.96}Mn_{0.04}$ has a Seebeck coefficient of $-90\,\mu\,V\,K^{-1}$, a resistivity of $31.3\,\mu\Omega\,m$, and a thermal conductivity of $3.1\,W\,m^{-1}\,K$. This gives it the reasonably high value for zT of 0.6 at 300 K. The alloy would also be useful as one component of a synthetic transverse thermoelement. The value of Z_ϕ for a composite consisting of fully dense sintered p-type bismuth telluride and $YbAl_{2.96}Mn_{0.04}$ is not as good as for one in which single-crystal bismuth is used, probably because advantage can be taken of the anisotropy of the latter. However, as shown by the broken curve in Fig. 11.5, there is a range of the porosity factor for which bismuth telluride and $YbAl_{2.96}Mn_{0.04}$ is the best combination. Even for lower porosities of the positive component, it may be preferable to use $YbAl_{2.96}Mn_{0.04}$ as the negative component since it does not need to be monocrystalline. The alloy has a cubic structure and can be prepared by sintering.

It will be seen that the concept of a porous or open-structured material of one conductivity type in conjunction with a dense material of the other conductivity type should lead to a great improvement in the performance of synthetic transverse thermoelements. It is expected that these improved composites will find application in cascade coolers, that will outperform conventional multi-stage thermoelectric refrigerators, and in fast-response radiation detectors.

11.4 Alternative Configurations

Although synthetic transverse thermoelements are easy to handle once they are made, their construction presents problems that are not encountered in ordinary thermoelectric modules. Thought has therefore been given to alternative arrangements for producing the transverse devices.

Investigators in the past always seem to have made use of the multi-layer principle to obtain a large anisotropy of the Seebeck coefficient. However, the same results should be obtained if one of the components is in the form of parallel wires that are embedded in the second component. The principle can be understood with reference to Fig. 11.8, which shows a single rod embedded in a bar of material. Suppose that the rod and the bar have Seebeck coefficients α_R and α_B that differ greatly from

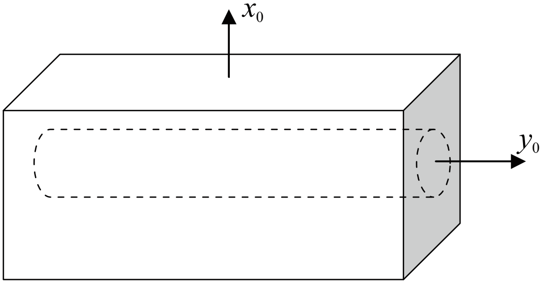


Fig. 11.8 Cylindrical conducting rod imbedded in another conductor

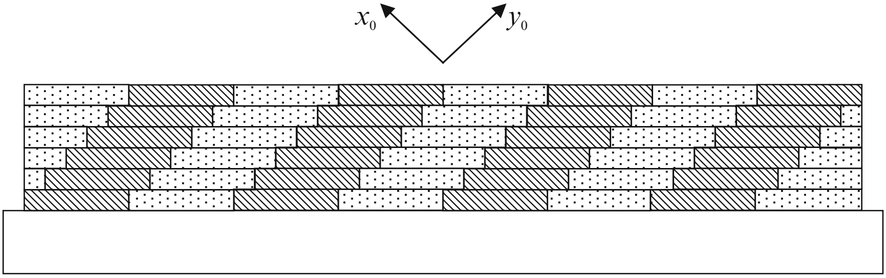


Fig. 11.9 Schematic representation of a synthetic anisotropic thermoelement, produced by thin-film technology. The mask is shifted in regular steps as each half-layer is deposited

one another. Also, let the electrical and thermal conductivities of the rod be much greater than those of the bar. Then, in the x_0 direction, the Seebeck coefficient will be equal to α_B since the rod will merely form an internal isothermal and equipotential surface. In the y_0 direction, the electrical and thermal flows will be controlled by the rod and the Seebeck coefficient will be equal to α_R.

It may not be necessary for the rods to extend from one end of the composite to the other. A substantial anisotropy of the Seebeck coefficient could result from the inclusion of shorter aligned rods within the matrix. If these rods were to take the form of nanowires, then one could, perhaps, combine the improvement in the figure of merit resulting from a nanostructure with the practical advantages of the transverse thermoelectric effects.

Even though the arrangement of rods or wires in a matrix might be simpler to manufacture than a layered structure, there still exists the need to cut a section from it at the preferred angle. A synthetic transverse device with the appropriate orientation might be made in a single process using thin-film technology. The principle is illustrated in Fig. 11.9.

References

1. V.P. Babin, T.S. Gudkin, Z.M. Dashevskii, L.D. Dudkin, E.K. Iordanishvili, V.I. Kaidanov, N.V. Kolomoets, O.M. Narva, L.S. Stil'bans, Sov. Phys. Semiconductors **8**, 478 (1974)
2. L.I. Anatychuk, *Thermoelectric Power Convertors* (Institute of Thermoelectricity, Kiev, 2005), p. 168
3. S.L. Korylyuk, I.M. Pilat, A.G. Samoilovich, V.N. Slipchenko, A.A. Snarski, E.F. Tsar'kov, Sov. Phys. Semiconductors **7**, 502 (1973)
4. T.S. Gudkin, E.K. Iordanishvili, E.E. Fiskind, Sov. Phys. Tech. Phys. Lett. **4**, 844 (1978)
5. A. Kyarad, H. Lengfellner, Appl. Phys. Lett **89**, 192103 (2006)
6. H.J. Goldsmid, J. Thermoelectricity N **1**, 7 (2008)
7. W.D. Kingery, J. Francl, R.L. Coble, T.J. Vasilos, J. Amer. Ceram. Soc. **37**, 107 (1954)
8. H.J. Goldsmid, Phys. Stat. Solidi A **12**, 2966 (2008)
9. T. He, J.J. Krajewski, M.A. Subramanian, U. S. Patent 7371960 B2 (2008)

Chapter 12
Properties of Nanostructured Materials

12.1 Theory of Nanostructures

Since the development of thermoelectric materials in the 1950s, most of the attempts to improve the figure of merit have centred on the reduction of the lattice conductivity. It is only since the advent of studies of nanostructures that there has been a real possibility of improving the electronic properties, i.e., an increase of the power factor.

The present interest in low-dimensional thermoelectric materials was prompted by the theoretical work of Hicks and Dresselhaus [1]. Their calculations were based on a conductor with a single parabolic band. The carriers were assumed to be electrons and the conduction band was used in the theory, but the valence band would have suited the purpose equally well. They first obtained an expression for the dimensionless figure of merit in the three-dimensional case in terms of the Fermi energy and a parameter that is virtually the same as Chasmar and Stratton's β, modified to allow for anisotropy of the effective mass.

The theory was then adapted to the situation in which the electrons are confined within a two-dimensional quantum well. This would be the case for a thin narrow-gap semiconductor sandwiched between two plates of a wide-gap material. The wide-gap semiconductor does not contribute to the transport processes. The dispersion relation for the electrons in three dimensions is

$$\varepsilon = \frac{\hbar^2 k_{\mathrm{x}}^2}{2m_{\mathrm{x}}} + \frac{\hbar^2 k_{\mathrm{y}}^2}{2m_{\mathrm{y}}} + \frac{\hbar^2 k_{\mathrm{z}}^2}{2m_{\mathrm{z}}}. \tag{12.1}$$

For the two-dimensional quantum well, this relation is changed to

$$\varepsilon = \frac{\hbar^2 k_{\mathrm{x}}^2}{2m_{\mathrm{x}}} + \frac{\hbar^2 k_{\mathrm{y}}^2}{2m_{\mathrm{y}}} + \frac{\hbar^2 \pi^2}{2m_{\mathrm{z}} a^2}, \tag{12.2}$$

where a is the width of the well in the z direction. The expressions for the Seebeck coefficient, the electrical conductivity and the electronic thermal conductivity, for a constant relaxation time, become

$$\alpha = -\frac{k}{e}\left(\frac{2F_1}{F_0} - \eta*\right), \tag{12.3}$$

$$\sigma = \frac{1}{2\pi a}\left(\frac{2kT}{\hbar^2}\right)\left(m_{\mathrm{x}}m_{\mathrm{y}}\right)^{1/2} F_0 e\mu_{\mathrm{x}}, \tag{12.4}$$

and

$$\lambda_{\mathrm{e}} = \frac{\tau\hbar^2}{4\pi a}\left(\frac{2kT}{\hbar^2}\right)^2\left(\frac{m_{\mathrm{y}}}{m_{\mathrm{x}}}\right)^{1/2} k\left(3F_2 - \frac{4F_1^2}{F_0}\right). \tag{12.5}$$

In (12.3), the quantity η^* is defined in terms of the reduced Fermi energy η by

$$\eta* = \eta - \frac{\hbar\pi^2}{2m_{\mathrm{x}}a^2kT}. \tag{12.6}$$

The dimensionless figure of merit is

$$Z_{2\mathrm{D}}T = \frac{\left(2F_1/F_0 - \eta*\right)^2 F_0}{1/\beta' + 3F_2 - 4F_1^2/F_0}, \tag{12.7}$$

where

$$\beta' = \frac{1}{2\pi a}\left(\frac{2kT}{\hbar^2}\right)\left(m_{\mathrm{x}}m_{\mathrm{y}}\right)^{1/2}\frac{k^2T\mu_{\mathrm{x}}}{e\lambda_{\mathrm{L}}}. \tag{12.8}$$

Equation (12.7) allows us to optimise η^*, just as η can be optimised for a three-dimensional material. Hicks and Dresselhaus found that η^*_{opt} remains fairly close to zero, as η_{opt} does in three dimensions, until β' becomes appreciably greater than unity. Figure 12.1 shows a schematic plot of η^*_{opt} against β'.

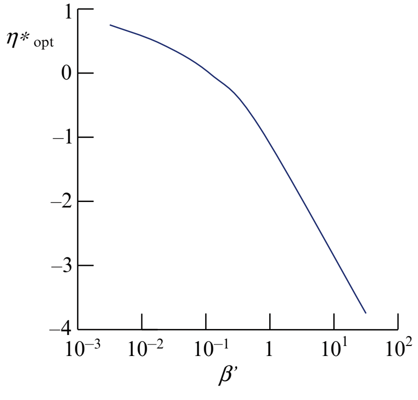


Fig. 12.1 Schematic plot of η^*_{opt} against β' for a 2D conductor. Diagram based on the data of Hicks and Dresselhaus [1]

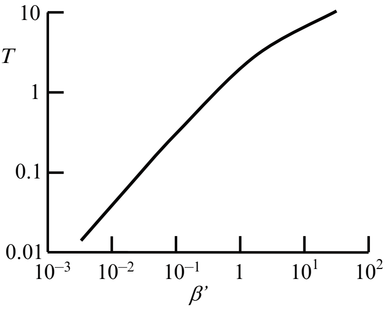


Fig. 12.2 Schematic plot of $z_{2D}T$ against β' for a 2D conductor. Diagram based on the data of Hicks and Dresselhaus [1]

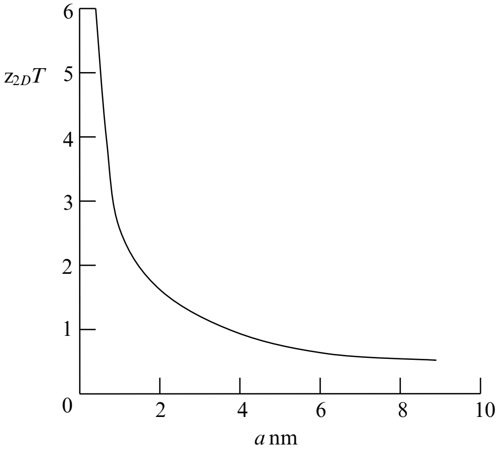


Fig. 12.3 Schematic plot of $z_{2D}T$ against thickness of confined region for bismuth telluride with current flow perpendicular to the c axis. Based on the data of Hicks and Dresselhaus [1]

Figure 12.2 shows the variation of the two-dimensional figure of merit with β' for the optimum value of η^*. It will be seen that if β' is of the order of unity or greater, then $z_{2D}T$ can be significantly larger than unity. Comparing β' with β we find that

$$\frac{\beta'}{\beta} \propto \frac{\left(m_x m_y\right)^{1/2}}{a m^{*3/2}}. \tag{12.9}$$

Since the respective figures of merit depend on β and β' in a similar fashion it follows that $z_{2D}T$ should eventually become larger than zT as a is made smaller.

Hicks and Dresselhaus applied their theory to predict the behaviour of two-dimensional bismuth telluride. They selected parameters for this material that corresponded to $z_{3D}T = 0.52$. They assumed that the mobility along the confined layer would remain the same for two-dimensional material since there should be no interface scattering of the electrons. Phonons are not confined in the same way so the lattice conductivity will probably become smaller for two-dimensional material, but in their calculations, Hicks and Dresselhaus did not take this into account. Nevertheless, they were able to predict a substantial increase for the figure of merit, over the value for three-dimensional samples, when the layer thickness a falls below about 5 nm, as illustrated in Fig. 12.3.

Having established the principle for two-dimensional conductors, Hicks and Dresselhaus [2] went on to extend their ideas to one-dimensional conductors or quantum wires. The simplest one-dimensional model has a square cross-section of side a. The two-dimensional dispersion relation (12.2) then changes to

$$\varepsilon = \frac{\hbar^2 k_{\mathrm{x}}^2}{2m_x} + \frac{\hbar^2 \pi^2}{2m_{\mathrm{y}} a^2} + \frac{\hbar^2 \pi^2}{2m_{\mathrm{z}} a^2}. \tag{12.10}$$

Once again one can obtain expressions for the transport properties. Thus,

$$\alpha = -\frac{k}{e}\left(\frac{3F_{1/2}}{F_{-1/2}} - \eta\right), \tag{12.11}$$

$$\sigma = \frac{1}{\pi a^2}\left(\frac{2kT}{\hbar^2}\right)^{1/2} (m_{\mathrm{x}})^{1/2} F_{-1/2} e \mu_{\mathrm{x}}, \tag{12.12}$$

and

$$\lambda_{\mathrm{e}} = \frac{2\tau}{\pi a^2}\left(\frac{2kT}{\hbar^2}\right)^{1/2} (m_{\mathrm{x}})^{-1/2} k^2 T \left(\frac{5}{2}F_{3/2} - \frac{9F_{1/2}^2}{2F_{-1/2}}\right). \tag{12.13}$$

η is the reduced chemical potential referred to the lowest bound state. The resultant expression for the one-dimensional figure of merit is

$$z_{1\mathrm{D}} T = \frac{\frac{1}{2}\left(3F_{1/2}/F_{-1/2} - \eta\right)^2 F_{-1/2}}{1/\beta'' + \frac{5}{2}F_{3/2} - 9F_{1/2}^2/2F_{-1/2}}, \tag{12.14}$$

where

$$\beta'' = \frac{2}{\pi a^2}\left(\frac{2kT}{\hbar^2}\right)^{1/2} \frac{k^2 T m_{\mathrm{x}}^{1/2} \mu_{\mathrm{x}}}{e \lambda_{\mathrm{L}}}. \tag{12.15}$$

Comparing β'' with β, we find that

$$\frac{\beta''}{\beta} \propto \frac{(m_{\mathrm{x}})^{1/2}}{a^2 m^{*3/2}}. \tag{12.16}$$

Since $z_{1\mathrm{D}}$ is dependent on β'' just as $z_{2\mathrm{D}}$ depends on β', we realise that the one-dimensional figure of merit should rise very rapidly as a decreases.

Hicks and Dresselhaus again applied their calculations to the specific case of bismuth telluride for quantum wires aligned parallel to the binary, bisectrix, and trigonal directions. Improvements in $z_{1\mathrm{D}}T$ were found for all orientations once the width of the specimen became less than about 10 nm. Comparing one- and two-dimensional quantum wells made from bismuth telluride, $z_{1\mathrm{D}}T$ was found to be 6 for a 1 nm wire whereas $z_{2\mathrm{D}}T$ was 2.5 for a thickness of 1 nm. There definitely seems to be an advantage in going from two dimensions to one dimension.

The obvious conclusion to be drawn from these findings is that a zero-dimension sample would be still better. The difficulty, of course, is to make use of a specimen of zero dimensions if it does, indeed, have a high figure of merit. As we shall see, the problem can be solved and the so-called quantum dots have led to the demonstration of zT values substantially greater than unity.

There are other low-dimensional configurations that have been studied. For example, there are quantum tubes in which both the wall thickness and the tube diameter can be on a nanoscale. Then, again quantum dots have been incorporated in nanowires to form superlattice nanowires [3]. Perhaps, the most important development has been the inclusion of nanodots in bulk material since this configuration seems to offer scope for exploitation in commercial thermoelectric devices. These and other directions for research on thermoelectric nanostructures have been reviewed by Dresselhaus and Heremans [4].

In ordinary thermoelectric materials, the power factor for a given Fermi energy depends on the carrier mobility and the density of electronic states. Since we assume that the mobility is the same for the nanostructure and the bulk material, it seems reasonable to assign the improvement in the low-dimensional configuration to an increase in the density of states. It is, in fact, correct to describe the quantum well effects as changes in the band structure. Beneficial changes can include both a greater density of states and an increase in the band gap. The latter could allow a semimetal, with favourable electronic properties other than a large enough band gap, to become a semiconductor. A schematic representation of the changes in the dependence on energy of the density of states is shown in Fig. 12.4.

In selecting materials that are likely to yield high figures of merit when fabricated as nanostructures, it is thought that they should already have reasonably

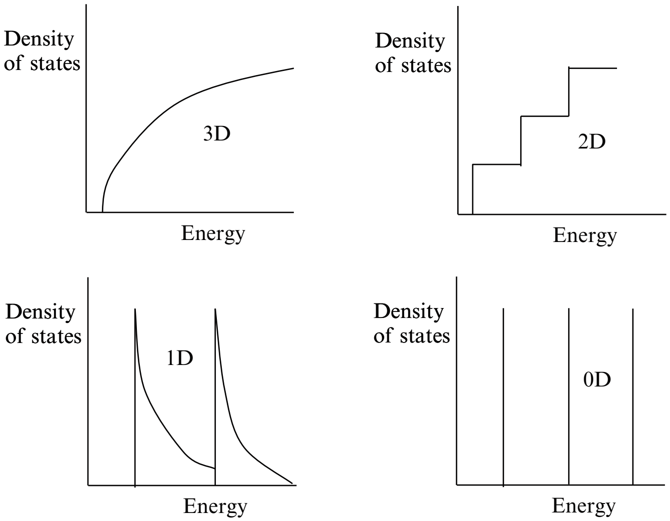


Fig. 12.4 Schematic diagrams for the density of states as a function of energy in bulk (3D), 2D, nanowires (1D), and quantum dot (0D) configurations. Based on the review by Dresselhaus and Heremans [4]

good thermoelectric properties as bulk samples. Bismuth is, particularly, promising since it has one of the highest known power factors. Nanostructured bismuth might display not only an increase in the density of states and a reduction in the lattice conductivity but also a transition from semimetal to semiconductor.

The thermoelectric properties of bismuth nanowires at 77 K have been calculated by Lin et al. [5]. It was found that zT in the trigonal direction should reach a value of 6 for an electron concentration of 10^{24} per cubic metre in a wire of 5-nm diameter. The Seebeck coefficient [6] would be $-400\,\mu\,\mathrm{V\,K^{-1}}$, a value that is much greater than that observed in single crystals in any direction, whatever be the doping. In spite of the high Seebeck coefficient, the electrical conductivity would be as large as $3.4 \times 10^6\ \Omega^{-1}\mathrm{m}^{-1}$ because of an increase in the density of states. The power factor would then, of course, be higher than for any known material. Lin et al. did not assume that the lattice conductivity would be reduced below its value for bulk material though the considerations in the following section suggest that further improvement should be possible. Rabin et al. [7] have extended the calculation of the electronic properties to Bi–Sb nanowires and these should, in any case, have a reduced thermal conductivity.

12.2 Thermal Conduction in Low-Dimensional Materials

There have been a number of experimental demonstrations of improved figures of merit in nanostructures and, in some cases, these have been due to modifications of the electronic properties. However, even when the electronic properties do not change, zT can still be larger than for bulk material. This benefit is associated with a reduction in the lattice conductivity. It is not surprising that the scattering of phonons should be stronger when the size of the sample in at least one direction is reduced. We have already discussed the effect of grain size on the lattice conductivity in bulk materials but there are some more subtle effects that become apparent in nanostructures.

Perhaps, we should discuss the first practical realisation of a figure of merit, $\mathrm{z}T$, that is substantially greater than unity at ordinary temperatures using a low-dimensional structure. In 2001, Venkatasubramanian et al. [8] observed a value of zT equal to 2.4 for a p-type $\mathrm{Bi_2Te_3/Sb_2Te_3}$ superlattice. There were certainly changes in the electronic properties when the superlattice period fell below 25 nm but these were not the major reason for the improved figure of merit. More significant was the reduction in the lattice conductivity. Measurements were made in the cross-plane direction which happened to be parallel to the c axis. The observed thermal conductivity was smaller by a factor of 2.2 than the value in the same direction for bulk crystals.

The thermal conductivity of $\mathrm{Bi_2Te_3/Sb_2Te_3}$ superlattices has been analysed by Touzelbaev et al. [9]. The thermal conductivity measurements were carried out using a thermoreflectance technique on superlattices with periods ranging from 4 to 12 nm. The results are shown schematically in Fig. 12.5.

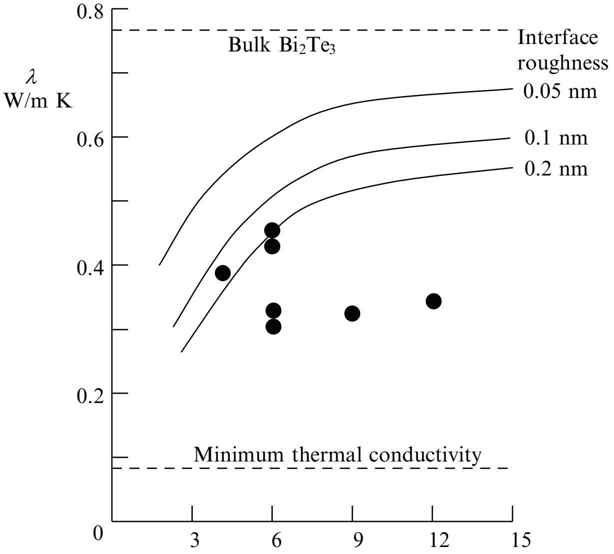


Fig. 12.5 Thermal conductivity of a Bi_2Te_3/Sb_2Te_3 superlattice plotted against the superlattice period. Schematic diagram based on the data of Touzelbaev et al. [9]. The three curves represent calculated values using the Boltzmann equation for different degrees of roughness at the interfaces. The broken lines show the thermal conductivities of bulk bismuth telluride and the minimum value predicted by Cahill's theory [10]. The dots are the experimental data

The thermal conductivity of all the superlattices was much less than the value for bulk bismuth telluride but did not become as low as one would expect for amorphous material, that is the minimum value calculated using the theory of Cahill et al. [10]. The effect of interface scattering was calculated using the Boltzmann theory for various degrees of surface roughness. Although these calculations for a roughness of 0.2 nm agreed with some of the experimental data, they did not account for the fact that the thermal conductivity failed to increase with increasing superlattice period l. Touzelbaev et al. speculated that disorder rather than interface scattering may have been the main reason for the reduction in thermal conductivity. They pointed out that previous work [11] on the thermal conductivity of Si–Ge superlattices had shown that, although there was an increase of λ with l as was expected for $l < 5$ nm, λ became appreciably smaller for $l > 13$ nm. It was supposed that the superlattices with the longer period were strongly disordered. If there were similar disorder in the Bi_2Te_3/Sb_2Te_3 superlattices, this could account for the observations.

Although the studies by Venkatasubramanian et al. [8] were made with current flowing in the cross-plane direction, superlattices can also be used with in-plane conduction. One would expect quite different phonon scattering mechanisms to apply in this situation. For one-dimensional nanostructures, the current always flows parallel to the interfaces and it is clear that one needs a treatment for the thermal conductivity behaviour in both directions.

The thermal conductivity of nanostructures has been reviewed by Dames and Chen [12]. They pointed out that one can change the lattice conductivity not only by altering the free path length of the phonons but also by modifying the specific heat or the speed of sound. It is, certainly, possible in principle to alter the dispersion relation for the phonons. However, there are practical problems if we wish to make use of such changes. If the specific heat is to be changed, the size of the nanostructure must be smaller than the phonon wavelength. It is also necessary that the phonons remain coherent. It should be somewhat easier to obtain changes in the group velocity of the phonons over part of the spectrum, particularly when the flow is perpendicular to the interfaces. It seems, though, that the reduction of the phonon free path length by additional scattering processes is a more profitable approach.

It would be easiest if one could use Matthiessen's rule in adding the reciprocal mean free paths for the different scattering processes, but one may question its validity because of the different phonon wavelengths that are involved. Nevertheless, Matthiessen's rule is still often used although one must be aware of its limitations.

The lattice conductivity of superlattices in the cross-plane direction has been considered by Chen [13]. He showed that the thermal resistance is the sum of a contribution from each layer and a thermal boundary resistance. This boundary resistance is not characteristic only of the interface but also depends on the layer thickness. When the superlattice period is small, the thermal boundary resistance is the dominant factor. Under this condition, the effective thermal conductivity may be quite different from that of a bulk sample. It is not necessarily true that those materials with the lowest lattice conductivity as bulk materials will also have the lowest value as nanostructures. A critical factor seems to be the acoustic mismatch at the interfaces. When there is a large acoustic contrast between the two materials on either side of the boundary, the reflection of phonons is strong. In this case, roughness of the interface may not be beneficial. Furthermore, if the confined region is very small, one can obtain interference effects and tunneling. It is not surprising that the variation with period of the thermal conductivity of the $\mathrm{Bi_2Te_3/Sb_2Te_3}$ superlattices in the cross-plane direction failed to satisfy the expectations of classical theory.

When the flow is parallel to the interfaces, there are significant differences. Specular reflection at the boundaries should not affect the thermal resistance, so interface roughness must be an important factor. As a general rule, the thermal conductivity in superlattices will be greater for in-plane flow than in the cross-plane direction. In superlattice nanowires, the flow is parallel to the outer surface but perpendicular to the internal boundaries.

The range of phonon wavelengths that makes a significant contribution to the thermal conductivity extends over 2 decades. It may well be that a different scattering mechanism is dominant for each part of that range. Thus, we know that Umklapp scattering is effective over a wide range of wavelengths but alloy scattering predominates at the short wavelength end. In nanostructures, we expect a major contribution to the scattering from the boundaries.

There are two simple models for the lattice vibrational spectrum that might be used. We have discussed the Debye model in Chap. 3. An alternative is the Born-von Karman model in which the frequency becomes more or less independent of wave vector near its limiting value. If this model is used, the short wavelength phonons have a low group velocity and do not contribute greatly to the thermal conductivity. At all longer wavelengths, the two models become equivalent. When the temperature is well below Θ_D, only the long wavelength modes are excited and it does not matter which model is used, but we are probably more interested in temperatures of the order of Θ_D or higher. Dames and Chen [12] have calculated that, even though the Born-von Karman model tends to favour heat conduction by the phonons of longer wavelength, 90% of the thermal conductivity at high temperatures is due to phonons with wavelength less than $2.94\,\Lambda_0$, where Λ_0 is the smallest possible phonon wavelength set by the size of the unit cell. It turns out that 90% of the heat is carried by phonons of wavelength Λ less than about 2 nm. This is small compared with the scale of all but the smallest nanostructures.

A consequence of the small wavelength of the phonons is that they will be diffusively scattered at the interfaces rather than specularly reflected. If we define the proportion of specular reflection as p, where $p = 0$ for completely diffuse scattering and $p = 1$ for perfect specular reflection,

$$p = \exp\left(\frac{-16\pi^3\varepsilon^2}{\Lambda^2}\right), \tag{12.17}$$

where ε is the surface roughness. It appears that the roughness of nanowires and superlattices is not less than 1 or 2 nm so that most interfaces are diffuse scatterers of phonons. This is favourable from the viewpoint of improving the figure of merit. It might be thought that there would also be a reduction of the carrier mobility but the electron wavelength is much larger. An interface that seems to be rough for the phonons can be specular for the electrons and holes.

Boundary scattering will be the more effective when most of the heat is carried by phonons that have a long mean free path. We have discussed this for bulk materials in Sect. 5.3. There, we showed that boundary scattering can affect the thermal conductivity even when the sample size is somewhat greater than the mean free path.

Dames and Chen expressed the lattice conductivity in terms of the free path length, l_t which itself is dependent on frequency ω. Thence,

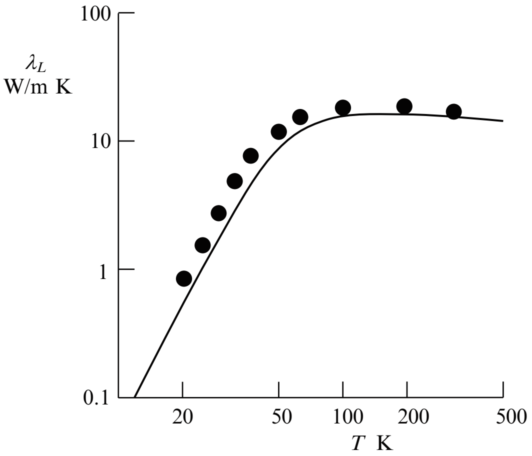
$$\lambda = \int_0^\infty \lambda_{l_t} d l_t, \tag{12.18}$$

where

$$\lambda_{l_t} = -\frac{1}{3}\hbar\omega \frac{q_L^2}{2\pi^2}\frac{\partial N_0}{\partial T} l_t \frac{d\omega}{d l_t}. \tag{12.19}$$

In this equation, N_0 is the Bose–Einstein distribution function. One can make use of (12.19) when the dispersion relation and the dependence of the mean free path on frequency are known. Dames and Chen were able to substitute appropriate values for two different types of material, namely silicon, which has a large lattice conductivity, and lead telluride for which the lattice conductivity is low. Thermal conductivity measurements have been made on both materials over a wide temperature range and, for silicon, these measurements have been extended to nanowires [14]. The experimental observations on silicon nanowires are compared with the theoretical predictions in Fig. 12.6.

Fig. 12.6 Schematic plot of lattice conductivity against temperature for a silicon nanowire of 56 nm diameter. The experimental points indicated by the dots are based on the data of Li et al. [14] and the curve represents the theoretical calculations of Dames and Chen [12]



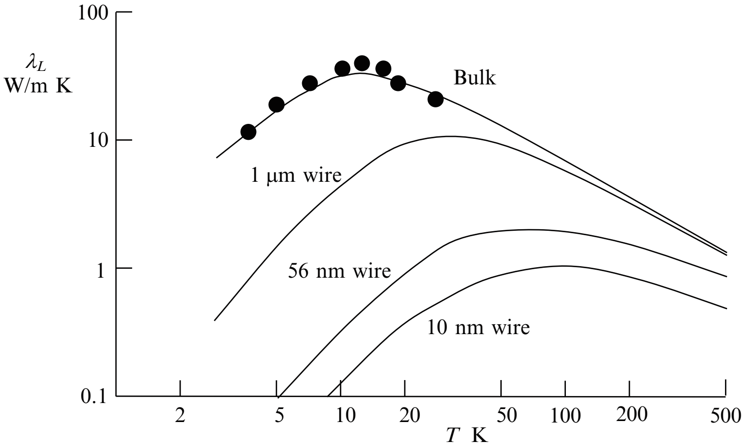


Fig. 12.7 Schematic plot of lattice conductivity against temperature for PbTe. The experimental points indicated by the dots are based on the data of Greig [16] for a bulk crystal and the curves represent the theoretical calculations of Dames and Chen [12]

Dames and Chen mentioned a surprising feature of their calculations, namely the significant contribution to the thermal conductivity of phonons with long wavelengths. To account for 90% of the heat conduction, wavelengths of up to 12.8 µm have to be considered. They pointed out, however, that this is consistent with the observation of boundary scattering in bulk silicon at high temperatures by Savvides and Goldsmid [15].

Dames and Chen predicted the lattice conductivity of nanowires of PbTe after showing that the theory was consistent with observations on large crystals of the compound. The experimental data of Greig [16] are shown in Fig. 12.7 together with theoretical curves for bulk material and nanowires down to 10 nm in diameter.

From the point of view of thermoelectric applications, the substantial fall in the lattice conductivity at high temperatures is, particularly, interesting.

The reduction of the lattice conductivity that results from the use of nanostructures is noticeable when the dimensions are too large for any changes in the electronic parameters to be apparent. This effect will be particularly valuable provided that the carrier mobility does not become significantly smaller than it is in bulk material.

12.3 Observations on Nanostructures

One of the predictions of Lin et al. [5] was that nanowires of bismuth should display large Seebeck coefficients. This is supported by the experimental work of Heremans et al. [17] who claim to have observed Seebeck coefficients of a magnitude of the order of 1 mV K^{-1} for bismuth nanowire composites. The composites consisted of wires of down to 7 nm diameter deposited by a vapour-phase technique in porous silica and alumina templates. The variation of electrical resistance with temperature was found to be consistent with semiconducting rather than semimetallic behaviour. A wire of 15 nm diameter appeared to have an energy gap of 0.18 eV and a 9 nm wire displayed a gap of 0.29 or 0.39 eV. However, the fact that the 9 nm wires seemed to have a Seebeck coefficient of -100 mV K^{-1} casts some doubt on the observations since it is difficult to explain such a high value. Heremans et al. mentioned that the measurements were difficult to perform since the sample resistances were in the range 1 MΩ–1G Ω, though they were able to use an electrometer with an impedance of 10^{14} Ω. Even if the measurements have to be accepted with some reservation, they do provide evidence for a substantial energy gap in bismuth nanowires. It seems that Seebeck coefficients in excess of -200 μV K^{-1} can be reached for $Bi_{0.95}Sb_{0.05}$ with an electron concentration of about 4×10^{22} per cubic metre when the wire diameter is less than 30 nm. There is evidence for neutral impurity scattering of the electrons but not boundary scattering, so it seems that high mobilities can still be maintained in the nanostructures.

The results of Heremans and his co-workers may be compared with those of Lin et al. [18]. In both cases, a semiconductor-like variation of the electrical resistivity with temperature was found though Lin et al. did not observe the same large Seebeck coefficients, admittedly for nanowires of a larger diameter. The overall impression seems to be that nanostructure dimensions of the order of 10 nm or less are needed if the potential of bismuth or Bi–Sb is to be fully realised.

Semiconducting behaviour has also been found for bismuth nanotubes with wall thicknesses of 15 nm though an increase of the wall thickness to 100 nm leads to metallic behaviour [19].

The improved figure of merit in bismuth telluride superlattices has been discussed in the previous section and other configurations based on this compound are being investigated. Here, we shall draw attention to some interesting experiments on point contacts to $Bi_{0.5}Sb_{1.5}Te_3$ and $Bi_2Te_{2.9}Se_{0.1}$. Ghoshal et al. [20]

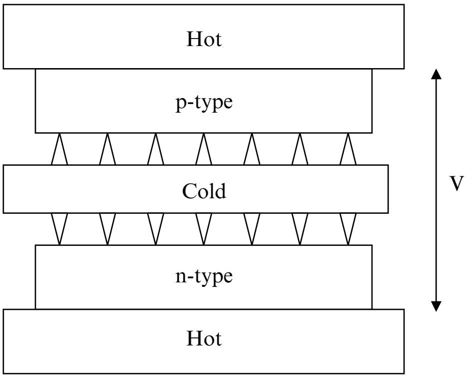


Fig. 12.8 Arrangement of point contacts to p-type $Bi_{0.5}Sb_{1.5}Te_3$ and n-type $Bi_2Te_{2.9}Se_{0.1}$ in the experiment of Ghoshal et al. [20]

used the arrangement shown in Fig. 12.8. They observed the cooling of the central plate when a voltage was applied between the outer plates. The cooling effect was consistent with a figure of merit *ZT* at room temperature as high as 1.7. The p-type and n-type materials were taken from batches used in commercial thermoelectric modules having a value of *ZT* below unity.

It is clear that the improved performance is due to the fact that the thermal and electrical gradients are concentrated in the thermoelectric materials in the region of the point contacts. Although the radius of the contact of each tip was stated to be 0.6 μm, the actual contact region was, no doubt, much smaller. Ghoshal et al., in fact, considered two extremes for the nature of the contacts. In one extreme, the electronic transport between the cone and the thermoelectric material takes place by tunneling and the phonons do not contribute in the contact region. At the other extreme, there is electrical and thermal contact over the entire tip width. It is probable that the actual contact takes place at a number of points each of which must be of much less than 1 μm radius. It may, therefore, be appropriate to regard the contact regions as nanostructures with a reduced thermal conductivity. It is significant that substantial cooling through over 60° was achieved using a current of some hundreds of mA, so the point contact arrangement may have practical applications.

One of the first demonstrations that low-dimensional thermoelements can have improved electronic properties as well as a reduced thermal conductivity was reported by Harman et al. [21]. These authors worked on a PbTe/Te superlattice and measured the Seebeck coefficient, Hall coefficient, and electrical resistivity. The specimens were prepared using molecular beam epitaxy (MBE) with superlattice periods between 15 and 30 nm. The tellurium layers were estimated to be between 0.8 and 1.5 nm in thickness. Measurements were made for the in-plane direction. The Seebeck coefficient at any particular carrier concentration was distinctly greater for the superlattices than for bulk PbTe, as shown in Fig. 12.9. The thermal conductivity was not measured but, using calculated values for this quantity, *zT* was found to have increased from 0.37 to 0.52 on changing from bulk material to superlattice.

Rather surprisingly, there does not seem to be any correlation between superlattice period and the enhancement of the Seebeck coefficient within the range covered

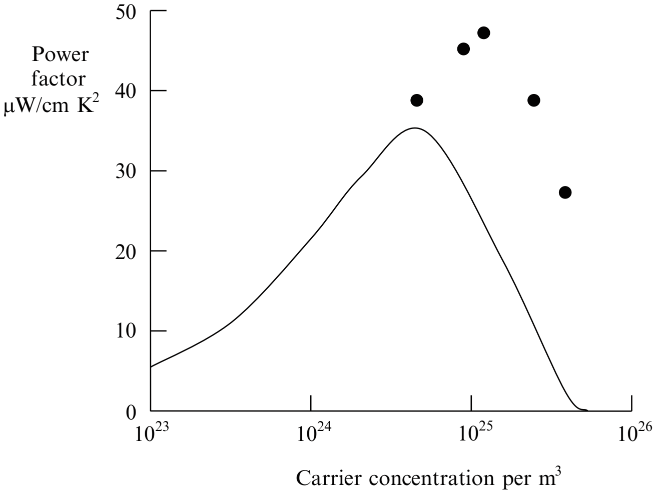
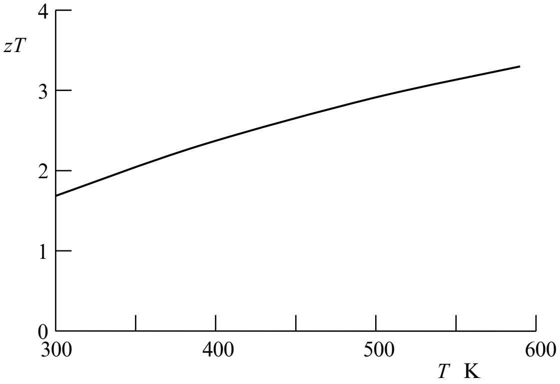


Fig. 12.9 Variation of power factor with carrier concentration. The curve represents the behaviour of bulk PbTe and the dots are the results for PbTe/Te superlattices. Schematic plot based on the data of Harman et al. [21] at 300 K

Fig. 12.10 Dimensionless figure of merit plotted against temperature for an n-type $PbSe_{0.98}Te_{0.02}/PbTe$ quantum dot superlattice. Based on the data of Harman et al. [22]



by the measurements. Harman et al. thought that a change in the scattering parameter might be affecting the Seebeck coefficient, but it is difficult to see how this could improve the power factor. Thus, the results are a good indication that the figure of merit can be increased in nanostructures by a modification of the density of states.

Outstanding results have been obtained by Harman et al. [22] for quantum dot superlattices made from Pb(Te–Se) and PbTe. Samples were produced on BaF_2 substrates using the MBE technique. A wide range of carrier concentration was investigated and values of zT in excess of 1.6 at 300 K were found for $PbSe_{0.98}Te_{0.02}/PbTe$ samples with Seebeck coefficients between −220 and $-240\,\mu V\,K^{-1}$. The dimensionless figure of merit rose with temperature, reaching a value in excess of 3 at about 550 K, as shown in Fig. 12.10.

Thermal conductivity measurements were performed on Pb(Te–Se) superlattices using the 3ω method at 300 K. The electronic component was calculated from the electrical conductivity so that the lattice conductivity could be found. The value at 300 K for the quantum dot superlattice was $0.33\,\mathrm{W\,m^{-1}\,K^{-1}}$ compared with $0.84\,\mathrm{W\,m^{-1}\,K^{-1}}$ for a simple superlattice. The results show that the high figure of merit is due to a reduction in the lattice conductivity as well as an increase in the power factor.

Most electronic devices that are in use at the present time make use of silicon and it would be an advantage if this element could be used as a thermoelectric material at ordinary temperatures. This is not feasible with bulk silicon, mainly because of its high thermal conductivity. It is, therefore, remarkable that recent work on silicon nanowires has led to high values of zT. Hochbaum et al. [23] made arrays of these nanowires by an aqueous electroless etching technique. The wires had a range of diameters between 20 and 300 nm and had a roughness of typically 1–5 nm. The improvement in the figure of merit is primarily due to a reduction in the thermal conductivity. Thus, a 52 nm silicon nanowire was found to have a thermal conductivity of $1.6\,\mathrm{W\,m^{-1}\,K^{-1}}$ of which $1.2\,\mathrm{W\,m^{-1}\,K^{-1}}$ is the lattice contribution. The nanowires have reasonable power factors that are comparable with the values found for bismuth telluride and zT reached 0.6 at 300 K.

The possibility of using nanostructures based on silicon is certainly a promising development but, perhaps, even more promising is the report of a high value of zT for a nanostructured bulk material. Poudel et al. [24] have been able to make such material by hot-pressing $(\mathrm{Bi{-}Sb})_2\,\mathrm{Te}_3$ alloys. They observed values of zT were equal to 1.2 at room temperature, 0.8 at 250°C, and a maximum of 1.4 at 100°C. These high values of the figure of merit were confirmed by the large temperature differences that were produced using the new material in Peltier coolers. Temperature differences of 86°, 106°, and 119° were observed with hot junction temperatures of 50°C, 100°C, and 250°C, respectively.

The material made by Poudel et al. differs from other sintered products in that nanosized powders were used and care was taken that a fine structure was maintained during sintering. They prevented oxidation of the powders, a problem that has been encountered by previous workers who attempted to use finely powdered bismuth–telluride alloys. Microstructure studies revealed that most of the grains are nanosized and larger grains contain nanodot regions. It is not sufficient to have nanoinclusions in a matrix through which the current flows. The high figure of merit is the result of flow within the nanoparticles. In other words, one cannot achieve the best performance merely by using nanoparticles as scattering centres.

12.4 Preparation of Nanostructures

The first practical demonstration of the outstanding properties of superlattices was presented by Venkatasubramanian et al. as mentioned in the Sect. 12.2. The superlattices were grown by a metallorganic chemical vapour deposition method [25, 26].

Films of Bi_2Te_3 and Sb_2Te_3 were grown on sapphire and fcc GaAs substrates in a vertical reactor at a pressure of 46.7 kPa and a temperature of 350°C. The growth of both the tellurides made use of pyrolytic reactions involving metallorganic compounds of bismuth, antimony, and tellurium. The crystal perfection of the films was ascertained using X-ray diffraction and low-energy electron diffraction. The other feature of significance was the stoichiometry and this was confirmed using X-ray photoemission spectrometry and Rutherford back-scattering. A key to the success of the deposition technique was the monitoring of the growth using spectroscopic ellipsometry [27]. Ellipsometry measures the change of polarisation of a beam of light on reflection from a surface and, in this work, allowed the accurate measurement and control of the growth rate and the thickness.

Nanowires are, usually, made by the deposition of the thermoelectric material within the pores of a template. One of the earliest techniques made use of the pressure injection of liquid bismuth into the nanochannels of a porous alumina sheet [28]. Arrays of nanowires of 65 nm diameter and 109 μm length were made by this method. Keyani et al. [29] were able to electrodeposit Bi–Sb from a non-aqueous solution into alumina that contained pores of 100 nm diameter. The template had been coated with a thin layer of nickel which acted as an electrode during the deposition process. The nanowires had the composition $Bi_{0.3}Sb_{0.7}$ which is probably not very favourable from the figure of merit viewpoint, but the authors were able to incorporate their nanowire array in a thermocouple with $Bi_{0.4}Sb_{1.6}Te_3$ to yield a Peltier cooling effect of 7 K.

Wang et al. [30] also used a porous alumina template for the precipitation of a Bi_2Te_3–Te heterostructure from supersaturated $Bi_{0.26}Te_{0.74}$. Nanowires of 60 nm diameter were produced. A pulsed electrodeposition technique was used by Trahey et al. [31] to produce nanowires of bismuth telluride with a diameter of 35 nm. These nanowires were crystalline with their length direction perpendicular to the c-axis, as is favourable for thermoelectric applications. Pulsed electrodeposition was also used by Dou et al. [32] to produce nanowires and nanotubes of Bi–Sb.

Nanotubes, rather than nanowires, of bismuth telluride were produced by a galvanic displacement technique that was described by Xiao et al. [33]. Ni nanotubes were displaced by bismuth telluride using the difference between the redox potentials to drive out bismuth and tellurium from a nitric acid solution. Nanotubes of bismuth telluride were also synthesised by Cao et al. [34] using a low-temperature aqueous chemical method. Their tubes were of about 100 nm diameter with a wall thickness of 30–40 nm and a length of 500 nm–1 μm. Li et al. [35] extended the production of nanotubes by an electrodeposition technique to $(Bi–Sb)_2 Te_3$ and $Bi_2 (Te–Se)_3$ and investigated the importance of the diameter of the channels in the alumina template, the reaction rate, and the thickness of the gold electrode that completed the electrolytic circuit.

An unusual method for creating the template for the electrodeposition of Bi_2Te_3 nanowires has been devised by Koukharenko et al. [36]. They introduced pores in polyimide by ion beam irradiation followed by etching. The pores were 30–80 nm in diameter and the thickness of the polyimide was 24 μm so the aspect ratio of the pores was 300:1 upwards.

Any enhancement of the electronic properties in nanostructures is likely to be greater as the dimensionality is reduced [4]. Thus, zero dimension should be superior to one dimension. It is this factor that encourages work on quantum dots. However, quantum dots present a problem in that it is not immediately apparent how electric current is to be made to flow through them. The quantum dots must, therefore, be embedded in another conducting medium. If the quantum dots are close together, one might hope for tunnelling directly between them. Alternatively, if tunnelling cannot occur, it is important that the surrounding medium should not detract significantly from the overall performance.

A successful quantum dot device was made by Harman et al. [37]. These authors used the MBE technique to grow $PbSe_{0.98}Te_{0.2}$ quantum dots in a PbTe matrix. A sample on which measurements were made was described as a thick film; its thickness was given as 104 μm with its other dimensions being 5 mm and 11 mm. Metallisation was applied so that current could be passed parallel to the 5 mm edges.

Quantum dot configurations, such as that of Harman et al., have the advantage over some other nanostructures in that they can be handled in the same way as bulk materials. They are, in this sense, not dissimilar to bulk materials that have built in nanostructures.

Hogan et al. [38] have discussed compounds in which there exist spontaneously formed endotaxially embedded nanostructures. They made samples of PbTe with 4% concentrations of Sb, Bi, and InSb. A reduction in the thermal conductivity of the samples containing Sb and InSb was attributed to the formation of nanostructured regions. Similar regions were observed in $AgPb_mSbTe_{2+m}$ and $Ag\,(Pb_{1-x}Sn_x)_m\,SbTe_{2+m}$.

Nanocomposites based on bismuth telluride have been prepared by Hu et al. [39]. These authors prepared powders of both nano-size and micro-size. The powders were mixed together and hot-pressed. The ratio of nano-powder to micro-powder lay in the range 0–15% with the optimum estimated at about 10%. Cao et al. [40] produced powders of bismuth–antimony telluride by a hydrothermal process and, when the powders were hot-pressed, it was found that the resultant samples possessed a nanostructure. Grain growth during the hot-pressing process seemed to have been avoided. Since the value of zT, equal to 1.28 at 303 K, is higher than is usual for bulk bismuth–antimony telluride, it seems reasonable to suppose that the structure has been changed and that nanostructured composites can be obtained by established sintering methods, if sufficiently small-sized powders are used. It is noted, however, that the improvement in figure of merit seems to come about through enhanced scattering of the phonons and it is much more difficult to develop nanostructures that are small enough to change the electron density of states.

An improvement in the figure of merit for the compound $Ag_{0.8}Pb_{22.5}SbTe_{20}$ when prepared by mechanical alloying and spark plasma sintering has been attributed to the formation of nanoscopic Ag/Sb-rich regions [41]. Once again, this supports the idea that nanostructured regions can be retained in bulk thermoelectric materials after sintering, if the composition and size of the starting powders are properly chosen.

Finally, we mention another technique for the production of a nanocomposite. It has been reported [42] that bismuth telluride with a nanostructure has been prepared by melt spinning. Melt spinning is a process that enables a molten material to be cooled at a very rapid rate. When applied to bismuth telluride, the flakes of the compound that were produced had rather low values of the Seebeck coefficient. This could possibly be improved by annealing but one cannot be sure that a nanostructure would then be retained.

References

1. L.D. Hicks, M.S. Dresselhaus, Phys. Rev. B. **47**, 12727 (1993)
2. L.D. Hicks, M.S. Dresselhaus, Phys. Rev. B. **47**, 16631 (1993)
3. Y.-M. Lin, M.S. Dresselhaus, Phys. Rev. B. **68**, 075304 (2003)
4. M.S. Dresselhaus, J.P. Heremans, in *Thermoelectrics Handbook: Macro to Nano*, ed. by D.M. Rowe, (CRC Taylor and Francis, Boca Raton, 2006), p. 39–1
5. Y.-M. Lin, X. Sun, M.S. Dresselhaus, Phys. Rev. B. **62**, 4610 (2000)
6. Y.-M. Lin, Private communication
7. O. Rabin, Y.-M. Lin, M.S. Dresselhaus, Appl. Phys. Lett. **79**, 81 (2001)
8. R. Venkatasubramanian, E. Silvola, T. Colpitts, B. O'Quinn, Nature **413**, 597 (2001)
9. M.N. Touzelbaev, P. Zhou, R. Venkatasubramanaian, K.E. Goodson, J. Appl. Phys. **90**, 763 (2001)
10. D.G. Cahill, S.K. Watson, R.O. Pohl, Phys. Rev. B. **46**, 6131 (1992)
11. S.-M. Lee, D. G. Cahill R. Venkatasubramanian, Appl. Phys. Lett., **70**, 2957 (1997)
12. C. Dames, G. Chen, in *Thermoelectrics Handbook: Macro to Nano*, ed. by D.M. Rowe (CRC Taylor and Francis, Boca Raton, 2006), p. 42–1
13. G. Chen, Phys. Rev. B. **57**, 14958 (1998)
14. D. Li, Y. Wu, P. Kim, L. Shi, P. Yang, A. Majumdar, Appl. Phys. Lett. **83**, 2934 (2003)
15. N. Savvides, H.J. Goldsmid, J. Phys. C: Solid State Phys. **6**, 1701 (1973)
16. D. Greig, Phys. Rev. **120**, 358 (1960)
17. J.P. Heremans, C.M. Thrush, D.T. Morelli, M.-C. Wu, Phys. Rev. Lett. **88**, 216801 (2002)
18. Y.-M. Lin, O. Rabin, S.B. Cronin, J.Y. Ying, M.S. Dresselhaus, Appl. Phys. Lett. **81**, 2403 (2002)
19. L. Li, Y.W. Yang, X.H. Huang, G.H. Li, R. Ang, L.D. Zhang, Appl. Phys. Lett. **88**, 103119 (2006)
20. U. Ghoshal, S. Ghoshal, C. McDowell, L. Shi, Appl. Phys. Lett. **80**, 3006 (2002)
21. T.C. Harman, D.L. Spears, M.P. Walsh, J. Electron. Mats. **28**, L1 (1999)
22. T.C. Harman, M.P. Walsh, B.E. LaForge, G.W. Turner, J. Electron. Mats. **34**, L19 (2005)
23. A.I. Hochbaum, R. Chen, R.D. Delgado, W. Liang, E.C. Garnett, M. Najarian, A. Majumdar, P. Yang, Nature. **451**, 163 (2008)
24. B. Poudel, Q. Hao, Y. Ma, Y.C. Lan, A. Minnich, B. Yu, X. Yan, D.Z. Wang, A. Muto, D. Vashae, X.Y. Chen, J.M. Liu, M.S. Dresselhaus, G. Chen, Z. Ren, Sciencexpress. 10.1126, 20 (2008)
25. R. Venkatasubramanian, E. Silvola, B. O'Quinn, in *Thermoelectrics Handbook: Macro to Nano*, ed. D.M. Rowe (CRC Taylor and Francis, Boca Raton, 2006), p. 49–1
26. R. Venkatasubramanian, T. Colpitts, E. Watko, M. Lamvik, N. El-Masry, J. Crystal Growth. **170**, 817 (1997)
27. H. Cui, I Bhat, B. O'Quinn, R. Venkatasubramanian, J. Electron. Mater. **30**, 1376 (2001)
28. Z. Zhang, X. Sun, M.S. Dresselhaus, J.Y. Ying, Appl. Phys. Lett. **73**, 1589 (1998)
29. J. Keyani, A.M. Stacy, J. Sharp, Appl. Phys. Lett. **89**, 233106 (2006)
30. W. Wang, X. Lu, T. Zhang, G. Zhang, W. Jiang, X. Li, J. Amer. Chem. Soc. **129**, 6702 (2007)

31. L. Trahey, C.R. Becker, A.M. Stacy, Nano Lett. **7**, 2535 (2007)
32. X. Dou, G. Li, X. Huang, L. Li, J. Phys. Chem. C, **112**, 8167 (2008)
33. F. Xiao, B. Yoo, K.H. Lee, N.V. Myung, J. Amer. Chem. Soc. **129**, 10068 (2007)
34. Y.Q. Cao, T.J. Zhu, X.B. Zhao, J. Alloys Compounds. **449**, 109 (2008)
35. X.-H. Li, B. Zhou, L. Pu, J.-J. Zhu, Crystal Growth Design, **8**, 771 (2008)
36. E. Koukharenko, X. Li, I. Nandhakumar, B. Schiedt, C. Trautmann, J. Speed, M.J. Tudor, S.P. Beeby, N.M. White, Electronics. Lett. **44**, 500 (2008)
37. T.C. Harman, P.J. Taylor, M.P. Walsh, B.E. LaForge, Science. **297**, 5590 (2002)
38. T.P. Hogan, A. Downey, J. Short, J. D'Angelo, C.-I. Wu, E. Quarez, J. Androulakis, P.F.P. Poudeu, J.R. Sootsman, D.-Y. Chung, M.G. Kanatzidis, S.D. Mahanti, E.J. Timm, H. Schock, F. Ren, J. Johnson, E.D. Case, J. Electronic Mater. **36**, 704 (2007)
39. J.Z. Hu, X.B. Zhao, T.J. Zhu, A.J. Zhou, Physica Scripta. **T129**, 120 (2007)
40. Y.Q. Cao, T.J. Zhu, X.B. Zhao, X.B. Zhang, J.P. Tu, Appl. Phys. A. **92**, 321 (2008)
41. M. Zhou, J.-F. Li, T. Kita, J. Am. Chem. Soc. **130**, 4527 (2008)
42. D.G. Ebling, A. Jacquot, M. Jägle, H. Böttner, U. Kühn, L. Kirste, Phys. Stat. Solid (RRL). **1**, 238 (2007)

Chapter 13
Thermionic Energy Conversion

13.1 Vacuum Thermoelements

Ioffe [1] drew attention to the fact that a major contributor to the losses in thermoelectric energy convertors, namely the heat conduction by the lattice, could be eliminated if the conduction of electricity took place in a vacuum. He gave the name *vacuum thermoelements* to vacuum diodes that are used in energy conversion. In such diodes, there would still, of course, be heat losses associated with radiation and these would not be negligible at high temperatures. In the 1950s, it was inconceivable that vacuum thermoelements could be operated at ordinary temperatures, since the thermionic emission of electrons from then known materials was too small for practical purposes below about 1,000 K. Although this seems to preclude the possibility of using vacuum diodes for refrigeration, they might still be employed as high-temperature generators. Hatsopoulos and Kaye [2] did, in fact, carry out experiments on thermionic diodes with this aim.

The anode and cathode of a vacuum diode may be connected externally via a resistive load. The potential distribution is shown in Fig. 13.1. It is supposed that the work functions at the cathode and anode are Φ_2 and Φ_1, respectively. The electrons in the space between the electrodes will oppose the flow of current between the electrodes, so a space charge potential δ is also included in the diagram.

If the temperature of the cathode is higher than that of the anode, there will be a tendency for electrons to pass from the former to the latter. The saturation current density from either electrode is given by Richardson's equation

$$i_{1,2} = A_0 T_{1,2}^2 \exp\left(-\frac{\Phi_{1,2}}{kT_{1,2}}\right), \tag{13.1}$$

where the subscripts 1 and 2 refer to the anode and cathode, respectively. A_0 is a constant that has the ideal value $1.2 \times 10^6 \,\mathrm{A\,m^{-2}K^{-2}}$.

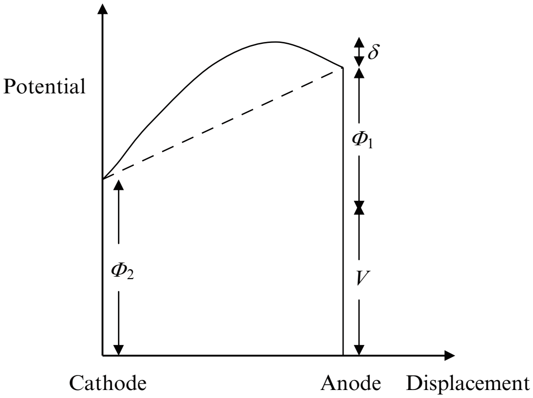


Fig. 13.1 Schematic plot showing potential as a function of displacement in a thermionic diode

The net current density, taking account of the emission from both electrodes and the space charge potential is

$$i = A_0\left[T_2^2 \exp\left(-\frac{\Phi_1}{kT_2}\right)\exp\left(-\frac{\delta}{kT_2}\right)\exp\left(-\frac{V}{kT_2}\right) - T_1^2 \exp\left(-\frac{\Phi_1}{kT_1}\right)\exp\left(-\frac{\delta}{kT_1}\right)\right]. \tag{13.2}$$

The current per unit area depends on the temperatures of the two electrodes, the height of the space charge barrier, and the work function of the cold anode. It does not depend on the work function of the hot cathode.

The space charge effect can be eliminated by making the space between the electrodes very small. In this case, the current density is approximately given by

$$i = A_0 T_2^2 \exp\left(-\frac{\Phi_1}{kT_2}\right)\exp\left(-\frac{V}{kT_2}\right). \tag{13.3}$$

The power output per unit area is

$$w = iV - w_L, \tag{13.4}$$

where w_L is the ohmic loss in the leads. The efficiency is

$$\eta = \frac{iV - w_L}{j}, \tag{13.5}$$

where j is the heat flux per unit area of the cathode. Part of this heat flux will be lost in radiation and there may be other losses in the supporting structure.

Hatsopoulos and Kaye's experiments were performed with both electrodes made from (Ba–Sr)CO_3. The cathode and the anode temperatures were 1,260 and 538°C, respectively. The efficiency was measured as a function of the output voltage and reached a maximum of about 13%.

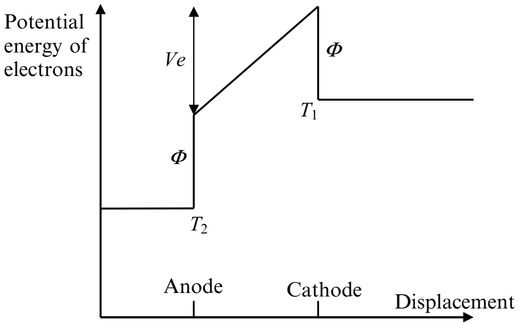


Fig. 13.2 Schematic plot of the potential energy of the electrons against displacement in a thermionic diode with $\Phi_1 = \Phi_2$

Mahan [3] showed that it might be possible to operate a thermionic diode as a refrigerator at a temperature of the order of 500 K without a major breakthrough in the development of emitters with low work functions. He considered that a work function of 0.7 eV would be sufficiently small. Even lower work functions have been reported for materials known as electrides and alkalides [4]. Measurable thermionic emission was observed [5] at a temperature as low as 193 K and a work function of 0.2 eV has been mentioned. Whether or not stable emitters with low work functions that can yield large electron current densities will be available is still an unanswered question. However, it does seem that one can do much better than using traditional emitter materials that have work functions of the order of 1 eV.

We follow Mahan's derivation of the coefficient of performance (COP) of a thermionic diode in the refrigeration mode. Figure 13.2 plots the potential energy of the electrons within the diode where the cathode is at a temperature T_1 and the anode is at a temperature T_2. The space between the anode and cathode represents the vacuum through which the electrons travel. We suppose that the work function has the same value, Φ, at both the anode and cathode though Mahan did not actually restrict his theory in this way.

We make use of Richardson's equation (13.1) to obtain the current densities from the electrodes. Part of the applied voltage, V, equal to V_0, is used to balance the different emissions from the anode and cathode due to the higher temperature of the former. The remaining part, $V - V_0$, drives the electron current. Since V_0 is the voltage that must be applied for the current to be equal to zero, it resembles the Seebeck voltage in a conventional thermoelement.

If we suppose that conventional current is positive when it flows towards the cathode (that is, when the electrons flow towards the anode), the total current density in the vacuum space is

$$i = i_1 - i_2 = A_0\left[T_1^2 \exp\left(-\frac{\Phi}{kT_1}\right) - T_2^2\left(-\frac{\Phi}{kT_2}\right)\right]. \tag{13.6}$$

The value of V_0 is found by setting i equal to zero. Thence,

$$eV_0 = \Phi\left(\frac{T_2}{T_1} - 1\right) + 2\ln\left(\frac{T_2}{T_1}\right). \tag{13.7}$$

Just as the energy of the carriers in a thermoelectric material consisting of a kinetic part and a potential contribution associated with the position of the Fermi level, so also does the energy of the electrons in a thermionic diode. The kinetic term has the value $2kT$ and the potential energy is Φ. Thus, as the electrons leave the cathode, they carry away heat at the rate j per unit area, where j is given by

$$j = i\frac{\Phi + 2kT_1}{e}. \tag{13.8}$$

Also, the rate of expenditure of electrical energy per unit area associated with the current flow is

$$w_\mathrm{i} = iV. \tag{13.9}$$

We have to take account of the transfer of heat between the electrodes by radiation at the rate

$$w_\mathrm{r} = -\varepsilon\sigma(T_2^4 - T_1^4), \tag{13.10}$$

where ε is the emissivity of the surfaces of the electrodes and σ is the Stefan–Boltzmann constant. The negative sign is consistent with the fact that thermal radiation opposes the cooling effect of the current flow.

It is necessary to provide a return path for the electrons just as one has to use two branches in a thermocouple [6]. The electrical resistance encountered by this return flow represents some loss of performance. We shall suppose that the return path makes use of a metallic conductor but Xuan [7] has shown that it is more effective under some circumstances to use a p-type thermoelectric material for this purpose. It will be assumed that the electrical resistance R and the thermal conductance K in the metal are linked by the Wiedemann–Franz law. R and K have values that are appropriate for unit area of the electrodes.

Taking into account the losses due to the thermal radiation and to the resistance and thermal conductance in the metal, the overall cooling effect per unit area is

$$j = j_\mathrm{i} + w_\mathrm{r} - i^2R/2 - K(T_2 - T_1), \tag{13.11}$$

The electrical power input per unit area is

$$w = w_\mathrm{i} + i^2R. \tag{13.12}$$

Mahan [3] has discussed the problem of space charge in the space between the electrodes. We are dealing with high current densities and the electrons already in the vacuum space will, undoubtedly, set up a barrier to further flow. The space charge effect becomes less as the distance between the electrodes is reduced. Mahan showed that an inter-electrode distance of 1 mm is the upper limit for operation at 700 K. The space charge problem becomes greater as the operating temperature is reduced, and may present severe practical difficulties, if low temperature thermionic refrigerators are ever constructed.

Mahan showed that the limiting space charge voltage is equal to $2kT \ln(pd/\pi)$ where p is defined as $\left(4\pi e^2/kT\,\Lambda^3\right)^{1/2}$ and $\Lambda^2 = h^2/2\pi mkT$. He assumed that the lowest possible work function is 0.71 eV but, as we shall see, a value of about 0.3 eV is necessary for a thermionic refrigerator working at ordinary temperatures. This means that $\ln(pd/\pi)$ would differ by a factor of 4.2 compared with the value derived by Mahan. This leads to a maximum value for d of no more than 0.4 μm and to attain this would certainly be a formidable problem. However, it is assumed that it can be solved.

Nolas and Goldsmid [6] considered the requirements for a thermionic refrigerator to be competitive with ordinary Peltier devices at 300 K. They assumed that a metallic connector would be used for the return path, though Xuan [7] has pointed out the advantage of using a thermoelement as stated above. It was thought that the cooling power would have to be at least 10^4 W m^{-2}. Figure 13.3 shows the variation of the thermionic cooling power q_i with work function assuming the heat sink to be at 300 K and the source at 260 K. Also shown, as a broken curve, is the cooling power minus the radiation loss. It will be seen that it is virtually impossible to obtain a cooling effect at these temperatures if the work function is greater than 0.5 eV. In order to achieve a cooling of 10^4 W m^{-2}, the work function should not be greater than about 0.3 eV, and this value will be assumed in the following calculations.

The cooling power has its maximum value when the Joule heating and thermal conduction losses are equal to one another. Thus,

$$R = [L(T_2 - T_1)(T_2 + T_1)]^{1/2}/I, \tag{13.13}$$

where L is the Lorenz number. However, this does not optimise the COP. The losses in the passive conductor do not have much effect on the cooling power but they have more influence on the COP. This would be more noticeable for a thermionic refrigerator than for present-day thermoelectric refrigerators because of the improvement

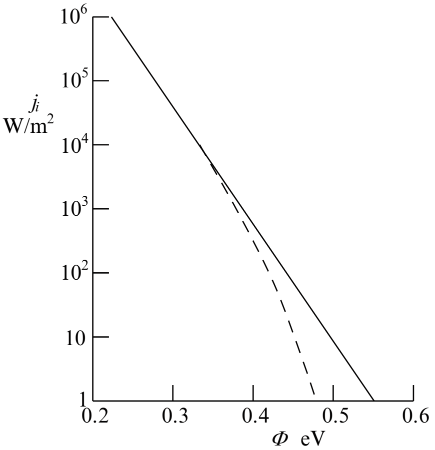


Fig. 13.3 Maximum cooling power of a thermionic diode against work function. The *solid line* gives values when there are no losses and the *broken curve* takes account of radiation. The source and sink are at 260 and 300 K, respectively

in efficiency. The modification to (13.13) that would be needed for optimum COP is somewhat complicated by the exponential terms in the relation between current and voltage.

A thermionic refrigerator is, probably, more suitable for improving the COP with small temperature differences between source and sink rather than for obtaining large vales of $\Delta T_{\max}$. Nevertheless, if the work function were as high as 0.3 eV, it should still be possible to reach a value of 100°C for $\Delta T_{\max}$, as shown in Fig. 13.4. In this diagram, $\Delta T_{\max}$ is plotted against Φ for a heat sink temperature of 300 K.

The COP for a thermionic refrigerator with Φ equal to 0.3 eV is plotted against the applied voltage in Fig. 13.5. The source and sink temperatures have again been set at 260 and 300 K, respectively. The cooling power for the same set of conditions is shown in Fig. 13.6.

The maximum in the plots of COP against voltage is due to the fact that there are radiation losses, small though they are. The radiation losses become significant only when the cooling power is extremely low. It is noteworthy that the cooling power remains above $10^4\,\mathrm{W\,m^{-2}}$ for a wide voltage range. The cut off in both

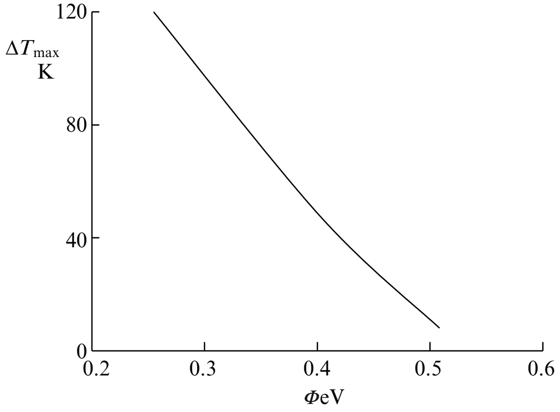


Fig. 13.4 Maximum temperature difference for a thermionic refrigerator plotted against the work function. The heat sink is at 300 K

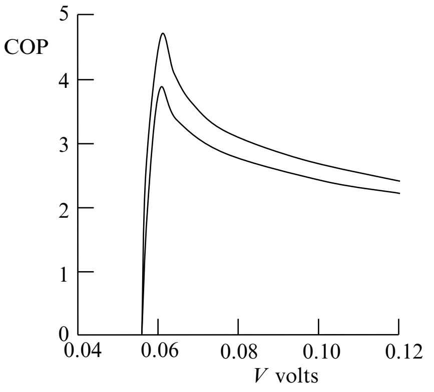
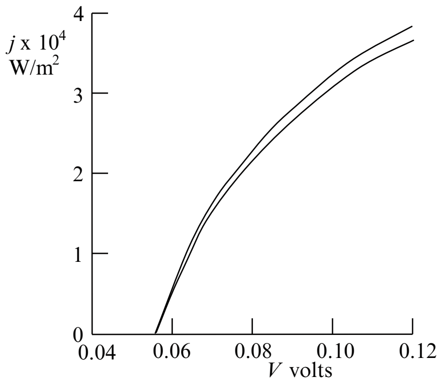


Fig. 13.5 Coefficient of performance as a function of voltage for a thermionic cooler with $\Phi = 0.3$. The source and sink are at 260 and 300 K, respectively. The upper curve is optimum COP and the lower for maximum cooling power

Fig. 13.6 Cooling power against voltage for conditions as in Fig. 13.5. The upper curve for maximum cooling power, the lower for optimum COP



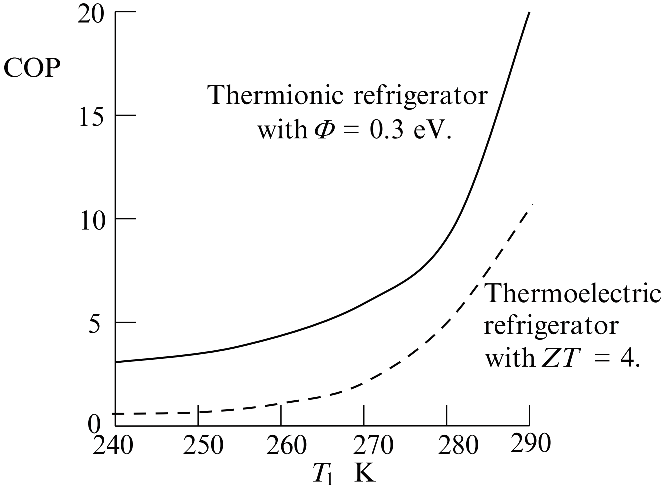


Fig. 13.7 Coefficient of performance plotted against source temperature for thermionic and thermoelectric refrigerators. The heat sink is at 300 K

the cooling power and the COP between 0.05 and 0.06 V occurs when the applied voltage is insufficient to compensate for the temperature difference between the source and sink.

The main objective in the study by Nolas and Goldsmid was to compare the performance of the projected thermionic refrigerator with that of an improved thermoelectric refrigerator that should be available in the future. It was supposed [8] that the projected thermoelectric cooler might be made from materials with ZT as high as 4. Figure 13.7 shows plots of COP for the thermionic and thermoelectric refrigerators against the temperature of the heat source. It is clear that the thermionic refrigerator is superior to the thermoelectric device but recent developments suggest that the latter is much closer to becoming available.

13.2 Thermionic Emission in Solids

The great attraction of vacuum thermionic energy convertors is the absence of heat conduction other than by the free electrons. However, there remains the question as to whether or not suitable electrode materials will become available. There is the alternative of making use of thermionic emission in solids. The difficulty of obtaining adequate emission currents can, then, be overcome but one no longer eliminates the conduction of heat through the medium in which the charge carriers move. Solid-state diodes have an advantage over vacuum devices in that both positive and negative carriers can be used, so the problems associated with the return path no longer exist.

The possibility of using the thermionic effects in solids instead of the thermoelectric effects has been discussed by Mahan et al. [9]. They first pointed out the basic difference between the two effects. In thermionic transport, the current flow is ballistic so that a carrier leaving one electrode arrives at the other with the same kinetic energy. On the other hand, in thermoelectric devices, the carriers undergo many collisions as they travel from one junction to the other, the motion then being described as diffusive. In some nanostructured configurations, the motion may be intermediate between ballistic and diffusive.

In a thermionic device, the applied voltage drives the more energetic electrons from the cold surface across a barrier to the hot surface. These electrons are replaced by others of lower energy through the external circuit. For the flow to be ballistic rather than diffusive, the mean free path, l_e, has to be greater than the barrier width, d. It is also necessary that the barrier width be greater than d_t, the maximum width for tunnelling to be possible. This width is given by the expression

$$d_t = \frac{\hbar}{2kT}\sqrt{\frac{e\Phi}{m^*}}, \tag{13.14}$$

where Φ is the barrier height. For most semiconductors, d_t is less than 10 nm and l_e is often greater than 100 nm so it is not difficult to satisfy the condition for ballistic flow.

In the absence of losses, the rate of cooling per unit area is still given by (13.8) and, allowing for heat conduction across the barrier,

$$j = i\frac{\Phi + 2kT_1}{e} - K\Delta T, \tag{13.15}$$

where K is the thermal conductance. It is obvious that the heat conduction loss is going to be very large, because of the small barrier width, unless ΔT is very small and, for this reason, Mahan et al. proposed the use of a multi-layer arrangement.

For a single layer, the theory is simplified by the fact that T_1 and T_2 are nearly equal. Thus, Richardson's equation is the same for both sides of the barrier,

$$i_{1,2} = A_0 T^2 \exp\left(-\frac{\Phi}{kT}\right), \tag{13.16}$$

and

$$i = \frac{eA_0T}{k}(V - V_0)\exp\left(-\frac{\Phi}{kT}\right), \tag{13.17}$$

where V is the applied voltage. V_0 is given by

$$V_0 = \frac{k\Delta T}{e}\left(\frac{\Phi}{kT} + 2\right). \tag{13.18}$$

In a semiconductor or metal, Φ can be small enough for Fermi–Dirac statistics to be necessary but, for the optimum COP in a device that is better than existing thermocouples, classical statistics are satisfactory.

The cooling power per unit area is

$$j = i_1\left(\frac{\Phi}{kT} + 2\right)(V - V_j), \tag{13.19}$$

where

$$V_j = \frac{k\Delta T}{e}\left(\frac{\Phi}{kT} + 2 + c\right). \tag{13.20}$$

The quantity c is defined as

$$c = \frac{2 + eK/ki_1}{2 + \Phi/kT}. \tag{13.21}$$

Thence the COP is

$$\phi = \frac{j}{w} = \frac{kT}{e}\left(\frac{\Phi}{kT} + 2\right)\frac{(V - V_j)}{V(V - V_0)}. \tag{13.22}$$

The COP has its maximum value when

$$V = V_j + \sqrt{V_j(V_j - V_0)}, \tag{13.23}$$

and then

$$\phi_{\max} = \frac{T}{\Delta T}\frac{\left(\frac{\Phi}{kT} + 2\right)}{\left(\sqrt{2 + \Phi/kT + c} + \sqrt{c}\right)^2}. \tag{13.24}$$

It is convenient to introduce a parameter T_{R}, with the dimensions of temperature, that is defined by

$$(kT_{\mathrm{R}})^2 = \frac{2\pi^2\hbar^3K}{m^*kT}. \tag{13.25}$$

This parameter is proportional to the square root of the thermal conductance of the barrier and should be as small as possible. In Fig. 13.8, the COP, divided by its

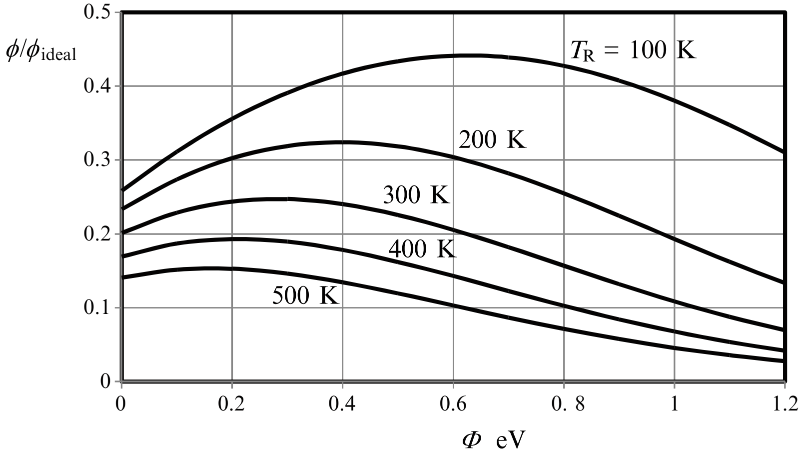
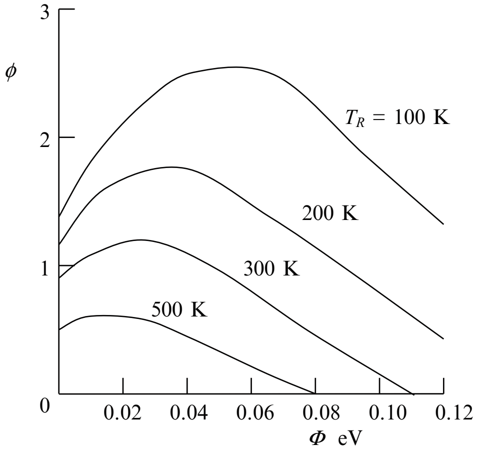


Fig. 13.8 Ratio of coefficient of performance to its value for a Carnot cycle plotted against barrier height. The parameter T_R represents the losses and is proportional to the square root of the thermal conductance of the barrier

Fig. 13.9 Coefficient of performance for a multi-layer thermionic refrigerator. A schematic plot based on the data of Mahan et al. [9]. The source is at 260 K and the sink at 300 K. Φ is the barrier height



value for a Carnot cycle, is plotted against the barrier height for selected values of T_R between 100 and 500 K. The optimum barrier height should be somewhat greater than kT.

Although the COP might be reasonably high for a single-barrier device, cooling across a reasonably large temperature difference requires a multi-layer arrangement. In other words, a practical device would operate as a multi-stage cascade. Mahan et al. have optimised the conditions for a multi-layer thermionic refrigerator and their results are shown in Fig. 13.9. In this diagram, the COP is plotted against the barrier height, for source and sink temperatures of 260 and 300 K. A thermoelectric refrigerator with ZT equal to unity would have a COP of 0.7 under these conditions.

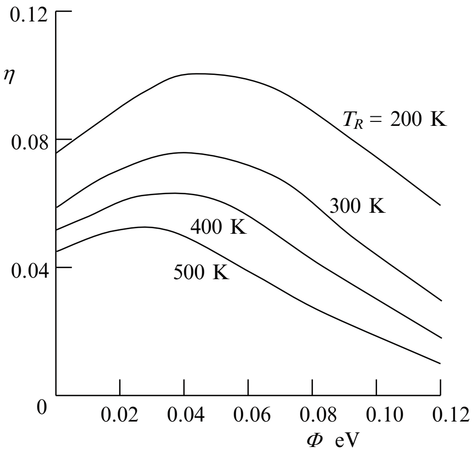


Fig. 13.10 Efficiency of a multi-layer thermionic generator plotted against the barrier height. A schematic plot based on the data of Mahan et al. [9]. The source is at 400 K and the sink at 300 K

The multi-layer thermionic device would give a similar performance, if T_R were equal to about 500 K. For the thermionic refrigerator to be superior, it would be necessary for T_R to be less than 500 K.

Mahan et al. [9] also considered solid-state thermionic devices as generators of electricity. The factor $(\Phi/kT + 2)\,[(\Phi/kT + 2 + c)^{1/2} + c^{1/2}]^{-2}$ that appears in (13.24) for the COP of a thermionic refrigerator is found in the expression for the efficiency of a single-barrier thermionic generator, in this case multiplied by $\Delta T/T$. A single barrier would generate a very small voltage and would, undoubtedly, be replaced by a multi-barrier system. The variation of the efficiency of a multi-layer generator with barrier height bears some resemblance to that of the COP for a multi-layer refrigerator, as shown in Fig. 13.10, except, of course, that the efficiency never falls to zero.

Mahan et al. pointed out that the equations for thermionic generators and refrigerators have the same form as those for thermoelectric energy convertors if the Seebeck coefficient, electrical conductivity, and thermal conductivity are replaced by three quantities, α_I, σ_I, and λ_I that are defined by

$$\alpha_I = \frac{k}{e}\,(\Phi/kT + 2)\,, \tag{13.26}$$

$$\sigma_I = \frac{e i_1 d}{kT}, \tag{13.27}$$

and

$$\lambda_I = \left(2\frac{k}{e}i_1 + K\right)d. \tag{13.28}$$

It is also possible to define a figure of merit for thermionic devices as

$$Z_I T = \frac{\Phi/kT + 2}{c}. \tag{13.29}$$

Vining and Mahan [10] have drawn attention to an analogy between thermionic and thermoelectric devices. They showed that, for thermionic materials, one can use a parameter $\beta_{\rm I}$ that plays more or less the same role as the parameter β that was discussed for thermoelectric materials in Chap. 4. The quantity $\beta_{\rm I}$ is defined as

$$\beta_{\rm I} = \frac{m^* k (kT)^2 d}{2\pi^2 \hbar^3 \lambda_{\rm L}}. \tag{13.30}$$

Then

$$Z_{\rm I}T = \frac{(\Phi/kT + 2)^2}{2 + [\exp(\Phi/kT)]/\beta_{\rm I}}. \tag{13.31}$$

Equation (13.31) is similar to the equation for zT in a thermoelectric material. A thermionic device will be the superior if $\beta_{\rm I} > \beta$.

The ratio of $\beta_{\rm I}$ to β is equal to $d/\left(l_{\rm e}\pi^{1/2}\right)$, but it has already been assumed that $d < l_{\rm e}$, if the device is to operate in the thermionic mode. Thus, as shown by Vining and Mahan, we might expect the thermoelectric mode to be the better. However, it is possible that $\lambda_{\rm L}$ will be much less for the thermionic structure and this may be sufficient compensation for any loss in the power factor.

There have, in fact, been experimental demonstrations of thermionic refrigeration in solids. For example, Shakouri et al. [11] made use of a (In–Ga–As)P barrier of 1 μm thickness between layers of n^+ (In–Ga)As. Although a cooling effect of only 0.5°C was observed in the preliminary experiment, it was expected that this would rise to about 10°C after improvements in the design.

More recently, it has been suggested that tunnelling effects might also be used in solid-state refrigeration. A device proposed by Chao et al. [12] takes advantage of both tunnelling and ballistic effects. Its operation is described with reference to Fig. 13.11.

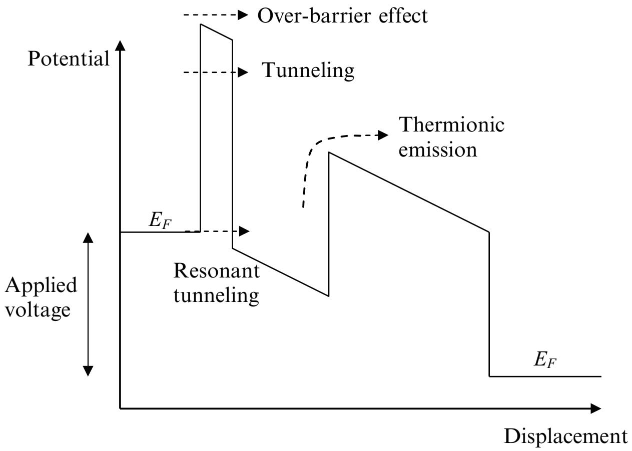


Fig. 13.11 Schematic potential diagram for a heterojunction refrigerator

The contribution from thermionic emission has already been described. The additional feature in the new device is a barrier that is thin enough to allow tunnelling. The predominant cooling is associated with resonant tunnelling. Chao et al. based their treatment on (Al–Ga)As heterostructures. They admitted that they are not good thermoelectric materials but are well characterised and allow accurate predictions to be made. In the embodiment of the device, the material at either end is n-type GaAs. Tunnelling takes place through $\text{Al}_y\text{Ga}_{1-y}\text{As}$ into GaAs and the carriers are emitted into $\text{Al}_x\text{Ga}_{1-x}\text{As}$. The calculations indicate that a temperature depression of between 5 and 7°C might be achieved using this structure. Although this is much less than the value for conventional devices made from bismuth telluride, it compares favourably with what could be obtained using GaAs as the thermoelectric material. Moreover, the structure could readily be produced using existing technology. Solid-state refrigerators based on ballistic rather than diffusive transport may well offer a viable alternative to existing cooling techniques, particularly, when integrated with semiconductor junction devices.

References

1. A.F. Ioffe, *Semiconductor Thermoelements and Thermoelectric Cooling* (Infosearch, London, 1957), p. 72
2. G.N. Hatsopoulos, J. Kaye, J. Appl. Phys. **29**, 1124 (1958)
3. G.D. Mahan, J. Appl. Phys. **76**, 4362 (1994)
4. J.L. Dye, Chemtracts: Inorg. Chem. **5**, 243 (1993)
5. R.H. Huang, J.L. Dye, Chem. Phys. Lett. **166**, 133 (1990)
6. G.S. Nolas, H.J. Goldsmid, J. Appl. Phys. **85**, 4066 (1999)
7. X.C. Xuan, J. Appl. Phys. **92**, 4746 (2002)
8. G.A. Slack, in *CRC Handbook of Thermoelectrics*, ed. by D.M. Rowe (CRC Press, Boca Raton, 1994), p. 407
9. G.D. Mahan, J.O. Sofo, M. Bartkowiak, J. Appl. Phys. **83**, 4683 (1998)
10. C.B. Vining, G.D. Mahan, Appl. Phys. Lett. **86**, 6852 (1999)
11. A. Shakouri, C. LaBounty, J. Piprek, P. Abraham, J.E. Bowers, Appl. Phys. Lett. **74**, 88 (1999)
12. K.A. Chao, M. Larsson, A.G. Mal'shukov, Appl. Phys. Lett. **87**, 022103 (2005)

Bibliography

1. L.I. Anatychuk, *Physics of Thermoelectricity* (Institute of Thermoelectricity, Kiev, 1998)
2. L.I. Anatychuk, *Thermoelectric Power Convertors* (Institute of Thermoelectricity, Kiev, 2005)
3. Albert C. Beer, *Galvanomagnetic Effects in Semiconductors* (Academic, New York, 1963)
4. C.M. Bhandari, D.M. Rowe, *Thermal Conduction in Semiconductors* (Wiley Eastern, New Delhi, 1988)
5. J.S. Blakemore, *Solid State Physics,* 2nd edn. (Cambrige University Press, Cambridge, 1985)
6. J.R. Drabble, H.J. Goldsmid, *Thermal Conduction in Semiconductors* (Pergamon, Oxford, 1961)
7. H.J. Goldsmid, *Applications of Thermoelecticity* (Methuen, London, 1960)
8. H.J. Goldsmid, *Thermoelectric Refrigeration* (Plenum, New York, 1964)
9. H.J. Goldsmid, *Electronic Refrigeration* (Pion, London, 1986)
10. T.C. Harman, J.M. Honig, *Thermoelectric and Thermomagnetic Effects and Applications* (McGraw-Hill, New York, 1967)
11. A.F. Ioffe, *Semiconductor Thermoelements and Thermoelectric Cooling* (Infosearch, London, 1957)
12. A.F. Ioffe, *Physics of Semiconductors* (Infosearch, London, 1960)
13. C. Kittel, *Introduction to Solid State Physics*, 5th edn. (Wiley, New York, 1976)
14. John P. McKelvey, *Solid State and Semiconductor Physics* (Harper and Row, New York, 1966)
15. B.R. Nag, *Electron Transport in Compound Semiconductors* (Springer, Berlin, 1980)
16. G.S. Nolas, J. Sharp, H.J. Goldsmid, *Thermoelectrics. Basic Principles and New Materials Developments* (Springer, Berlin, 2001)
17. E.H. Putley, *The Hall Effect and Related Phenomena* (Butterworths, London, 1960)
18. D.M. Rowe (ed.), *CRC Handbook of Thermoelectrics* (CRC, Boca Raton, 1994)
19. D.M. Rowe (ed.), *Thermoelectrics Handbook. Macro to Nano* (CRC Taylor and Francis, Boca Raton, 2006)
20. D.M. Rowe, C.M. Bhandari, *Modern Thermoelectrics* (Holt, Rinehart and Winston, London, 1983)
21. G.A. Slack, in *The Thermal Conductivity of Nonmetallic Crystals*. ed. by H. Ehrenreich, F. Seitz, D. Turnbull. Solid State Physics, vol 34 (Academic, New York, 1979), pp. 1–71
22. Arthur C. Smith, James F. Janak, Richard B. Adler, *Electronic Conduction in Solids* (McGraw-Hill, New York, 1967)
23. Terry M. Tritt (ed.), *Semiconductors and Semimetals*, vol 69. Recent Trends in Thermoelectric Materials Research I (Academic, San Diego, 2001)
24. Terry. M. Tritt (ed.), *Thermal Conductivity. Theory, Properties and Applications* (Kluwer, New York, 2004)
25. Terry M. Tritt, Mercouri G. Kanatzidis, Gerald D. Mahan, Hylan B. Lyon Jr., *Thermoelectric Materials 1998 – The Next Generation Materials for Small-Scale Refrigeration and Power Generation Applications* (Materials Research Society: Warrendale, Pennsylvania, 1999)

26. L.M. Tsidil'kovskii, *Thermomagnetic Effects in Semiconductors* (Infosearch, London, 1962)
27. J.M. Ziman, *Electrons and Phonons* (Clarendon, Oxford, 1960)
28. J.M. Ziman, *Theory of Solids* (Cambridge University Press, Cambridge, 1964)

Index

Index of Elements, Compounds and Alloys