THERMIONIC ENERGY CONVERSION

Volume II: Theory, Technology, and Application

G. N. Hatsopoulos
and E. P. Gyftopoulos

Prepared under the auspices of the
U.S. Department of Energy Division
of Power Systems

Thermionic Energy Conversion

Volume II: Theory, Technology, and Application

Thermionic Energy Conversion

Volume II: Theory, Technology, and Application

G. N. Hatsopoulos and
E. P. Gyftopoulos

Prepared under the auspices of the
U. S. Department of Energy
Division of Power Systems

MASTER

DISCLAIMER

This book was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

The MIT Press
Cambridge, Massachusetts, and London, England

Release for Announcement in
Energy Research Abstracts

Copyright © 1979 by
The Secretary of the United States Department of Energy

All rights reserved. No part of this book may be reproduced
in any form or by any means, electronic or mechanical, including
photocopying, recording, or by any information storage
and retrieval system, without permission in writing.

All royalties from the sale of this book accrue to the
United States Government.

Printed and bound in the United States of America.

Library of Congress Cataloging in Publication Data

Hatsopoulos, George N
Thermionic energy conversion.

Vol. 2 prepared under the auspices of the U. S. Dept.
of Energy, Division of Power Systems.
Includes bibliographical references.
CONTENTS: v.1. Processes and devices.
–v.2. Theory, technology, and application.
1. Thermionic converters. I. Gyftopoulos, E. P.,
joint author. II. Title.
TK2955.H37 621.312'139 72-13718
ISBN 0-262-08059-1

Contents

Preface, Volume II xvii

Preface, Volume I xviii

List of Contributors xxii

MASTER

5 Elements of Quantum Mechanics

5.1 General Remarks 1

5.2 Classical Vs. Quantum Mechanics 2
5.2.1 Discrete Energy Levels 2
5.2.2 Wave–Particle Duality 3
5.2.3 Indeterminacy, Uncertainty, Complementarity, and Correspondence 4

5.3 Postulates of Quantum Mechanics 5
5.3.1 General Remarks 5
5.3.2 The State and the Equation of Motion in Classical Mechanics 5
5.3.3 The State in Quantum Mechanics: Postulate 1 7
5.3.4 The Relation Between Probability Amplitudes: Postulate 2 9
5.3.5 Operators Corresponding to Observable Quantities: Postulate 3 10
5.3.6 The Schroedinger Equation for Change in State with Time: Postulate 4 14

5.4 Some Theorems of Quantum Mechanics 15
5.4.1 Conservation of Total Probability 15
5.4.2 Probability Current 15
5.4.3 The Principle of Superposition 16
5.4.4 Stationary States: Energy Eigenvalues and Eigenfunctions 17
5.4.5 Eigenstates in General 18
5.4.6 Expansions in Terms of Eigenfunctions 19
5.4.7 Conservation of Energy in Quantum Mechanics 20

5.5 One-Dimensional Motion of a Free Particle 22
5.5.1 General Formulation 22
5.5.2 Transmission and Reflection at a Barrier 24
5.5.3 Transmission and Reflection Through a Rectangular Barrier 27
5.5.4 The Rectangular Potential Well 29

5.6 Particle in a Container 32

5.7 Two Interacting Particles 37
5.7.1 General Remarks 37
5.7.2 Electrostatic or Central-Force Interaction Potential Energy 40
5.7.3 Electron Spin 42
5.7.4 The Hydrogen Atom 43

5.8 Symmetry Considerations 47
5.8.1 Symmetric and Antisymmetric Solutions for Pairs of Identical Particles 47
5.8.2 Symmetric and Antisymmetric Solutions for Systems with Many Identical Particles 50
5.8.3 Pauli Exclusion Principle 51
5.8.4 Atoms with Many Electrons—The Self-Consistent-Field Method 53

5.9 Diatomic Molecules 56
5.9.1 The Energy Eigenvalue Equation for a Diatomic Molecule 57
5.9.2 The Born–Oppenheimer Approximation for the Nuclear and Electron Motions 57
5.9.3 The Hydrogen Molecular Ion, H_2^+ 58
5.9.4 Linear Combination of Atomic Orbitals and the Overlap Charge for H_2^+ 63
5.9.5 The Hydrogen Molecule, H_2 67
5.9.6 The Hartree–Fock Equation and the Molecular-Orbital Method 71
5.9.7 The Valence-Bond Method 72
5.9.8 Heteronuclear Molecules 73
5.9.9 Atomic Radii 74

5.10 Electrons in Metals 75
5.10.1 General Remarks 75
5.10.2 The Crystal Lattice 76
5.10.3 The Energy Eigenvalue Equation for the Electronic Structure 80
5.10.4 One-Electron Approximations 82
5.10.5 Periodic Boundary Conditions 82
5.10.6 The Free-Electron Gas Model 83
5.10.7 The Theory of Energy Bands in Crystals 84
5.10.8 Charge Distribution and Potential Energy for the Self-Consistent Field in a Crystal Lattice 84
5.10.9 The Reciprocal Lattice—Brillouin Zones 86
5.10.10 Fourier Expansions of Periodic Functions 88
5.10.11 The Eigenvalue Equation for a Periodic-Potential Problem 89
5.10.12 Energy Bands in the One-Dimensional Lattice 91
5.10.13 Energy Bands in Three-Dimensional Lattices—The Fermi Surface 94
5.10.14 Discussion of One-Electron Approximations 97

References 100

6 Principles of Thermodynamics 102

6.1 General Remarks 102

6.2 Preliminary Concepts 104
6.2.1 Systems, States, and Properties 104
6.2.2 Work and Nonwork 105

6.3 First Law 106
6.3.1 Uniqueness of Work Values 106
6.3.2 Definition of Property Energy 106
6.3.3 Additivity of Energy 106

6.3.4 Conservation of Energy 107
6.3.5 Impossibility of a Perpetual-Motion Machine of the First Kind 107
6.3.6 The First Law and Relativity 107

6.4 Types of States 107

6.5 Second Law 108
6.5.1 State Principle 108
6.5.2 Reversible and Irreversible Processes 109
6.5.3 Impossibility of a Perpetual-Motion Machine of the Second Kind 109
6.5.4 Mutual Stable Equilibrium and Reservoirs 109
6.5.5 Work Done Reversibly by a System in Combination with a Reservoir 110

6.6 Entropy 111
6.6.1 Available Work 111
6.6.2 Definition of Property Entropy 111
6.6.3 Entropy Changes in Adiabatic Processes 112
6.6.4 Principle of Increase of Entropy 112

6.7 Concepts Associated with Stable Equilibrium 112
6.7.1 Criterion for Stable Equilibrium 112
6.7.2 Equations Relating Properties for Stable Equilibrium States 113
6.7.3 Temperature 113
6.7.4 Total Potentials 114
6.7.5 Generalized Body Forces 114
6.7.6 Heat 114
6.7.7 Heat and Flow of Entropy 116

6.8 Work-Producing Systems and Devices 117
6.8.1 Sources of Available Work 117
6.8.2 Heat Engines 118

6.9 Statistical Thermodynamics 119
6.9.1 Brief Review 119
6.9.2 Some Concepts from Gibbsian Statistics 120

6.10 Quantum Thermodynamics 121
6.10.1 Operator and Matrix Representation of Properties 121
6.10.2 State in the Language of Quantum Mechanics 122
6.10.3 Values of Properties 124
6.10.4 State in Quantum Mechanics and in Quantum Thermodynamics 125
6.10.5 Reversible Equation of Motion 125
6.10.6 General Expression for Entropy 126
6.10.7 Stable Equilibrium States 127

6.11 Generalizations to All Systems 128

7
Thermodynamics of Stable Equilibrium States 130

7.1 General Remarks 130

7.2 Ordinary Closed Systems 130
7.2.1 Equations of Stable Equilibrium States 130
7.2.2 Quasistable States 131
7.2.3 A Condition for Mutual Stable Equilibrium 132
7.2.4 Thermodynamic Temperature Scales 134
7.2.5 The Kelvin Temperature Scale 134
7.2.6 Analytical Descriptions of Work and Heat Interactions 135
7.2.7 Reversible Work and Heat in the Language of Quantum Mechanics 137
7.2.8 Work Plus Heat in a General Interaction 138
7.2.9 Explicit Forms of Equations of Stable Equilibrium States 139
7.2.10 Simple Systems 139

7.3 Applications 141
7.3.1 Efficiency of a Thermionic Converter—Carnot Efficiency 141
7.3.2 Maximum Useful Work 142
7.3.3 Equations of Stable Equilibrium States for One Structureless Particle 144
7.3.4 The Maxwellian Velocity Distribution 145
7.3.5 The One-Dimensional Harmonic Oscillator 146

7.4 Ordinary Open Systems 147
7.4.1 General Remarks 147
7.4.2 The Generalized Gibbs Equation 149
7.4.3 Conditions for Mutual Stable Equilibrium 150

7.5 One Atom as an Ordinary Open System 151

7.6 Homogeneous Systems with Volume as an Independent Constraint 154
7.6.1 General Remarks 154
7.6.2 Extensive and Intensive Properties 154
7.6.3 Homogeneous and Heterogeneous Systems 155
7.6.4 The Gibbs Equation for Systems with Two Constraints 156
7.6.5 Internal Energy—Chemical Potential 158
7.6.6 The T–p Equations 159
7.6.7 Partial Properties 160
7.6.8 Partial Gibbs Free Energy 162
7.6.9 The Partition Function 163
7.6.10 The Gibbs–Duhem Equation 164

7.7 On Total Potentials 165
7.7.1 General Remarks 165
7.7.2 Total Potentials and Partial Properties 165
7.7.3 Total Potentials in Terms of Specific Properties 166
7.7.4 Total Potentials in Terms of Work 167
7.7.5 Qualitative Interpretation 167

7.8 One-Particle Approximations 168
7.8.1 General Remarks 168
7.8.2 The One-Particle Approximation for One-Component Open Systems 168
7.8.3 The Partition Function 171
7.8.4 The Expectation Value of One-Particle Distribution ν_j 172
7.8.5 The p–V–T–β–μ Relation for a Dilute Gas 172

7.8.6 A Fundamental Equation for a One-Component Dilute Gas 173
7.8.7 The Boltzmann Approximations 174
7.8.8 The Constant-Potential Approximation 175
7.8.9 The Point-Particle and Continuous Eigenvalue-Spectrum Approximations 176
7.8.10 The Sommerfeld Gas 177
7.8.11 The Semiperfect Gas 178
7.8.12 Specific Heats 180
7.8.13 The Perfect Gas 181

7.9 Mixtures of Gases 183
7.9.1 General Remarks 183
7.9.2 The Gibbs–Dalton Law 183

7.10 Mixtures of Ionized Gases 185
7.10.1 The Saha Equation 185
7.10.2 Mutual Stable Equilibrium of a Mixture of Ionized Gases with a Solid 188
7.10.3 Random-Flux Densities 189

7.11 Surfaces 190
7.11.1 General Remarks 190
7.11.2 Basic Relations 190
7.11.3 Dilute Surfaces 191
7.11.4 Chemical Potential of Nondilute Surfaces 192

7.12 Lattice Vibrations 193
7.12.1 General Remarks 193
7.12.2 Partition Function 194
7.12.3 Gibbs Free Energy 197

References 197

8

Thermodynamics of Steady States 198

8.1 General Remarks 198

8.2 Steady-Rate Processes 199
8.2.1 Steady States 199
8.2.2 Fluxes and Flux Densities 199
8.2.3 Fluxes and Thermodynamic Potentials 200
8.2.4 Alternate Form of Thermodynamic Potentials 201
8.2.5 Entropy Generation 202

8.3 Boundary-Value Processes 203
8.3.1 Definition 203
8.3.2 Linear and Nonlinear Processes 204

8.4 One-Dimensional Rate Processes 205
8.4.1 Continuous Rate Processes 205
8.4.2 Entropy Generation in an Elementary Process 206
8.4.3 Conjugate Fluxes and Forces 207

8.4.4 Flux Densities as Functions of a State and Forces 207
8.4.5 Onsager Reciprocal Relations 209

8.5 Applications 209
8.5.1 General Remarks 209
8.5.2 Flow of a Substance Through a Porous Barrier 209

8.6 Thermoelectric Effects 215
8.6.1 General Remarks 215
8.6.2 One-Dimensional Electron Flow 215
8.6.3 Phenomenological Equations for a Homogeneous Conductor 217
8.6.4 Special Case 1: Heat Conduction for $J_1 = 0$ 218
8.6.5 Special Case 2: Isothermal Electron Flow for $\tau' = 0$ 218
8.6.6 Electric-Conduction Equation 220

8.7 Chemical Reactions 221
8.7.1 General Remarks 221
8.7.2 Three-Component Three-Temperature Ionized Gases 221
8.7.3 Rate of Entropy Generation 223
8.7.4 Uncoupling of Spatial and Local Phenomena 225
8.7.5 Spatial Phenomenological Equations 226
8.7.6 Local Phenomenological Equations 228
8.7.7 Approximate Calculation of Chemical Potentials 228

8.8 Kinetic Theory of Semiperfect Gases 232
8.8.1 General Remarks 232
8.8.2 General Definition of Semiperfect Gases 233
8.8.3 The Boltzmann Equation 234
8.8.4 Discussion of the Boltzmann Equation 239
8.8.5 Steady-State Boltzmann Equations for a Mixture of Ionized Gases 240
8.8.6 Hydrodynamic Equations 242
8.8.7 Solutions of Linearized Boltzmann Equations 244

References 248

9
Emission Phenomena 249

9.1 General Remarks 249

9.2 Electron Emission from Pure Uniform Surfaces 252
9.2.1 General Remarks 252
9.2.2 Electron-Random-Current Density 254
9.2.3 Electron-Saturation-Current Density 263
9.2.4 Velocity Distribution 266
9.2.5 Kinetic Energy Flux 266
9.2.6 Schottky Effect 266
9.2.7 Apparent Emission Constant 269
9.2.8 Experimental Results 270

9.3 Calculating the Work Function of a Pure Uniform Surface 270
9.3.1 General Remarks 270
9.3.2 Rigorous Program for Calculating ϕ 271
9.3.3 Interfacial Electron Motive ψ^*—The Double Layer 274
9.3.4 Electrochemical Potential μ_e^* 278
9.3.5 Wigner–Bardeen Analysis of Total Potentials 279
9.3.6 Calculating ϕ^* 281
9.3.7 A Relation Between ϕ and Other Thermophysical Properties 283
9.3.8 Effect of Surface Structure on ϕ^* 289
9.3.9 An Empirical Correlation for ϕ^* 291
9.3.10 Temperature Coefficient α 295

9.4 Electron Emission from a Pure Nonuniform Surface 301
9.4.1 General Remarks 301
9.4.2 Effects of Patch Fields on the Motive 303
9.4.3 Effects of Patch Fields on Electron Emission 305
9.4.4 Effects of Patch Fields on the Velocity Distribution of Emitted Electrons 308

9.5 Composite Surfaces—Definitions 310

9.6 Emission Characteristics of Electropositive Surfaces 311
9.6.1 General Remarks 311
9.6.2 Model for Moderate Coverages ($\theta < 0.5$) 315
9.6.3 Equations for Stable Equilibrium States 319
9.6.4 Equations for Steady States 329
9.6.5 Comparing Experimental and Theoretical Results 333
9.6.6 Nonuniform Electropositive Surfaces 334

9.7 Electronegative Surfaces and Mixed Surfaces 334

Appendix 9.A Emission Properties 335
9.A.1 Calculating Work Functions as Functions of Temperature 335
9.A.2 Equivalence of Multivalent and Monovalent Atoms in a Metal 337
9.A.3 Comparing Various Emission Rates from Refractory Metal Surfaces 339

Appendix 9.B Calculating Enthalpy of Desorption 339

References 341

10

Collisionless Transport Phenomena 345

10.1 General Remarks 345

10.2 The Child–Langmuir Theory 346

10.3 The Langmuir Theory 350
10.3.1 Interelectrode Motive 350
10.3.2 Output-Current Characteristics of Vacuum Diodes 357

10.4 Electron-Space-Charge Theory in the Presence of Back Emission 360

10.5 Space-Charge Theories in the Presence of Both Electrons and Positive Ions 363
10.5.1 Planar Electron and Ion Sources 363
10.5.2 Planar Electron and Ion Sources at Identical Positions—Half-Maxwellian Distributions of Initial Velocities 371
10.5.3 On Limiting Values of β for Negative-Space-Charge Neutralization 378
10.5.4 Planar Electron and Ion Sources at Two Positions with Half-Maxwellian Distributions of Velocities 379

References 382

11

Collisional Transport Phenomena 383

11.1 General Remarks 383

11.2 Stable Equilibrium States of the Cesium Vapor System 385

11.3 Collisional Phenomena 390
11.3.1 General Remarks 390
11.3.2 Dynamics of Elastic Binary Collisions 391
11.3.3 Cross Section, Free Path, and Collision Frequency 394
11.3.4 Hard-Sphere Elastic Collisions 396
11.3.5 Coulomb Scattering 396
11.3.6 Experimental and Computed Cesium Cross Sections 401
11.3.7 Average Values, Relaxation Times, and Mean Free Paths 402

11.4 General Formulation of Steady-State Transport Phenomena 405

11.5 Solutions of Linearized Boltzmann Equations 409
11.5.1 Linearized Boltzmann Equations 409
11.5.2 Sonine Polynomial Expansion for Distribution Functions 412
11.5.3 Transport Differential Equations 414
11.5.4 Evaluation of Transport Coefficients 416

11.6 Range of Validity of Linear Approximations for Cesium Vapor Systems 421
11.6.1 Conditions on T_e' and μ_e' 421
11.6.2 Failure of Linear Approximations in Sheath Regions 422
11.6.3 Method for Preliminary Sheath Analysis—Plasma Boundary Conditions 424
11.6.4 Validity of Linear Approximations in the Plasma Region 427

11.7 Plasma Boundary Conditions—Monotonic Sheath 428
11.7.1 Boundary Conditions for Small Sheath Heights 428
11.7.2 Distribution Functions at Sheath–Plasma Interfaces 430

11.7.3 Effects of Barriers on Non-Maxwellian Distributions 433
11.7.4 Boundary Conditions for Large Sheath Heights and Strong Interfacial Electrostatic Fields 434
11.8 Double-Valued Sheaths 437
11.8.1 Criterion for the Formation of Double-Valued Sheaths 437
11.8.2 Fundamental Difficulty with Analyses of Collisionless Sheaths 438
11.9 Program for Collisional Analyses of Sheaths 439
11.9.1 General Remarks 439
11.9.2 General Expression for Distribution Functions 440
11.9.3 Collisional Sheath Analysis 443
References 445

12
Analysis of Performance Characteristics of High-Pressure Diodes 447
12.1 General Remarks 447
12.2 Analytical Model 448
12.3 Dimensionless Model 452
12.4 Review of Plasma and Performance Analyses 453
12.5 Numerical Results 454
References 470

13
Experimental Techniques 472
13.1 General Remarks 472
13.2 Characterization of Materials 472
13.3 Characterization of Bulk 473
13.3.1 Determining Bulk Composition 473
13.3.2 Determining Bulk Structure 476
13.4 Characterization of Surfaces 476
13.4.1 Mass Spectrometry 478
13.4.2 Auger Emission 480
13.4.3 Optical and Electron Microscopy 480
13.4.4 Profilometry 481
13.4.5 X-Ray Diffraction 481
13.4.6 Low-Energy Electron Diffraction 481
13.4.7 Etch-Pit Analysis 482
13.4.8 Light-Reflection Goniometry 482
13.5 Analytical Relations for Electron-Emission Properties 483

13.5.1 General Remarks 483
13.5.2 Analysis of Electron Space-Charge-Limited Measurements 483
13.5.3 Analysis of Electron-Emission Measurements in the Accelerating Range 486
13.5.4 Determining Work Functions 487
13.5.5 Determining Work Functions of Cesiated Emitter Surfaces 488

13.6 Measurements of Electron Emission and Collection Properties 488
13.6.1 Diode Arrangement 489
13.6.2 Triode Arrangement 491
13.6.3 Emission Microscopy 492
13.6.4 Plasma Anode (Collector) Device 493
13.6.5 Electron Beam Collector 494
13.6.6 Field-Emission Retarding-Potential Device 494
13.6.7 Kelvin Contact-Potential Detector 495

13.7 Temperature Measurements by Means of Thermocouples 495
13.7.1 Principle of Operation 495
13.7.2 Practical Considerations 496
13.7.3 Recommended Procedures 500

13.8 Optical Pyrometry 501
13.8.1 General Remarks 501
13.8.2 Vignetting 501
13.8.3 Optical Resolution 503
13.8.4 Blackbody Cavity 503
13.8.5 Calibration of Optical Pyrometers 506

13.9 Voltage and Current Measurements 509
13.9.1 Electrical Connections 509
13.9.2 Guard Rings 510
13.9.3 Transient Tests 510
13.9.4 Instruments for Transient Tests 511
13.9.5 Computer-Controlled Data Acquisition 512

References 515

14

Experimental Studies and Correlations of Characteristics 517

14.1 General Remarks 517

14.2 Characteristics 518
14.2.1 $J-V$ Characteristics for Low Values of pd ($pd < 1$ mil-torr) 519
14.2.2 $J-V$ Characteristics for High Values of pd ($pd > 1$ mil-torr) 526

14.3 Correlations 529
14.3.1 Emitter-Work-Function Correlations 529
14.3.2 Collector-Work-Function Correlations 533
14.3.3 Effect of Collector Temperature 534
14.3.4 Voltage-Loss Correlations 540
14.3.5 Correlation of $J-V$ Characteristics with Parameters β and pd 546

14.4 Use of Electrical Characteristics in Design and Diagnostics 547
14.4.1 Optimization of Converter Variables 547
14.4.2 Performance Comparisons 556
14.4.3 Converter Diagnostics 559
References 569

15
Metallurgy of Electrode Materials 572
15.1 General Remarks 572
15.2 Structure of Perfect Crystals 572
15.3 Directions, Planes, and Orientations in Crystals 573
15.4 Structure of Real Metals 575
15.4.1 Impurities 576
15.4.2 Vacancies 577
15.4.3 Grain Boundaries 577
15.4.4 Dislocations 578
15.5 Refractory Metals for Electrodes 579
15.6 Electrode Surfaces 582
15.6.1 Surface Free Energy 582
15.6.2 Effects of Heating on Metal Surfaces 584
15.7 Metallurgical Preparation of Electrode Surfaces 586
References 590

16
Design and Fabrication of Practical Converters 591
16.1 General Remarks 591
16.2 Functional Requirements on Converters 591
16.2.1 General Requirements 591
16.2.2 Specific Requirements 592
16.3 Component Design 593
16.3.1 Emitter Electrode 593
16.3.2 Emitter Support 594
16.3.3 Insulating Seal 597
16.3.4 Cesium Reservoir 598
16.4 Materials 599
16.4.1 Thermophysical Properties 599
16.4.2 Mechanical Considerations 603
16.4.3 Chemical Considerations 603
16.4.4 Metallurgical Considerations 604

16.5 Fabrication Processes 606
16.5.1 Material-Forming Processes 606
16.5.2 Cleaning and Handling Processes 607
16.5.3 Joining Processes 608
16.5.4 Converter Outgassing Processes 611
16.5.5 Cesium-Charging Processes 612
References 614

17
Thermionic Power System Engineering 615
17.1 General Remarks 615
17.2 Solar Thermionic Generators 615
17.2.1 General Configuration 616
17.2.2 Solar-Cavity Design 618
17.2.3 The Solar-Energy Thermionic Generator 624
17.2.4 Solar Probes 625
17.3 Radioisotopic Thermionic Generators 627
17.3.1 Radioisotopes as Heat Sources or Fuels 627
17.3.2 Handling of Radioisotopes 630
17.3.3 Module Designs 631
17.4 Nuclear Reactor Thermionic Generators 635
17.5 Out-of-Core Thermionic Reactors 637
17.5.1 Core-Surface or Conduction Designs 637
17.5.2 Forced-Convection Designs 637
17.5.3 Heat-Pipe Designs 639
17.6 In-Core Thermionic Reactors 639
17.6.1 Flashlight Designs 639
17.6.2 Pancake Designs 641
17.6.3 Externally Fueled Designs 641
17.6.4 Control of In-Core Thermionic Reactors 642
17.7 Chemical Thermionic Generators 642
17.7.1 General Configuration 643
17.7.2 Development of Emitter Structure 644
17.8 Topping Thermionic Generators 645
References 646
Index 649

Preface

Volume II

The preface to Volume I of *Thermionic Energy Conversion*, repeated for the reader in this volume on the following pages, briefly discusses the background of thermionic energy conversion and states the purpose, the aspirations, the audience, and the scope, outlined by chapter, of the two-volume publication.

Since the Volume I preface was written, however, a significant change has occurred. Consequently, this additional preface, which updates the thinking on the potential of thermionic conversion, is provided.

We are becoming increasingly more aware of the end of the era of abundant and inexpensive energy sources. Among other consequences, this awareness has rekindled interest in increasing the efficiency of converting primary fuels into electricity.

Presently most of our electricity is generated from oil and coal in steam power plants. Although the temperature of the products of combustion is typically about 2000°K, the fundamental properties of the water and structural materials constrain the peak steam temperature to less than 800°K. The transfer of energy from the products of combustion to the water across such a large temperature difference (2000 to less than 800°K) is irreversible and, therefore, reduces the efficiency potential for converting the energy to electricity.

Some of the efficiency potential can be captured by extracting useful work from the hot products of combustion as energy is transferred to the water, namely, by using a so-called topping system.

Thermionic conversion and magnetohydrodynamics are attractive methods for topping applications. Since the early seventies, the emphasis of the research and development work in thermionic conversion has shifted from special applications (such as power supplies for remote areas and space vehicles) to topping systems for steam power plants. Successful development of such systems holds the promise of improving the efficiency of steam power plants by about 25% at a total energy cost lower than that of new energy supplies. Topping thermionic generators are described briefly in Chap. 17.

We hope that the experience and analytical methods summarized in this volume will assist in the success of this new application of thermionic conversion.

George N. Hatsopoulos
Elias P. Gyftopoulos

Waltham, Massachusetts
February 1978

Preface

Volume I

Over the past few decades many efforts have been devoted to the development of new methods of generating electrical power. These efforts were stimulated both by the discovery of new sources of energy, such as nuclear reactions, and by the increasing demand for electric power supplies that can meet special requirements of portability, compactness, low noise, long life, unattended operation, and so forth.

In these efforts much emphasis was placed on the development of direct energy conversion devices, such as thermoelectric generators, photovoltaic cells, fuel cells, and thermionic converters, because the principles of their operation do not depend on the motion of mechanical parts. Compared with conventional turboalternators that involve rotating components, direct-conversion devices were expected to be more reliable, longer-lived, and more compact.

Serious investigation of thermionic energy conversion began in the middle fifties when vacuum diodes and triodes were tested and analyzed. Although high efficiency of energy conversion was demonstrated in late 1956 by means of a close-spaced diode, vacuum converters proved to be impractical.

A major advance occurred in 1957 when it was shown experimentally that the introduction of cesium vapor in the interelectrode space eliminated the need for close spacing and resulted in substantial improvements of performance. Today cesium converters are the only thermionic converters of practical importance.

The purpose of this volume and of Volume II is to describe the experience and theoretical understanding of thermionic energy conversion. The two volumes are written both as texts for use in energy conversion studies and as references for scientists and engineers engaged in research and development of thermionic conversion.

The present volume consists of four chapters summarizing information which does not depend on any highly specialized theories and which will not be substantially altered by future developments.

Descriptions of the thermionic converter as a heat engine and of the emission and transport phenomena encountered in thermionic conversion are given in the introductory chapter. These descriptions are intended to provide the reader with a brief overview of the field from principle to application so that he can place the subsequent discussions in their proper context.

A partial but fairly important analysis of converters is presented in Chap. 2, "The Ideal Performance of Diode Thermionic Converters." The results of this analysis are extremely valuable for these reasons: (1) they provide upper

performance limits for any diode converter; (2) they bring forth the processes that are absolutely essential to thermionic conversion; and (3) they provide a basis for the assessment of the degree and the direction of improvement to be expected from any approach to the problems of thermionic conversion.

A summary of analytical and experimental results on vacuum converters is given in Chap. 3, "Vacuum Thermionic Converters." Although they are not practical, vacuum converters are discussed because an understanding of their performance characteristics is helpful in the analysis of vapor converters.

Cesium and other vapor converters are discussed in Chap. 4, which includes a representative set of experimental performance characteristics and qualitative descriptions of the major phenomena resulting from the presence of cesium in the interelectrode space. The discussions are intended to provide a summary of the state of the art of cesium converters and an outline of the theoretical questions concerning their operation.

Finally a number of graphs and tables are given in the Appendix. These have been found useful in engineering calculations of thermionic conversion systems.

Theoretical results in Chaps. 1 to 4 are given without proof where specialized knowledge of physics is necessary or else in detail when only simple mathematical manipulations are involved. Rigorous theoretical analyses of the phenomena of thermionic conversion are reserved for Volume II, in which the selection of the material has been influenced primarily by two observations: First, many phenomena of thermionic conversion are both quantum mechanical and subject to the limitations imposed by the laws of thermodynamics. Second, the performance of thermionic converters is extremely sensitive to the values of certain electrode and plasma properties. Accordingly, analyses of thermionic converters require mastery of both quantum mechanics and thermodynamics, and accuracy in results better than that necessary in usual engineering calculations.

The first four chapters of Volume II, Chaps. 5 to 8, are devoted to a novel presentation of the basic principles and applications of quantum thermodynamics. They are far more extensive than the reader might have expected to find in a book on thermionic energy conversion. The principal authors have developed a way of looking at problems of quantum mechanics and thermodynamics that differs from that found in the literature. This method is used in the analyses presented in subsequent chapters, and, therefore, it is necessary to describe it in some detail in Chaps. 5 to 8.

A thorough review of the theoretical understanding of emission properties of different types of refractory metal surfaces is given in Chap. 9. The emphasis of the review is placed both on the available tools for the analysis of thermionic emission and on the problems that remain to be solved.

Transport of low-energy particles through a vapor, of the type encountered in thermionic conversion, is discussed in Chaps. 10 and 11. The transport of charged particles in the absence of collisions is analyzed in Chap. 10, whereas a systematic development of the theory of transport phenomena in the presence of collisions in low-energy plasmas is given in Chap. 11. These two chapters provide the principal tools for the analysis of the performance characteristics of cesium converters. This analysis is given in Chap. 12, together with a discussion of areas that require further analytical work for a better understanding.

Experimental techniques employed in the chemical and metallurgical characterization of converter materials and in the measurement of various converter parameters are discussed in Chap. 13.

Experimental performance characteristics of a large number of thermionic converters under a great variety of conditions of operation are compiled and correlated in Chap. 14. These characteristics provide valuable aid for both design and theoretical studies and illustrate the effects of the many interrelated factors that influence the performance of thermionic converters.

The last three chapters, Chap. 15, "Metallurgy of Electrode Materials," Chap. 16, "Design and Fabrication of Practical Converters," and Chap. 17, "Thermionic Power System Engineering," are devoted to a number of engineering problems whose resolution is essential to the successful development of thermionic power systems. Some of these problems have been resolved. Others have not, and for that reason thermionic conversion remains a promising method of energy conversion for the future.

The quality of the book has been enriched by the participation of many individuals. We are indebted to the contributors for supplying and organizing the material for the chapters listed here. Roland Breitwieser wrote the material for Chap. 13, Firooz Rufeh for Chap. 14, Laurence van Someren for Chap. 15, Pierre J. Brosens for Chap. 16, and Pierre J. Brosens, Jerry P. Davis, John B. Dunlay, and Lazaros J. Lazaridis for Chap. 17. Although the material was substantially edited to achieve uniformity of terminology and style and to assure responsiveness to the needs of the other chapters, the book could not be written without the expertise of our contributors.

We are thankful to Ned S. Rasor for his active participation in the initiation of the project and for his many ideas concerning Chaps. 1 and 2, to James C. Keck for his contributions to Chap. 11, to Joseph H. Keenan for many helpful discussions, and to David Lieb for writing and executing computer programs for a number of calculations. Many thanks also go to Judy A. Bellizia and to D. Lee Carchia for their meticulous help in the preparation of the manuscript, to Grace Rowe for her art work, and to Margaret L. Givens and Dee Jared for their excellent editorial assistance. We are appreciative of helpful comments that we

received from many friends during the review of a number of the chapters, and particularly the comments of John M. Houston, Charles Warner, and Monte V. Davis. Finally, we would like to express our sincere thanks to Donald S. Beard and Robert F. Pigeon of the AEC for their patience and understanding over the many years that we took to finish our writings.

George N. Hatsopoulos
Elias P. Gyftopoulos

Waltham, Massachusetts
July 1972

List of Contributors

Roland Breitwieser, NASA-Lewis Laboratory
Pierre J. Brosens, Thermo Electron Corporation
Jerry P. Davis, Thermo Electron Corporation
John B. Dunlay, Thermo Electron Corporation
James C. Keck, Department of Mechanical Engineering, M.I.T.
Lazaros J. Lazaridis, Thermo Electron Corporation
Ned S. Rasor, Consultant
Firooz Rufeh, Thermo Electron Corporation
Laurence van Someren, Thermo Electron Corporation

Thermionic Energy Conversion

Volume II: Theory, Technology, and Application

5

Elements of Quantum Mechanics

5.1 General Remarks

Understanding the basic phenomena of thermionic conversion and analyzing experimental results obtained from thermionic converters involve two steps: (1) a study of systems of particles (electrons, ions, and others) under conditions of stable equilibrium and (2) a study of systems of particles in steady states. Relations concerning stable equilibrium states and steady states are given in Chaps. 7 and 8, respectively.

In general, the properties of a system, such as the pressure of a vapor or the saturation current from a thermionic emitter, are given by suitably defined averages of characteristics of its constituent particles. The analysis of the characteristics of particles and the procedures for averaging over these characteristics are the subjects of mechanics and thermodynamics. Thus the starting point of a rigorous analysis of thermionic converters requires knowledge of the principles of these two parts of physics. Accordingly a brief review of mechanics is given here, and a review of thermodynamics is given in Chap. 6.

Because many phenomena of interest to thermionic conversion involve important quantum effects and because the relations between properties and characteristics of particles of a system are based on quantum mechanical concepts, this chapter summarizes the principles of quantum mechanics, with emphasis on the aspects of the theory which are especially relevant to thermionic conversion. The theory is applied to simple examples chosen so that the results are directly or indirectly useful to the discussions of Chaps. 6 to 12.

The majority of the problems of thermionic conversion cannot be presently analyzed exactly by means of quantum mechanics, primarily because of insurmountable computational difficulties. Nevertheless, approximate quantum-mechanical studies provide a variety of qualitative results that help to establish the validity and the limitations of different semiempirical analyses and to suggest new ways of interpreting experiments.

In this chapter there is, first, a brief account of the physical and conceptual problems that resulted in the development of quantum mechanics. Second, the postulates of quantum mechanics are summarized and discussed. Third, the theory is applied to simple systems simulating, approximately, problems of interest to this book. Finally, a variety of qualitative results from the quantum-mechanical theory of molecules and solids is presented because of their usefulness in the discussions of surface phenomena in Chap. 9.

5.2 Classical Vs. Quantum Mechanics

The addition of quantum mechanics to the theories necessary for the interpreting and predicting physical phenomena is the result of many observations that cannot be explained by means of the classical ideas of physics. A complete account of these observations is beyond the scope of the present elementary discussion of quantum mechanics. The reader can find it in several textbooks on the subject.[1,2] Nevertheless, to place the discussion in the proper perspective, we describe briefly the steps that led physical theory from classical to quantum mechanics.

5.2.1 Discrete Energy Levels. In 1900 Planck derived a relation for the observed distribution of thermal radiation as a function of frequency and temperature inside a hollow enclosure in stable equilibrium.[3] The derivation of this relation is based on the assumption that radiation is in interaction with a set of electric oscillators of all different frequencies which are themselves capable of absorbing and emitting energy only in discrete quanta. For a given frequency, ν, this definite amount of energy equals $h\nu$, where h is Planck's constant ($h = 6.6256 \times 10^{-27}$ erg-sec or 4.140×10^{-15} ev-sec).

This development introduced into physics the new idea that systems, such as electric oscillators, have a set of discrete energy levels separated by differences in energy which are related to the frequency of the radiation emitted and absorbed. The energy difference from one level to the other is given by the relation

$$\varepsilon_2 - \varepsilon_1 = h\nu \tag{5.1}$$

where ε_1 and ε_2 are the energies of levels 1 and 2, respectively.

The idea of quantized energy levels received further support from the works of Einstein,[4] Debye,[5] and others on the specific heat of solids. In these works the investigators assumed that the modes of vibration of a crystalline solid are characterized by discrete energy levels $h\nu$ apart. Thus they were able to predict the observed specific heats of crystals as functions of temperature.

The concept of discrete energy levels was even more importantly extended by the work of Bohr[6] on the radiation spectrum of the hydrogen atom. He assumed that the atom of hydrogen consists of a single electron rotating around a central proton. The centrifugal force on the electron is balanced by the electrostatic attraction between the electron and the proton. He gave a simple rule for the selection of discrete quantized orbits from the continuum of possible orbits of classical mechanics. The energies of these orbits are discrete and consistent with the different spectral lines of the spectrum of the hydrogen atom.

The recognition that electric oscillators and atoms are characterized by discrete values of energy is one of the key ideas introduced into physics by quantum mechanics. This idea constitutes an important departure from the

fundamental principles of classical mechanics. In classical mechanics all energy levels are acceptable, and no criteria exist for the selection of specific possible energy levels such as the levels of an electron around a positive nucleus.

5.2.2 Wave–Particle Duality. In 1905 Einstein[7] suggested that not only the radiation emitted or absorbed by electric oscillators but also the electromagnetic radiation itself is quantized. Furthermore, he introduced the idea that the quantized radiation can sometimes be regarded as having a corpuscular character. The energy ϵ of each corpuscle or *photon* associated with a radiation frequency ν is given by the relation

$$\epsilon = h\nu \tag{5.2}$$

and the momentum p of the photon in the direction of propagation is given by the expression

$$p = \frac{h\nu}{c} = \frac{h}{\lambda} \tag{5.3}$$

where c is the speed of light ($c = 2.99793 \times 10^{10}$ cm/sec) and λ is the wavelength. The equations (5.3) are consistent with the relativistic relations between mass, energy, and momentum.

The new idea of the corpuscular interpretation of radiation was successfully applied to the explanation of the photoelectric effect, namely, the phenomenon of electron emission from a metal surface by means of electromagnetic radiation. Later other physical phenomena gave further support to the view that radiation could be regarded either as a wave or as a particle.

In 1924 the idea of wave–particle duality was extended to matter by Louis de Broglie.[8] As Einstein associated particles with waves of electromagnetic radiation, de Broglie proposed that waves should be associated with each of the known particles of matter. This association was done by means of Eqs. 5.2 and 5.3. Specifically, a particle of energy ϵ and momentum p can be regarded as a wave of frequency ν given by the relation

$$\nu = \frac{\epsilon}{h} \tag{5.4a}$$

and wavelength

$$\lambda = \frac{h}{p} \tag{5.4b}$$

The wavelike behavior of matter was made evident by the electron-diffraction experiments of Davisson and Germer,[9] who studied the reflection of slow electrons by a nickel crystal surface. They found that the results could be justified if the incoming electrons were regarded as waves of wavelength λ given by the relation

$$\lambda = \frac{h}{p} = \frac{h}{m_e v} \tag{5.5}$$

where m_e and v are the mass and the speed of the electron, respectively.

The assignment of wave and particle properties to both radiation and matter is another key idea introduced into physics by quantum mechanics. It is one example of a large number of dualities that can be established between physical properties. Classically the wave–particle duality appears to involve the characterization of one entity by contradictory properties. As will become evident later, however, the contradiction is resolved by the basic postulates of quantum mechanics.

5.2.3 Indeterminacy, Uncertainty, Complementarity, and Correspondence. Upon review of the experimental evidence that cannot be understood by means of classical mechanics, Heisenberg[10] and Bohr[11] developed a number of general principles that have guided the formulation of the postulates of quantum mechanics. These principles are discussed briefly with hindsight rather than in terms of the original arguments of Heisenberg and Bohr.

In classical mechanics it is assumed that at each instant of time a particle is at a definite position x. Review of experiments, however, reveals that each of many measurements of position of identical particles in identical conditions does not yield the same result. In general, each measurement of position yields a different result. In addition, and more importantly, the result of each measurement is unpredictable. Similar remarks can be made about measurements of other dynamic variables of any system, e.g., momentum and energy. In quantum mechanics the result of a measurement of a variable is predicted probabilistically. The assignment of probabilities to measurement results reflects the quantum-mechanical principle of *indeterminacy.*

Measurement results and their probabilities can be used to compute a measure, such as the standard deviation, of the spread or dispersion of these results. If a measure of the spread in measurement results of position x is denoted by Δx and of momentum p_x along the x-axis by Δp_x, we find that the product of Δx and Δp_x is greater than a lower limit. In quantum mechanics this lower limit is taken equal to Planck's constant, h. Similar comments can be made about position y and momentum p_y along the y-axis and position z and momentum p_z along the z-axis so that

$$\Delta x \cdot \Delta p_x > h \quad \Delta y \cdot \Delta p_y > h \quad \Delta z \cdot \Delta p_z > h \tag{5.6}$$

In general, relations 5.6 prescribe limits to the spreads or dispersions of values that can be simultaneously assigned to the characterization of the position and momentum of a particle. The limits apply always to conjugated pairs of variables such as position and momentum. Relations 5.6 are special examples of a general principle known as *Heisenberg's uncertainty principle.*

The relation between two variables whose values are subject to interconnected dispersions has been called *complementarity.*[11] Complementarity of two variables implies that reduction in the spread of values associated with one is accompanied by increase in the spread of values associated with the other.

Finally, although the interpretation of many physical phenomena requires quantum-mechanical ideas, many others can be successfully interpreted by means of classical mechanics. This apparently conflicting evidence can be rationalized by recognizing that quantum mechanics conforms to Bohr's *correspondence principle.* This principle stipulates that classical and quantum-mechanical treatments of any system must converge to the same results when the degree of ambiguity arising from the dispersion of measurement results is negligible compared with the average value of the results.

5.3 Postulates of Quantum Mechanics

5.3.1 General Remarks. The foundations of quantum mechanics can be expressed in terms of four postulates[12] that make certain assertions concerning the results of measurements that might be made on a physical system and the change of the results with time. Before we present the postulates, let us review the concept of state and the equation of motion of classical mechanics.

5.3.2 The State and the Equation of Motion in Classical Mechanics. In classical mechanics the state of a system is specified by the instantaneous values of the positions and velocities of all of its particles. For example, if a system consists of particles without internal structure and if the particles are moving under the influence of their mutual interactions and of other external forces, at any instant of time the state of the system is specified by the values of the Cartesian coordinates x, y, and z for the position of each particle and by the values of the corresponding velocity components $\dot{x}$, $\dot{y}$, and $\dot{z}$, which give the time rates of change of the coordinates, namely,

$$\dot{x} = \frac{dx}{dt} \quad \text{etc.} \tag{5.7}$$

In general, the state of a system need not be represented in Cartesian coordinates. Instead, for a system having N *degrees of freedom* (i.e., N

independent ways in which the space configuration of the system may change), the state can be specified by means of a set of *N generalized coordinates,*[13]

$$q_1, q_2, \ldots, q_N \tag{5.8}$$

and a set of *N generalized velocities,*

$$\dot{q}_1, \dot{q}_2, \ldots, \dot{q}_N \tag{5.9}$$

where

$$\dot{q}_k = \frac{dq_k}{dt} \quad \text{for } k = 1, 2, \ldots, N \tag{5.10}$$

The systematic procedure for determining these two sets is given in Ref. 13. For brevity the two sets are denoted by the pair $(q,\dot{q})$ without the use of subscripts.

The definition of the state implies that the values of both positions and velocities of all the particles can be described simultaneously by unambiguous values and that these values can be verified experimentally. In addition, it will become evident from the subsequent discussion of the law of motion of classical mechanics that knowledge of the values $(q,\dot{q})$ at an initial time and of the nature of the system are sufficient for predicting the exact states at future times.

The law of motion of classical mechanics can be written in different equivalent forms. The law is presented here in the form of Hamilton's canonical equations and only for a nonrelativistic, conservative, ordinary mechanical system. For such a system with N degrees of freedom, the following quantities can be readily defined:

1. The *kinetic energy, T,* is defined as the sum of the kinetic energies of all the particles; each particle has a kinetic energy equal to $(\frac{1}{2})\, m_j v_j^2$, where m_j and v_j are the mass and the speed of the jth particle; if generalized coordinates are used, the kinetic energy should be expressed in terms of the generalized velocities by means of the procedure described in Ref. 13.

2. The *potential energy, V,* is defined as a function of all the coordinates q_k and is such that the *generalized particle force* f_k associated with the coordinate q_k is given by the relation

$$f_k = -\frac{\partial V}{\partial q_k} \quad \text{for } k = 1, 2, \ldots, N \tag{5.11}$$

3. The *Lagrangian function,* $L(q,\dot{q})$, is defined by the relation

$$L = T - V \tag{5.12}$$

4. The *momentum, p_k, canonically conjugated to q_k* is defined by the relation

$$p_k = \frac{\partial L}{\partial \dot{q}_k} \quad \text{for } k = 1, 2, \ldots, N \tag{5.13}$$

5. The *Hamiltonian function, $H(q,p)$*, is defined by the relation

$$H(q,p) = \sum_{k=1}^{N} p_k \dot{q}_k - L(q,\dot{q}) \tag{5.14}$$

In terms of these definitions, the motion of the system is described by *Hamilton's canonical equations* given by the relations

$$\dot{p}_k = -\frac{\partial H}{\partial q_k} \tag{5.15a}$$

and

$$\dot{q}_k = \frac{\partial H}{\partial p_k} \quad \text{for } k = 1, 2, \ldots, N \tag{5.15b}$$

Equations 5.15a and 5.15b define a family of trajectories in the $2N$-dimensional (p,q) space, which is called the *phase space.* As time goes on, the system moves along one of the trajectories of phase space. If the values of the q_k's and p_k's are specified at an initial time, Hamilton's equations define the exact trajectory in phase space and provide an exact prediction of the configuration and the momenta possessed by the system at any given time. The numerical values of the momenta are proportional to those of the velocities. It follows that at each instant of time the values of the N coordinates q_k and the N momenta p_k are equivalent to the values $(q,\dot{q})$ and define the state of the system. All dynamic properties of the system, such as the kinetic energy and potential energy, can be computed when the values of the q_k's and the p_k's are known.

5.3.3 The State in Quantum Mechanics: Postulate 1. In quantum mechanics the state of a system of particles with N degrees of freedom cannot be specified in terms of instantaneous values of the q_k's and p_k's, because both sets of numbers cannot be assigned unambiguous values simultaneously. Specifically, if the values of the q_k's are specified, the values of the p_k's are described probabilistically; if the values of the p_k's are specified, the values of the q_k's are described probabilistically; or the values of both the q_k's and p_k's are described

probabilistically. The essence of these general remarks is contained in the following quantum-mechanical definition of state.

Postulate 1. For a given system of particles that classically would be characterized by N generalized coordinates $(q_1, q_2, \ldots, q_N)$ and N generalized momenta $(p_1, p_2, \ldots, p_N)$, the quantum-mechanical state† is defined by two *complex probability amplitudes* or *wave functions*, $\psi(q,t)$ and $\phi(p,t)$, such that at time t: (1) if the coordinates of the system are measured, the probability that they will be found within the ranges q_1 and $q_1 + dq_1$, q_2 and $q_2 + dq_2, \ldots$, is given by the expression

$$\psi^*(q,t)\,\psi(q,t)\,dq_1 dq_2 \ldots dq_N \tag{5.16}$$

and (2) if the momenta of the system are measured, the probability that they will be found within the ranges between p_1 and $p_1 + dp_1$, p_2 and $p_2 + dp_2, \ldots$, is given by the expression

$$\phi^*(p,t)\,\phi(p,t)\;dp_1\;dp_2 \ldots dp_N \tag{5.17}$$

where q and p denote all the q's and p's, respectively, and superscript * denotes the complex conjugate of the quantity.‡

The positive quantities $W_q(q,t)$ and $W_p(p,t)$, defined by the relations

$$W_q(q,t) = |\psi(q,t)|^2 = \psi^*(q,t)\,\psi(q,t) \tag{5.18}$$

and

$$W_p(p,t) = |\phi(p,t)|^2 = \phi^*(p,t)\,\phi(p,t) \tag{5.19}$$

are called *probability distribution functions* or *probability densities.*

Since at each time each coordinate must be found to have some value within its total range, the integral of $W_q(q,t)$ over all *coordinate space*, the *total probability*, must be equal to unity. In terms of symbols

$$\int \ldots \int W_q(q,t)\;dq_1\;dq_2 \ldots dq_N = 1 \tag{5.20}$$

†A more general definition of state of a system is given in Chap. 6.

‡For a vector or matrix, the asterisk (*) denotes the complex conjugate of the transpose of the vector or matrix.

For convenience the multiple integral in Eq. 5.20 can be denoted by a single integral sign, and the product $dq_1\, dq_2 \ldots dq_N$ by $d^N q$ so that

$$\int W_q(q,t)\, d^N q = 1 \tag{5.21}$$

Similarly, for the momenta

$$\int W_p(p,t)\, d^N p = 1 \tag{5.22}$$

where

$$dp_1\, dp_2 \ldots dp_N \equiv d^N p \tag{5.23}$$

the single integral sign stands for the multiple integral, and all the integrals are taken over all possible values of the momenta, i.e., over all *momentum space*.

For results that are physically meaningful, both wave functions ψ and ϕ must be bounded and single-valued. In addition, both ψ and ϕ must approach zero for values of the coordinates and momenta approaching $-\infty$ and $+\infty$, and the squares of their magnitudes must be integrable over all coordinate and momentum spaces, respectively.

5.3.4 The Relation Between Probability Amplitudes: Postulate 2. The first postulate of quantum mechanics asserts the existence of the probability amplitudes ψ and ϕ that together specify a quantum mechanical state of a system. In view of the uncertainty and complementarity principles discussed earlier, these two functions are not independent of each other. For canonical coordinates and momenta,† the relation between the two amplitude functions is the subject of the second postulate of quantum mechanics.

Postulate 2. The probability amplitudes ψ and ϕ are connected by the relations

$$\phi(p,t) = h^{-N/2} \int \psi(q,t) \exp\left(-i \sum_{k=1}^{N} \frac{p_k q_k}{\hbar}\right) d^N q \tag{5.24}$$

and

$$\psi(q,t) = h^{-N/2} \int \phi(p,t) \exp\left(+i \sum_{k=1}^{N} \frac{p_k q_k}{\hbar}\right) d^N p \tag{5.25}$$

†For example, Cartesian coordinates and the corresponding momenta of a system of particles furnish a canonical set.

where

$$\hbar = \frac{h}{2\pi} \quad \text{and} \quad i = \sqrt{-1} \tag{5.26}$$

In words, ϕ is the multidimensional Fourier transform of ψ, and ψ is the multidimensional inverse Fourier transform of ϕ, and, therefore, either ϕ or ψ is adequate to describe the quantum-mechanical state.

The second postulate is consistent with the notion that, if the coordinates are specified by values without ambiguity, then the momenta are specified by ranges of values, and vice versa. For example, if ϕ is a sharply localized function containing only momenta near some value p_0, then ψ is a long wave train.

The possibility of describing a state of a system in alternative ways by use of one or another of various probability amplitudes is a characteristic feature of quantum mechanics. The different kinds of probability amplitudes are themselves functions of different sets of variables. Thus it is sometimes convenient to say that a particular kind of language is used, such as *q-language* or *p-language.*

5.3.5 Operators Corresponding to Observable Quantities: Postulate 3. In accordance with the foregoing, at any time of interest a quantum-mechanical state of a system is completely specified by a single complex probability amplitude, ψ or ϕ. This means that all information about the properties of the system is contained in either of these two functions.

One type of information of practical interest is the prediction of the *expectation* or *average value* of a dynamic variable (quantum-mechanical observable) which in classical mechanics corresponds to a function of the form $F(q,p)$. Experimentally this value would be found by means of a large number of measurements of $F(q,p)$ made on the system in the same state. Theoretically this value can be calculated as follows:

When the dynamic variable is of the form $F(q)$ (function of the coordinates only), the expectation value can be readily calculated if the state of the system is specified in q-language. By virtue of the probabilistic interpretation of $|\psi|^2$, the expectation value $\langle F \rangle$ of $F(q)$ is given by the relation

$$\begin{aligned} \langle F \rangle &= \int F(q)\, W_q(q,t)\, d^N q \\ &= \int \psi^*(q,t)\, F(q)\, \psi(q,t)\, d^N q \end{aligned} \tag{5.27}$$

Similarly, when the dynamic variable is of the form $F(p)$ (function of momenta only), $\langle F \rangle$ of $F(p)$ can be computed fairly readily if the state is expressed in p-language. Specifically

$$\langle F \rangle = \int \phi^*(p,t)\, F(p)\, \phi(p,t)\, d^N p \tag{5.28}$$

By virtue of the second postulate, expectation values of dynamic variables in the form of rational polynomials in the q's and the p's can be computed by using p-language and q-language, respectively. For example, the expectation value of a particular momentum, say p_k, is given by the relation (Eq. 5.28)

$$\langle p_k \rangle = \int \phi^*(p,t)\, p_k\, \phi(p,t)\, d^N p \tag{5.29}$$

Using Eq. 5.24 and some mathematical manipulations in Eq. 5.29 yields

$$\langle p_k \rangle = \int \psi^*(q,t)\, \frac{\hbar}{i}\, \frac{\partial \psi(q,t)}{\partial q_k}\, d^N q \tag{5.30}$$

Comparison of Eqs. 5.27 and 5.30 indicates that $\langle p_k \rangle$ can be computed in q-language if p_k is transformed into a q-dependent quantity. The transformation is achieved by replacing p_k by the differential operator $(\hbar/i)\, \partial/\partial q_k$.

By similar mathematical manipulations, we can readily show[12] that the expectation value of any rational product of momenta $p_1^\alpha\, p_2^\beta \ldots p_N^\mu$ can be calculated in q-language by transforming this product into the differential operator

$$\left(\frac{\hbar}{i}\frac{\partial}{\partial q_1}\right)^\alpha \left(\frac{\hbar}{i}\frac{\partial}{\partial q_2}\right)^\beta \cdots \left(\frac{\hbar}{i}\frac{\partial}{\partial q_N}\right)^\mu$$

and by using this operator in the right-hand side of Eq. 5.27:

$$\begin{aligned} \langle p_1^\alpha\, p_2^\beta \ldots p_N^\mu \rangle &= \int \phi^*(p,t)\, p_1^\alpha\, p_2^\beta \ldots p_N^\mu\, \phi(p,t)\, d^N p \\ &= \int \psi^*(q,t) \left(\frac{\hbar}{i}\right)^\rho \frac{\partial^\rho \psi(q,t)}{\partial q_1^\alpha\, \partial_2^\beta \ldots \partial q_N^\mu}\, d^N q \end{aligned} \tag{5.31}$$

where

$$\rho = \alpha + \beta + \ldots + \mu \tag{5.32}$$

It follows from Eqs. 5.31 and 5.32 that the expectation value of any quantity that is a rational polynomial of the momenta can be found in q-language by transforming the polynomial into a differential operator according to the relation

$$\langle F\rangle = \int \phi^*(p,t)\, F(p_1,p_2,\ldots,p_N)\, \phi(p,t)\, d^Np$$

$$= \int \psi^*(q,t)\, \hat{F}\left(\frac{\hbar}{i}\frac{\partial}{\partial q_1}, \frac{\hbar}{i}\frac{\partial}{\partial q_2}, \ldots, \frac{\hbar}{i}\frac{\partial}{\partial q_N}\right) \psi(q,t)\, d^Nq \tag{5.33}$$

where the symbol $\hat{}$ denotes an operator.

In an entirely analogous manner, for a given rational polynomial $F(q)$ of the coordinates, the expectation value $\langle F\rangle$ can be calculated[12] in p-language by transforming $F(q)$ into a p-dependent differential operator and using this operator in an expression of the form given by Eq. 5.28. The transformation proceeds according to the rule that the general rational product

$$q_1^\alpha\, q_2^\beta \ldots q_N^\mu$$

corresponds to the p-dependent differential operator of the form

$$\left(-\frac{\hbar}{i}\frac{\partial}{\partial p_1}\right)^\alpha \left(-\frac{\hbar}{i}\frac{\partial}{\partial p_2}\right)^\beta \cdots \left(-\frac{\hbar}{i}\frac{\partial}{\partial p_N}\right)^\mu$$

and $\langle F\rangle$ is given by the relation

$$\langle F\rangle = \int \psi^*(q,t)\, F(q_1,q_2,\ldots,q_N)\, \psi(q,t)\, d^Nq$$

$$= \int \phi^*(p,t)\, \hat{F}\left(-\frac{\hbar}{i}\frac{\partial}{\partial p_1}, -\frac{\hbar}{i}\frac{\partial}{\partial p_2}, \ldots, -\frac{\hbar}{i}\frac{\partial}{\partial p_N}\right) \phi(p,t)\, d^Np \tag{5.34}$$

For dynamic variables of the form $F(q,p)$, functions of both q's and p's, the results represented by relations 5.33 and 5.34 suggest that $\langle F\rangle$ could be calculated in either q-language or p-language if an appropriate operator could be found to represent $F(q,p)$ in expressions of the forms

$$\langle F\rangle = \int \psi^*(q,t)\, \hat{F}_q\, \psi(q,t)\, d^Nq$$

and

$$\langle F\rangle = \int \phi^*(p,t)\, \hat{F}_p\, \phi(p,t)\, d^Np$$

The third postulate shows how either the operator $\hat{F}_q$ in q-language or the operator $\hat{F}_p$ in p-language is determined.

Postulate 3. The expectation value $\langle F\rangle$ of any dynamic variable $F(q,p)$ can be evaluated by using either

$$\langle F \rangle = \int \psi^*(q,t) \hat{F}_q \, \psi(q,t) \, d^N q \tag{5.35}$$

or

$$\langle F \rangle = \int \phi^*(p,t) \hat{F}_p \, \phi(p,t) \, d^N p \tag{5.36}$$

where $\hat{F}_q$ is a *linear Hermitian operator* obtained by replacing p_k by $(\hbar/i)\,\partial/\partial q_k$ in $F(q,p)$ and $\hat{F}_p$ is a linear Hermitian operator obtained by replacing q_k by $(-\hbar/i)\,\partial/\partial p_k$ in $F(q,p)$.

Linearity and Hermiticity imply certain important properties of operators. If an operator $\hat{F}$ and two arbitrary complex functions X and Y are expressed in the same language—for example the q-language—and if X and Y satisfy all the boundedness and integrability requirements imposed on wave functions, then the operator $\hat{F}$ is linear if the principle of superposition holds, namely,

$$\hat{F}(X + Y) = \hat{F}X + \hat{F}Y \tag{5.37}$$

and

$$\hat{F}(c_1 X) = c_1 \, \hat{F}X \tag{5.38}$$

where c_1 is an arbitrary constant multiplier. Moreover, the operator $\hat{F}$ is Hermitian if it satisfies the relation

$$\int X^* \, \hat{F}Y \, d^N q = \int Y \, (\hat{F}X)^* \, d^N q \tag{5.39}$$

Operators associated with dynamic variables are required to be linear so that the principle of superposition will hold for the different solutions of the quantum-mechanical equation of motion (Postulate 4) and for the boundary conditions of the system. On the other hand, the requirement of Hermiticity guarantees that expectation values of observables are real. Indeed, the expectation value $\langle F \rangle$ and its complex conjugate $\langle F \rangle^*$ are given by the relations

$$\langle F \rangle = \int \psi^*(q,t) \, \hat{F}_q \, \psi(q,t) \, d^N q \tag{5.40}$$

and

$$\langle F \rangle^* = \int \psi(q,t) \, [\hat{F}_q \, \psi(q,t)]^* \, d^N q \tag{5.41}$$

When $\hat{F}_q$ is Hermitian (Eq. 5.39), the right-hand sides of Eqs. 5.40 and 5.41 are equal and, therefore,

$$\langle F \rangle = \langle F \rangle^* = \text{real} \tag{5.42}$$

Finally, it is noteworthy that the expectation value of any power of a measurable quantity can be obtained by procedures similar to those outlined. It follows from probability theory that a complete prognosis of the probabilities of finding different values of a quantity can thus be established, regardless of the language used for the description of the state of the system.

Many other important implications resulting from the linear and Hermitian character of operators representing dynamic variables (observables) are not discussed here.

For convenience, the symbol $\langle\rangle$ is omitted from expectation values in the remainder of this book.

5.3.6 The Schroedinger Equation for Change in State with Time: Postulate 4. The discussion so far has centered around the descriptions of states and expectation values of observables of a physical system in terms of wave functions ψ and ϕ. The question, "How do the functions ψ and ϕ change with time?" (i.e., how does the system move?) is the subject of the fourth postulate.

Postulate 4. The wave functions ψ and ϕ change with time according to the equations

$$\hat{H}_q \psi = -\frac{\hbar}{i} \frac{\partial \psi}{\partial t} \tag{5.43}$$

and

$$\hat{H}_p \phi = -\frac{\hbar}{i} \frac{\partial \phi}{\partial t} \tag{5.44}$$

where $\hat{H}_q$ and $\hat{H}_p$ are the Hamiltonian operators for the system in q-language and p-language, respectively. These operators are obtained from the classical Hamiltonian expression for the energy as a function of coordinates and momenta, in the manner described in Postulate 3.

Each of Eqs. 5.43 and 5.44 is called the *Schroedinger equation*. Only one of these two equations is needed for each problem since the other follows from the second postulate. Nevertheless, both equations are given to indicate the complete symmetry between coordinates and momenta which exists throughout quantum mechanics. This symmetry, which parallels that of Hamilton's equations of motion in classical mechanics (Eqs. 5.15a and 5.15b), has many important implications, some of which are discussed later.

The four postulates provide a basis for the nonrelativistic quantum-mechanical treatment of systems of particles whose configurations can be specified in terms

of coordinates and momenta in the classical sense. It is noteworthy, however, that certain observable quantities, such as *spin* and *parity*, have no analogs in classical mechanics. Introducing these quantities into the formalism requires the consideration of additional assumptions.[1,2]

In view of the complete equivalence between different languages of representation, only the q-language is used in the subsequent discussions unless otherwise stated. Accordingly the subscript q is omitted from operators and probability densities.

5.4 Some Theorems of Quantum Mechanics

5.4.1 Conservation of Total Probability. The complex probability amplitudes are functions of time. For consistency the time dependence must be such as to conserve the total probability. That this is indeed the case can be verified by examining the time rate of change of the total probability. For example, the time rate of change of the total probability of finding the values of the coordinates somewhere in their respective ranges is given by the relation

$$\frac{d}{dt}\int \psi^*\psi \, d^N q = \frac{i}{\hbar}\int [\psi(\hat{H}\psi)^* - \psi^*(\hat{H}\psi)] \, d^N q = 0 \tag{5.45}$$

where use has been made of Eq. 5.43. The value zero results from the Hermitian character of the Hamiltonian operator (see Eq. 5.39). Conversely, it can be argued that the Hamiltonian operator must be Hermitian for the total probability to be time independent.

5.4.2 Probability Current. Although the total probability remains constant, the probability density at a particular point in space changes with time. The change can be illustrated by a single particle of mass m in a field corresponding to a potential energy $V(x,y,z)$. The Hamiltonian function, H, for this particle is given by the relation

$$H = \frac{1}{2m}(p_x^2 + p_y^2 + p_z^2) + V \tag{5.46}$$

The Hamiltonian operator, $\hat{H}$, corresponding to H is given by the expression

$$\hat{H} = -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + V \tag{5.47}$$

Consequently the Schroedinger equation and its complex conjugate are given by the relations

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}\right)+V\psi=-\frac{\hbar}{i}\frac{\partial\psi}{\partial t} \tag{5.48}$$

and

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi^*}{\partial x^2}+\frac{\partial^2\psi^*}{\partial y^2}+\frac{\partial^2\psi^*}{\partial z^2}\right)+V\psi^*=\frac{\hbar}{i}\frac{\partial\psi^*}{\partial t} \tag{5.49}$$

If Eq. 5.48 is multiplied by ψ^*, Eq. 5.49 multiplied by ψ, and the results subtracted from each other,

$$\frac{\partial}{\partial t}(\psi^*\psi)+\frac{\hbar}{2mi}\left(\psi^*\frac{\partial^2\psi}{\partial x^2}+\psi^*\frac{\partial^2\psi}{\partial y^2}+\psi^*\frac{\partial^2\psi}{\partial z^2}-\psi\frac{\partial^2\psi^*}{\partial x^2}-\psi\frac{\partial^2\psi^*}{\partial y^2}-\psi\frac{\partial^2\psi^*}{\partial z^2}\right)=0 \tag{5.50a}$$

or equivalently

$$\frac{\partial W}{\partial t}+\operatorname{div}\mathbf{S}=0 \tag{5.50b}$$

where $\mathbf{S}$ is the *probability current density* defined by the vector relation

$$\mathbf{S}=\frac{\hbar}{2mi}(\psi^*\operatorname{grad}\psi-\psi\operatorname{grad}\psi^*) \tag{5.51}$$

and W is the probability density $|\psi|^2$. It is seen from Eq. 5.50b that the change in probability in a given region of space can be regarded as due to probability flow through the boundary of the region. Analogous relations can be derived for other systems.

5.4.3 The Principle of Superposition. The Hamiltonian operator and the derivative with respect to time in Schroedinger's equation are linear operators. Consequently any linear combination of individual solutions of this equation is also a solution of the same equation. For example, if ψ_1 and ψ_2 are solutions, the linear combination given by the relation

$$\psi=c_1\psi_1+c_2\psi_2 \tag{5.52}$$

is also a solution, for all values of the arbitrary constants c_1 and c_2. This result is called the *principle of superposition of quantum mechanics.* The possibility for superposition and decomposition of states is a quantum-mechanical result for which there is no analogy in classical mechanics.

5.4.4 Stationary States: Energy Eigenvalues and Eigenfunctions. A system is said to be in a *stationary* or *equilibrium state* when both the probability density W and the probability current density $\mathbf{S}$ are independent of time. It can be easily verified that both requirements are satisfied by a wave function $\psi(q,t)$ of the form given by the relation

$$\psi(q,t) = u(q) \exp\,[-i\,f(t)] \tag{5.53}$$

where $f(t)$ is a real function of time and both $u(q)$ and $f(t)$ are such that $\psi(q,t)$ is a solution of Schroedinger's equation.

For a system to be in a stationary state, its Hamiltonian operator $\hat{H}$ must be time independent. Thus it can be shown that $\psi(q,t)$ can be written in the form

$$\psi(q,t) = u(q) \exp\left[-\left(\frac{i}{\hbar}\right) Et\right] \tag{5.54}$$

where E is a real constant and $u(q)$ satisfies the relation

$$\hat{H}\,u(q) = E\,u(q) \tag{5.55}$$

Equation 5.55 is called the *energy eigenvalue equation.*

The description of the stationary states of a system by means of Eqs. 5.54 and 5.55 leads to important general conclusions, some of which are discussed here.

1. Substituting Eq. 5.54 into Eq. 5.35, and using Eq. 5.55, we find that the expectation value of any integral power n of the Hamiltonian operator, $\hat{H}$, is given by the relations

$$H^n = \int u^*(q)\,\hat{H}^n\,u(q)\,d^Nq = E^n \qquad \text{for } n = 1, 2, \ldots \tag{5.56}$$

provided that u is normalized to unity, i.e.,

$$\int u^*\,u\,d^Nq = 1$$

Since the expectation value of the Hamiltonian operator equals the average energy of the system, Eq. 5.56 implies that this energy is identically equal to the value of the constant E. Conversely, a system in a stationary state is characterized by an unambiguous (dispersion-free) value for its energy.

2. Not all the solutions $u(q)$ of Eq. 5.55 are acceptable for a given system in a stationary state with a given set of boundary conditions. For example, some of the solutions are not acceptable because they are not bounded throughout the range of q. In other words, only certain values of E lead to satisfactory solutions $u(q)$ for the system and boundary conditions in question. Each acceptable value of E is called an *energy eigenvalue* of the system, and the corresponding function $u(q)$ is called an *energy eigenfunction* of the system. Each pair of an eigenvalue and an eigenfunction is denoted by a suffix, such as E_n and $u_n(q)$, and determines an *energy eigenstate.*

3. Depending on the system, energy eigenvalues may be discrete or continuous, or both. This is consistent with the discussion presented in Sec. 5.2.

4. Energy eigenvalues are real.

5. Given two energy eigenvalues

$$E_m \neq E_n \qquad \text{for } m \neq n$$

the corresponding energy eigenfunctions, u_m and u_n, are orthogonal, i.e.,

$$\int u_m^* u_n \, d^N q = 0 \qquad \text{for } m \neq n$$

6. The same energy eigenvalue may correspond to more than one eigenfunction. For example,

$$E_n = E_m \qquad u_n \neq u_m$$

When this happens, the state is called *degenerate.* If g eigenfunctions correspond to the same eigenvalue, then the state is called *g-fold degenerate.* Eigenfunctions of a degenerate state are not necessarily orthogonal. Nevertheless, a given set of g linearly independent eigenfunctions can always be combined in such a way as to construct a set of g functions that are mutually orthogonal. Hence all eigenfunctions can be considered as orthogonal to each other.

7. For a given system the energy eigenfunctions form a complete, orthonormal set; this set can be used to represent any square integrable function of the q's in the form of a Fourier series expansion. If the energy eigenvalues are both discrete and continuous, the expansion consists of a series over the energy eigenfunctions with the discrete eigenvalues plus an integral over the energy eigenfunctions corresponding to the continuum.

5.4.5 Eigenstates in General. The discussion of eigenstates of the Hamiltonian operator can be readily generalized to other operators. If, at a given time t_0, the quantum-mechanical state $\psi(q,t_0)$ of a system is such that the observable quantity $F(q,p)$ has an unambiguous value, F_n, then the state is called an

eigenstate for the observable in question and can be derived from an eigenvalue problem. Specifically, if the normalized $\psi(q,t_0)$ is given by the relation

$$\psi(q,t_0) = v_n(q)$$

where $v_n(q)$ satisfies the eigenvalue equation

$$\hat{F}\, v_n(q) = F_n v_n(q) \tag{5.57}$$

and $\hat{F}$ is the quantum-mechanical operator corresponding to the observable $F(q,p)$, then the expectation value F^m of the mth power of $\hat{F}$ is given by the relations

$$F^m = \int v_n^*(q)\, \hat{F}^m\, v_n(q)\, d^N q = F_n^m \qquad \text{for } m = 1, 2, \ldots \tag{5.58}$$

where use has been made of Eq. 5.57. It follows that, if $v_n(q)$ satisfies the eigenvalue relation represented by Eq. 5.57, the observable $F(q,p)$ has a dispersion free value and the system is in an eigenstate of $\hat{F}$.

In view of the analogy between Eqs. 5.57 and 5.55, we can make the following remarks about eigenstates of a general linear Hermitian operator $\hat{F}$:

1. Eigenstates are specified by the unambiguous value of an observable.
2. The value of the observable and the wave function describing the state of the system are the eigenvalue and the eigenfunction, respectively, of the eigenvalue problem defined by the operator $\hat{F}$ corresponding to the observable.
3. Conclusions 1 to 7 in Sec. 5.4.4 are also applicable to the eigenvalues and eigenfunctions of $\hat{F}$.

5.4.6 Expansions in Terms of Eigenfunctions. Square integrable functions can be represented by series expansions in terms of energy or other eigenfunctions. Such expansions play a very important role in quantum mechanics. For example, if, for a conservative system with discrete energy eigenstates, $\{u_k(q)\}$ denotes the orthonormal, complete set of energy eigenfunctions and $\{E_k\}$ the corresponding set of energy eigenvalues, then it can be shown that the wave function $\psi(q,t)$ describing the states of the system can be expanded into a series of the form

$$\psi(q,t) = \sum_k a_k(t)\, u_k(q)$$

$$= \sum_k c_k u_k(q) \exp\left(-iE_k \frac{t}{\hbar}\right) \tag{5.59}$$

where each c_k is a constant independent of time. When $\psi(q,t)$ is normalized to unity, each c_k is given by the relations

$$c_k = \int u_k^*(q) \exp\left(iE_k \frac{t}{\hbar}\right) \psi(q,t)\, d^N q \tag{5.60}$$

and

$$\sum_k |c_k|^2 = 1 \tag{5.61}$$

The expectation value E of the energy of the system (the expectation value of $\hat{H}$) is given by the relation

$$\begin{aligned} E &= \int \psi^* \hat{H} \psi \, d^N q \\ &= \sum_k |c_k|^2 E_k \end{aligned} \tag{5.62}$$

Consistent with Eq. 5.62, the coefficients c_k are given the following interpretation. At time t, if a measurement is carried out on the system to determine the value of its energy, the measurement yields the energy eigenvalue E_k with a probability $W(E_k)$ given by the relation

$$W(E_k) = |c_k|^2 \tag{5.63}$$

if the kth energy eigenstate is nondegenerate, and by

$$W(E_k) = \sum_{l=k}^{l=k+g-1} |c_l|^2 \tag{5.64}$$

if the kth energy eigenstate is g-fold degenerate.

The interpretation of expansion coefficients with respect to any set of eigenstates in terms of probabilities and the relation of the coefficients to results of measurements are discussed again in Chap. 6. Here it suffices to indicate that, because of the generality of Eq. 5.59, the study of states of systems having a time independent Hamiltonian is often restricted to the study of the solutions of the energy eigenvalue equation represented by Eq. 5.55. Once the energy eigenstates are found, any state ψ can be described by means of Eq. 5.59.

5.4.7 Conservation of Energy in Quantum Mechanics. For a conservative system in a state ψ, the expectation value F of an observable associated with the operator $\hat{F}$ is given by the relation

$$F = \int \psi^* \hat{F} \psi \, d^N q \tag{5.65}$$

If $\hat{F}$ is not an explicit function of time, the time rate of change of F is given by the expressions

$$\frac{dF}{dt} = \int \left(\frac{\partial \psi^*}{\partial t} \hat{F} \psi + \psi^* \hat{F} \frac{\partial \psi}{\partial t} \right) d^N q$$

$$= \frac{i}{\hbar} \int [(\hat{H}\psi)^* \hat{F} \psi - \psi^* \hat{F} (\hat{H} \psi)] \, d^N q$$

$$= \frac{i}{\hbar} \int \psi^* (\hat{H}\hat{F} - \hat{F}\hat{H}) \, \psi \, d^N q \tag{5.66}$$

where use has been made of the Schroedinger equation (Eq. 5.43) and of the Hermitian properties of the operators $\hat{F}$ and $\hat{H}$.† It is noteworthy that, in general,

$$\hat{H}\hat{F} \neq \hat{F}\hat{H} \tag{5.67}$$

and, therefore, dF/dt is different from zero. The last integral expression in Eq. 5.67 is called a *commutator.* Moreover, it is said that, when relation 5.67 is satisfied, the operators $\hat{F}$ and $\hat{H}$ *do not commute* with each other and, when it is not satisfied, the operators $\hat{H}$ and $\hat{F}$ *commute* with each other.

If $\hat{F}$ is taken equal to $\hat{H}$, the expectation value is E, and Eq. 5.66 yields

$$\frac{dE}{dt} = 0 \tag{5.68}$$

since $\hat{H}$ commutes with itself. In addition, since $\hat{H}$ commutes with any power of itself, it is also readily verified that

$$\frac{dE^n}{dt} = 0 \tag{5.69}$$

It follows from Eqs. 5.68 and 5.69 that the expectation value E of the energy and the probabilities of finding different energy eigenvalues of a system with a fixed $\hat{H}$ do not change with time. These conclusions are the quantum-mechanical equivalent of the *principle of conservation of energy*. Other conservation laws can be derived in an analogous manner.

†If $\hat{F}$ is an explicit function of time, the term $\partial F/\partial t$ must be added to the right-hand side of Eq. 5.66.

For a number of important relations and conclusions that can be obtained from the postulates of quantum mechanics, the reader is referred to other texts.[1,2]

In the following sections, the postulates and methods of quantum mechanics are applied to some specific systems chosen to illustrate special results obtained when the motion of particles is analyzed quantum mechanically and to provide information for problems of interest to thermionic conversion.

5.5 One-Dimensional Motion of a Free Particle

5.5.1 General Formulation. If a system consists of n noninteracting, structureless particles, each of mass m and each moving along the x-axis, its Hamiltonian function, H, is given by the relation

$$H = \sum_i^n \left[\frac{p_i^2}{2m} + V(x_i) \right] \tag{5.70}$$

and its Hamiltonian operator, $\hat{H}$, by

$$\hat{H} = \sum_i^n \left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x_i^2} + V(x_i) \right] \tag{5.71}$$

where x_i is the position, p_i is the momentum, and $V(x_i)$ is the potential energy of the ith particle.

For the quantum-mechanical analysis of the motion of these particles, it is sufficient to investigate the solutions of the energy eigenvalue equation given by the relation

$$\sum_i^n \left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x_i^2} + V(x_i) \right] u_m = E_m u_m \tag{5.72}$$

Because of lack of interactions between the particles, no potential energies in Eq. 5.72 depend on the coordinates of more than one particle. For this reason each energy eigenfunction u_m for the system can be written as a product of energy eigenfunctions given by the relation

$$u_m = u(x_1)\, u(x_2) \ldots u(x_n) \tag{5.73}$$

where the ith term, $u(x_i)$, in the product satisfies the eigenvalue equation

$$\left[-\frac{\hbar^2}{2m} \frac{d^2 u}{dx_i^2} + V(x_i) \right] u = \epsilon_i u \tag{5.74a}$$

or, more simply,

$$\left[-\frac{\hbar^2}{2m}\frac{d^2u}{dx^2} + V(x)\right]u = \epsilon u \tag{5.74b}$$

and the energy eigenvalue, E_m, of the system is given by the relation

$$E_m = \sum_{i=1}^{n} \epsilon_i \tag{5.75}$$

In words, for a system of n noninteracting particles, the energy eigenfunctions are products of single-particle energy eigenfunctions (Eq. 5.73), and the energy eigenvalues are sums of single-particle energy eigenvalues (Eq. 5.75). The single-particle energy eigenfunctions and eigenvalues are given by the single-particle eigenvalue problem described by Eqs. 5.74a and b. By virtue of these results, the subsequent discussion is devoted to the single-particle equation.

If the potential energy in Eq. 5.74b is independent of position, the equation is said to describe a *free particle.* The eigenfunctions for a free particle are of two general types, depending on the value of the difference between ϵ and V.

When the energy ϵ is smaller than the constant potential energy V, the general solution of Eq. 5.74b is of the form

$$u(x) = A \exp(\alpha x) + B \exp(-\alpha x) \tag{5.76}$$

where

$$\alpha = \frac{[2m(V-\epsilon)]^{1/2}}{\hbar} \tag{5.77}$$

and A and B are constants determined by the requirements that must be satisfied by all physically meaningful wave functions and by boundary conditions relevant to the system under investigation.

When ϵ is greater than the constant V, the general solution of Eq. 5.74b is of the form

$$u(x) = C \exp(i\beta x) + \mathrm{D} \exp(-i\beta x) \tag{5.78}$$

where

$$\beta = \frac{[2m(\epsilon - V)]^{1/2}}{\hbar} \tag{5.79}$$

and C and D are constants determined in a manner analogous to that for A and B.

Although each of the solutions given by Eqs. 5.76 and 5.78 holds for values of V independent of x, it is possible to use the same solutions in problems in which V undergoes sudden changes, e.g., from a constant value V_1 in one range of x to another constant value V_2 in a neighboring range of x. In such problems the appropriate expression for the general solution is written for each range of x, and then the constants appearing in these expressions are determined so that the eigenfunction u has the proper behavior at each discontinuity of V, namely, the expressions for u in the different ranges of x are matched at the discontinuities. The matching of different solutions proceeds as follows: In passing from one region of constant V to another, both u and du/dx must be continuous. In the problem under consideration, the discontinuities of V are neither sources nor sinks of particles, and therefore both the probabilities and the probability current densities must be continuous functions of x. In other words, u and du/dx must be continuous. Since u is continuous, it follows from the form of Eq. 5.74b that d^2u/dx^2 suffers a discontinuous change in value at each stepwise change in the value of V.

Stepwise-changing potential energy functions can be used to approximate different continuous functions $V(x)$. Thus approximate solutions for a variety of practical problems can be found. This point is illustrated by the following examples.

5.5.2 Transmission and Reflection at a Barrier. A collection of noninteracting, uniform energy particles impinging upon a potential barrier or "wall" of height V, as shown in Fig. 5.1, could be used to model approximately the motion of electrons of a given energy from the inside of a semiinfinite block of metal to the adjacent free space.

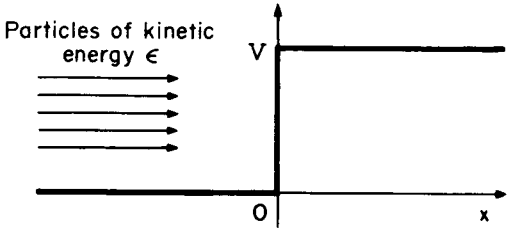


Fig. 5.1 Schematic of potential barrier of height V.

If the origin of the coordinate x is at the potential discontinuity (Fig. 5.1), then for x either negative or positive, the particles behave as free. Whenever a particle impinges upon the wall at x equal to zero, however, it suffers an impulsive force directed along the negative x direction.

In each region where the motion of the particles is free, the solution of the single-particle energy eigenvalue equation has either the form given by Eq. 5.76 or that given by Eq. 5.78. The exact form depends on the relative magnitudes of ϵ and V. Three cases can be distinguished.

1. $\epsilon < 0$. Here, for both negative and positive values of x, the solution is of the form given by Eq. 5.76 and can be written as

$$u(x) = \begin{cases} A \exp(\alpha' x) + B \exp(-\alpha' x) & \text{for } x < 0 \\ C \exp(\alpha x) + D \exp(-\alpha x) & \text{for } x > 0 \end{cases} \tag{5.80}$$

where

$$\alpha' = \frac{(-2m\epsilon)^{1/2}}{\hbar} \tag{5.81}$$

The constants B and C must be taken equal to zero to satisfy the requirements that the wave function vanishes both at x equal to $+\infty$ and at x equal to $-\infty$. On the other hand, the constants A and D must be taken equal to zero to satisfy the requirements that both u and du/dx are continuous at x equal to zero. The problem does not admit a solution.

2. $0 < \epsilon < V$. In this case the solution is of the form

$$u(x) = \begin{cases} A \exp(i\beta' x) + B \exp(-i\beta' x) & \text{for } x < 0 \\ C \exp(\alpha x) + D \exp(-\alpha x) & \text{for } x > 0 \end{cases} \tag{5.82}$$

where

$$\beta' = \frac{(2m\epsilon)^{1/2}}{\hbar} \tag{5.83}$$

For the wave function to vanish at $x = \pm\infty$ and for u and du/dx to be continuous at $x = 0$, the coefficients in Eq. 5.82 must satisfy the relations

$$C = 0 \qquad A + B = D \qquad i\beta'(A - B) = -\alpha D \tag{5.84}$$

It follows that

$$\frac{B}{A} = \frac{i\beta' + \alpha}{i\beta' - \alpha} \qquad \frac{D}{A} = \frac{2i\beta'}{i\beta' - \alpha} \tag{5.85}$$

and

$$u(x) = \begin{cases} A\left[\exp(i\beta' x) + \dfrac{i\beta' + \alpha}{i\beta' - \alpha}\exp(-i\beta' x)\right] & \text{for } x < 0 \\ A\dfrac{2i\beta'}{i\beta' - \alpha}\exp(-\alpha x) & \text{for } x > 0 \end{cases} \tag{5.86}$$

The value of the constant A can be determined from the normalization condition for the total probability. For the present problem, however, the normalization requires some special considerations that are not discussed here.

The quantity $|B/A|^2$, called the *reflection coefficient,* is a measure of the degree to which particles impinging on a potential-energy wall are reflected back. The quantity $|D/A|^2$, called the *transmission coefficient,* is a measure of the degree to which particles impinging on a potential-energy wall penetrate through it.

The results represented by Eq. 5.86 are different from those expected from a classical treatment of the problem. In classical mechanics a particle with kinetic energy ϵ impinging on a potential-energy barrier of height V greater than ϵ is totally reflected at the barrier. Quantum mechanically, the particle is also totally reflected since $|B/A| = 1$ (Eq. 5.85), but it does so after it penetrates the region that is classically forbidden since $|D/A| \neq 0$ (Eq. 5.85). The penetration is such that the probability of finding a particle in the forbidden region (x positive) is an exponentially decreasing function of the distance from the barrier. As the barrier height tends to infinity, the penetration tends to zero and the barrier becomes perfectly reflecting. For x negative the eigenfunction has the form of a standing wave owing to interference of waves moving in opposite directions. These results can be comprehended only in the light of the wave–particle duality introduced by quantum mechanics.

3. $\epsilon > V$. In this case the eigenfunction has the form

$$u(x) = \begin{cases} A\exp(i\beta' x) + B\exp(-i\beta' x) & \text{for } x < 0 \\ C\exp(i\beta x) + D\exp(-i\beta x) & \text{for } x > 0 \end{cases} \tag{5.87}$$

In the absence of particles originally approaching the wall from right to left (Fig. 5.1), the coefficient D must be equal to zero. Moreover, the continuity requirements yield

$$A + B = C \qquad \beta'(A - B) = \beta C \tag{5.88}$$

Therefore

$$\frac{B}{A}=\frac{\beta'-\beta}{\beta'+\beta} \quad \frac{C}{A}=\frac{2\beta'}{\beta'+\beta} \tag{5.89}$$

and

$$u(x)=\begin{cases} A\left[\exp(i\beta' x)+\dfrac{\beta'-\beta}{\beta'+\beta}\exp(-i\beta' x)\right] & \text{for } x<0 \\ A\dfrac{2\beta'}{\beta'+\beta}\exp(i\beta x) & \text{for } x>0 \end{cases} \tag{5.90}$$

Again the results are different from those derived from classical mechanics. For example, the particles are reflected, i.e., $|B/A| \neq 0$ (Eq. 5.89), in spite of the fact that the barrier height V is smaller than the kinetic energy ϵ.

We note that in both cases 2 and 3, no restrictions have been established for ϵ. All positive eigenvalues ϵ and all the corresponding eigenfunctions u are acceptable. In other words, the spectrum of energy eigenvalues forms a continuum.

5.5.3 Transmission and Reflection Through a Rectangular Barrier. Another example of a one-dimensional free-particle problem involving stepwise discontinuous potential energies is that of the motion of noninteracting particles

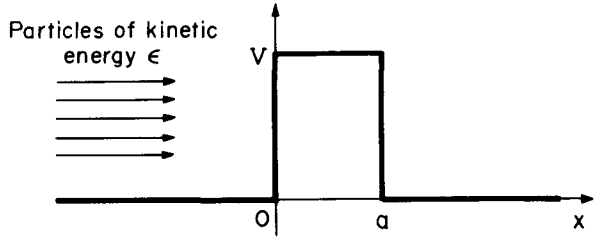


Fig. 5.2 Schematic of rectangular potential barrier of height V.

impinging on a rectangular potential barrier, as shown in Fig. 5.2. The particles move from left to right. This type of barrier is an approximation of the conditions prevailing at a metal–vacuum interface when electron thermionic emission occurs under the influence of an externally applied strong electric field.

Here again it is found that only positive energies ϵ are acceptable. For these energies two cases can be distinguished.

1. $0 < \epsilon < V$. The eigenfunction is given by the relation

$$u(x) = \begin{cases} A \exp(i\beta' x) + B \exp(-i\beta' x) & \text{for } x < 0 \\ C \exp(\alpha x) + D \exp(-\alpha x) & \text{for } 0 < x < a \\ F \exp(i\beta' x) & \text{for } \alpha < x \end{cases} \tag{5.91}$$

From the continuity conditions on u and du/dx at x equal to zero and a (Fig. 5.2), all the coefficients in Eq. 5.91 can be expressed in terms of one coefficient, say, A. Thus the transmission coefficient is given by the relation

$$\left|\frac{F}{A}\right|^2 = \left[1 + \frac{V^2 \sin h^2 \alpha a}{4\epsilon(V - \epsilon)}\right]^{-1} \tag{5.92}$$

Although the other ratios of coefficients can be readily computed, they are not given here, because the expressions are somewhat complicated.

2. $V < \epsilon$. In this case the eigenfunction is given by the relation

$$u(x) = \begin{cases} A \exp(i\beta' x) + B \exp(-i\beta' x) & \text{for } x < 0 \\ C \exp(i\beta x) + D \exp(i\beta x) & \text{for } 0 < x < a \\ F \exp(i\beta' x) & \text{for } a < x \end{cases} \tag{5.93}$$

and from the continuity conditions the transmission coefficient is found to be

$$\left|\frac{F}{A}\right|^2 = \left[1 + \frac{V^2 \sin^2 \beta a}{4\epsilon(\epsilon - V)}\right]^{-1} \tag{5.94}$$

The two transmission coefficients, Eqs. 5.92 and 5.94, describe the degree to which particles of different energies ϵ penetrate the rectangular barrier. We see from Eq. 5.92 that particles do penetrate the barrier even when their energy ϵ is smaller than V. This result is contrary to that expected from a classical analysis of the problem. Similarly, Eq. 5.94 indicates that, even if ϵ is greater than V, not all the particles go past the barrier, because $|F/A| < 1$ for $\epsilon > V$, except for specific values of ϵ for which $|F/A|$ equals unity.

The exact value of the transmission coefficient depends on the values of ϵ and V and of the thickness a of the barrier. A graph of $|F/A|^2$ vs. ϵ/V is shown in Fig. 5.3. It corresponds to a barrier for which[14] $V = 8\hbar^2/ma^2$. This figure shows that the transmission coefficient is less than unity for all energies up to about $1.5V$. Beyond that energy the coefficient oscillates between a maximum value of unity and gradually increasing minima.

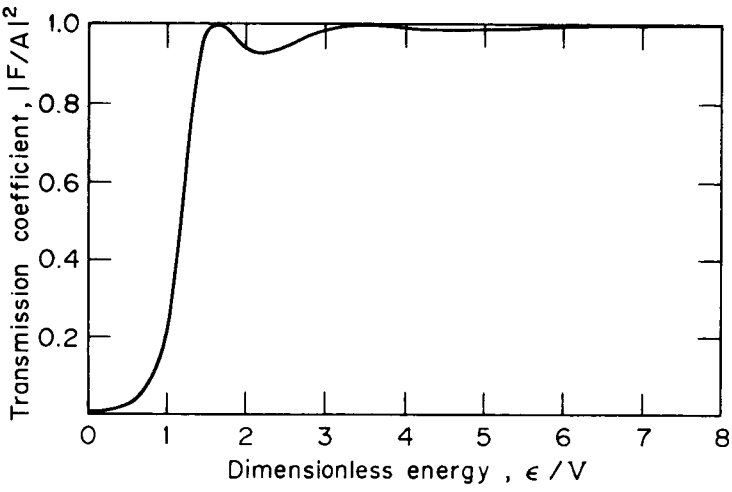


Fig. 5.3 Transmission coefficient $|F/A|^2$ for a rectangular barrier (adapted from Ref. 14) vs. dimensionless energy ϵ/V and for $V = \hbar/ma^2$.

Finally, we note that all positive-energy eigenvalues and corresponding eigenfunctions are acceptable, as they were for the barrier problem discussed in Sec. 5.5.2. An example of a system with both discrete- and continuous-energy eigenvalues is discussed in Sec. 5.5.4.

5.5.4 The Rectangular Potential Well. If a particle moves in a field derived from a potential energy represented by the potential "well" shown in Fig. 5.4, the well can be thought of as a stepwise approximation to a one-dimensional atom.

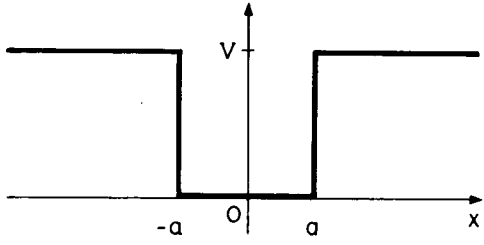


Fig. 5.4 Schematic of a potential well.

For simplicity the potential energy is taken equal to zero within a distance a on either side of the origin of the coordinate x and equal to $+V$ elsewhere. For particle energies smaller than zero, no solution exists. For energies $0 < \epsilon < V$, the general eigenfunction is given by the relation

$$u(x) = \begin{cases} A \exp(\alpha x) + B \exp(-\alpha x) & \text{for } x < -a \\ C \exp(i\beta' x) + D \exp(-i\beta' x) & \text{for } -a < x < a \\ F \exp(\alpha x) + G \exp(-\alpha x) & \text{for } a < x \end{cases} \tag{5.95}$$

The boundary conditions at $x = \pm\infty$ are satisfied if

$$B = F = 0 \tag{5.96}$$

On the other hand, the continuity requirements at x equal to $\pm a$ result in the following relations.

$$A \exp(-\alpha a) = C \exp(-i\beta' a) + D \exp(i\beta' a) \tag{5.97a}$$

$$\alpha A \exp(-\alpha a) = i\beta' [C \exp(i\beta' a) - D \exp(i\beta' a)] \tag{5.97b}$$

$$G \exp(-\alpha a) = C \exp(i\beta' a) + D \exp(-i\beta' a) \tag{5.98a}$$

$$-\alpha G \exp(-\alpha a) = i\beta' [C \exp(i\beta' a) - D \exp(-i\beta' a)] \tag{5.98b}$$

Eliminating A from Eqs. 5.97a and 5.97b and G from Eqs. 5.98a and 5.98b yields

$$C = \pm D \tag{5.99}$$

Thus, from Eqs. 5.97a and 5.98a, we find that

$$A = G \quad \text{for } C = D \tag{5.100a}$$

and

$$A = -G \quad \text{for } C = -D \tag{5.100b}$$

The meaning of Eqs. 5.100a and 5.100b is that each eigenfunction exhibits the even or odd symmetry or *parity* with respect to x which is always associated with a Hamiltonian operator that is symmetric in x.

Equations 5.97a and 5.97b (or Eqs. 5.98a and 5.98b) are compatible with each other if and only if certain conditions are satisfied by the exponents. Specifically,

$$\cot \beta' a = \frac{\beta'}{\alpha} \quad \text{for } C = D \tag{5.101}$$

and

$$\tan \beta' a = -\frac{\beta'}{\alpha} \quad \text{for } C = -D \tag{5.102}$$

Either Eq. 5.101 or 5.102 is satisfied only for discrete values of the energy. This fact can be verified by rewriting these two equations in the forms

$$\cot \xi = \frac{\xi}{(\zeta^2 - \xi^2)^{1/2}} \tag{5.103}$$

and

$$-\tan \xi = \frac{\xi}{(\zeta^2 - \xi^2)^{1/2}} \tag{5.104}$$

where

$$\xi = \frac{a}{\hbar}(2m\epsilon)^{1/2} \qquad \zeta = \frac{a}{\hbar}(2mV)^{1/2} \tag{5.105}$$

For given numerical values of the quantities V and a, solutions of the transcendental Eqs. 5.103 and 5.104 can be found graphically. Such solutions are shown in Fig. 5.5 for the particular values V, m, and a, which obey the relation[15]

$$V = \frac{49h^2}{128ma^2}$$

It is seen from this figure that only four values of ξ, and hence of the energy ϵ, satisfy Eqs. 5.103 and 5.104. All these values of ξ are smaller than ζ; i.e., all the energy eigenvalues are smaller than V. Graphs of the corresponding eigenfunctions vs. x are shown in Fig. 5.6.

Figure 5.5 also indicates that as ζ approaches infinity (i.e., as the walls of the well become infinitely high), the solutions of Eqs. 5.103 and 5.104 are uniformly spaced at intervals of $\pi/2$. Thus the energy eigenvalues are given by the general relation

$$\epsilon_n = \frac{n^2\pi^2\hbar^2}{8ma^2} = \frac{n^2h^2}{32ma^2} \quad \text{for } n = 1, 2, \ldots, \text{ and } V = \infty \tag{5.106}$$

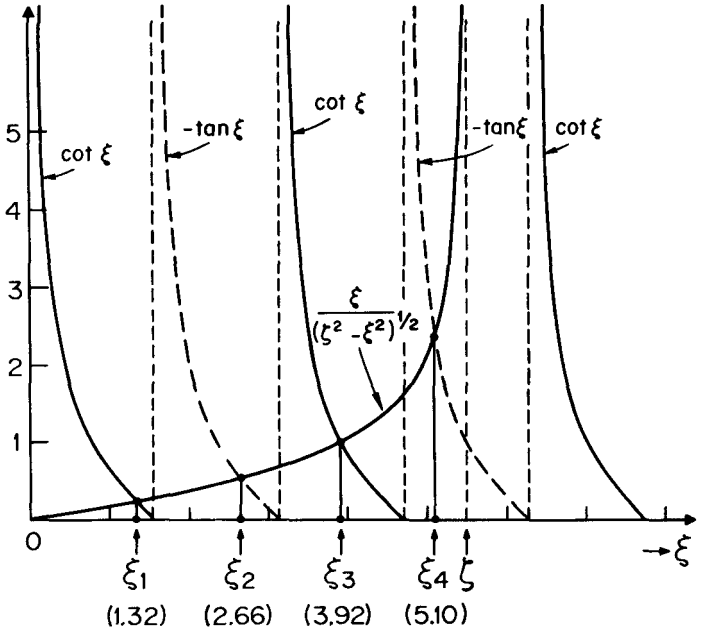


Fig. 5.5 Solutions of transcendental equations (from Ref. 15) $\cot \xi = \xi/(\zeta^2 - \xi^2)^{1/2}$ and $-\tan \xi = \xi/(\zeta^2 - \xi^2)^{1/2}$.

For energies ϵ greater than V, both discrete and continuous energy values are found to be acceptable. Thus it is concluded that the stationary states of a particle in the presence of a potential well consist of a set of discrete energy levels plus a continuum.

5.6 Particle in a Container

The investigation of the quantum-mechanical states of a particle of mass m moving inside a container constitutes the first step for many subsequent analyses of interest to thermionic conversion.

In terms of Cartesian coordinates, the Hamiltonian function, H, and the corresponding Hamiltonian operator, $\hat{H}$, for the particle are given by Eqs. 5.46 and 5.47, which are repeated here for convenience:

$$H = \frac{1}{2m}\left(p_x^2 + p_y^2 + p_z^2\right) + V(x,y,z) \tag{5.107}$$

$$\hat{H} = -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial}{\partial z^2}\right) + V(x,y,z) \tag{5.108}$$

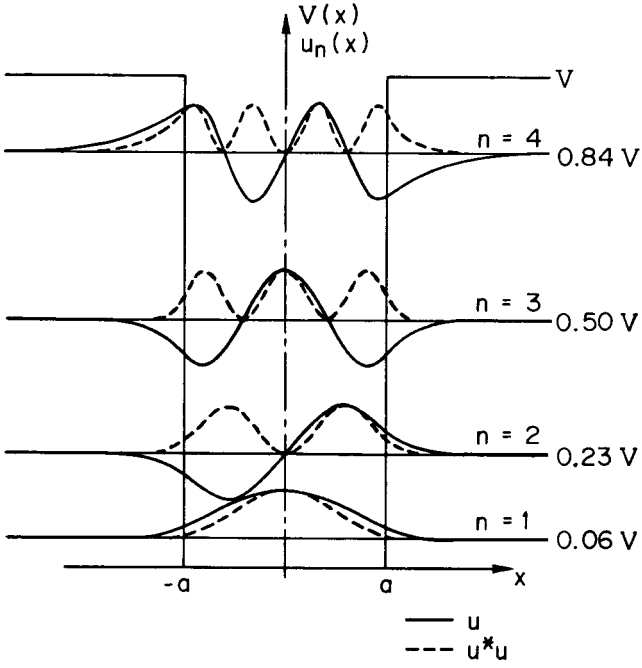


Fig. 5.6 Energy eigenfunctions corresponding to the discrete eigenvalues in Fig. 5.5.

Therefore the stationary states of the particle are described by the energy eigenvalue equation given by the relation

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right)+V(x,y,z)\,u=\epsilon u \tag{5.109}$$

For mathematical simplicity, the container may be regarded as a rectangular box, with sides l_1, l_2, and l_3, located in the first octant of the Cartesian system of coordinates (Fig. 5.7), and the potential energy can be taken equal to zero inside the box and to infinity at the walls. The infinite potential energy at the walls implies that the particle is restricted to remain in the box.

For V equal to zero, the general solution of Eq. 5.109 is given by the relation

$$u_k = A \sin(k_x x + \delta_x) \sin(k_y y + \delta_y) \sin(k_z z + \delta_z) \tag{5.110}$$

where A, k_x, k_y, k_z, δ_x, δ_y, δ_z are constants to be subsequently determined. This solution can also be written in the equivalent form of an exponential function, the so-called *plane-wave* form.

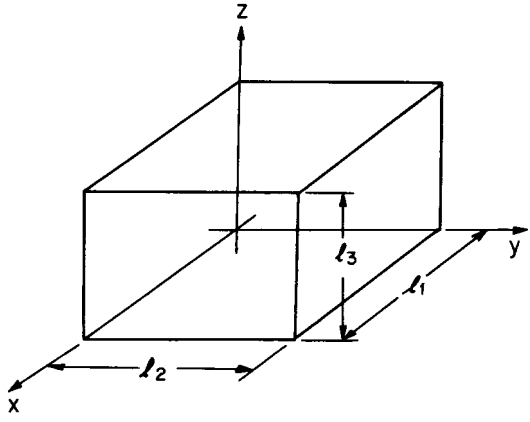


Fig. 5.7 Schematic of cubical container.

Substituting Eq. 5.110 into Eq. 5.109 yields

$$|\mathbf{k}|^2 = k_x^2 + k_y^2 + k_z^2 = \frac{2m}{\hbar^2}\epsilon_k \tag{5.111}$$

where $\mathbf{k}$ is a vector with components k_x, k_y, and k_z and ϵ_k is the energy eigenvalue. In addition, since the potential energy is infinite at the walls, the wave function must be equal to zero there. Otherwise, a finite probability of finding the particle outside the box would exist. The requirement is satisfied if the constants in the arguments of the sinusoidal solution are given by the relations

$$\delta_x = \delta_y = \delta_z = 0 \tag{5.112}$$

and

$$k_x = \frac{n_1\pi}{l_1} \qquad k_y = \frac{n_2\pi}{l_2} \qquad k_z = \frac{n_3\pi}{l_3} \tag{5.113}$$

where n_1, n_2, and n_3, are positive integers. Finally, the normalization of the energy eigenfunction u_k yields that

$$A = \left(\frac{8}{l_1 l_2 l_3}\right)^{1/2} \tag{5.114}$$

In summary, for a particle in a box with zero potential energy inside and infinite potential energy at the walls, the normalized energy eigenfunctions u_k are given by the relation

$$u_k = \left(\frac{8}{l_1 l_2 l_3}\right)^{1/2} \sin\frac{n_1 \pi x}{l_1} \sin\frac{n_2 \pi y}{l_2} \sin\frac{n_3 \pi z}{l_3} \tag{5.115}$$

The corresponding energy eigenvalues ϵ_k are given by the expression

$$\epsilon_k = \frac{\hbar^2 |\mathbf{k}|^2}{2m} = \frac{\pi^2 \hbar^2}{2m}\left(\frac{n_1^2}{l_1^2} + \frac{n_2^2}{l_2^2} + \frac{n_3^3}{l_3^2}\right) = \frac{\hbar^2}{8m}\left(\frac{n_1^2}{l_1^2} + \frac{n_2^2}{l_2^2} + \frac{n_3^2}{l_3^2}\right) \tag{5.116}$$

Each of the orthonormal energy eigenfunctions $u_k(x,y,z)$ corresponds to a different choice of *quantum numbers* n_1, n_2, n_3 or of $\mathbf{k}$ (Eq. 5.114) and to a different discrete energy eigenvalue ϵ_k. The complete set of these functions can be used to represent any quantum-mechanical state ψ. When any two of the lengths l_1, l_2, and l_3 are commensurable, the energy eigenstates are degenerate because then different eigenfunctions may correspond to the same energy eigenvalue.

In subsequent discussions the need arises for the number of energy eigenstates corresponding to eigenvalues lying within the energy range between ϵ and $\epsilon + d\epsilon$. This number can be determined with the help of an orthogonal coordinate system with axes for plotting the values of the quantum numbers n_1, n_2, and n_3 (Fig. 5.8). In a cubic lattice of points whose coordinates are integral values of the quantum numbers n_1, n_2, and n_3 and in the absence of degeneracies, each point corresponds to one eigenstate of the system. Moreover, the interior of the lattice can be regarded as being made of unit cubicals, each of which can be associated with one lattice point. If the energy ϵ is not too small, it follows from this simple geometric interpretation that the total number of energy eigenstates corresponding to eigenvalues smaller than ϵ, i.e., the number of discrete values of $\mathbf{k}$, can be found by computing the volume of the lattice over the range of quantum numbers that satisfy the relation

$$\frac{n_1^2}{l_1^2} + \frac{n_2^2}{l_2^2} + \frac{n_3^2}{l_3^2} \leq \frac{8m}{h^2}\epsilon \tag{5.117}$$

Inequality 5.117 can be regarded as representing the inside of an ellipsoid in (n_1, n_2, n_3) space, with semiaxes $l_1(8m\epsilon/h^2)^{1/2}$, $l_2(8m\epsilon/h^2)^{1/2}$, and $l_3(8m\epsilon/h^2)^{1/2}$. The volume of such an ellipsoid is

$$\frac{4}{3}\pi\left(\frac{8m\epsilon}{h^2}\right)^{3/2} l_1 l_2 l_3$$

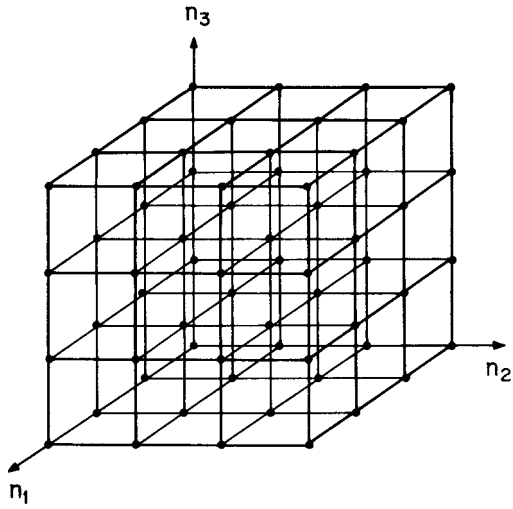


Fig. 5.8 Cubic lattice.

The quantum numbers of interest here are positive and therefore correspond to one octant of the ellipsoid with a volume equal to

$$\frac{1}{6}\pi\left(\frac{8m\epsilon}{h^2}\right)^{3/2} l_1 l_2 l_3$$

This volume is also equal to the number of quantum states with energies smaller than ϵ, and its differential with respect to ϵ gives the number $N(\epsilon)\,d\epsilon$ of states lying within the energy range between ϵ and $\epsilon + d\epsilon$, i.e.,

$$N(\epsilon)\,d\epsilon = \frac{4\pi m V}{h^3}(2m\epsilon)^{1/2}\,d\epsilon \tag{5.118}$$

The volume V of the container, given by the relation

$$V = l_1 l_2 l_3 \tag{5.119}$$

should not be confused with the potential energy V. The quantity $N(\epsilon)$ is called the *density of states per unit energy.*

As discussed later, when the particle under consideration has a spin, the states are twofold degenerate and the density of states is twice that given by Eq. 5.118, namely,

$$N(\epsilon) = \frac{8\pi m V}{h^3}(2m\epsilon)^{1/2} \tag{5.120}$$

5.7 Two Interacting Particles

5.7.1 General Remarks. The foregoing applications refer to identical, noninteracting particles in different force fields. The present section is concerned with the quantum-mechanical states of two nonidentical, interacting particles of masses m_1 and m_2. The results provide useful information for the understanding of the structure of the elements of the periodic table and for approximate analyses of many-particle problems, such as those related to molecules and solids.

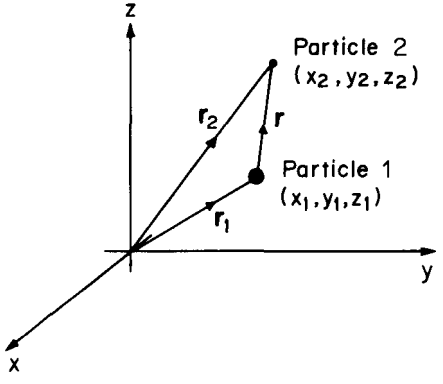


Fig. 5.9 Schematic of Cartesian coordinates of two structureless particles.

A system of two interacting structureless particles, shown as particles 1 and 2 in Fig. 5.9, has six degrees of freedom. In accordance with the preceding discussions, the Hamiltonian operator for the system is given by the relation

$$\hat{H} = -\frac{\hbar^2}{2m_1}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial z_1^2}\right) - \frac{\hbar^2}{2m_2}\left(\frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial y_2^2} + \frac{\partial^2}{\partial z_2^2}\right) + V(x_1, y_1, z_1, x_2, y_2, z_2) \tag{5.121}$$

and the energy eigenstates are determined by the eigenvalue equation

$$-\frac{\hbar^2}{2m_1}\left(\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial y_1^2} + \frac{\partial^2 u}{\partial z_1^2}\right) - \frac{\hbar^2}{2m_2}\left(\frac{\partial^2 u}{\partial x_2^2} + \frac{\partial^2 u}{\partial y_2^2} + \frac{\partial^2 u}{\partial z_2^2}\right) + Vu = Eu \tag{5.122}$$

For a potential energy V that is an arbitrary function of the coordinates of both particles, Eq. 5.122 cannot be solved explicitly. Even for this simple two-particle system, only approximate results are possible. These results are obtained by a variety of techniques.

The mathematical complexity of Eq. 5.122 is reduced if the potential energy is the sum of two parts, one depending only on the coordinates of the center of mass of the two particles and the other depending only on the relative distance between particles 1 and 2. In terms of the coordinates X, Y, and Z of the center of mass,

$$X = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2} \qquad Y = \frac{m_1 y_1 + m_2 y_2}{m_1 + m_2} \qquad Z = \frac{m_1 z_1 + m_2 z_2}{m_1 + m_2} \tag{5.123a}$$

and of the differences x, y, and z of the coordinates of the two particles,

$$x = x_2 - x_1 \qquad y = y_2 - y_1 \qquad z = z_2 - z_1 \tag{5.123b}$$

it can be shown that Eq. 5.122 assumes the form

$$-\frac{\hbar^2}{2(m_1 + m_2)}\left(\frac{\partial^2 u}{\partial X^2} + \frac{\partial^2 u}{\partial Y^2} + \frac{\partial^2 u}{\partial Z^2}\right) - \frac{\hbar^2}{2m_r}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) + [V_{\text{ext}}(X,Y,Z) + V_{\text{int}}(x,y,z)]\,u = Eu \tag{5.124}$$

where m_r is the *reduced mass of the two particles* defined by the relation

$$m_r = \frac{m_1 m_2}{m_1 + m_2} \tag{5.125}$$

V_{ext} is the part of the potential energy which depends only on the coordinates of the center of mass, and V_{int} is the part of the potential energy which depends only on the relative distance between the two particles. In this form the Hamiltonian operator and the corresponding energy eigenvalue equation are said to be *separable in the two sets of coordinates* (X,Y,Z) and (x,y,z). Separability means that the solutions u of Eq. 5.124 can be written in the form

$$u = u_{\text{ext}}(X,Y,Z)\,u_{\text{int}}(x,y,z) \tag{5.126}$$

where the functions u_{ext} and u_{int} satisfy the eigenvalue relations

$$\hat{H}_{\text{ext}} u_{\text{ext}} = -\frac{\hbar^2}{2(m_1 + m_2)}\left(\frac{\partial^2 u_{\text{ext}}}{\partial X^2} + \frac{\partial^2 u_{\text{ext}}}{\partial Y^2} + \frac{\partial^2 u_{\text{ext}}}{\partial Z^2}\right) + V_{\text{ext}} u_{\text{ext}}$$

$$= \epsilon_{\text{ext}} u_{\text{ext}} \tag{5.127}$$

and

$$\hat{H}_{\text{int}} u_{\text{int}} = -\frac{\hbar^2}{2m_r}\left(\frac{\partial^2 u_{\text{int}}}{\partial x^2} + \frac{\partial^2 u_{\text{int}}}{\partial y^2} + \frac{\partial^2 u_{\text{int}}}{\partial z^2}\right) + V_{\text{int}} u_{\text{int}} = \epsilon_{\text{int}} u_{\text{int}} \tag{5.128}$$

$$\hat{H}_{\text{ext}} + \hat{H}_{\text{int}} = \hat{H} \tag{5.129}$$

and

$$\epsilon_{\text{ext}} + \epsilon_{\text{int}} = E \tag{5.130}$$

Each of the quantities ϵ_{ext} and ϵ_{int} is a constant independent of any coordinates. Thus the solution of the original eigenvalue problem, Eq. 5.122, is reduced to that of two eigenvalue problems, Eqs. 5.127 and 5.128. The two eigenvalue problems may be easier to solve for given V_{ext} and V_{int} than the original problem for the total V.

Equations 5.126 to 5.130 illustrate a general result of quantum mechanics. If the Hamiltonian operator for any given system is separable into two independent parts (Eq. 5.129), each of which is expressed in terms of coordinates not encountered in the expression for the other part, the energy eigenfunctions can be expressed as a product of two separate eigenfunctions (Eq. 5.126), and the energy eigenvalues can be expressed as a sum of two separate eigenvalues (Eq. 5.130). The partial eigenfunctions and eigenvalues are related, as illustrated by Eqs. 5.127 and 5.128. This result can be readily extended to any number of parts of the Hamiltonian operator, namely, any number of sets of separable degrees of freedom (coordinates or quantum-mechanical observables). An example of this generalization is discussed in Sec. 5.5.1 in connection with the system of n identical, noninteracting particles.

Since Eq. 5.127 has the same form as Eq. 5.109 for a particle with mass $m_1 + m_2$, it can be treated by the same methods as those used for a single particle with potential energy V_{ext}. The allowable energy eigenvalues ϵ_{ext} can be either continuous or discrete or both depending on the exact form of V_{ext}.

On the other hand, Eq. 5.128 can be viewed as the equation that describes the stationary states associated with the relative motion of particle 2 with respect to particle 1 under the influence of their mutual interaction. For this motion, particle 1 can be regarded as immobile, resulting only in a force on particle 2, whereas particle 2 behaves as if it had a reduced mass m_r, given by Eq. 5.125, and such that

$$m_r \approx m_2 \qquad \text{for } m_1 \gg m_2 \tag{5.131}$$

Explicit solutions of Eq. 5.128 can be found only for certain functions $V_{\text{int}}(x,y,z)$. A class of such solutions is discussed in the next section.

5.7.2 Electrostatic or Central-Force Interaction Potential Energy. The exact solutions of Eq. 5.128 can be found when the two particles interact via a central or electrostatic force.

For electrostatic forces, V_{int} can be expressed as

$$V = -\frac{Ze^2}{4\pi\varepsilon_0(x^2+y^2+z^2)^{1/2}} \tag{5.132}$$

where ε_0 is the dielectric constant of vacuum ($\varepsilon_0 = 8.854 \times 10^{-14}$ f/cm). The subscript "int" is omitted here and from all subsequent expressions. Such a potential energy arises when a particle with a negative charge $-e$ moves in the vicinity of another particle with a positive charge $+Ze$.

When Eq. 5.132 is substituted into Eq. 5.128, the exact eigenfunctions and eigenvalues are determined by a series of mathematical steps. These steps are somewhat involved[2,12] and are not given here. Instead, only some of the important features of the results are summarized in the following paragraphs.

Two classes of energy eigenstates are found. In the first class, which corresponds to the continuum of all positive energy eigenvalues, the states are called *ionized states.* In the second class, which corresponds to discrete negative energy eigenvalues, the states are called *bound states.* The discrete energy eigenvalues are given by the relation

$$\epsilon_n = -\frac{m_r Z^2 e^4}{32\pi^2\varepsilon_0^2\hbar^2 n^2} \tag{5.133}$$

for different positive values of the integral quantum number n ($n \geq 1$).

In general, several eigenfunctions correspond to each eigenvalue. These eigenfunctions are characterized by three integral quantum numbers, n, l, and m. Each integral quantum number n defines one energy eigenvalue (Eq. 5.133) and can be combined with integral values l varying from zero to $n-1$. Furthermore, each value of l can be combined with integral values of m varying from $-l$ to l. It follows that each eigenstate is n^2-fold degenerate.

Each quantum number l corresponds to an eigenvalue of the magnitude L_0 of the angular momentum of the two particles. Specifically, a given energy eigenfunction is also an eigenfunction of the quantum-mechanical operator that corresponds to the square of the magnitude of the angular momentum. The quantized values of the angular momentum L_0 are given by the relation

$$L_0^2 = l(l+1)\hbar^2 \tag{5.134}$$

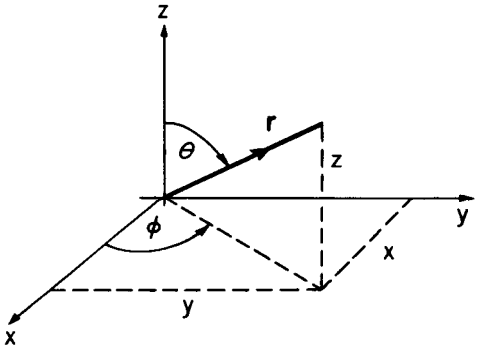


Fig. 5.10 Spherical coordinates r, ϕ, and θ.

Each quantum number m corresponds to an eigenvalue of the z-component of the angular momentum. The quantized value of this component is equal to $\pm m\hbar$. This result implies that a given energy eigenfunction is also an eigenfunction of the quantum-mechanical operator that corresponds to the z-component of the angular momentum.

Each eigenfunction is denoted by three subscripts, such as u_{nlm}, to reflect the facts that it depends on three quantum numbers and that it is an eigenfunction of three different operators. A complete set of orthonormal energy eigenfunctions, including each of the n^2 functions of the n^2-fold degenerate eigenstates, can be constructed. This set is expressed in terms of spherical coordinates (r, ϕ, θ) shown in Fig. 5.10 instead of Cartesian coordinates (x, y, z). The first few members of the set are given by the relations

$$u_{100} = (\pi)^{-1/2} \left(\frac{Z}{a_0'}\right)^{3/2} \exp\left(-\frac{Zr}{a_0'}\right)$$

$$u_{200} = (32\pi)^{-1/2} \left(\frac{Z}{a_0'}\right)^{3/2} \left(2 - \frac{Zr}{a_0'}\right) \exp\left(-\frac{Zr}{2a_0'}\right)$$

$$u_{210} = (32\pi)^{-1/2} \left(\frac{Z}{a_0'}\right)^{3/2} \left(\frac{Zr}{a_0'}\right) \exp\left(-\frac{Zr}{2a_0'}\right) \cos\theta \qquad (5.135)$$

$$u_{21\pm1} = (64\pi)^{-1/2} \left(\frac{Z}{a_0'}\right)^{3/2} \left(\frac{Zr}{a_0'}\right) \exp\left(-\frac{Zr}{2a_0'}\right) \sin\theta \exp(\pm i\phi)$$

$$u_{300} = \frac{(Z/a_0')^{3/2}}{81\,(3\pi)^{1/2}} \left(27 - 18\frac{Zr}{a_0'} + 2\frac{Z^2r^2}{a_0'^2}\right) \exp\left(-\frac{Zr}{3a_0'}\right)$$

where

$$a_0' = a_0 \left(1 + \frac{m_2}{m_1}\right) \tag{5.136}$$

and

$$a_0 = \frac{4\pi\varepsilon_0\hbar^2}{m_2 e^2} \tag{5.137}$$

The quantity a_0 is called the *Bohr radius* when m_2 is the mass of the electron; it is then equal to 5.29171×10^{-9} cm or 0.529171 Å.

In many applications each of the energy eigenfunctions u_{nlm} is said to define an *orbit* in which particle 2 is found when it is in the stationary state characterized by the quantum numbers n, l, and m. The quantum-mechanical meaning of an orbit, however, differs substantially from the ordinary usage of the word. Each u_{nlm} is defined throughout coordinate space, and its value at a particular position is related to the probability of finding the particle at this position as a result of a measurement. This probability may be zero in certain regions of space. For example, for the u_{210}-orbit the probability that the particle will be found along the directions θ equal to $\pm\pi/2$ is identically zero, and the probability along the directions θ equal to 0 or π is large; in other words, orbits may have a "lobe" structure throughout coordinate space.

The foregoing results for Z equal to unity, m_1 the mass of the proton, and m_2 the mass of the electron describe approximately the energy eigenstates of the hydrogen atom. The description is approximate because spin and relativistic effects are not included in the analysis. Some of the important consequences of the consideration of spin effects are discussed in Sec. 5.7.3.

5.7.3 Electron Spin. A number of experimental observations and theoretical results suggest that electrons possess a *spin* associated with an *intrinsic angular momentum* and an *intrinsic magnetic moment*. The magnitude of the spin angular momentum parallel to any coordinate axis can exhibit only the eigenvalues $m_s\hbar$, where

$$m_s = \pm\tfrac{1}{2} \tag{5.138}$$

Moreover, the total magnitude of this momentum has the eigenvalue $[s(s+1)]^{1/2}\ \hbar$, where $s = \frac{1}{2}$. The spin magnetic moment is proportional to the spin angular momentum. The constant of proportionality is equal to $-e/m_e$, where m_e is the mass of the electron. It follows that the component of the

magnetic moment along any axis is quantized and can have only two eigenvalues, $\pm e\hbar/2m_e$, each of which is called a *Bohr magneton.*

The intrinsic properties of the electron have several unique features worth emphasizing. First, in contrast to the integral values of the quantum numbers l and m for the orbital angular momentum discussed in Sec. 5.7.2, the quantum numbers s and m_s are half integers. Second, in contrast to the usual case of an infinite number of discrete and continuous eigenvalues associated with other quantum-mechanical observables, spin exhibits only two eigenvalues. Third, spin has no analog in classical mechanics. And, finally, the constant of proportionality between spin magnetic moment and spin angular momentum is twice as large as that which would be found for the corresponding quantities of an orbiting electron.

Discussion of the procedure used for constructing operators that account for the spin properties of electrons and of the methods used for including these operators in quantum-mechanical studies is beyond the scope of the present elementary exposition of quantum mechanics. Only some of the important spin effects are summarized here.

Spin effects must be included in studies of systems with many electrons, e.g., atoms. For this purpose several considerations must be taken into account. First, the total angular momentum and its conservation in fields of suitable symmetry must be expressed in terms of both spin and orbital momentum. Second, eigenstates of atoms involving more than a single electron correspond to eigenfunctions that are antisymmetric both in the coordinate and in the spin variables. This topic, which is related to the Pauli exclusion principle, is discussed in detail in Sec. 5.8. Third, the Hamiltonian operator for the electrons must include terms that classically correspond to the interactions of the intrinsic magnetic moment with an imposed magnetic field and with the internal electromagnetic field the electron may encounter in its orbital motion. The latter is called the *spin–orbit interaction.* The classical forces resulting from differentiation of the energy of the spin–orbit interaction with respect to position coordinates are called *spin–orbit forces.*

5.7.4 The Hydrogen Atom. When spin and relativistic effects are neglected, the bound states of the hydrogen atom are described by the results of Sec. 5.7.2. The Hamiltonian operator $\hat{H}$ is given by the relation

$$\hat{H} = -\frac{\hbar^2}{2m_r}\nabla^2 - \frac{e^2}{4\pi\varepsilon_0 r} \tag{5.139}$$

where m_r is almost equal to the mass of the electron. This operator and related quantities are qualified as unperturbed.

Spin and relativistic effects can be included in the analysis by adding two other operators to $\hat{H}$, one for spin–orbit interactions and one for relativistic corrections. The exact eigenstates of the resulting overall operator, however, are difficult to compute. The difficulty is avoided by treating the two additional operators as small perturbations of the unperturbed Hamiltonian operator $\hat{H}$. Thus approximate energy eigenstates having the following properties are obtained by means of perturbation theory.

1. The degeneracy of the unperturbed states is removed. The perturbed and unperturbed eigenvalues, however, differ only slightly.
2. The number of allowable energy levels is two times that of the unperturbed system.
3. The dependence of the eigenfunctions on the spatial coordinates is almost the same as that of the unperturbed u_{nlm} discussed in Sec. 5.7.2; in other words, the energy eigenfunctions are given by relations 5.135 for Z equal to unity and a_0' approximately equal to the Bohr radius a_0.

Each orbit u_{nlm} is called an *atomic orbital.* An atomic orbital occupied by an electron with a definite spin is called a *spin-orbital*, and the corresponding probability density can be given a useful, although unrealistic, pictorial interpretation.[16] The electron is regarded as a charge cloud spread around the nucleus. The probability density associated with the spin-orbital is interpreted as being proportional to the charge density ρ of the charge cloud. In locations where $|u_{nlm}|^2$ is large, the charge density is also large. This interpretation is helpful for the qualitative understanding of many physical phenomena.

Atomic orbitals can be represented graphically in a variety of ways. On a given plane going through the nucleus, the following quantities can be shown: (1) u or $|u|^2$ vs. r; (2) contours of constant values of the magnitude of u; (3) the radial charge density $4\pi r^2 \rho$ vs. r; (4) a boundary surface (a boundary surface is one of the contours referred to in item 2 in which the total charge outside the contour is a definite small percentage, say 10%, of the total electronic charge).

For example, to a good approximation the lowest energy orbital u_{100} of the hydrogen atom is given by

$$u_{100} = (\pi a_0^3)^{-\frac{1}{2}} \exp\left(-\frac{r}{a_0}\right)$$

and its energy is given by

$$\epsilon_1 = -\frac{me^4}{32\pi^2 \epsilon_0^2 \hbar^2} = -13.6 \text{ ev}$$

Different graphical representations of u_{100} are shown in Fig. 5.11. Diagrams a, b, and c are detailed and require an exact knowledge of the dependence of u_{100}

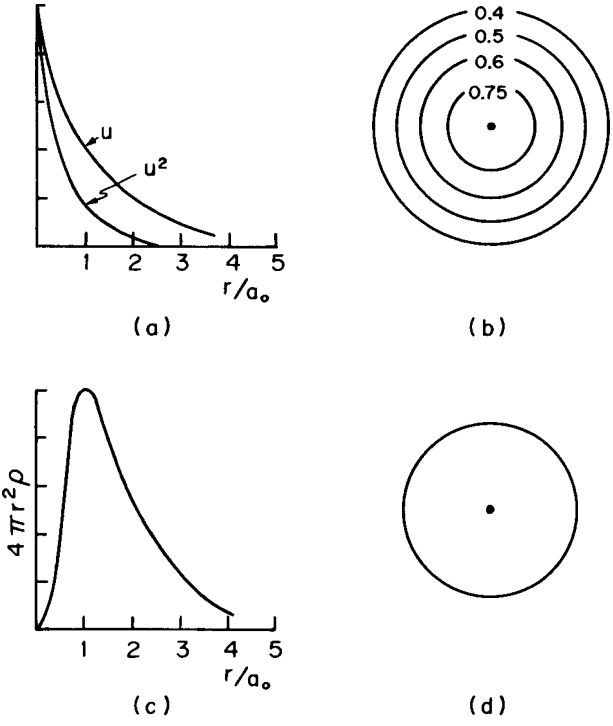


Fig. 5.11 Graphic representation of the u_{100} orbital of the hydrogen atom. (a) Graph of u and u^2 vs. r. (b) Contours of constant values of $u/(\pi a_0^3)^{-\frac{1}{2}}$. (c) Graph of $4\pi r^2\rho$ vs. r. (d) A boundary surface.

on r. On the other hand, diagram d is very simple and for many purposes provides surprisingly adequate information. In essence it represents the atom as a sphere with a well-defined radius; the radius may be taken as the distance at which the radial charge density is maximum. It is interesting to note in Fig. 5.11c that the maximum radial charge density occurs at a distance of one Bohr radius. The quantities u_{100} and ϵ_1 are said to define the *ground state* of the hydrogen atom, and the negative of ϵ_1 is called the *ionization energy*.

The atomic orbital of the ground state is spherically symmetric, and the diagrams of Fig. 5.11 are independent of the plane on which they are plotted. The other atomic orbitals are not spherically symmetric, and therefore the shape of, say, a boundary surface depends on the orientation of the plane through the nucleus. Boundaries of some of the more frequently used atomic orbitals are shown in Fig. 5.12.

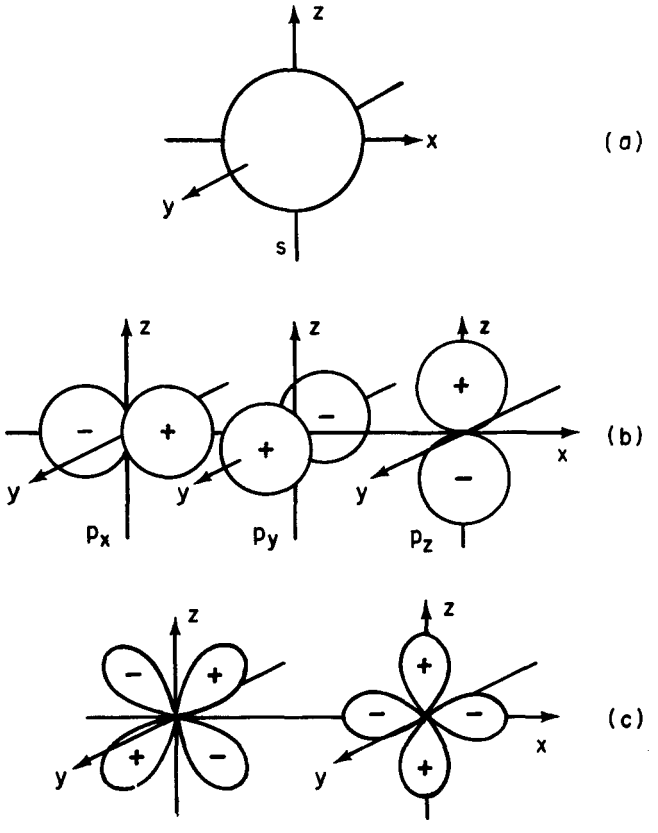


Fig. 5.12 Approximate boundary surfaces for different types of orbitals. (a) s orbitals. (b) p orbitals. (c) d orbitals.

Atomic orbitals are classified into different types denoted by the letters s, p, d, f, etc. Examples of such orbitals are shown in Fig. 5.12. Type s orbitals are spherically symmetric (Fig. 5.12a); in other words, the charge-cloud density is a function of r only. All orbitals of the form u_{n00} (Eq. 5.135) are type s. The distinction between different values of n is made by characterizing u_{100} as $1s$, u_{200} as $2s$, and u_{n00} as ns.

Atomic orbitals other than s are not spherically symmetric. For example, for each of the three type p orbitals, the boundary surface consists of two regions that together resemble a dumbbell. In one half of the dumbbell, the eigenfunction is positive, and in the other it is negative, as indicated by the ± signs in Fig. 5.12b. The important feature of type p orbitals is their directional character. The three p orbitals are entirely equivalent but linearly independent.

Therefore any other similar p orbital can be regarded as the superposition of certain amounts of the basic orbitals shown in Fig. 5.12b.

For type p orbitals the value of the quantum number l equals unity, and the lowest value of n is two. Each of these orbitals is threefold degenerate since the quantum number m may assume one of the values -1, 0, or 1. The distinction between different values of m is made by characterizing u_{21m} as $2p$, u_{31m} as $3p$, etc.

The degeneracy and the dependence on different quantum numbers of different types of orbitals are summarized in Table 5.1. Also included in the

Table 5.1 Characterization of Some Atomic Orbitals in Terms of Quantum Numbers

Type of orbital	Degeneracy	l value	Lowest n value	Spin number m_s
s	1	0	1	$\pm 1/2$
p	3	1	2	$\pm 1/2$
d	5	2	3	$\pm 1/2$
f	7	3	4	$\pm 1/2$

table is the fact that any given orbital may correspond to an electron with a spin number of $\pm 1/2$. Other important features of atomic orbitals are not discussed here because they are not of interest to the purposes of this book.

5.8 Symmetry Considerations

This section presents some considerations that must be reflected in the quantum-mechanical analysis of systems consisting of two or more indistinguishable particles.[17] These considerations result from experimental observations of intrinsic properties of different kinds of particles.

As an illustration of the need for such considerations, let us consider two identical particles in a box in a stationary state. Since the particles are identical, it follows that the formalism for treating the given stationary state should be devised so as to be entirely symmetrical in its predictions regarding the behavior of the two particles. Otherwise, indistinguishable particles would be assigned distinguishing features. As discussed later, the type of symmetry that is acceptable depends on the kind of particles in question.

5.8.1 Symmetric and Antisymmetric Solutions for Pairs of Identical Particles.

In a system consisting of two identical particles with a Hamiltonian operator $\hat{H}$, the Schroedinger equation is given by the relation

$$\hat{H}\psi = -\frac{\hbar}{i}\frac{\partial \psi}{\partial t} \tag{5.140}$$

and a possible solution of this equation can be written as

$$\psi = \psi(q_1, q_2, t) \tag{5.141}$$

where q_1 denotes the quantum-mechanical observables, such as coordinates and spin, of particle 1 and q_2 the quantum-mechanical observables of particle 2.

A possible solution does not necessarily exhibit any symmetry in the observables q_1 and q_2 for the two particles. For example, for two identical, noninteracting particles in a box, a formally correct solution could be given by the relation

$$\psi(q_1, q_2, t) = u_k(q_1)\, u_l(q_2) \exp\left[-i\frac{(\epsilon_k + \epsilon_l)t}{\hbar}\right] \tag{5.142}$$

where u_k and u_l are the single-particle eigenfunctions corresponding to the eigenvalues ϵ_k and ϵ_l, respectively. This solution does not exhibit any symmetry, and, in fact, it asserts that particles 1 and 2 have energies ϵ_k and ϵ_l, respectively. Such an assertion is not amenable to observation, however, for, even if it is found that one particle has energy ϵ_k and the other has energy ϵ_l, no experiment exists for telling which particle is which since the particles are, by definition, inherently indistinguishable.

It follows from such considerations that a possible solution of Schroedinger's equation is not necessarily a physically acceptable solution, in the same sense that not all eigenvalues correspond to acceptable wave functions for a system with specified boundary conditions. Nevertheless, starting with any formally possible solution, we can construct an acceptable solution with suitable symmetry properties.

For a pair of identical particles, the Hamiltonian operator is invariant to an interchange of q_1 and q_2. Consequently, if $\psi(q_1, q_2, t)$ is a correct solution of the Schroedinger equation, then $\psi(q_2, q_1, t)$ is also a correct solution. Moreover, by virtue of the principle of superposition, any linear combination of these two ψ's is also a correct solution. Two linear combinations are of special interest. The first of these, called the *symmetric solution* ψ_s, is given by the relation

$$\psi_s = (2)^{-1/2}\left[\psi(q_1, q_2, t) + \psi(q_2, q_1, t)\right] \tag{5.143}$$

where ψ_s and ψ are normalized to unity. The symmetric solution is characterized by complete invariance to an interchange of the observables q_1 and q_2. In symbols,

$$\psi_s(q_1,q_2,t) \equiv \psi_s(q_2,q_1,t) \tag{5.144}$$

The state corresponding to ψ_s is called a *symmetric state.*

The second linear combination of interest, called the *antisymmetric solution,* ψ_a, is given by the relation

$$\psi_a = (2)^{-\frac{1}{2}}[\psi(q_1,q_2,t) - \psi(q_2,q_1,t)] \tag{5.145}$$

This solution changes sign when the observables of the two particles are interchanged. Specifically,

$$\psi_a(q_1,q_2,t) \equiv -\psi_a(q_2,q_1,t) \tag{5.146}$$

The state determined by ψ_a is called an *antisymmetric state.*

The symmetric and antisymmetric solutions do have the desirable symmetry properties. They yield symmetrical probabilities with regard to finding the observables q_1 and q_2 since, by virtue of Eqs. 5.144 and 5.146, the probability densities

$$W_s = \psi_s^*(q_1,q_2,t)\,\psi_s(q_1,q_2,t) \tag{5.147}$$

and

$$W_a = \psi_a^*(q_1,q_2,t)\,\psi_a(q_1,q_2,t) \tag{5.148}$$

are invariant to an interchange in q_1 and q_2.

Another important property of symmetric and antisymmetric solutions is that they always maintain their original symmetry character as time proceeds. As a result a pair of identical particles originally in a symmetric or an antisymmetric state remains in a symmetric or an antisymmetric state for as long as the Hamiltonian operator remains undisturbed.

Experiments indicate that each kind of particle exists in either a symmetric or an antisymmetric state. For example, electrons, protons, and neutrons exist only in antisymmetric states. On the other hand, photons exist only in symmetric states. Again complex material nuclei and atoms, when treated as structureless particles, are found to exist in symmetric and antisymmetric states, respectively, depending on whether they are constructed from even and odd numbers of particles such as electrons, protons, and neutrons. These experimental findings are referred to as the *exchange-symmetry rules.* Particles that exist in antisymmetric states are called *fermions,* and particles that exist in symmetric states are called *bosons.*

5.8.2 Symmetric and Antisymmetric Solutions for Systems with Many Identical Particles. In a system consisting of n identical particles, the question arises as to which states are physically acceptable.

In the light of the discussion of Sec. 5.8.1, it is evident that symmetric and antisymmetric solutions would be satisfactory. For n particles symmetric solutions are those which satisfy the general relation

$$P\psi_s(q_1,q_2,\ldots,q_n;t) = \psi_s(q_1,q_2,\ldots,q_n;t) \tag{5.149}$$

where $\hat{P}$ represents an operator signifying any permutation of the particle indices, and antisymmetric solutions are those which satisfy the general relation

$$\hat{P}\psi_a(q_1,q_2,\ldots,q_n;t) = \mp\psi_a(q_1,q_2,\ldots,q_n;t) \tag{5.150}$$

where the negative sign applies when the number of permutations of indices is odd and the plus sign applies when the number of permutations is even.

States other than symmetric and antisymmetric may result in probability densities that are invariant to permutations of the n indices. Nevertheless, many experimental results can be adequately described by means of completely symmetric or completely antisymmetric states that obey the exchange-symmetry rules mentioned earlier in connection with two particles. For example, n electrons can be described by an antisymmetric state.

Symmetric and antisymmetric states can be constructed from any formally correct solution of the Schroedinger equation for a system of n particles. If $\psi(q_1,q_2,\ldots,q_n;t)$ is a normalized solution, the normalized symmetric and antisymmetric states are given by the relations

$$\psi_s = (n!)^{-1/2} \sum_{\hat{P}} \hat{P}\,\psi(q_1,q_2,\ldots,q_n;t) \tag{5.151}$$

and

$$\psi_a = (n!)^{-1/2} \sum_{\hat{P}} (\mp)\,\hat{P}\,\psi(q_1,q_2,\ldots,q_n;t) \tag{5.152}$$

where the summations are taken over all possible permutations of the indices 1, 2, . . ., n, including that of identity. In the symmetric state all $n!$ terms of the summation have the same sign. In the antisymmetric solution a term has a negative or a positive sign depending on whether the number of permutations in ψ is odd or even.

5.8.3 Pauli Exclusion Principle. Practically all systems consisting of more than one particle are very difficult to analyze exactly. For this reason they are approximated by models with known eigenstates. The approximate eigenfunctions are then used to construct symmetric or antisymmetric solutions, as the case may be, to represent approximately the states of the system. The procedure can be illustrated by considering a system consisting of n identical particles that do not interact strongly with each other, i.e., particles for which the energy of mutual interaction is a small fraction of the potential energy. If the energy of mutual interaction is neglected from the Hamiltonian operator of the system, the resulting approximate operator, called the *unperturbed operator,* can be separated into n single-particle operators each of which defines an eigenvalue problem of the form of Eq. 5.74b, and the system can be approximated by a model of n, identical, noninteracting particles. These particles can be viewed as existing in the various energy eigenstates $u_k(q)$, $u_l(q)$, $u_m(q)$, ... of the single-particle problem so that a formally correct overall eigenfunction of the unperturbed operator is given by the product

$$u_k(q_1)\, u_l(q_2) \ldots u_r(q_n)$$

At any instant of time, different products can be superimposed to provide an approximate representation for the state of the system. For example, if the particles are bosons, a symmetric function, U_s, is given by the relation

$$U_s = \sum_{\hat{P}} \hat{P} u_k(q_1)\, u_l(q_2) \ldots u_r(q_n) \tag{5.153}$$

where the summation is taken over all $n!$ permutations of the indices. This function is determined by the choice of n states $k, l, \ldots, r$ of the single-particle operator. It is an energy eigenfunction of the unperturbed operator for all the particles and is symmetric since

$$\hat{P} U_s \equiv U_s(q_1, q_2, \ldots, q_n) \tag{5.154}$$

for any arbitrary permutation of the q_i. A linear combination of the U_s with time-dependent coefficients $c_{kl \ldots r}$ is also symmetric in all the q_i and can be used to represent approximately the symmetric states of the n identical particles in question. In principle, the time-dependent coefficients $c_{kl \ldots r}$ can be computed from the Schroedinger equation by means of, say, perturbation theory. The computation leads to an answer within a constant of proportionality which is determined by the normalization condition.

On the other hand, if the particles are fermions, an antisymmetric function U_a is given by the relation

$$U_a = \sum_{\hat{P}} \mp \hat{P}\, u_k(q_1)\, u_l(q_2) \ldots u_r(q_n) \tag{5.155}$$

where the $\mp$ signs are used as discussed previously. This function is indeed antisymmetric since

$$\hat{P}\, U_a(q_1, q_2, \ldots, q_n) = \mp U_a(q_1, q_2, \ldots, q_n) \tag{5.156}$$

A linear combination of U_a with time-dependent coefficients can be used to represent approximately any time-dependent, antisymmetric state that satisfies the exact Schroedinger equation in a manner analogous to that discussed for symmetric states.

Antisymmetric solutions U_a can be expressed in the form of the *Slater determinant* given by the relation

$$U_a = \begin{vmatrix} u_k(q_1) & u_l(q_1) & \ldots & u_r(q_1) \\ u_k(q_2) & u_l(q_2) & \ldots & u_r(q_2) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ u_k(q_n) & u_l(q_n) & \ldots & u_r(q_n) \end{vmatrix} \tag{5.157}$$

In this form it is easy to verify that U_a changes sign for any single interchange of a pair of indices of quantum-mechanical observables since this is equivalent to the interchange of two rows of the determinant.

In the products $u_k(q_1)\, u_l(q_2) \ldots u_r(q_n)$ used for the construction of symmetric functions (Eq. 5.153), the same one-particle eigenfunction $u_m(q)$ may occur more than once. This implies that, in the initial allocation of particles in the states $u_k(q)$, more than one particle can occupy the same state.

In the products used for antisymmetric functions, however, two particles cannot occupy the same quantum state. If the same individual solution appeared more than once in the original product, two or more columns in the Slater determinant for U_a (Eq. 5.157) would be identical, and the determinant would be identically equal to zero. Hence two or more antisymmetric particles can never occupy the same unperturbed quantum state. This result holds for all fermions. It is a generalization of the original *Pauli exclusion principle* that was first advanced for electrons in an atom. The Pauli exclusion principle states that in an atom it is not possible to find more than one electron with the same

quantum numbers, such as n, l, m, and m_s discussed for the hydrogen atom. In other words, each spin-orbital is occupied by one electron only. The antisymmetric property of solutions implies the exclusion principle; and, conversely, the exclusion principle is satisfied only by antisymmetric solutions. Only antisymmetric solutions have the property of vanishing when two particles are in the same unperturbed state.

5.8.4 Atoms with Many Electrons—The Self-Consistent-Field Method. The study of the stationary states of atoms with many electrons proceeds in a manner somewhat analogous to that used for the hydrogen atom. For these states the heavy nucleus may be thought of as immobile and the electrons as moving relative to the nucleus. It is easy to write down the energy eigenvalue equation that describes the states of the electrons relative to the nucleus, but this equation cannot be solved exactly. The major difficulty arises from the fact that each electron interacts not only with the nucleus but also with every other electron in the atom. The Hamiltonian operator includes terms that depend not only on distances of electrons from the nucleus but also on distances of each electron from all the others. This operator is not separable. The nonseparability introduces real complexity into the problem. Solutions can be found only by means of approximations.

For example, if each electron of an atom with Z electrons occupies an orbital, each orbital may be found as the solution of an approximate one-electron eigenvalue equation for which the potential energy is determined by the charge of the nucleus and the space charge of all the other orbitals. The space-charge distribution can be taken as spherically symmetric (the angular dependence of the orbitals can be averaged out) and its contribution to the potential energy can be computed by means of Poisson's equation. Thus the potential energy $V(r)$ in the one-electron equation is a function of the distance r of the electron in question from the nucleus; i.e., $V(r)$ is of the central-field type. The resulting one-electron eigenfunctions and eigenvalues are somewhat similar to those discussed for the hydrogen atom.

Hartree[18] suggested that this way of thinking could be used for an approximate analysis of the electronic structure of an atom with Z electrons. Each of the Z electrons is described by a plausible spin orbital (charge cloud). These orbitals are not hard to guess. On the basis of the set of the plausible orbitals, the potential energy $V(r)$ for one of the electrons can be computed as described previously. The corresponding one-electron energy eigenvalue equation can be solved numerically, and a computed orbital is obtained for the chosen electron. The computed orbital is consistent with the remaining orbitals of the set and consequently represents an "improved" description of the electron in question. Next, the improved instead of the plausible orbital is used in the

calculation of $V(r)$ for a second electron in a manner analogous to that employed for the first. Thus an improved orbital is established for the second electron. The procedure is continued and is repeated for all the electrons until successive iterations do not make appreciable difference to the orbitals. Then it is said that *self-consistent atomic orbitals and eigenvalues* have been established. Thus the charge cloud of any electron is precisely that which results from a one-electron energy eigenvalue equation with a $V(r)$ due to the charge of the nucleus and to the space charge of all the other electrons. The space-charge distribution is assumed to be spherically symmetric, and its potential energy is assumed to obey (to be consistent with) Poisson's equation. The method is called the *Hartree* or the *self-consistent-field method.*

Once the self-consistent-field calculation is completed, the electrons are distributed on the self-consistent orbitals in accordance with the Pauli exclusion principle. When the orbitals with the lowest energies are occupied by the electrons, the atom is said to be in the *ground state.*

In the context of the Hartree approximation, the ground state of an atom with many electrons can be described as follows:[16]

1. Each electron is represented by an energy eigenfunction, u, called an atomic orbital. The orbital is determined from the one-electron eigenvalue equation in accordance with the Hartree method. The quantity $|u|^2$ is a measure of the charge density of the charge cloud of the electron.

2. Each atomic orbital is associated with a set of quantum numbers. First is the principal quantum number n that determines the energy and the size of the orbital. Electrons with the same value of n are said to be in the same *shell.* The value $n = 1$ defines the K-shell, $n = 2$ the L-shell, $n = 3$ the M-shell, etc. Second is the number l that determines the geometric shape, or the s, p, d, . . ., classification of orbitals. Orbitals of type p are threefold degenerate; type d are fivefold degenerate; etc. Orbitals with the same values of n and l are said to belong to the same *subshell.* Last is the spin number m_s that has the values $\pm\frac{1}{2}$. When spin is included in the characterization of the orbital, the orbital is called a spin-orbital.

3. Each atomic orbital has its own energy eigenvalue. The usual order of the eigenvalues is given by the relation

$$1s < 2s < 2p < 3s < 4d \sim 4s \ldots < 0 \tag{5.158}$$

This relation is shown schematically in Fig. 5.13.

4. Each atomic orbital can accommodate two electrons with opposite spins. More generally, it follows from the Pauli exclusion principle that the K-shell can accommodate two electrons in the $1s$ level; the L-shell 8 electrons, two in the $2s$ level and six in the $2p$ level; the M-shell eighteen electrons, etc.

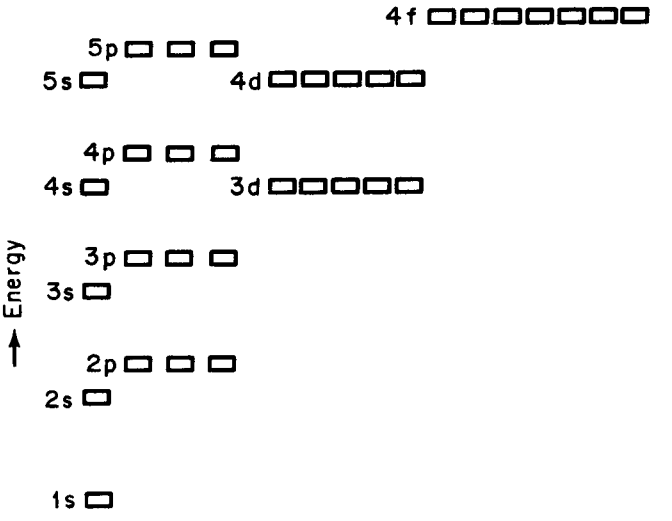


Fig. 5.13 Schematic order of energy eigenvalues of different atomic orbitals.

5. In the ground state the allotment of electrons to shells proceeds as follows. First the K-shell is filled, then the L-shell, etc., until all Z electrons of the atom have been accommodated. These electron allotments constitute the *atomic configuration.* For example, the atomic configuration of hydrogen is $1s$; that of helium is $(1s)^2$, indicating that both electrons are in the $1s$ state; and that of argon is $(1s)^2\ (2s)^2\ (2p)^6\ (3s)^2\ (3p)^6$.

6. In a given atomic configuration, the smaller a subshell's eigenvalue is, the closer it is thought to be to the nucleus. Thus subshells are distinguished as inner and outermost. The outermost may be completely or partially filled. For example, the atomic configuration of oxygen is $(1s)^2\ (2s)^2\ (2p)^4$. In this configuration the two inner subshells, $1s$ and $2s$, are completely filled, and the outermost subshell $2p$, is partially filled.

7. The chemical properties of atoms and their *chemical valence* are determined by the degree to which the outermost subshells are filled and by the types of subshells that are or are not completely filled. In other words, the chemical activity of elements is determined primarily by the electrons with the largest principal quantum number in conjunction with the appropriate s, p, d, . . ., classification, as suggested by relation 5.158. It should be noted, however, that, for atoms with many electrons and for the outermost subshells, the exact order in this relation is not always well definable.

8. The negative of the energy eigenvalue of each spin-orbital (atomic orbital occupied by an electron with a given spin) is interpreted as approximately equal

to the work required to remove the electron from the spin-orbital and bring it to rest at a large distance. This work is called the *ionization energy of the orbital.* The interpretation is approximate because of the following facts. Quantum mechanically, an ionization energy should be computed as a difference of two energies derived from the solution of two problems. The first problem describes the energy eigenstates of the neutral atom. The second problem describes the energy eigenstates of the ion, the atom minus one electron. If both problems are approached by the Hartree method, different spin-orbitals would be found for each problem. If the difference between the spin-orbitals of the atom and of the ion were considered, it would be difficult to interpret each eigenvalue of the one-electron problem that describes the atom in terms of work of removal of an electron from an atomic spin-orbital. On the other hand, if we assume that the spin-orbitals of the one-electron problem for the ion are approximately the same as those of the atom, we can easily show that the negative of the energy of each spin-orbital equals the ionization energy of the orbital.

9. In spite of the approximations, ionization energies as defined here are in good agreement with spectroscopically observed values of these quantities.

10. In the ground state the sum of the energy eigenvalues of the occupied spin-orbitals is appreciably different from the energy of the ground state of the atom as a whole. In other words, although the approximations involved in the Hartree method do not affect appreciably the computed ionization energy of an individual electron, these approximations are poor when the atom is considered as a whole. Nevertheless, the Hartree method is used for many studies of the electronic structure of atoms.

11. The Hartree method implies that, if the occupied spin-orbitals are u_1, $u_2, \ldots, u_Z$, the energy eigenfunction of the ground state of the atom should be given by the product $u_1(q_1)\,u_2(q_2)\ldots u_Z(q_Z)$, where q_i stands for the observables, coordinates and spin, of the ith electron. This product is not consistent with the symmetry requirements of quantum mechanics. To avoid the inconsistency, we should represent the energy eigenfunction of the atom by an antisymmetric determinantal function of spin-orbitals, a Slater determinant

When the energy eigenfunction of the atom is represented by an antisymmetric determinantal function, the best spin-orbitals to use in this function are not those determined by means of the Hartree method. The best spin-orbitals are obtained from a one-electron energy eigenvalue equation called the *Hartree–Fock* or the *self-consistent-field plus exchange equation.* This approach to the problem, characterized as the *Hartree–Fock* method, is discussed later.

5.9 Diatomic Molecules

After atoms the next most complicated type of system is two atoms in the form of a diatomic molecule. The exact quantum-mechanical study of molecules, with

the exception of H_2^+, presents serious computational difficulties. Very few exact results are available. Nonetheless, approximate analyses have yielded a variety of very informative qualitative answers. Without such answers it would be hard to understand even why molecules are formed at all.

The purpose of this section is to present briefly some of the approximate methods used in quantum-mechanical studies of molecules and some of the qualitative results obtained by these methods. The results provide useful guidelines for the development of semiempirical theories for surface phenomena occurring in thermionic converters.

5.9.1 The Energy Eigenvalue Equation for a Diatomic Molecule. In a diatomic molecule consisting of two nuclei of masses m_1 and m_2 and n electrons each of mass m_e, let the positions of the nuclei be denoted by X_i for $i = 1, 2$ (three coordinates, X, Y, and Z, for each nucleus in the molecule), those of the electrons by x_j for $j = 1, 2, \ldots, n$ (three coordinates, x, y, and z, for each electron in the molecule), and the total potential energy of the molecule by $V(X_i,x_j)$. In general, the potential energy depends on the coordinates of both the nuclei and the electrons since it consists of the Coulomb electrostatic interactions between all pairs of charges, nuclei and electrons.

If spin-orbit interactions, relativistic effects, and external fields are neglected, the Hamiltonian operator, $\hat{H}$, for the molecule is given by the relation

$$\hat{H} = -\sum_i \frac{\hbar^2}{2m_i} \nabla_i^2 - \sum_j \frac{\hbar^2}{2m} \nabla_j^2 + V(X_i,x_j) \tag{5.159}$$

where ∇_i^2 is the Laplacian operator with respect to the Cartesian coordinates of the ith nucleus, ∇_j^2 is the Laplacian operator with respect to the coordinates of the jth electron, and the two summations are taken over all the nuclei (i) and all the electrons (j) in the molecule. Thus the energy eigenvalue equation is given by the expression

$$\left[-\sum_i \frac{\hbar^2}{2m_i} \nabla_i^2 - \sum_j \frac{\hbar^2}{2m} \nabla_j^2 + V(X_i,x_j)\right] U(X_i,x_j) = E\, U(X_i,x_j) \tag{5.160}$$

Even for the simplest molecule, H_2^+, the general solution of Eq. 5.160 is almost impossible to find. For this reason a number of approximations must be introduced. The first of these, which pertains to the separation of the nuclear and the electronic motions, is called the *Born–Oppenheimer approximation.*[19]

5.9.2 The Born–Oppenheimer Approximation for the Nuclear and Electron Motions. The first step in the Born–Oppenheimer approximation consists in

solving Eq. 5.160 without the operators corresponding to the kinetic energy of the nuclei but for a fixed distance between the nuclei. Thus energy eigenvalues $\epsilon(X_i)$ and eigenfunctions $u(X_i,x_j)$ are found from the allowable solutions of the relation

$$\left[-\sum_j \frac{\hbar^2}{2m}\nabla_j^2 + V(X_i,x_j)\right] u(X_i,x_j) = \epsilon(X_i)\, u(X_i,x_j) \tag{5.161}$$

Equation 5.161 is said to describe the *electronic structure* of the molecule for fixed positions of the nuclei. This structure is a fair approximation to the actual one because the nuclei are much heavier than the electrons, and therefore the nuclear motions have a minor effect on the motions of the electrons.

The second step in the Born–Oppenheimer approximation consists in the use of the eigenvalues $\epsilon(X_i)$ as potential energy functions in an eigenvalue equation that describes the motion of the nuclei. This equation is independent of the coordinates of the electrons and is given by the relation

$$\left[-\sum_i \frac{\hbar^2}{2m_i}\nabla_i^2 + \epsilon(X_i)\right] v(X_i) = \epsilon'\, v(X_i) \tag{5.162}$$

By virtue of the small ratio of the mass of the electron and that of any nucleus, it can be shown that an eigenfunction, $U(X_i,x_j)$, of the complete molecule is given approximately by the relation[19]

$$U(X_i,x_j) = u(X_i,x_j)\, v(X_i) \tag{5.163}$$

The approximation can be illustrated as follows: If Eq. 5.163 is replaced in Eq. 5.160 and use is made of Eqs. 5.161 and 5.162, a residual is found. The maximum of the absolute value of this residual is closer to zero (the residual for an exact eigenfunction U), the smaller the ratios m_e/m_i.

The solutions of Eq. 5.162 represent the vibrational and rotational energy eigenstates of the molecule. An approximate form of $\epsilon(X_i)$ for the ground state of the electronic structure is given by the Morse curve.[20]

On the other hand, the general solutions of Eq. 5.161 are much more difficult to establish and require use of additional approximations. The understanding of these approximations is facilitated through some familiarity with the electronic structures of H_2^+ and H_2.

5.9.3 The Hydrogen Molecular Ion, H_2^+. The electronic structure of the hydrogen molecular ion is that of one electron moving under the influence of two protons at fixed positions. In molecular physics, understanding this

structure plays the same role as understanding the hydrogen atom in atomic physics.

If distances are measured as multiples of the Bohr radius

$$a_0 = \frac{4\pi\varepsilon_0\hbar^2}{m_e e^2} = 5.29171 \times 10^{-9} \text{ cm}$$

and energies are measured in Rydberg units, Ry, where

$$1 \text{ Ry} = \frac{m_e e^4}{32\pi^2\varepsilon_0^2\hbar^2} = 13.6 \text{ ev}$$

the energy eigenvalue equation for the electronic structure of H_2^+ is given by the relation

$$\left(-\nabla^2 + \frac{2}{R} - \frac{2}{r_a} - \frac{2}{r_b}\right)u = \epsilon u \tag{5.164}$$

where the dimensionless distances R, r_a, and r_b are as shown in Fig. 5.14. The eigenvalues and eigenfunctions of this equation have been accurately computed

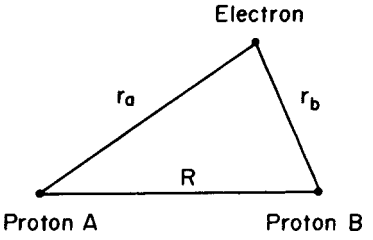


Fig. 5.14 Schematic of interparticle distance in H_2^+.

for the entire range of R from zero to infinity. The results can be summarized as follows:[21]

1. The H_2^+ molecule exhibits rotational symmetry around the axis joining the two nuclei. If the angle of rotation is ϕ, the energy eigenfunctions depend on this angle through the function $\exp(i\lambda\phi)$, where λ is a positive or negative integer.

2. Each value of λ leads to a different energy eigenvalue, except that equal and opposite values of λ are associated with the same energy. Thus with respect to λ the energy levels are discrete and doubly degenerate. The exception to this result is the nondegenerate level that corresponds to λ equal to zero.

3. Each energy eigenfunction is said to define a *molecular orbital* (MO), i.e., a bicentric orbital characteristic of the molecule as a whole rather than of either of the nuclei in the molecule.

4. Each molecular orbital can be interpreted as a charge cloud in a manner entirely analogous to that used for atomic orbitals.

5. Different molecular orbitals are denoted by the Greek letters $\sigma, \pi, \delta, \ldots$ corresponding to values of λ equal to $0, \pm 1, \pm 2, \ldots$. The $\sigma, \pi, \delta, \ldots$ notation for molecular orbitals is analogous to the $s, p, d, \ldots$ notation for atomic orbitals.

6. The H_2^+ molecule is symmetrical with respect to the midpoint between the nuclei. The energy eigenfunctions, excluding spin, are either unchanged or merely change sign when an inversion is performed with respect to this midpoint. Functions that are even with respect to spatial coordinates are denoted by the subscript g and those that are odd, which change sign on inversion, by the subscript u. Thus the molecular orbitals are classified as σ_g, σ_u, π_g, π_u, etc.

7. At infinite separation the eigenvalues converge to one of the eigenvalues of the hydrogen atom. This is to be expected since at this distance the system behaves like a hydrogen ion and a hydrogen atom.

8. At zero separation the eigenvalues, excluding the Coulombic energy $2/R$ of the interaction between the nuclei, converge to those of He^+. This is also to be expected. When the two nuclei come together, they appear to the electron as a doubly charged nucleus ($Z = 2$), and consequently in the Born–Oppenheimer approximation the energies under consideration must be the same as those of the helium ion.

The two molecular orbitals with the lowest energy eigenvalues are called the *bonding* and *antibonding* orbitals and are denoted by $1\sigma_g$ and $1\sigma_u$, respectively. These molecular orbitals are characterized as 1σ because at infinite internuclear separation both have the same energy as the $1s$ orbital of the hydrogen atom. The term "bonding" is used because only the orbital $1\sigma_g$ corresponds to a stable H_2^+, i.e., two protons bonded together by the electron. Finally, the term "antibonding" is used because the $1\sigma_u$ orbital does not correspond to a molecular combination of the two protons. These points will become clearer in the following discussion.

Without or with the Coulombic energy $2/R$ of the interaction between the nuclei, the eigenvalues of the states $1\sigma_g$ and $1\sigma_u$ vs. the dimensionless internuclear distance R are shown in Figs. 5.15a and 5.15b, respectively. These figures show that at very large distances the energies of the two states approach the value of -1 Ry. This is the energy of the ground state of the hydrogen atom or, equivalently, the sum of the energies of the ground state (-1 Ry) plus the ionized state (0 Ry) of the hydrogen atom.

At zero internuclear distance the energy of $1\sigma_g$ less $2/R$ (Fig. 5.15a) is -4 Ry, namely, the energy of the $1s$ state of He^+. The energy of $1\sigma_u$ less $2/R$ is -1 Ry,

namely, the energy of the $2p$ state of He^+. As already mentioned, these results are exactly as expected.

At the equilibrium internuclear distance of 2 Bohr radii, the energy of $1\sigma_g$ is minimum (Fig. 5.15b) and is below that corresponding to infinite separation. On the other hand, the energy of $1\sigma_u$ is a monotonically decreasing function of R and is always greater than -1 Ry. When the two nuclei are at the internuclear distance $2a_0$, the minimum of the energy of $1\sigma_g$ implies that a stable molecule is formed and that the ground state of this molecule is described by the molecular orbital $1\sigma_g$.

Further insight into the reason for the existence of the hydrogen molecular ion is achieved through the study of the form of the eigenfunctions for the orbitals $1\sigma_g$ and $1\sigma_u$ at different internuclear distances. The calculated functions for these orbitals are shown in Fig. 5.16. The graphs in this figure are the values of the energy eigenfunctions vs. position along the axis joining the two nuclei.

At large distances R (Fig. 5.16a), the values of the energy eigenfunctions and, correspondingly, the charge-cloud densities are large only around each nucleus. In fact, it can be shown that the eigenfunctions have approximately the same dependence on spatial coordinates as the $1s$ eigenfunction for the hydrogen

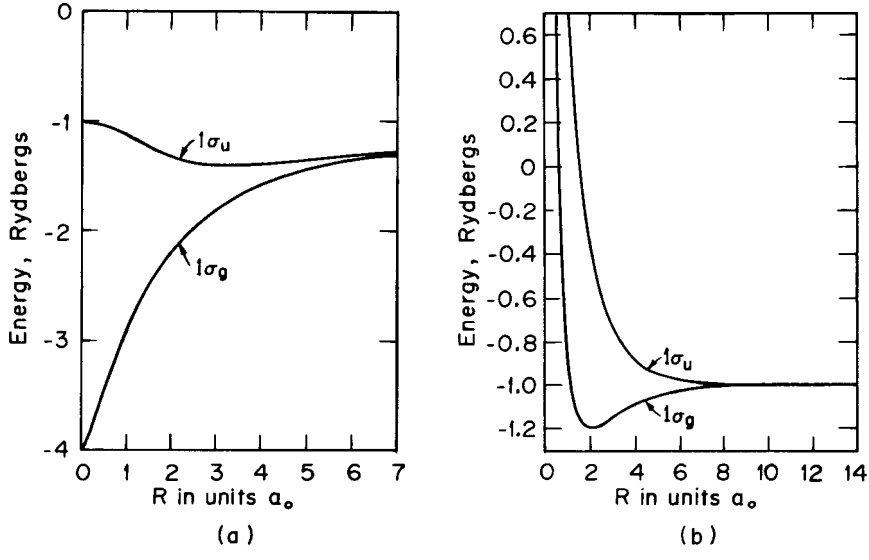


Fig. 5.15 Energy of bonding $1\sigma_g$ and antibonding $1\sigma_u$ orbitals of H_2^+ vs. internuclear distance R. (a) Without coulombic energy $2/R$. (b) With $2/R$. (Adapted from J. C. Slater, *Quantum Theory of Molecules and Solids*, Vol. 1, Chaps. 1 and 2, McGraw-Hill Book Company, Inc., New York, 1963.)

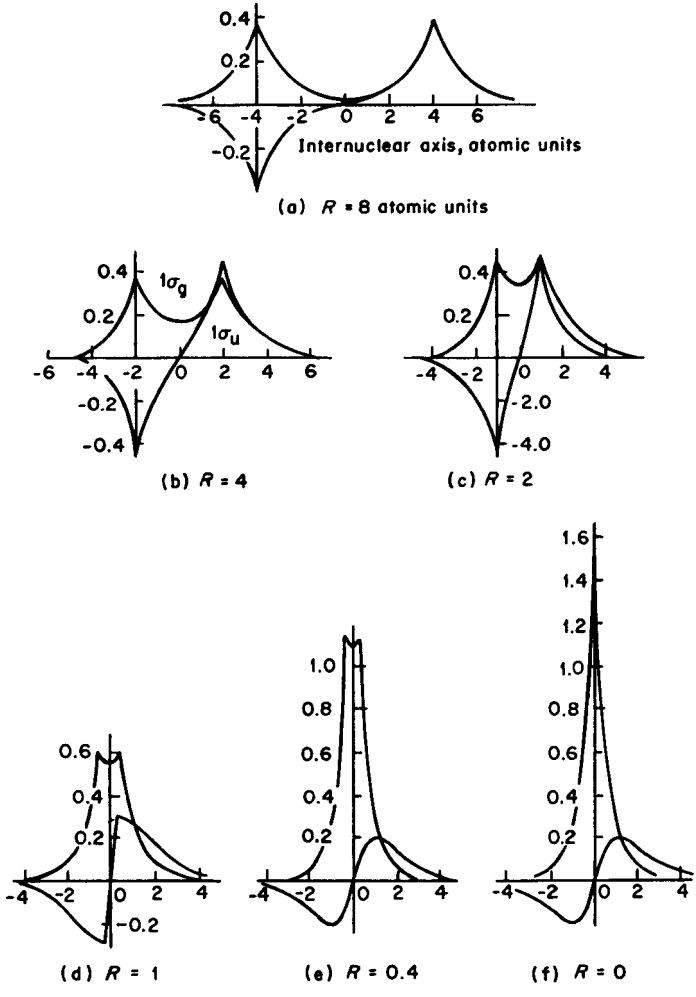


Fig. 5.16 Normalized wave functions for $1\sigma_g$ and $1\sigma_u$ orbitals of H_2^+ along the internuclear axis, for various internuclear distances R. (Adapted from J. C. Slater, *Quantum Theory of Molecules and Solids*, Vol. 1, Chaps. 1 and 2, McGraw-Hill Book Company, Inc., New York, 1963.)

atom. Specifically, normalization factors notwithstanding, the calculations show that

$$u(1\sigma_g) \propto \exp(-r_a) + \exp(-r_b) \quad \text{for large } R \tag{5.165}$$

and

$$u(1\sigma_u) \propto \exp(-r_a) - \exp(-r_b) \quad \text{for large } R \tag{5.166}$$

For intermediate values of R (Figs. 5.16b and 5.16c), the values of the energy eigenfunctions are again relatively large around each nucleus. Here, however, the values of $u(1\sigma_g)$ at positions between the nuclei are appreciable, whereas those of $u(1\sigma_u)$ are small.

For very small values of R (Fig. 5.16d, 5.16e, and 5.16f), only $u(1\sigma_g)$ peaks around the nuclei. The orbital $u(1\sigma_u)$ assumes very small values near the nuclei and larger values away from them. This behavior is to be expected since at zero distance $u(1\sigma_g)$ and $u(1\sigma_u)$ coincide with the $1s$ and $2p$ orbitals of He^+, respectively.

Substantial evidence suggests that the formation of H_2^+ and the fact that its ground state is described by $u(1\sigma_g)$ are due to the behavior of $u(1\sigma_g)$ between the two nuclei. In other words, the formation of H_2^+ can be explained by the fact that the charge density, $u^2(1\sigma_g)$, is appreciable between the two nuclei. Because such charge densities are responsible for the formation of many molecules and aggregates of matter, they are discussed in some detail in the next section.

5.9.4 Linear Combination of Atomic Orbitals and the Overlap Charge for H_2^+. At large distances R the energy eigenfunctions $u(1\sigma_g)$ and $u(1\sigma_u)$ are represented by Eqs. 5.165 and 5.166, respectively. In other words, they are *linear combinations of atomic orbitals* (LCAO) of the hydrogen atom. The LCAO is a good approximation even at distances near the equilibrium distance. The adequacy of the approximation is illustrated by the graphs shown in Fig. 5.17. At very small distances the LCAO fails. At these distances the two molecular orbitals in question tend to the $1s$ and $2p$ orbitals of He^+ and cannot be approximated by the $1s$ orbitals of the hydrogen atom.

In view of these results, it can be argued that $u(1\sigma_g)$ is adequately represented by an expression of the form

$$u(1\sigma_g) \propto \exp(-c_1 r_a) + \exp(-c_1 r_b) \tag{5.167}$$

where c_1 is an adjustable constant that equals unity at large internuclear distances and equals two at zero distance since the $1s$ orbital of He^+ is

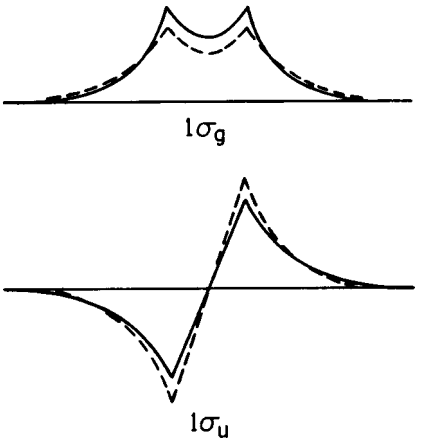


Fig. 5.17 Comparison of wave functions $1\sigma_g$ and $1\sigma_u$ orbitals for H_2^+ for $R = 2a_0$. —, exact. - - - -, linear combination of atomic orbitals. (Adapted from J. C. Slater, *Quantum Theory of Molecules and Solids*, Vol. 1, Chaps. 1 and 2, McGraw-Hill Book Company, Inc., New York, 1963.)

proportional to $\exp(-2r)$. At intermediate distances the value of c_1 is not known a priori. It can be determined by a combination of experimental and theoretical considerations as follows:

In general, for a given energy eigenvalue equation, the energy eigenvalue E and the corresponding exact eigenfunction u satisfy the relation

$$E = \frac{\int u^* \hat{H} u \, d^N q}{\int u^* u \, d^N q} \tag{5.168}$$

where the integrals are over all coordinate space. This relation is found by multiplying both sides of the eigenvalue equation by u^* and integrating the result over all space.

It can be shown that the quantity on the right-hand side of Eq. 5.168 is stationary (its variation equals zero) with respect to small variations of u. In fact, this quantity not only is stationary but also is an extremum when $\hat{H}$ is Hermitian. It follows that, when u is represented by an approximate function with one or more adjustable parameters, such as c_1 in Eq. 5.167, the best choice of the adjustable parameters is that which renders the variation of Eq. 5.168 equal to zero. Moreover, given two sets of parameters each of which yields a zero variation, the most appropriate set is the one that results in the smallest value of E. For the particular molecule H_2^+, if the equilibrium internuclear distance is known experimentally, then the "best" value of c_1 in Eq. 5.167 can be

determined by applying the variational procedure outlined to the $\hat{H}$ corresponding to the equilibrium distance.

Regardless of the method used for calculating c_1, the representation of $u(1\sigma_g)$ by a linear combination of two atomic orbitals (Eq. 5.167) affords the following interpretation. The electron charge density associated with the molecular orbital is given by the relation

$$\rho = u^2(1\sigma_g)$$

$$\propto \exp(-2c_1 r_a) + 2\exp[-c_1(r_a + r_b)] + \exp(-2c_1 r_b) \tag{5.169}$$

If the *overlap integral* S is defined by the relation

$$S = \frac{c_1^3}{\pi}\int \exp[-c_1(r_a + r_b)]\, dv \tag{5.170}$$

the normalized molecular orbital and charge density are given by the expressions

$$u(1\sigma_g) = \left[\frac{c_1^3}{2\pi(1+S)}\right]^{1/2} [\exp(-c_1 r_a) + \exp(-c_1 r_b)] \tag{5.171}$$

and

$$\rho = \frac{c_1^3}{2\pi(1+S)}\left\{\exp(-2c_1 r_a) + 2\exp[-c_1(r_a + r_b)] + \exp(-2c_1 r_b)\right\} \tag{5.172}$$

The first term in the right-hand side of Eq. 5.172 may be interpreted as a spherically symmetric charge density surrounding proton A, the total fractional charge being $1/[2(1+S)]$. The third term may be interpreted as a similar charge density surrounding proton B, and again the total fractional charge equals $1/[2(1+S)]$. Finally, the second term may be regarded as an *overlap charge density* with a total *overlap charge fraction* equal to $S/(1+S)$. Of course, the sum total of these three charges equals unity, which corresponds to one electron. For the dimensionless distance R equal to 2, the three charge densities vs. position along the internuclear axis are plotted in Fig. 5.18.

The overlap integral approaches zero when the internuclear distance is large. Under this condition the atomic orbitals in the LCAO are such that, when one assumes large values, the other assumes almost zero values. In other words, these orbitals do not overlap.

As the internuclear distance approaches zero, S approaches unity and the overlap charge increases. For R equal to zero, the overlap charge approaches one-half of the total charge. Similarly, each of the spherically symmetric charges

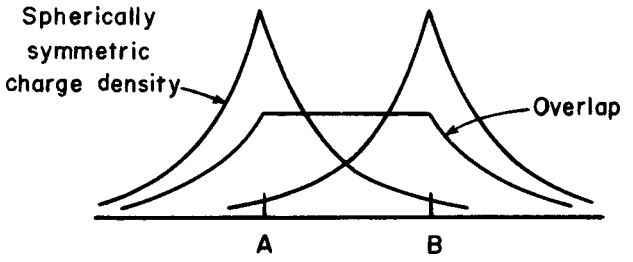


Fig. 5.18 Graphs of spherically symmetric charge densities and overlap charge density along internuclear axis for H_2^+, $R = 2a_0$. (Adapted from J. C. Slater, *Quantum Theory of Molecules and Solids*, Vol. 1, Chaps. 1 and 2, McGraw-Hill Book Company, Inc., New York, 1963.)

around each proton equals one-half of the total charge when the distance is very large and shrinks to one-fourth of the total charge when the distance is zero.

By virtue of this interpretation, the formation of H_2^+ can be readily understood. For a given distance R, consider the force exerted on one of the protons, say B in Fig. 5.14, by all the other charges. First, note that the force due to the spherically symmetric charge around proton B equals zero. Second, proton A is shielded by the spherically symmetric negative charge that surrounds it but, nevertheless, always exerts a repulsive force on B. It is clear that the degree of shielding depends on the location of B within the spherically symmetric charge cloud around A. Finally, the overlap charge exerts an attractive force on B. The attractive force is relatively large because the mean distance of the overlap charge from B equals one-half the internuclear distance.

When R is greater than the equilibrium distance, the attraction due to overlap charge outweighs the repulsion of the nuclei. When R is smaller than the equilibrium distance, the repulsion outweighs the attraction. Hence H_2^+ is formed at the internuclear distance at which the attractive and repulsive forces balance each other. This distance corresponds to that for the minimum energy eigenvalue.

This way of thinking about the formation of H_2^+ leads to the following conclusions:

1. The binding in this *homonuclear molecule* (a molecule where A and B are identical) is caused by the overlap charge. The explanation of binding by means of overlap charges is a purely quantum-mechanical result with no analog in classical mechanics.

2. Molecular orbitals can be approximated by LCAO. This possibility is suggested by the behavior of the molecular orbitals at infinite internuclear separation.

3. Adjustable constants in LCAO can be determined by means of variational methods applied to Eq. 5.168. In the case treated earlier, the only adjustable constant is c_1 (Eq. 5.167) because polarization effects are omitted. Such effects could be included by considering a different form for $u(1\sigma_g)$.

5.9.5 The Hydrogen Molecule, H_2. The electronic structure of the hydrogen molecule is that of two electrons interacting with each other and with two protons. Its understanding brings forth several points that are useful in analyzing any many-electron system.

Two hydrogen atoms can be represented by two nuclei (protons), A and B, and two electrons, 1 and 2. If the two atoms are very far from each other, they do not interact. Thus electrons 1 and 2 can be thought of as belonging to A and B, respectively. This situation is shown schematically in Fig. 5.19. In the ground state the electrons occupy the normalized atomic orbitals given by the relations

$$u_A(1) \propto \exp(-r_{1a}) \qquad u_B(2) \propto \exp(-r_{2b}) \tag{5.173}$$

where both r_{1a} and r_{2b} (Fig. 5.19) are measured in terms of the Bohr radius. In addition, the energy of each of these orbitals is $\epsilon_1 = 1$ Ry, i.e., the energy of the ground state of the hydrogen atom.

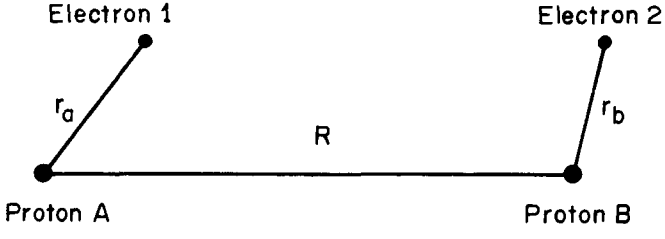


Fig. 5.19 Schematic of hydrogen atoms at large distances R from each other. No interactions occur between any of the charges of the two atoms.

If the two atoms are sufficiently near each other, interactions between their respective positive and negative charges are possible. As a result of these interactions, the valence electron orbitals (charge clouds) are different from those prevailing at infinite separation. One important difference is that the charge clouds overlap each other in the region between the nuclei.

To a first approximation, the energy eigenfunctions of the molecule can be expressed as linear combinations of atomic orbitals. The combinations must be constructed in accordance with the symmetry requirements presented in Sec. 5.8. For the ground state two possible approximations are given by the relations

$$u_+ = u_A(1)\,u_B(2) + u_A(2)\,u_B(1) \tag{5.174}$$

and

$$u_- = u_A(1)\,u_B(2) - u_A(2)\,u_B(1) \tag{5.175}$$

Of course, each of the functions u_+ and u_- must be associated with the appropriate spin combination so that the resulting overall expressions are antisymmetric in all the observables, coordinates and spins.

One of the functions u_+ and u_- is a better approximation for the ground state than the other. Which is better can be decided by using each of these functions in Eq. 5.168 to compute an approximate value for the energy. The Hamiltonian operator, $\hat{H}$, for the electronic structure of H_2 is given by the relation

$$\hat{H} = -\nabla_1^2 - \nabla_2^2 - \frac{2}{r_{1a}} - \frac{2}{r_{1b}} - \frac{2}{r_{2a}} - \frac{2}{r_{2b}} + \frac{2}{r_{12}} + \frac{2}{R} \tag{5.176}$$

where all symbols are as defined in Fig. 5.20, distances are expressed in terms of the Bohr radius, and energies are measured in Rydbergs. Substituting Eq. 5.176 and either Eq. 5.174 or Eq. 5.175 in Eq. 5.168 yields

$$\epsilon_\pm = -2 + \frac{H_0 \pm H_1}{1 \pm S^2} \text{ Ry} \tag{5.177}$$

where

$$H_0 = \int\int u_A^2(1)\,u_B^2(2)\left[-\frac{2}{r_{1b}} - \frac{2}{r_{2a}} + \frac{2}{r_{12}} + \frac{2}{R}\right] dv_1\,dv_2 \tag{5.178}$$

$$H_1 = \int\int u_A(1)\,u_B(1)\,u_A(2)\,u_B(2) \times \left[-\frac{2}{r_{1b}} - \frac{2}{r_{2a}} + \frac{2}{r_{12}} + \frac{2}{R}\right] dv_1\,dv_2 \tag{5.179}$$

$$S = \int u_A(1)\,u_B(1)\,dv_1 = \int u_A(2)\,u_B(2)\,dv_2 \tag{5.180}$$

and the $\pm$ signs correspond to u_+ and u_-, respectively. The quantity H_0 is called the *Coulomb integral*; the quantity H_1 is called the *exchange integral*; the quantity S is the *overlap integral*, and the integrand of S is the overlap or exchange charge density. The constant term –2 Ry in Eq. 5.177 arises from

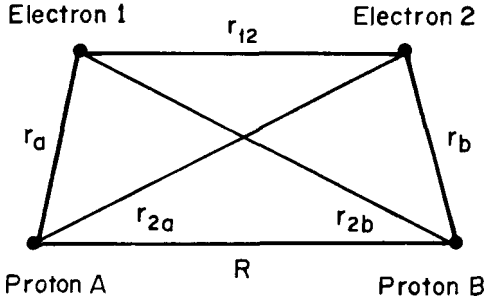


Fig. 5.20 Schematic of two interacting hydrogen atoms. Coulomb interactions occur between all pairs of charges.

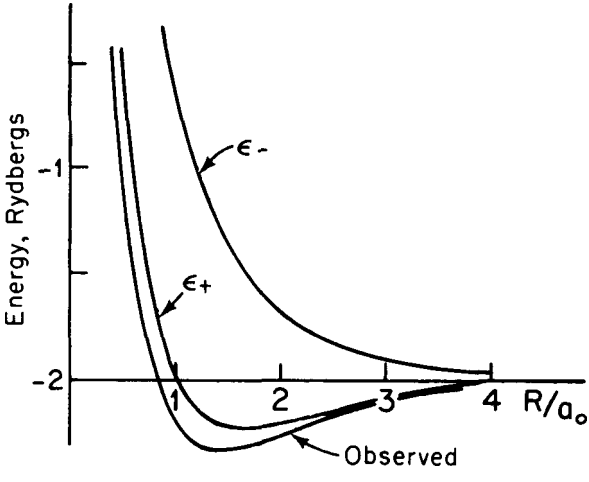


Fig. 5.21 Energies ϵ_+ and ϵ_- vs. R/a_0 for H_2. (Adapted from J. C. Slater, *Quantum Theory of Molecules and Solids*, Vol. 1, Chap. 3, McGraw-Hill Book Company, Inc., New York, 1963.)

integrals related to the condition of the atoms at infinite separation. The other terms represent the effect of the interactions. These terms vanish when the internuclear distance is very large because then r_{1b}, r_{2a}, r_{12}, and R are very large.

The energies ϵ_+ and ϵ_- computed for different values of R are plotted vs. R in Fig. 5.21. The figure shows that only ϵ_+ exhibits a minimum below -2 Ry, i.e., below the energy of the ground state of the separated atoms. Consequently u_+ is a better approximation than u_- for the ground state of H_2. The minimum value

of ϵ_+ is found to be -2.228 Ry (-30.34 ev). This value corresponds to an approximate binding energy of $+0.228$ Ry (3.14 ev).

Figure 5.21 also includes a graph of the observed energy of the ground state of H_2 vs. R. The experimental binding energy is 4.75 ev. It is seen from the figure that ϵ_+ is not a satisfactory approximation to the observed energy. This in turn implies that u_+ is an inadequate approximation for the actual ground-state eigenfunction and that it might be further improved. Improvements can be introduced in ways analogous to those used in LCAO for H_2^+. For present purposes it is sufficient to discuss some general qualitative conclusions emerging from the use of u_+.

The Coulomb integral H_0 (Eq. 5.178) consists of the sum of four terms. The first term can be interpreted as the Coulomb attraction exerted on proton B by an electron in an orbital of A. The second term can be interpreted as the identical attraction on A by an electron in an orbital of B. Finally, the third and fourth terms can be thought of as representing the Coulomb repulsion between the two orbitals and the two protons, respectively. From the numerical values of each of these terms, we conclude that at large internuclear distances the attraction is larger than the repulsion, whereas the opposite is true at small distances. This simple electrostatic interpretation is consistent with the notion of the interpenetrating charge clouds. When the atoms are very far apart, the nucleus of each atom is perfectly shielded by its electronic charge cloud, and there is no interaction between the atoms. At smaller internuclear distances the charge cloud of one atom penetrates that of the other, and therefore it can be attracted by the nucleus. Similar charges repel each other, but the attraction outweighs the repulsion, except at very small distances, where the internuclear repulsion term $2/R$ predominates and outweighs all attractions.

The exchange integral H_1 (Eq. 5.179) can be interpreted in a manner analogous to that used for H_0. The first term in H_1 can be thought of as the attraction exerted on a point charge of amount S located on B by a charge distribution of density $u_A(1)\,u_B(1)$ and of total amount S. The second term represents an identical attraction exerted on a point charge of amount S located on A by a charge distribution of density $u_A(2)\,u_B(2)$ and of total amount S. The third term can be interpreted as a repulsion between the two charge distributions $u_A(1)\,u_B(1)$ and $u_A(2)\,u_B(2)$. Finally, the fourth term in H_1 can be regarded as representing the repulsion between the two overlap charges S on A and B. Except for very small distances, the attractive terms in H_1 outweigh the repulsive ones and are greater than the attractive terms in H_0.

The conclusion emerging from the preceding interpretation is that binding in H_2 is due primarily to overlap or exchange charge that piles up between the atoms. This charge is strongly attracted by the nuclei because it tends to be localized between them.

5.9.6 The Hartree–Fock Equation and the Molecular-Orbital Method. As mentioned earlier, Eq. 5.161 for the electronic structure of a diatomic molecule with n electrons cannot be solved exactly. Approximate solutions can be established by different techniques.

One of these techniques consists in representing an energy eigenfunction of the molecule by means of an antisymmetric, determinantal function of n orthonormal spin-orbitals. The representation is approximate because it implies separability of observables (coordinates and spins). Strictly speaking, separability is not possible because of the mutual interactions between the electrons.

In the context of the approximation, the optimum set of spin-orbitals is determined by means of a variational procedure. Specifically, the determinantal function is replaced in Eq. 5.168 to compute the energy. Each spin-orbital is then varied so as to render the energy a minimum, the variations being limited to those for which the spin-orbitals remain orthonormal.

The result of the variational procedure is an optimum set of spin-orbitals for the approximate representation of the energy eigenfunction of the electronic structure of the molecule by means of one Slater determinant. Each spin-orbital of the optimum set is a solution of a one-electron energy eigenvalue equation known as the *Hartree–Fock equation.*[22] This equation is not given here. For present purposes it suffices to indicate the following:

1. The Hartree–Fock equation is satisfied by an infinite number of eigenfunctions u_i, each of which has its own eigenvalue ϵ_i. Each eigenvalue is called a *one-electron energy.*

2. Each eigenfunction u_i is called a molecular orbital (MO). Each molecular orbital is bicentric; i.e., it belongs to the molecule as a whole and is not monocentric as an atomic orbital is.

3. Each MO can accommodate, at most, two electrons with opposite spins. A molecular orbital occupied by an electron with a given spin is a spin-orbital.

4. The square of the eigenfunction of a spin-orbital is interpreted as a charge-cloud density in the same sense as that discussed for atomic orbitals. Graphic representations of MO can be given in a manner similar to that used for atomic orbitals.

5. In a given state the electrons of the molecule are regarded as arranged in n spin-orbitals in accordance with the Pauli exclusion principle and energetic requirements. For example, in the ground state the electrons occupy the n spin-orbitals with the lowest one-electron energies.

6. The eigenvalue of a spin-orbital is interpreted as the negative of the ionization energy of this orbital. The interpretation is based on the assumption that removal of the electron from the spin-orbital does not affect the eigenfunctions representing the remaining electrons. The interpretation is known as Koopmans' theorem.[23]

7. One-electron energies are in fair agreement with experimentally observed ionization energies. For a given state, however, the sum of the one-electron energies of the occupied spin-orbitals is a poor approximation for the energy of the molecule in that state. A method for improving the approximation is discussed later.

This way of thinking about the electron structure of a homonuclear molecule is referred to as the *molecular-orbital method* because of the allocation of the electrons to molecular orbitals.

In practice the molecular-orbital method cannot be used directly, because the Hartree–Fock equation cannot be solved explicitly. Further approximations are necessary to proceed with the analysis. One of these approximations is Roothaan's method,[24] which will not be discussed in this book.

Several homonuclear molecules have been analyzed by the molecular-orbital method with more or less satisfactory results. The calculations are lengthy. The results indicate that binding in these molecules can be interpreted by means of overlap charges in a manner analogous to that used for H_2^+. This type of binding is called *homopolar binding.*

As already mentioned, representing a given energy eigenfunction of a molecule by a single determinantal function constitutes an approximation. Determination of the spin-orbitals from the Hartree–Fock equation yields the optimum approximate solution possible in the form of a single Slater determinant but not necessarily the best approximation for the energy eigenstate in question. A better approximation can be achieved by considering a linear combination of determinantal functions. The coefficients of the linear combination are determined by minimizing the energy. Each determinant represents a configuration, and this approximate calculation is called the *method of configuration interactions.* For important details on configuration interactions, the reader should consult the literature.[25]

The method of configuration interactions may yield more-accurate results than the molecular-orbital method. The results, however, cannot be interpreted in terms of orbitals occupied by individual electrons with individual ionization energies. Instead, the energy eigenstates refer to the electronic structure of the molecule as a whole.

5.9.7 The Valence-Bond Method.[16] The valence-bond (VB) method is an alternate approximate technique for analyzing the electronic structure of homonuclear diatomic molecules with many electrons. The distinguishing feature of the valence-bond method is that it considers each of the combining atoms as a whole. A molecule is thought to form by the bringing together of complete atoms whose valence electrons interact and form pairs. In this regard, the VB method differs from the molecular-orbital method. In the MO method,

only the nuclei and possibly the inner electron shells are first brought together, and valence electrons are then allotted to suitable molecular orbitals. The VB method description is more closely related to the chemical picture of the formation of a molecule, in which unpaired valence electrons of one atom combine with unpaired valence electrons of another, thus forming a stable molecule. This is the picture conveyed by the discussion of H_2. Electron-pair binding is called *covalent binding,* and the bond is a *covalent bond.*

Strictly speaking, the VB method is more a way of thinking about molecules than a technique for the analysis of the electronic structure of homonuclear molecules. In fact, when used properly, the VB and the MO methods are indistinguishable. For details the reader should consult the literature.[16,25]

5.9.8 Heteronuclear Molecules. Up to this point, only diatomic molecules with identical atoms A and B have been considered. Most of the molecules, however, are heteronuclear; i.e., atom A is different from B. Molecular orbitals of heteronuclear molecules do not have the even–odd type of symmetry with respect to spatial coordinates. They have permanent dipole moments. As two dissimilar atoms are brought together, the electronic charge is redistributed in a way that results in the appearance of the permanent dipole moment. The redistribution, which is known as *charge transfer,* is closely related to the problems of the electronegativity of various atoms; to Volta differences of potentials between different types of solids; to other related questions; and to the question of whether a given molecule is held by a covalent bond (such as that of H_2), by a *polar bond* (such as that existing between two oppositely charged ions), or by both.

The electronic structure of heteronuclear molecules can be studied by the MO method. A given molecular orbital can be approximated by a linear combination of atomic orbitals of the form

$$u \propto u_A + c_2 u_B \tag{5.181}$$

with $|c_2|$ a constant different from unity. In principle, this constant is found by means of variational techniques. The novelty in heteronuclear MO's is that if atom B is more electronegative than atom A then the value of $|c_2|$ is greater than unity. This implies that the amount of negative charge in the vicinity of atom B is larger than that in the vicinity of atom A, and thus the molecule exhibits a dipole moment.

Calculations for heteronuclear molecules by means of the MO method are neither as numerous nor as complete as those for homonuclear molecules. Moreover, quantitative results for dipole moments and their polarities are not reliable.

Qualitatively, the bond between two dissimilar atoms is regarded as *partially covalent* and *partially ionic*. The covalent part is due to the effects of overlap or exchange charges that pile up between the nuclei when a stable molecule is formed. The formation of overlap charge and its contribution to binding arise from the interpretation of molecular orbitals in terms analogous to those discussed for homonuclear molecules. On the other hand, the ionic part is a consequence of the transfer of charge from one atom to the other, namely, from the vicinity of the electropositive to the vicinity of the electronegative atom. This charge transfer makes one region of the molecule appear to have a net positive charge, whereas the other appears to have a net negative charge. These two charges of opposite sign result in an attraction that is referred to as *ionic binding*.

The interpretation of bonds in heteronuclear molecules as partially covalent and partially ionic is artificial and ambiguous. The division of the binding into two parts is both crude and arbitrary. Nevertheless, this way of thinking about bonds has proved helpful in some approximate molecular studies.

5.9.9 Atomic Radii. Many experimental studies of diatomic molecules and crystal structures have shown that an atom can be assigned an approximate radius independent of the presence or absence of similar or dissimilar atoms chemically bound to it. This is called the *atomic radius*.

The concept of an atomic radius implies that, in any molecule or aggregate of atoms, the equilibrium internuclear distances are equal to approximately the sum of the radii of the atoms involved. In other words, in all compounds or crystals, atoms behave like spheres of definite size. The list of atomic radii given in Table 5.2 is the result of correlations of internuclear distances from a very large number of experiments.[26]

The long history of the development of radii associated with atoms is not discussed here. For present purposes it suffices to say that such concepts as *ionic radius* and *covalent radius* have been used to describe atoms in different types of molecules or crystals with different degrees of success. Some ionic radii are given in Table 5.2. Nonetheless, as more and more experimental data on internuclear distances were accumulated, it became evident that one radius for each atom is adequate to describe the majority of experimental results.

Atomic radii can be given a simple physical interpretation. Calculations show that each atomic radius represents the radial distance at which the charge density in the outermost shell of the ground state of the atom is maximum. For example, for the hydrogen atom this distance is the Bohr radius a_0. There are, however, exceptions to this interpretation. For these and other aspects of atomic radii, the reader is referred to the literature.[27]

Table 5.2 Atomic Radii

Element	Atomic radius, Å	Ionic radius, Å	Element	Atomic radius, Å	Ionic radius, Å
Li	1.45	0.60	Ni	1.35	
Be	1.05		Cu	1.35	0.96
C	0.70		Zn	1.35	
N	0.65		As	1.15	
O	0.60		Se	1.15	
F	0.50		Br	1.15	
Na	1.80	0.95	Rb	2.35	
Mg	1.50		Sr	2.00	
Al	1.25		Ag	1.60	1.25
Si	1.10		Cd	1.55	
S	1.00		Sn	1.45	
Cl	1.00		Sb	1.45	
K	2.20	1.33	Te	1.40	
Ca	1.80		I	1.40	
Ti	1.40		Cs	2.60	1.69
Cr	1.40		Ba	2.15	
Mn	1.40		Tl	1.90	
Fe	1.40		Pb	1.80	
Co	1.35		Bi	1.60	

5.10 Electrons in Metals

5.10.1 General Remarks. In this section the quantum-mechanical treatment of the behavior of electrons in metals is discussed.[26] This behavior is the key to understanding the essential phenomena of thermionic conversion and one of the most difficult phenomena to analyze exactly. Only approximate results have been derived to date, and much more work needs to be done.

Ideally the study of a solid, and more specifically of a metal, should begin with an analysis of the energy eigenstates defined by the energy eigenvalue equation for all the particles in the solid. Under identical assumptions the equation is similar to that for diatomic molecules (Eq. 5.160), except that here the number of electrons and nuclei is much larger. Clearly, there is no hope of finding exact solutions for such a many-particle problem.

The first step in reducing the complexity of the problem consists in the use of the Born–Oppenheimer approximation. As in the case of molecules discussed in Sec. 5.9.2, all nuclei of the atoms of the solid are assumed to occupy fixed positions. Thus the electronic structure of the solid is described by an equation similar to Eq. 5.161.

For a given solid the equilibrium positions and the equilibrium configuration of the nuclei should be computed theoretically, as, for example, in the case of H_2^+, but no such computations have been made thus far. Instead, equilibrium distances and configurations are considered as given parameters consistent with experimental observations. For example, as previously mentioned, atoms in solids behave like touching spheres with radii approximately equal to the atomic radius.

5.10.2 The Crystal Lattice. The most important experimental fact about the nature of solids is that by far the largest number are crystalline. This is particularly true for the metals of interest to thermionic conversion. The characteristic feature of a crystal is not so much the regularity of its outer shape when it is grown or cleaved as it is the regularity of the atomic arrangement. This regularity has been the subject of many experimental studies by X-ray diffraction.

Another fact about a crystalline structure on an atomic scale is the existence of *unit cells* in the crystal. These cells are small volumes, all of identical shape, size, orientation, and constitution, which taken together fill all the crystal. Each unit cell contains an integral number of units of the molecular composition of the crystal. For example, in the simplest crystal structures of the elements, each unit cell contains an amount of matter equivalent to one atom, but there are exceptions.

The geometric structure approximating the arrangement of atoms in a crystal is called a *crystal lattice.* The crystal lattice is a regular array of points or sites in a three-dimensional space. Each point is occupied by an atom or a molecule of the stable solid. Of course, the crystal lattice is an idealized picture of a real crystal in which there may be a variety of imperfections. The idealization, however, provides a frame of reference for the systematic description of crystal properties.

Crystal lattices have a number of symmetries with respect to translation, rotation, and reflection. For example, each of two identical sets of lattice sites at different locations in the crystal lattice defines a small volume. An arbitrary point within the first volume corresponds to a geometrically similar point within the second volume. These two points define a translation vector, $\mathbf{R}_n$. Translation through any of the vectors $\mathbf{R}_n$ is a symmetry operation for the crystal lattice because, when an infinite crystal lattice is translated through that vector, an identical crystal lattice results. Similar remarks can be made with regard to specific rotation and reflection operations whose type depends on the particular type of crystal lattice under consideration. These symmetry operations are called *space groups* because they can be easily studied by means of *group theory.* They are discussed in Ref. 26.

The analytical prescription for a lattice is somewhat arbitrary. Usually a lattice is subdivided into equal, geometrically similar polyhedra that represent unit cells. Each unit cell contains only one lattice point, and all unit cells together fill all space. When this is done, we speak of a *Bravais lattice*. The unit cell in Bravais lattices can be defined in a number of ways. The most expedient method results in a unit cell that exhibits the same symmetry properties as the lattice itself. Such a cell is called the *Wigner–Seitz unit cell* for the Bravais lattice in question.

For example, in a simple cubic Bravais lattice (i.e., a lattice in which neighboring sites define small cubes with sides equal to a), a possible unit cell is one of the small cubes nested in a Cartesian coordinate system. Such a cube is shown in Fig. 5.22a. Three vectors, $\mathbf{t}_1$, $\mathbf{t}_2$, and $\mathbf{t}_3$, can be defined by the relations

$$\mathbf{t}_1 = \mathbf{i}_x a \qquad \mathbf{t}_2 = \mathbf{i}_y a \qquad \mathbf{t}_3 = \mathbf{i}_z a \tag{5.182}$$

where $\mathbf{i}_x$, $\mathbf{i}_y$, and $\mathbf{i}_z$ are unit vectors along the x, y, and z axes, respectively. These vectors, which are determined by an arbitrary lattice point and three of its nearest neighbors, are called *primitive vectors* and can be used to represent any space point as a translation of the points of the unit cell. For example, lattice points are found at the locations described by the relation

$$\mathbf{R}_n = n_1 \mathbf{t}_1 + n_2 \mathbf{t}_2 + n_3 \mathbf{t}_3 \tag{5.183}$$

where n_1, n_2, and n_3 are integers. Similarly, at any given point $\mathbf{r}$ within the unit cell in Fig. 5.22a, a corresponding point in any other unit cell is determined by

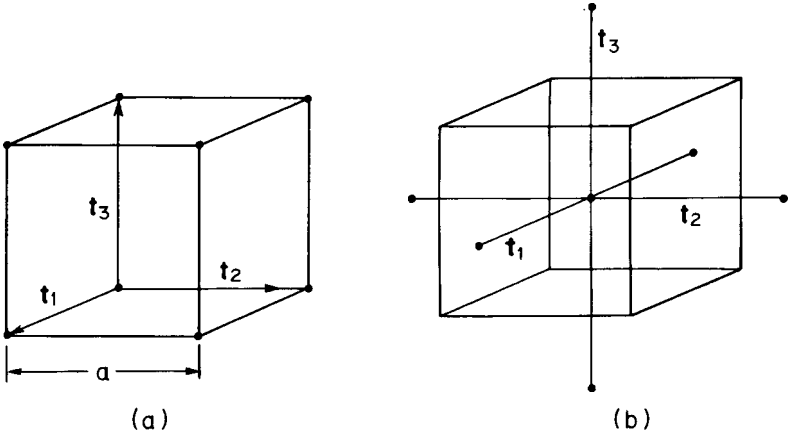


Fig. 5.22 Unit cells in a cubic lattice. (a) Arbitrary. (b) Wigner–Seitz unit cell.

the expression $\mathbf{r} + \mathbf{R}_n$. Finally, the unit cell has a volume equal to a^3 and contains only one lattice point. Indeed, if lattice points are regarded as small spheres, each corner of the cube encloses only one octant of each sphere, and hence there is one lattice point per unit cell.

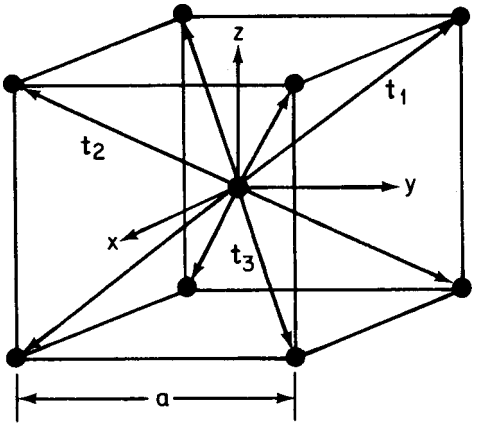
Although the unit cell of Fig. 5.22a gives a full description of the cubic lattice, it is not as convenient a choice as that shown in Fig. 5.22b. In this figure the cube of side a completely surrounds one lattice point. It can be shown that this unit cell has all the symmetry properties of the cubic lattice. Hence it is the Wigner–Seitz unit cell for the cubic Bravais lattice.

The construction of the unit cell in Fig. 5.22b can be given a simple geometric interpretation. Each site in a cubic lattice has six nearest neighbors. For example, if the origin of coordinates is at one lattice point, its six nearest neighbors are along the directions $\pm\mathbf{i}_x$, $\pm\mathbf{i}_y$, and $\pm\mathbf{i}_z$, called the ⟨1 0 0⟩ directions. The unit cell of Fig. 5.22b is formed by six planes perpendicular to the ⟨1 0 0⟩ directions halfway between nearest neighbors.

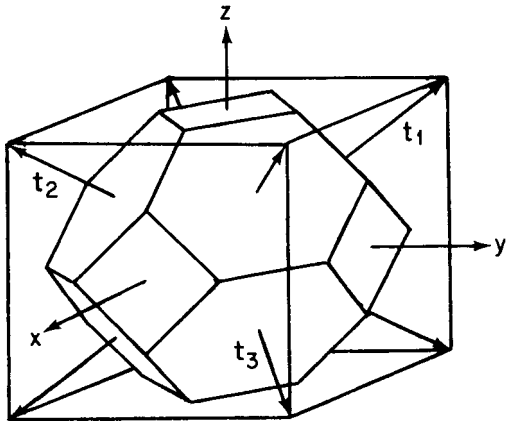
For the other two cubic Bravais lattices, the body-centered cubic (bcc) and the face-centered cubic (fcc), the Wigner–Seitz unit cells can be constructed in a manner analogous to that described for the simple cubic lattice. Specifically, for the bcc lattice each lattice site has eight nearest neighbors and six next-nearest neighbors. The distances between a lattice point and three of its nearest neighbors define three linearly independent primitive vectors given by the relations

$$\mathbf{t}_1 = \frac{a}{2}(-\mathbf{i}_x + \mathbf{i}_y + \mathbf{i}_z) \qquad \mathbf{t}_2 = \frac{a}{2}(\mathbf{i}_x - \mathbf{i}_y - \mathbf{i}_z) \qquad \mathbf{t}_3 = \frac{a}{2}(\mathbf{i}_x + \mathbf{i}_y - \mathbf{i}_z) \tag{5.184}$$

These are three vectors along the diagonals of the cube shown in Fig. 5.23a. The length of these vectors is $(\sqrt{3}/2)a$. It follows that the eight nearest neighbors are along the directions $\pm\mathbf{i}_x$ $\pm\mathbf{i}_y$ $\pm\mathbf{i}_z$, called the ⟨1 1 1⟩ directions, at a distance $(\sqrt{3}/2)a$. On the other hand, the six next-nearest neighbors are along the ⟨1 0 0⟩ directions $\pm\mathbf{i}_x$, $\pm\mathbf{i}_y$, and $\pm\mathbf{i}_z$ and at a distance a. The Wigner–Seitz unit cell is formed by the 14 planes perpendicular to the directions to the eight nearest and the six next-nearest neighbors at halfway between the corresponding distances. The resulting polyhedron is shown in Fig. 5.23b. Its faces consist of eight hexagons and six squares. This cell contains only one lattice point, and it is proven to have all the symmetry properties of the bcc lattices. The volume, given by $\mathbf{t}_1 \cdot (\mathbf{t}_2 \times \mathbf{t}_3)$, equals $a^3/2$, i.e., half the volume of the cube of side a. This result is as it should be since the cube of Fig. 5.23a contains two lattice points. Finally, the primitive vectors given by Eqs. 5.184 can be used to represent any space point as a translation of a point of the Wigner–Seitz unit cell.



(a)



(b)

Fig. 5.23 Construction of Wigner–Seitz unit cell for body-centered cubic structure. (a) Primitive vectors $\mathbf{t}_1$, $\mathbf{t}_2$, and $\mathbf{t}_3$ pointing from atom in the center of the cube to the three nearest neighbors at the cube corners. (b) Wigner–Seitz unit cell.

In the fcc lattice each lattice point has 12 nearest neighbors along the directions $\pm\mathbf{i}_x \pm\mathbf{i}_y$, $\pm\mathbf{i}_y \pm\mathbf{i}_z$, and $\pm\mathbf{i}_z \pm\mathbf{i}_x$, as shown in Fig. 5.24a. Three convenient primitive vectors are determined by the relations

$$\mathbf{t}_1 = \frac{a}{2}(\mathbf{i}_x + \mathbf{i}_y) \qquad \mathbf{t}_2 = \frac{a}{2}(\mathbf{i}_y + \mathbf{i}_z) \qquad \mathbf{t}_3 = \frac{a}{2}(\mathbf{i}_z + \mathbf{i}_x) \tag{5.185}$$

The volume, $\mathbf{t}_1 \cdot (\mathbf{t}_2 \times \mathbf{t}_3)$, of the parallelepiped bounded by these vectors equals $a^2/4$, as it should. The Wigner–Seitz unit cell, which is formed by the 12 planes perpendicular to the directions at half the distances between a lattice point and its 12 nearest neighbors, is shown in Fig. 5.24b. All other planes forming perpendicular bisectors of vectors to more-distant lattice points lie outside the cell in Fig. 5.24b.

Wigner–Seitz unit cells constructed for other Bravais lattices, such as the hexagonal close-packed lattice (hcp), are not discussed here.

5.10.3 The Energy Eigenvalue Equation for the Electronic Structure. In terms of the Born–Oppenheimer approximation, the physical picture associated with a crystal lattice is somewhat as follows: For the solid of an element, e.g., tungsten, each lattice site is occupied by the nucleus of the element. Each nucleus is surrounded by its inner closed-shell electrons and the charge clouds of its valence electrons. The inner closed shells are so tightly bound to the nucleus that they remain practically unaffected by the presence of the other nuclei and electrons in the lattice. On the other hand, the valence electrons move under the influence of the fields created by the nuclei plus the inner electrons plus all the valence electrons. Because of this influence, distribution of the valence electron charges is much different from, say, that prevailing around a single atom. This conclusion is analogous to that derived for diatomic molecules by the MO method. For example, the valence electron orbitals of a solid are spread throughout the solid and are not monocentric as they are for atoms.

In view of these remarks, the energy eigenvalue equation for the electronic structure of a solid is written for the valence electrons only. The remaining charges, nuclei and inner-shell electrons, are considered only as contributors to the electrostatic potential energy of the valence electrons. If spin-orbit and relativistic effects are neglected, the energy eigenvalue equation is given by the relation

$$\left[-\sum_j \frac{\hbar^2}{2m_e}\nabla_j^2 + \sum_j V(\mathbf{r}_j) + \sum_{ij} \frac{e^2}{4\pi\,\epsilon_0 |\mathbf{r}_j - \mathbf{r}_i|}\right] u = Eu \tag{5.186}$$

where the first two summations are over all the valence electrons (e.g., n valence electrons in a volume V), and the third double summation is over all pairs of

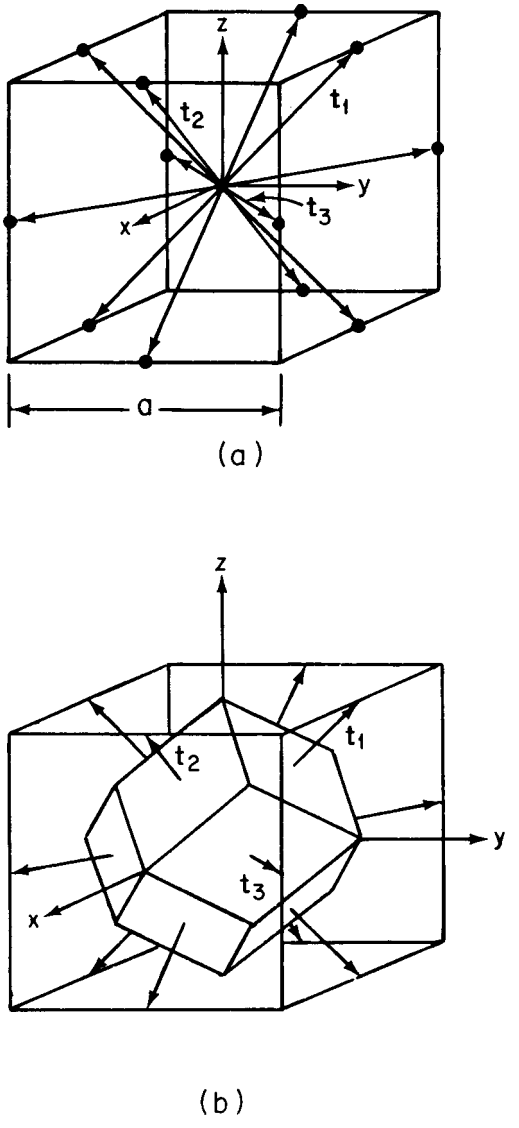


Fig. 5.24 Construction of Wigner–Seitz unit cells for face-centered cubic structure. (a) Primitive vectors $\mathbf{t}_1$, $\mathbf{t}_2$, and $\mathbf{t}_3$ pointing from atom in the center of the cube to three nearest neighbors at the midpoints of the edges of the cube. (b) Wigner–Seitz unit cell.

valence electrons. Each term in the first summation represents the operator for the kinetic energy of a valence electron at position $\mathbf{r}_j$. Each of the terms $V(\mathbf{r}_j)$ in the second sum is the electrostatic potential energy for the electron at $\mathbf{r}_j$ due to all the nuclei and inner-shell electrons. If each lattice site is represented by $\mathbf{R}_n$, then $V(\mathbf{r}_j)$ is given by the relation

$$V(\mathbf{r}_j) = \sum_{\mathbf{R}_n} V_l(\mathbf{r}_j - \mathbf{R}_n) \tag{5.187}$$

where $V_l(\mathbf{r} - \mathbf{R}_n)$ is the electrostatic potential energy at $\mathbf{r}$ due to a nucleus and corresponding inner shells at $\mathbf{R}_n$. Finally, each term in the third summation on the left-hand side of Eq. 5.186 represents the Coulomb energy of interaction between a valence electron at $\mathbf{r}_j$ and another at $\mathbf{r}_i$. The eigenfunction u depends on all n coordinates $\mathbf{r}_j$ of the valence electrons. Finding an exact solution of Eq. 5.186 is an impossible task, and drastic approximations are needed.

5.10.4 One-Electron Approximations. One class of approximations used in studying the electronic structure of crystals consists in considering the force on each valence electron and, therefore, the potential energy of each electron as a function of the position coordinate of the valence electron only. In this approximation the Hamiltonian operator in Eq. 5.186 is separable. Each of the separated Hamiltonian operators defines a *one-electron eigenvalue problem* given by the relation

$$\left[-\frac{\hbar^2}{2m_e}\nabla^2 + V(\mathbf{r})\right]u = \epsilon u \tag{5.188}$$

where $V(\mathbf{r})$ is the one-electron potential energy depending on the position $\mathbf{r}$ of the electron only. The energy eigenfunctions of Eq. 5.186 are represented by suitable antisymmetric combinations of one-electron energy eigenfunctions. The energy eigenvalues of Eq. 5.186 are given by sums of one-electron energy eigenvalues. When this method of treatment of the electronic structure of crystals is used, it is said that the problem is solved by means of the *one-electron approximation.*

If the one-electron potential energy $V(\mathbf{r})$ is taken to be a constant independent of $\mathbf{r}$, the approximation is called *Sommerfeld's free-electron model* or the *free-electron gas model.* When $V(\mathbf{r})$ is a periodic function of position in a Wigner–Seitz unit cell, the approximation is called the *theory of energy bands in crystals.* These two approximations are discussed in Secs. 5.10.6 and 5.10.7.

5.10.5 Periodic Boundary Conditions. Physically a crystal can be regarded as a finite piece of matter bounded by exterior surfaces outside of which is empty

space. The very complicated analysis of the electronic structure in the vicinity of the bounding surfaces is discussed separately in Chap. 9. In this chapter the discussion is restricted to phenomena occurring away from bounding surfaces.

The effects of a bounding surface on the electronic structure of a crystal do not extend very deeply into the bulk. For this reason the bulk of the crystal away from the bounding surfaces can be thought of as part of an infinite crystal unaffected by surface perturbations.

Analyzing the electronic structure in the bulk does not require considering an infinite crystal. The reason is that, if the bulk has ordinary rather than atomic dimensions and contains a large number of unit cells, its electronic structure will be identical to that of any other similar piece of the infinite crystal. In other words, when the number of unit cells is very large, the electronic structure repeats periodically in identical finite crystals. It follows that this structure can be found by solving the energy eigenvalue equation with so-called *periodic boundary conditions.* Periodic boundary conditions are used throughout the remainder of this discussion.

5.10.6 The Free-Electron Gas Model. The most drastic approximation used in solving Eq. 5.186 consists in replacing the potential energy caused by the nuclei and the inner electron shells and that caused by interactions between the valence electrons by a constant independent of position. This approximation, which implies that the valence electrons are regarded as noninteracting gas particles, is called *Sommerfeld's free-electron model* or simply the *free-electron gas model.* Without loss of generality the free-electron gas can be thought to be enclosed in a rectangular container of dimensions l_1, l_2, and l_3 and volume V. Moreover, the periodic boundary conditions can be applied to the walls of this container.

Since the zero-energy reference level is arbitrary, the constant potential energy can be taken equal to zero. From the discussion in Sec. 5.6, it follows that the one-electron energy eigenvalues are given by the relation

$$\epsilon_k = \frac{\hbar|\mathbf{k}|^2}{2m_e} \tag{5.189}$$

where the wave vector $\mathbf{k}$ satisfies relations 5.111 and 5.113. In $\mathbf{k}$-space Eq. 5.189 indicates that surfaces of constant one-electron energy are spheres and that the momentum of the electrons equals $\hbar\mathbf{k}$. Finally, the one-electron energy eigenfunction corresponding to ϵ_k is given by Eq. 5.115.

For a given energy eigenstate of the free-electron gas, the n-valence electrons occupy a group of one-electron energy eigenstates. Because of the Pauli exclusion principle, no more than two electrons can occupy the same one-electron energy eigenstate.

In the ground state the valence electrons are distributed among the $n/2$ lowest energy eigenstates. The maximum one-electron energy level occupied can be determined as follows: The density of states, $N(\epsilon)$, is given by relation 5.120, which is repeated here for convenience

$$N(\epsilon) = \frac{8\pi m_e V}{h^3}(2m_e\epsilon)^{1/2}$$

The maximum energy, ϵ_F, called the *Fermi energy*, is such that

$$n = \int^{\epsilon_F} N(\epsilon)\, d\epsilon = \frac{8\pi V}{3h^3}(2m_e\epsilon_F)^{3/2} \tag{5.190a}$$

or

$$\epsilon_F = \frac{1}{2m_e}\left(\frac{3h^3}{8\pi}\right)^{2/3}\left(\frac{n}{V}\right)^{2/3} \tag{5.190b}$$

In **k**-space the Fermi energy defines a surface (here a spherical surface) called the *Fermi surface.* Other properties of the ground state can be computed from the one-electron energy eigenstates.

5.10.7 The Theory of Energy Bands in Crystals.[26] The free-electron gas model is a valuable guide to some properties of a metal but a very crude approximation for others. An improved approximation is achieved by the use of the one-electron equation (Eq. 5.188) with a one-electron potential energy $V(\mathbf{r})$ that is a periodic function of position in a Wigner–Seitz unit cell. As already mentioned, this one-electron approximation is called the *theory of energy bands in crystals.* The discussion of the theory can be conveniently subdivided into two parts: (1) the procedures used for determining a reasonable one-electron potential energy $V(\mathbf{r})$ and (2) the methods used for solving the one-electron eigenvalue problem.

5.10.8 Charge Distribution and Potential Energy for the Self-Consistent Field in a Crystal Lattice. Macroscopically a solid is electrically neutral in the same sense that an atom has no net electric charge. On a microscopic scale, however, the charge density in a solid is not uniformly equal to zero. By virtue of this nonuniformity and by Poisson's equation, it follows that the electrostatic potential in the solid is not constant but varies as a function of position. If this potential were known, it might be used as a first approximation in the one-electron equation. We must recognize, however, that the results of the

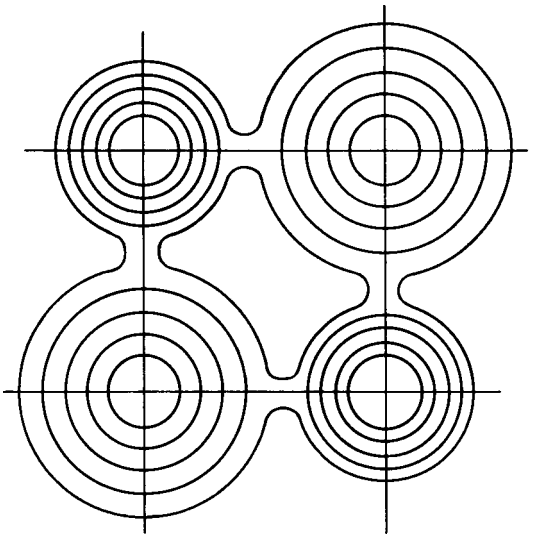


Fig. 5.25 Contours of constant charge density for an ionic crystal.

approximation would be poor since, among other things, this approach does not account for the fact that each valence electron does not act on itself.

If a good self-consistent-field solution for solids were available, the charge distribution would be easy to calculate. But, since such a solution is not available, the charge distribution caused by all electrically charged particles in the solid is usually postulated on the basis of intuitive arguments. This distribution is then used in Poisson's equation to derive a potential energy $V(\mathbf{r})$ for the one-electron equation. It is clear that each postulated charge distribution and associated $V(\mathbf{r})$ result in different one-electron energy eigenstates.

The ionic crystal illustrates the difficulties regarding the description of the charge distribution in a solid. For example, for NaCl a "reasonable" way to make a first approximation to the charge distribution is to start with the known self-consistent-field distributions for each of the isolated ions Na^+ and Cl^- and to superimpose these distributions when the ions occupy the equilibrium internuclear positions in the crystal lattice of NaCl. The resulting contours of constant charge density on a typical plane going through four neighboring ions are as shown in Fig. 5.25. Now, the question arises as to whether this picture indicates that sodium and chlorine behave as positive and negative ions, respectively. The answer to this question is ambiguous. For example, if space is divided into volumes surrounding the sodium sites and other volumes surrounding the chlorine sites, then we see that the total amount of charge in each of these volumes changes with the size of the arbitrarily chosen volumes. If the

sodium volume is too small and the chlorine volume too large, the sodium appears highly positively charged and the chlorine negatively charged, and vice versa. It follows that whether the sodium and chlorine atoms are viewed as neutral or ionized is a matter of arbitrary choice. A similar conclusion is reached if the charge distribution in the lattice is approximated by the superposition of the self-consistent-field solutions for sodium and chlorine atoms.

From the preceding remarks it follows that the question of whether a given crystal is ionic is almost unanswerable or meaningless when considered from the standpoint of starting with the charge density and proceeding from there. In other words, any postulated charge distribution is somewhat arbitrary. Arbitrariness also exists in the choice of charge distributions for crystals of a pure metal.

The preceding difficulty notwithstanding, the potential energy $V(\mathbf{r})$ in Eq. 5.188 is obtained by the following procedure. Charge distributions are assumed to be somewhat like the distribution shown in Fig. 5.25. For ionic crystals the self-consistent-field charge distributions for the individual ions are superimposed. For nonionic or covalent crystals of an element, the self-consistent-field charge distribution of the atom of the element is used at each lattice site, and a symmetric charge distribution results. In general, the charge density is spherically symmetric within a sphere around each atom or ion and assumes rather small values between the spheres. The potential energy $V(\mathbf{r})$ is obtained from such a charge density by means of Poisson's equation.

More often than not, Poisson's equation is not solved explicitly. Instead a $V(\mathbf{r})$ function is postulated which reflects the essential features of what might be expected from the solution of Poisson's equation. For example, for pure metals $V(\mathbf{r})$ is taken to be a spherically symmetric function of $\mathbf{r}$ in spheres touching each other and surrounding each lattice site and a constant independent of $\mathbf{r}$ between the spheres. The values of $V(\mathbf{r})$ are identical in corresponding points of Wigner–Seitz unit cells. In other words, $V(\mathbf{r})$ is invariant or periodic with respect to the translation operation defined by the vector $\mathbf{R}_n$. In terms of symbols this invariance is described by the relation

$$V(\mathbf{r} + \mathbf{R}_n) = V(\mathbf{r}) \tag{5.191}$$

For periodic potential energies that satisfy relation 5.191, the solutions of Eq. 5.188 can be found by representing both $u(\mathbf{r})$ and $V(\mathbf{r})$ by series expansions in terms of plane waves. The nomenclature and the procedures used for these expansions are discussed in the following sections.

5.10.9 The Reciprocal Lattice—Brillouin Zones. In preparation for the representation of $V(\mathbf{r})$ and $u(\mathbf{r})$ in terms of plane waves, it is convenient to define the reciprocal lattice.

Vectors $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$ reciprocal to the primitive vectors $\mathbf{t}_1$, $\mathbf{t}_2$, and $\mathbf{t}_3$ of the Bravais lattice under consideration are determined by the relations

$$\mathbf{t}_i \cdot \mathbf{b}_j = \delta_{ij} \quad \text{for } \mathbf{i}, \mathbf{j} = 1, 2, 3 \tag{5.192}$$

where δ_{ij} is the Kronecker delta. In words, $\mathbf{b}_j$ is perpendicular to two $\mathbf{t}_i$'s with different indices ($i \neq j$), and $\mathbf{b}_j$ and $\mathbf{t}_j$ are reciprocal to each other since their scalar product equals unity. The vectors $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$ constitute a set of primitive vectors for a reciprocal space, in the same way as the vectors $\mathbf{t}_1$, $\mathbf{t}_2$, and $\mathbf{t}_3$ constitute primitive vectors for ordinary space. The vectors

$$h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3$$

where h_1, h_2, and h_3 are integers, reach from the origin to a lattice of points, just as the vectors

$$n_1\mathbf{t}_1 + n_2\mathbf{t}_2 + n_3\mathbf{t}_3$$

reach from the origin to the lattice points of the crystal lattice. The lattice of points defined by the vectors

$$h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3$$

is called the *reciprocal lattice.*

Reciprocal lattices have symmetry properties very similar to those of ordinary crystal lattices. In fact, the reciprocal lattice of a Bravais lattice is also a Bravais lattice but is not necessarily of the same type. Unit cells can be defined in reciprocal lattices in the same way as in ordinary lattices. For example, given the three primitive vectors $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$, a Wigner–Seitz unit cell can be constructed in exactly the same manner as that discussed in connection with the primitive vectors $\mathbf{t}_1$, $\mathbf{t}_2$, and $\mathbf{t}_3$.

A *wave vector lattice* and a **k**-*space* are associated with each reciprocal lattice. These are defined as follows: The points of the wave vector lattice are defined by the vector $\mathbf{K}_h$ given by the relation

$$\mathbf{K}_h = 2\pi(h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3) \tag{5.193}$$

The space of the wave vector lattice is called **k**-space, and any point in that space is denoted by a vector **k**.

The primitive vectors in **k**-space are $2\pi\mathbf{b}_1$, $2\pi\mathbf{b}_2$, and $2\pi\mathbf{b}_3$, and Wigner–Seitz unit cells in **k**-space are determined in a manner analogous to that used for ordinary or reciprocal spaces.

The unit-cell volume surrounding the origin of **k**-space is called the *central Brillouin zone.* Each **k** vector corresponding to a point inside the central Brillouin zone is called a *reduced-wave vector.* Any point in **k**-space can be represented as a sum of a reduced-wave vector **k** plus a wave vector $\mathbf{K}_h$. Consideration of polyhedra larger than that enclosing the central Brillouin zone leads to the definitions of the *second, third, etc., Brillouin zones.*

5.10.10 Fourier Expansions of Periodic Functions. It is well known that a periodic function can be expanded into a Fourier series. Accordingly, since the potential energy $V(\mathbf{r})$ is a periodic function in ordinary lattice space, it can be expanded into a series of a suitably chosen set of functions that are also periodic in the same space.

The set of exponential functions

$$\{\exp(i\mathbf{K}_h \cdot \mathbf{r})\}$$

is periodic in ordinary lattice space. Each exponential function is called a *plane wave.* In addition, it can be shown that plane waves for all values of $\mathbf{K}_h$ form a complete set of orthogonal functions over the volume of one unit cell. It follows that a periodic function $F(\mathbf{r})$, such that

$$F(\mathbf{r}) = F(\mathbf{r} + \mathbf{R}_n) \tag{5.194}$$

can be represented by the Fourier series given by the relation

$$F(\mathbf{r}) = \sum_h G(\mathbf{K}_h) \exp(i\mathbf{K}_h \cdot \mathbf{r}) \tag{5.195}$$

where $G(\mathbf{K}_h)$ are the Fourier expansion coefficients. More specifically, the periodic function $V(\mathbf{r})$ can be represented by the expression

$$V(\mathbf{r}) = \sum_h U(\mathbf{K}_h) \exp(i\mathbf{K}_h \cdot \mathbf{r}) \tag{5.196}$$

The Fourier expansion coefficients in either Eq. 5.195 or Eq. 5.196 are determined by standard techniques. Expansions of the form of Eq. 5.195 are called *expansions in terms of plane waves.*

In addition to $V(\mathbf{r})$, the eigenfunctions of Eq. 5.188 can be represented as sums of plane waves. If $V(\mathbf{r})$ is a periodic function in the crystal lattice, then the corresponding one-electron Hamiltonian operator is also periodic. By virtue of Floquet's theorem, the eigenfunctions $u(\mathbf{r})$ of such an operator are given by the general expression

$$u(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r})\, w(\mathbf{r}) \tag{5.197}$$

where $w(\mathbf{r})$ is a periodic function given by the relation

$$w(\mathbf{r} + \mathbf{R}_n) = w(\mathbf{r}) \tag{5.198}$$

and $\mathbf{k}$ is a vector with real components. In accordance with the preceding discussion, $w(\mathbf{r})$ can be written as a Fourier series of the form of Eq. 5.195. When this is done, the eigenfunctions $u(\mathbf{r})$ can be represented by the relation

$$u(\mathbf{r}) = \sum_h v(\mathbf{K}_h + \mathbf{k}) \exp[i(\mathbf{K}_h + \mathbf{k}) \cdot \mathbf{r}] \tag{5.199a}$$

where the expansion coefficients $v(\mathbf{K}_h + \mathbf{k})$ depend on the values of the vector $\mathbf{K}_h + \mathbf{k}$. This series is defined without any ambiguity when the magnitude of the vector $\mathbf{k}$ is smaller than that of any of the vectors $\mathbf{K}_h + \mathbf{k}$, i.e., when $\mathbf{k}$ is the reduced-wave vector. Under this condition the eigenfunction u can be regarded as a function of both $\mathbf{r}$ and $\mathbf{k}$ and for clarity can be written in the form

$$u(\mathbf{k},\mathbf{r}) = \sum_h v(\mathbf{K}_h + \mathbf{k}) \exp[i(\mathbf{K}_h + \mathbf{k}) \cdot \mathbf{r}] \tag{5.199b}$$

It is seen from the form of Eq. 5.199b that, if the eigenfunctions are known in the central Brillouin zone ($\mathbf{K}_h = 0$), then they are known for any other zones ($\mathbf{K}_h \neq 0$).

5.10.11 The Eigenvalue Equation for a Periodic-Potential Problem. Substituting Eqs. 5.196 and 5.199b into Eq. 5.188, multiplying the result by the function

$$\exp[-i(\mathbf{K}_j + \mathbf{k}') \cdot \mathbf{r}]$$

where $\mathbf{k}'$ is an allowed reduced-wave vector and $\mathbf{K}_j$ is arbitrary, and integrating the multiplied equation over an integral number of Wigner–Seitz unit cells of coordinate space yield the relation

$$|\mathbf{K}_j + \mathbf{k}|^2 \frac{\hbar^2}{2m} v(\mathbf{K}_j + \mathbf{k}) + \sum_h U(\mathbf{K}_h)\, v(\mathbf{K}_j - \mathbf{K}_h + \mathbf{k}) = \epsilon(\mathbf{k})\, v(\mathbf{K}_j + \mathbf{k}) \quad \text{for all } \mathbf{K}_j \tag{5.200}$$

where the eigenvalue $\epsilon(\mathbf{k})$ is written as a function of the reduced-wave vector $\mathbf{k}$ to indicate that it corresponds to the eigenfunction $u(\mathbf{k},\mathbf{r})$. Equation 5.200

represents a set of simultaneous and homogeneous algebraic equations for the amplitudes $\nu(\mathbf{K}_j + \mathbf{k})$; each member of the set corresponds to a specific value $\mathbf{K}_j$. This set is called the one-electron eigenvalue equation in k-space for a periodic-potential problem. In other words, introducing the series expansions for $V(\mathbf{r})$ and $u(\mathbf{k},\mathbf{r})$ reduces the original differential Eq. 5.188 to a set of algebraic equations which is much easier to study than the original differential equation. When $V(\mathbf{r}) = 0$, i.e., when all the amplitudes $U(\mathbf{K}_h)$ are zero, Eq. 5.200 is equivalent to the equation derived for the energy eigenstates of the free-electron gas model.

A set of homogeneous algebraic equations admits nontrivial solutions only when the determinant of the coefficients vanishes. Applying this requirement to the set represented by Eq. 5.200 results in a secular equation for the eigenvalue $\epsilon(\mathbf{k})$. For a given value of k and an infinite number of values for $\mathbf{K}_j$, the secular equation is satisfied by an infinite number of values of $\epsilon(\mathbf{k})$. These values can be denoted by $\epsilon_n(\mathbf{k})$ and numbered in order of ascending magnitude. Associated with each $\epsilon_n(\mathbf{k})$ is a set of amplitudes $\nu_n(\mathbf{K}_j + \mathbf{k})$ and an eigenfunction $u_n(\mathbf{k},\mathbf{r})$. The eigenfunction $u_n(\mathbf{k},\mathbf{r})$ involves a constant multiplier whose magnitude is obtained by means of the normalization condition. It is clear that if an infinite number of functions

$$\exp\,[-i(\mathbf{K}_j + \mathbf{k}') \cdot \mathbf{r}]$$

is used in the derivation of Eq. 5.200, the corresponding secular equation cannot be solved in a practical manner. For this reason, only a finite number of $\mathbf{K}_j$'s can be considered.

The lowest eigenvalues $\epsilon_1(\mathbf{k})$ associated with the allowable values of the reduced-wave vector k are taken together. Since the allowable values of k are very close to each other in k-space, these eigenvalues can be represented by a function $\epsilon_1(\mathbf{k})$ that is almost continuous with respect to k. Thus a range of energies is obtained which is called the *first* or *lowest energy band.* Similarly, the next lowest energies $\epsilon_2(\mathbf{k})$ are taken together, and another range of energies, called the *second energy band,* is determined.

The amplitudes $\nu(\mathbf{K}_j + \mathbf{k})$ can be given an interesting physical interpretation. In a Bravais lattice consisting of N_c unit cells each of which has a volume Ω_c, the total integrated probability that an electron in the eigenstate $u(\mathbf{k},\mathbf{r})$ is somewhere in the lattice is given by the relation

$$\int_{(N_c\Omega_c)} u^*(\mathbf{k},\mathbf{r})\, u(\mathbf{k},\mathbf{r})\, d\mathbf{r} = 1 \qquad (5.201a)$$

where the integral is taken over the entire lattice. Substituting Eq. 5.199b into Eq. 5.201a yields

$$(N_c\Omega_c)\sum_j \nu^*(\mathbf{K}_j + \mathbf{k})\,\nu(\mathbf{K}_j + \mathbf{k}) = 1 \qquad (5.201b)$$

This relation implies that the quantity $(N_c\Omega_c)\,\nu^*(\mathbf{K}_j + \mathbf{k})\,\nu(\mathbf{K}_j + \mathbf{k})$ measures the probability that the electron characterized by the eigenfunction $u(\mathbf{k},\mathbf{r})$ will, as a result of a measurement, be found to have the momentum $(\mathbf{K}_j + \mathbf{k})\hbar$. For this reason $\nu(\mathbf{K}_j + \mathbf{k})$ is often interpreted as a momentum eigenfunction.

The results of energy-band calculations are substantially different from those derived from the free-electron gas model. Some features of these results are illustrated by the analysis of the one-dimensional example discussed in the next section.

5.10.12 Energy Bands in the One-Dimensional Lattice. In a lattice extending only along the x-axis, if the lattice points are a distance a apart, the potential energy $V(x)$ is periodic with a period equal to the lattice spacing a. In the other spaces, the periods are as listed in Table 5.3.

Table 5.3 Periods for Various Spaces

Space	Period
Reciprocal space	$1/a$
k-space	$2\pi/a$
Momentum space	$2\pi\hbar/a = h/a$

If distances are measured in units of a, energies in units of $h^2/2ma^2$, and momentum in units h/a, then Eqs. 5.196, 5.199b, and 5.200 reduce to the relation

$$V(x) = \sum_l U(l)\exp(2\pi i l x) \qquad \text{for } l = \text{integer} \qquad (5.202)$$

$$u(p,x) = \sum_q \nu(q+p)\exp[2\pi i(q+p)x] \qquad \text{for } q = \text{integer} \qquad (5.203)$$

and

$$(q+p)^2\,\nu(q+p) + \sum_l U(l)\,\nu(q+p-1) = \epsilon(p)\,\nu(q+p) \qquad (5.204)$$

where x is distance divided by a, ϵ is energy divided by $h^2/2ma^2$, and p and q are momentum divided by h/a. In k-space the central Brillouin zone extends

from $-\pi/a$ to π/a, whereas in p-space this zone extends from $-\frac{1}{2}$ to $\frac{1}{2}$. Finally, $\epsilon(p)$ is periodic with the period being equal to unity.

To avoid numerical complications, we will let the potential energy be a cosinusoidal function of x. In symbols

$$U(1) = U(-1) = -U$$

$$U(l) = 0 \qquad \text{for } l \neq \pm 1 \tag{5.205}$$

and

$$V(x) = -2U \cos 2\pi x \tag{5.206}$$

For this potential energy Eq. 5.204 reduces to the expression

$$(q+p)^2\, \nu(q+p) - U\nu(q+p-1) - U\nu(q+p-1) = \epsilon(p)\, \nu(q+p) \tag{5.207a}$$

or equivalently

$$U\nu(q+p+1) = [(q+p)^2 - \epsilon(p)]\ \nu(q+p) - U\nu(q+p-1) \tag{5.207b}$$

In the form of Eq. 5.207b, the one-electron problem appears as a recursion formula for the momentum eigenfunctions $\nu(q+p)$. Specifically, the values of $\nu(q+p)$ as a function of the integer q can be found if the values for two q's differing by unity are known. Not all pairs of initially assumed values result in allowable eigenfunctions. In other words, not all energy values are acceptable. Typical examples of $\epsilon(p)$ and $\nu(p)$ vs. p (or k) are shown in Figs. 5.26 and 5.27, respectively. Figure 5.26 shows that the energy is periodic in k-space, as expected. Moreover, acceptable energy levels occur for certain *energy bands*, but no energy levels occur in the *energy gaps* between the energy bands.

The occurrence of energy gaps is a direct consequence of the periodic-potential energy and can be explained by means of the following quantum-mechanical arguments. An electron can be regarded as a wave propagating along the x direction. This wave is scattered by the peaks of the cosinusoidal potential energy. The scattered waves travel backward along the x direction and are able to reinforce each other if they are in phase. This means that the path difference from one peak to the next must be a wavelength or an integral multiple of a wavelength. In general, the de Broglie wavelength is $2\pi/k$, where k is in actual units. At the end of the central Brillouin zone, the wavelength equals $2a$. Since the wave scattered from each successive peak must travel a distance of $2a$ greater than from the preceding peak, it follows that at the end of this zone the scattered waves reinforce each other. The superposition of the forward and

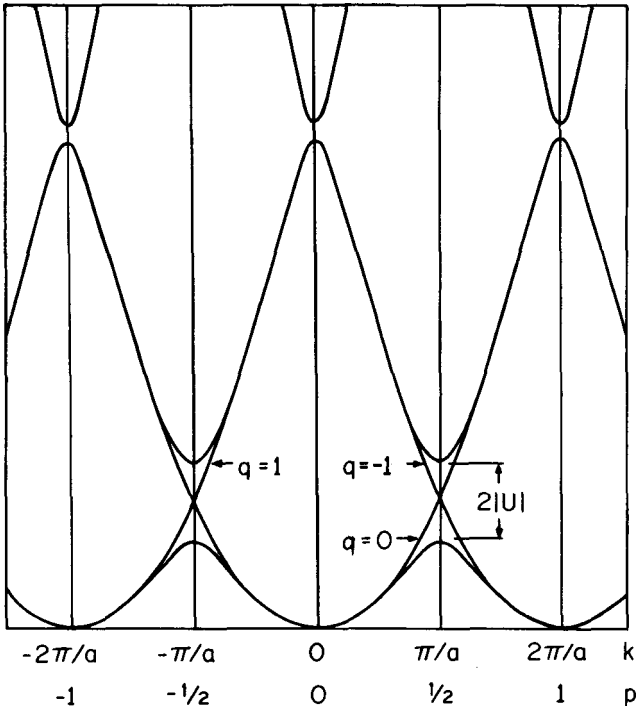


Fig. 5.26 Energy bands for one-dimensional periodic potential. (Adapted from J. C. Slater, *Quantum Theory of Molecules and Solids*, Vol. 2, McGraw-Hill Book Company, Inc., New York, 1965.)

backward waves results in a standing wave. The standing wave can have its maxima either at the minima of the potential energy, in which case the energy of the wave is depressed, or at the maxima of the potential energy, in which case the energy is raised. These two energies correspond to the two limits of the energy gap at the end of the central Brillouin zone.

The momentum eigenfunctions shown in Fig. 5.27 fit this wave interpretation. To find the eigenfunction $u(p,x)$ associated with a given energy $\epsilon(p)$, we must consider the set of points $p, p \pm 1, p \pm 2, \ldots$ and obtain the values of $\nu(q+p)$ at these points from the $\nu(p)$ vs. p curve. If p is about zero, it is seen from Fig. 5.27 that, for the lowest energy band (l = zero), only the momentum eigenfunction for p about zero is appreciable. All the others, for p about ± 1, $\pm 2, \ldots$ have magnitudes almost equal to zero. It follows that the eigenfunction $u(p,x)$ (Eq. 5.203) is almost a simple plane wave of the form $\exp(2\pi ipx)$. On the other hand, as p increases to, say, $\frac{1}{2}$, then not only $\nu(\frac{1}{2})$ but also $\nu(-\frac{1}{2})$ yields considerable amplitudes. At p equal to $\pm\frac{1}{2}$, the momentum eigenfunctions for

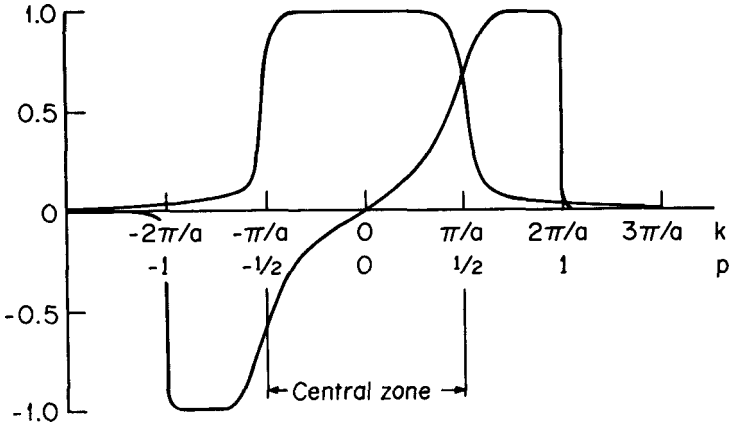


Fig. 5.27 Momentum eigenfunctions for the two lowest energy bands in Fig. 5.26. (Adapted from J. C. Slater, *Quantum Theory of Molecules and Solids*, Vol. 2, McGraw-Hill Book Company, Inc., New York, 1965.)

the lowest energy band are equal, each being approximately $\sqrt{1/2}$. This implies a superposition of a direct and a reflected wave. The superposition results in an eigenfunction $u(1/2,x)$ proportional to cos πx and in a maximum electron charge density at the location of the minimum of the potential energy. On the other hand, for the second energy band (l equal to unity) and p equal to $\pm 1/2$, the two values of v are of opposite sign. Therefore a sinusoidal eigenfunction with a maximum charge density at the point of the maximum of the potential energy results. The behavior of v and the corresponding u for other values of p can be readily deduced from Fig. 5.27.

5.10.13 Energy Bands in Three-Dimensional Lattices—The Fermi Surface. When the plane-wave-expansion method is used for calculating one-electron energies in three-dimensional crystal lattices, energy bands and energy gaps are again obtained. Because the computations become extremely tedious, alternative computational methods have been developed in an effort to reduce the mathematical complexity of the problem.[26] Description of these methods is beyond the scope of this chapter. For completeness here it suffices to state the following.

1. Potential energies and eigenfunctions are expanded in terms of other basis functions that are more convenient to compute than are the plane waves discussed previously. Each choice of basis functions results in a different method of analysis. For example, two such methods are qualified as the *tight binding and orthogonalized plane-wave method* (OPW), and the *cellular and augmented plane-wave method* (APW). Different materials are treated by different methods.

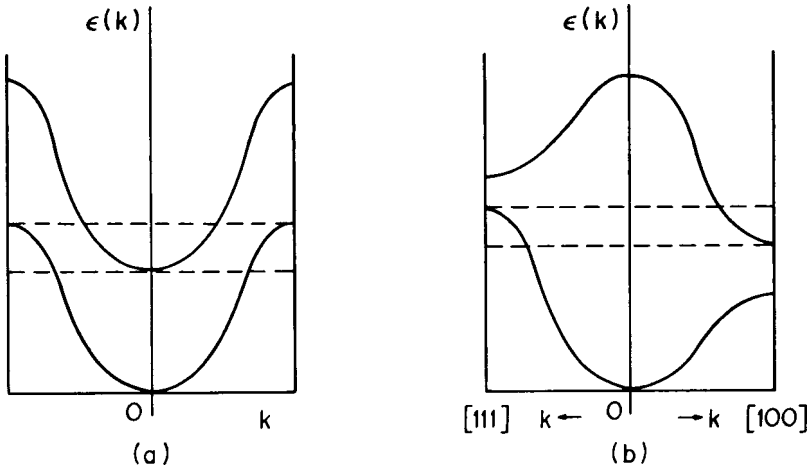


Fig. 5.28 Overlapping energy bands. (a) In one dimension. (b) In three dimensions.

2. Regardless of method, the one-electron equation (Eq. 5.188) is reduced to a set of homogeneous algebraic equations for the expansion coefficients of the eigenfunctions. The need for this set to admit nontrivial solutions results in a secular equation for the energy eigenvalues.

3. Energy bands and energy gaps are obtained because of the periodic character of the potential energy. The gaps can be explained by the superposition of forward and scattered waves.

4. Not all energy bands are always separated by energy gaps. In general, for a given value **k** the one-electron energy eigenvalues belong to different energy bands. This, however, does not exclude the possibility of overlapping bands, as illustrated by the graphs of schematic energy vs. **k** in Fig. 5.28.

5. The number of one-electron states in each band equals the number of distinct values of **k** in the central Brillouin zone. The latter in turn equals the number of unit cells in the crystal. Each state can accommodate only two electrons, and these are of opposite spin. If the order in which the bands occur, their widths in energy, and the occurrence of gaps are known, it is easy to decide whether the ground state of the electronic structure of the crystal consists of exactly filled bands or not.

6. Many properties of a given solid depend on the degree to which bands are occupied by the valence electrons. A full band is nonconducting, and the solid to which it corresponds may be an electrical insulator. If a band is only partially filled (e.g., when there is one valence electron per atom to be put into two spin–orbitals per unit cell), then the solid is a good conductor, a metal. If energy bands overlap, the same effect may exist. Although the lower bands could

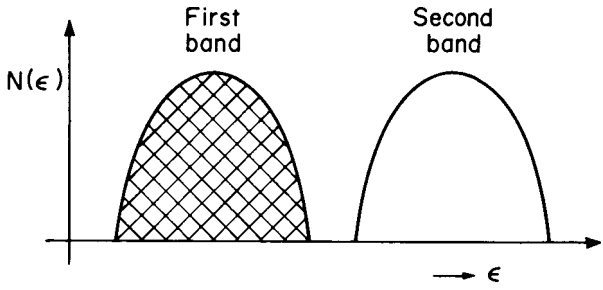


Fig. 5.29 Density of states $N(\epsilon)$ vs. ϵ for air insulators.

contain all the electrons, some electrons prefer to spill into the lower levels of the upper band. Thus the solid behaves like one with a partially filled band.

7. Important information is contained in the functional dependence of the density of states $N(\epsilon)$ on the energy ϵ. The density of states for an insulator may be as shown in Fig. 5.29. Here the lower band is filled, but the upper band is empty. In a metal the bands may overlap, and the density of states may be a complicated function of energy. Two graphs of $N(\epsilon)$ vs. ϵ for tungsten and rhenium are shown in Figs. 5.30 and 5.31, respectively.

8. Another way of representing the information obtained from energy-band calculations is by means of maps of constant-energy surfaces in k-space. An example of such a map is given in Fig. 5.32.

9. A constant-energy surface of particular interest is that corresponding to the maximum energy level occupied by the valence electrons when the solid is in the ground state. As already stated, this surface is defined as the Fermi surface, and its energy is called the Fermi energy.

Fermi surfaces resulting from energy-band calculations have much more complicated shapes than the spherical shape obtained by means of the free-electron gas model. An example of a Fermi surface is shown in Fig. 5.33. Other examples correspond to much more intricate surface geometries.

This completes the brief and incomplete survey of the approximate theory of energy bands in crystals. In summary, this theory involves: (1) a postulated one-electron periodic-potential energy $V(\mathbf{r})$ for regions in the crystal which are not affected by surface perturbations; (2) expansion techniques, such as expansions in terms of plane waves, to reduce the one-electron differential equation of the crystal into a set of algebraic equations; (3) approximate solutions of this set for the one-electron energy eigenvalues and eigenfunctions; and (4) the classification of eigenvalues into energy bands. For a given state of the electronic structure, certain bands may be partially or completely filled by the valence electrons. The degree to which bands are filled is related to

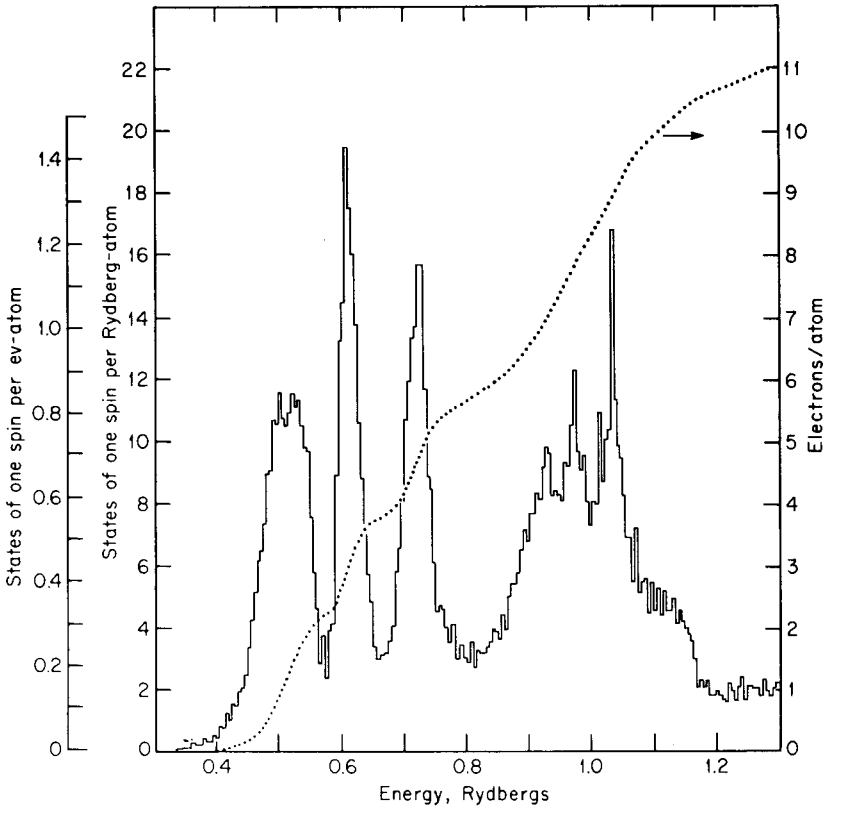


Fig. 5.30 Density of states $N(\epsilon)$ vs. ϵ for tungsten. [Adapted from L. F. Mattheis, Fermi Surface in Tungsten, *Physical Review*, **139**(6A): A1893-1904 (1965).]

important properties of the crystal, such as the conducting and insulating properties.

5.10.14 Discussion of One-Electron Approximations. In the preceding sections, an attempt was made to acquaint the reader with some of the ideas, terminology, and results of the one-electron approximation for solids. The essence of the approximation is that it reduces the many-electron problem for the electronic structure of the crystal to an approximate one-electron problem. In other words, the reduction implies that the energy of a given state of the crystal is approximated by the sum of the energies of the occupied one-electron eigenfunctions.

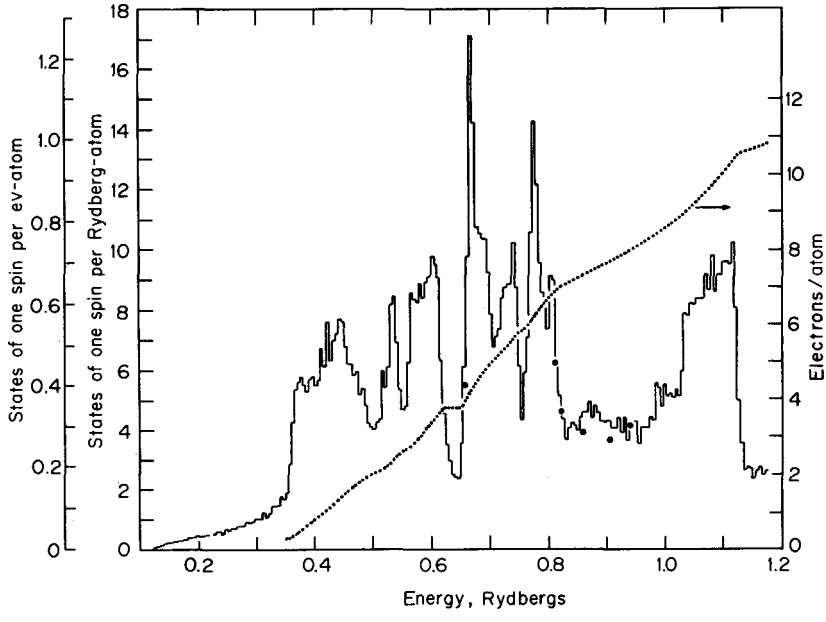


Fig. 5.31 Density of states $N(\epsilon)$ vs. ϵ for rhenium. [Adapted from L. F. Mattheis, Band Structure and Fermi Surface for Rhenium, *Physical Review*, **151**(2): 450-464 (1966).]

The one-electron equation (Eq. 5.188) is the result of two approximations. First, the nuclear motions are separated from those of the valence electrons by means of the Born–Oppenheimer approximation. It follows that results derived from Eq. 5.188 do not provide information about lattice vibrations, effects of thermal expansion on the energy eigenstates, and related questions. Second, the valence electron coordinates are assumed separable. The potential energy for each valence electron is approximated by a function of position that is more or less consistent with a "reasonable" charge distribution in the crystal. The one-electron energies that result from this potential energy represent only the first step toward a self-consistent-field approximation. To achieve a truly self-consistent-field solution, we should continue the calculation by an iterative procedure. The charge density of the valence electrons should be computed on the basis of the occupied one-electron states obtained in the first step. Then this charge density should be used to compute another one-electron potential energy and new one-electron energy eigenstates. The calculation must be repeated until a truly self-consistent answer is established. Extensive calculations along this direction have not been performed to date.

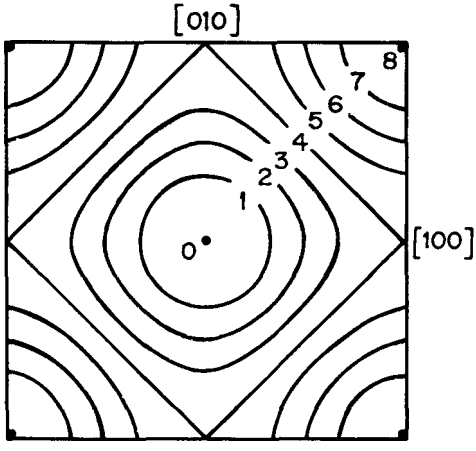


Fig. 5.32 Constant-energy contours in a simple cubic lattice.

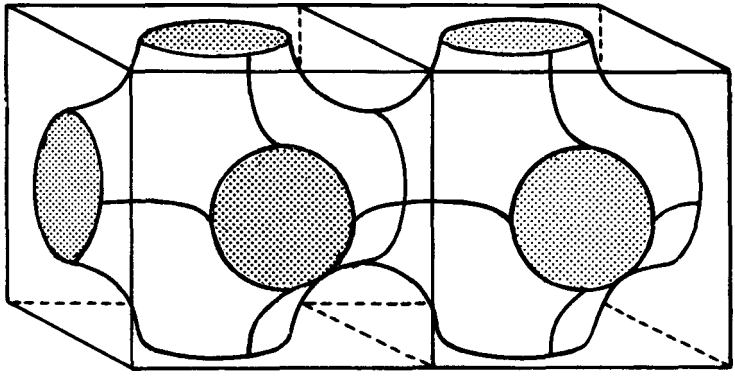


Fig. 5.33 Schematic of a multiply connected Fermi surface.

Even if such calculations were available, however, the self-consistent-field results would be quantitatively inadequate. They must be corrected for exchange effects. A reasonable way of accounting for exchange effects would be to use the Hartree–Fock method for reducing the many-electron problem into a one-electron problem, in a manner similar to that employed for atoms or molecules. If this were done, the negative of the one-electron energies would be approximately equal to the ionization energies of the crystal orbitals (Koopmans' theorem).

Unfortunately even the Hartree–Fock method yields results in poor agreement with experiments. The reasons for the inadequacy are as follows: First, the Hartree–Fock approach does not account for *correlation effects*, namely, the fact that the presence of a valence electron at a particular position affects the charge distribution of the other valence electrons. Correlation effects have an important influence on the total energy of the system and cannot be neglected. Without their consideration, the agreement between one-electron calculations of any type and experimental results on metal binding energies is very poor. Correlation effects are also important near boundary surfaces. They are discussed in Chap. 9.

The second reason stems from the difficulties associated with solving the Hartree–Fock equation for a crystal. Because of this difficulty, exchange effects are included in the calculation in an approximate manner.[28] As a consequence of the approximation, the exchange energy correction for each one-electron energy turns out to be dependent on the value of the wave vector **k**. The dependence on **k** in turn yields results contrary to experiment.

To avoid the inadequacies of the Hartree–Fock approximation, Slater suggests another prescription for the reduction of the many-electron equation into a one-electron equation,[28] but it is not repeated here. For the purposes of this book, we shall assume that a good one-electron eigenvalue equation is available and that good one-electron energy eigenstates can be found. These eigenstates are used in connection with thermodynamic analyses discussed later.

References

1. F. K. Richtmyer and E. H. Kennard, *Introduction to Modern Physics*, 4th ed., McGraw-Hill Book Company, Inc., New York, 1947.
2. R. B. Leighton, *Principles of Modern Physics*, McGraw-Hill Book Company, Inc., New York, 1959.
3. M. Planck, *The Theory of Heat Radiation*, translated by M. Masius, The Blackiston Division, McGraw-Hill Book Company, Inc., New York, 1914.
4. A. Einstein, Planck's Theory of Radiation and the Theory of Specific Heat, *Ann. Phys.*, **22**: 180 (1907).
5. P. Debye, Zur Theorie der Specifischen Wärmen, *Ann. Phys.*, **39**: 789 (1912).
6. N. Bohr, On the Constitution of Atoms and Molecules, *Phil. Mag.*, **26**: 1 (1913).
7. A. Einstein, Über einem die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt, *Ann. Physik*, **17**: 132 (1905).
8. L. deBroglie, A Tentative Theory of Light Quanta, *Phil. Mag.*, **47**: 446 (1924).
9. C. Davisson and L. H. Germer, Diffraction of Electrons by a Crystal of Nickel, *Phys. Rev.*, **30**: 705 (1927).
10. W. Heisenberg, Über der anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik, *Z. Physik*, **43**: 172 (1927).
11. N. Bohr, The Quantum Postulate and the Recent Development of Atomic Theory, *Nature*, **121**: 580 (1928); Das Quantenpostulat und die neuere Entwicklung der Atomistik, *Naturwissenschaften*, **16**: 245 (1928).
12. R. C. Tolman, *The Principles of Statistical Mechanics*, Clarendon Press, Oxford, (1938).

13. H. Goldstein, *Classical Mechanics*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1950.
14. L. I. Schieff, *Quantum Mechanics*, 2nd ed., Vol. 1, p. 9, McGraw-Hill Book Company, Inc., New York, 1955.
15. Ref. 4, p. 154.
16. C. A. Coulson, *Valence*, 2nd ed., Oxford University Press, Inc., New York, 1961.
17. Ref. 12, Section 76.
18. D. R. Hartree, The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods, pp. 89–110; Part II. Some Results and Discussion, pp. 111–112, *Proc. Cambridge Phil. Soc.*, **24** (1928).
19. J. C. Slater, *Quantum Theory of Molecules and Solids*, Vol. 1, p. 9, McGraw-Hill Book Company, Inc., New York, 1963.
20. P. M. Morse, Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels, *Phys. Rev.*, **34**: 57 (1929).
21. Ref. 19, Chaps. 1 and 2.
22. Ref. 19, Chap. 3, pp. 93-97.
23. T. Koopmans, Ordering of Wave Functions and Eigenenergies to the Individual Electrons of an Atom, *Physica*, **1**: 104 (1933); also Ref. 19, Chap. 3, p. 96.
24. C. C. J. Roothaan, New Developments in Molecular Orbital Theory, *Rev. Mod. Phys.*, **23**: 69–89 (1951).
25. See Ref. 19.
26. J. C. Slater, *Quantum Theory of Molecules and Solids*, Vol. 2, McGraw-Hill Book Company, Inc., New York, 1965.
27. Ref. 26, p. 101.
28. J. C. Slater, *Theory of Molecules and Solids*, Vol. 3, McGraw-Hill Book Company, Inc., New York, 1967.

6

Principles of Thermodynamics

6.1 General Remarks

In Volume I many thermodynamic concepts and relations are presented which illustrate the importance of thermodynamics in the understanding and analysis of thermionic conversion and thermionic converters. For example, the processes of conversion of heat to work, condensation, evaporation, adsorption, desorption, electron and ion thermionic emission from a solid, and production of charged particles in a gaseous phase which are involved in thermionic conversion require thermodynamics for their analyses. Similarly, analyzing performance characteristics of thermionic converters, such as efficiency and power characteristics, requires the use of thermodynamics. Because of the importance of the subject, the principles of thermodynamics and some of the results that can be deduced from these principles are discussed here and in Chaps. 7 and 8.

Thermodynamics is a part of physical science which is concerned with the conditions material systems may assume and the changes in conditions which may occur either spontaneously or as a result of interactions between systems, including interactions such as heat, which cannot be described in terms of mechanics. It is basic to the distinction between mechanics and thermodynamics that mechanics deals only with states for which all the energy of a system is convertible to work, whereas thermodynamics deals with states for which a fraction of the energy of a system is not convertible to work without other effects. This distinction is clarified later.

The scope of thermodynamics has been steadily broadening over the years as a result of better understanding of its underlying principles. The beginnings of thermodynamics are found in the early nineteenth century in the study of the motive power of heat, i.e., the capability of hot bodies to produce mechanical work. These studies revealed several underlying principles, the most important of which are the first and second laws of thermodynamics. The discovery and clear enunciation of the two laws by Clausius in 1850 broadened the scope of the subject greatly. Over the following decades thermodynamics developed into a science, now known as classical thermodynamics, concerned primarily with physical systems in or passing through stable equilibrium states. More recently thermodynamics has been extended to include physical systems in nonequilibrium states.

Because the development of thermodynamics, in contrast with other parts of physics, has progressed without regard to the details of the microscopic structures of physical systems, many scientists, beginning with Boltzmann, have

felt that the laws of thermodynamics might be justified and explained by means of detailed microscopic descriptions of matter, by laws of mechanics (classical or quantum), and by statistics. Work in this direction has resulted in the development of the branch of thermodynamics known as statistical thermodynamics, which, though it plays an important role in evaluating some properties of systems, has not achieved one of its original goals, namely, the derivation of the laws of thermodynamics from the laws of mechanics. In particular, it now appears that the second law is an independent law of nature. This point will be clarified in this chapter.

The discussion that follows is divided into two parts. First, the general principles of thermodynamics which apply to any physical system in any state and the corollaries of these principles are presented without reference to the detailed microscopic structure of matter. This presentation, therefore, encompasses both classical and nonequilibrium thermodynamics. Second, the detailed microscopic structures of physical systems are considered, and the states of such systems are discussed in the light of the laws of both quantum mechanics and thermodynamics. The discussion reveals that quantum mechanics and thermodynamics are complementary parts of physical science and are neither separable nor derivable from one another.

The presentation of thermodynamics given in this book makes no distinction between the so-called macroscopic and microscopic states of a system, irrespective of whether the system is large or small. We will see that the state of any system can be described in terms of either a set of expectation values of independent properties or a set of independent quantum-mechanical probabilities. These two sets are completely equivalent, and either describes fully any state, including the states encountered in quantum mechanics (which are usually represented by means of wave functions) and the states encountered in thermodynamics.

The methods of this chapter are applicable to systems having any number of degrees of freedom. It is true that systems with enormous numbers of degrees of freedom differ substantially from those with few degrees of freedom. The differences are only quantitative and not qualitative, however. For example, in general, measurement indicates that the relative dispersion in values of the energy of a system with few degrees of freedom is much larger than that of a system with many degrees of freedom. Nevertheless, conceptually both types of systems have energies that are not dispersion free. Contrary to common opinion, not only are all the methods of thermodynamics applicable to systems with few degrees of freedom and with many degrees of freedom but also the exact thermodynamic analysis of a specific system may prove far easier when the number of degrees of freedom is small. This is demonstrated by several examples given in Chaps. 7 and 8.

Special attention is given in this chapter to definitions, even of terms that appear to be well understood. We believe that many unjustified conclusions are reached in the literature because of erroneous interpretation of terms.

6.2 Preliminary Concepts

6.2.1 Systems, States, and Properties. The description of physical phenomena is based on the concept of state of a system and the changes of state that occur either spontaneously or because of interactions with other systems. Knowledge of the states of a system is equivalent to knowledge of the outcomes of all possible observations on the system, i.e., observations of all its properties.

The term "system" means any identifiable collection of matter that can be separated from everything else by a well-defined surface so that changes in everything else need not affect the condition of the collection. Examples of systems are a number of water molecules confined within a watertight container, a thermonuclear plasma confined by a suitably designed magnetic field, and the vapor of a solid within a cavity in the solid.

The particles of a system may have a finite rest mass (as do molecules, atoms, or electrons) or zero rest mass (as do photons). The number of particles of a system can be large (as in a cubic centimeter of tungsten) or small (as in a hydrogen atom).

At any instant of time a system is in a condition called a *state,* which encompasses all that can be said about the results of any measurements or observations that can be performed on the system at that time. The state at a given instant of time determines the properties of the system. A *property* is any quantity whose value depends on the state but not on the history of the system. For a given state, the value of a property can be determined by some type of measurement (some physical operation or test). When the values of *all* properties of two states are identical the *two states are identical;* otherwise they are different.

Because of unavoidable quantum-mechanical considerations, a simple measurement, no matter how perfect, will not yield the value for a property. In general, a very large number of measurements of the same kind is required; each measurement is performed on a duplicate of the system in the given state, and each yields a different result. The value of a property is defined, therefore, as the sum of all the results of the measurements of the property divided by the number of measurements, i.e., as the arithmetic average or expectation value of all the results of the measurements of the property. We will show here that a state can be described by a set of weighting factors, or probabilities, that enter into the evaluation of all properties.

Although an indefinite number of properties can be associated with a state, the values of these properties are not all independent of each other. The number of

independent properties is finite and is either small or large depending on the type of state in question, i.e., on whether or not it is a stable equilibrium state. The meaning of the terms defining the various types of states is given in the following discussion.

6.2.2 Work and Nonwork. An interaction between two systems such that whatever happens in each of the interacting systems could have been brought about while the sole effect external to each system was the change in level of a mass in a gravity field is defined as *work.* That is, work is an interaction that must satisfy a test: Each of the interacting systems is separately required to experience the same changes as in the actual interaction, but, through alteration of circumstances external to that system, the effect outside of it is solely the rise or fall of a mass. If this test can be satisfied with both systems, then the original interaction is work.

Experimental evidence shows that the following interactions conform to this definition and are, therefore, work interactions: (1) The displacement of a point on the boundary of a system at which point a force is exerted in the direction of the displacement by another system. This is essentially the definition of work in mechanics (it is less general than the definition given previously, covering only interactions considered in mechanics) and implies that the force applied on a body by a stationary force field, such as gravity or an electrostatic field, results in no work even though the body is in motion—e.g., freely falling. (2) The electromagnetic interactions between primary and secondary coils of a transformer. (3) The exchange of radiation emitted by lasers. Such radiation consists of wave-particles, which are in order relative to each other as compared with those from an incandescent lamp, for example, which are randomly arranged. On the other hand, it can be shown by virtue of the second law that the exchange of radiation between a hot and a cold system (such as blackbody radiation) does not conform to the requirement stated in the definition of work and is not, therefore, a work interaction.

Work is one of several distinguishable types of interactions that can occur between two systems. Any interaction that is not work will be called *nonwork.* A special type of nonwork interaction discussed here in detail is a *heat interaction.* In general, an interaction is fully specified if the exact sequence of states, called the *path,* of each of the interacting systems is specified. In turn, the path of one of the systems and the interactions that occur specify the *process* experienced by the system. In an interaction that satisfies the definition of work, each system is said to be experiencing an *adiabatic process.*

The state and, therefore, the values of the properties of a system can change either spontaneously or as a result of interactions with other systems. During changes of state the values of certain properties are subject to limitations imposed by the laws of physics. The limitations imposed by thermodynamics are

consequences of the first and second laws, which will be stated here in an unusually general way.

6.3 First Law

The first law is a statement of existence of a property called *energy*. Based on the concept of work, it can be stated as follows: For any process involving no effects external to the system except displacement of a mass between specified levels in a gravity field, the magnitude of that mass is fixed by the end states of the system and is independent of the details of the process.

This law has many implications or corollaries, some of which are discussed in the following subsections.

6.3.1 Uniqueness of Work Values. By virtue of the first law, a unique number can be assigned to a work interaction between two systems. This number is n when n arbitrarily selected units of mass are displaced between two arbitrarily specified levels in a gravity field. The sign of work is selected positive when *work is done by a system,* i.e., when the sole external effect would be to cause the rise of a weight.

6.3.2 Definition of Property Energy. By virtue of the first law, the work in an adiabatic process depends on the initial and final, or end, states only. It follows that a property of a system, *energy,* can be defined such that its change of value between states A_1 and A_2 is equal to the work W^a_{12} involved in an adiabatic process that has A_1 and A_2 as end states. In symbols, if E_1 and E_2 denote the energies of states A_1 and A_2, respectively, and W^a_{12} is the work done in any adiabatic process connecting A_1 and A_2, then

$$E_2 - E_1 = -W^a_{12} \tag{6.1}$$

Energy is only partially defined by Eq. 6.1 because the difference between E_1 and E_2 does not specify the values of E_1 and E_2. An arbitrary value, such as zero, however, can be assigned to E_0 of some reference state A_0 so that the energies of all other states are fixed by Eq. 6.1.

We will show in this discussion that, by virtue of the second law, certain interactions that cause change of energy cannot be described as work. The corresponding process is, by definition, a *nonadiabatic* process. Because energy is a property, the energy change associated with any nonadiabatic process will equal numerically the work in any adiabatic process between the same end states.

6.3.3 Additivity of Energy. Because work values are additive, energy is an additive or *extensive* property; i.e., the energy of a whole is equal to the sum of the energies of its parts.

6.3.4 Conservation of Energy. The magnitude of the work in an adiabatic process can be positive, negative, or zero depending on the end states of the process, but zero adiabatic work does not necessarily imply identical end states. An adiabatic, zero-work change of state may occur spontaneously in a system, e.g., when an electrical capacitor discharges through an internal resistor. During such a process, Eq. 6.1 requires that the energy of the system remain invariant. This is the so-called *law of conservation of energy*, which is a consequence of the first law of thermodynamics. The first law, on the other hand, cannot be deduced from the law of conservation of energy.

A system that experiences an adiabatic zero-work process is said to be *isolated* from all other systems in the environment.

6.3.5 Impossibility of a Perpetual-Motion Machine of the First Kind. A perpetual-motion machine of the first kind (PMM1) is any system that undergoes a cycle and has no external effect except the rise of a weight. The term *cycle* (or cyclic process) is defined as a process in which the initial and final states of the system are identical.

By virtue of the first law, the work of an adiabatic cycle must be equal to zero, and therefore a PMM1 is impossible.

6.3.6 The First Law and Relativity. According to the general theory of relativity, the energy E of a system is equal to the product of its mass m and the square of the speed of light c in vacuo, namely,

$$E = mc^2 \tag{6.2}$$

By virtue of the first law, the energy E of an isolated system is conserved. Therefore Eq. 6.2 requires that the mass m of the system is also conserved, regardless of the processes occurring within the system (such as chemical reactions, nuclear reactions, and creation and annihilation of particles). Moreover, if the energy of a system is altered by an adiabatic or a nonadiabatic interaction, the mass of the system must also be altered in accordance with Eq. 6.2. In most applications the energy transferred to or from a system is so much smaller than the total energy of the system that the change in mass is negligible.

6.4 Types of States

Various types of states can be distinguished. The *equilibrium state* is a state that does not change with time while the system is isolated from all other systems in the environment. Equilibrium state is sometimes confused with steady state. A *steady state* does not change with time even though the system, rather than being isolated, is interacting with other systems. All states that are not equilibrium are called *nonequilibrium* states.

If a system is in an equilibrium state, a finite change of state may be caused by interactions with the environment, including those that leave no finite net effects in the environment. Several types of equilibrium states can be distinguished. Among these is the *stable equilibrium state,* for which a finite change of state cannot occur despite interactions that leave no net effects in the environment.

Consistent with its nature, a system might assume any one of a set of possible states, but some of the possible states may be prohibited by restrictions imposed by the environment. These restrictions are called *constraints.* For example, a given amount of gas may assume any of a large number of states, but the same amount of gas confined in a gastight container of fixed volume is restricted to states with volumes smaller than or equal to the volume of the container. The possible states of the system which are consistent with the constraints are called *allowed states.*

6.5 Second Law

The second law is a statement of existence of stable equilibrium states. More than the first law, the second law distinguishes thermodynamics from other parts of physics. Because it has far-reaching implications concerning the properties and behavior of matter, it is frequently invoked in discussions of philosophy as well as physics.

Of the many statements of the second law, those of Clausius, Planck, and Caratheodory are the most notable. They all imply the existence of a stable equilibrium state for given values of energy, number of particles, and constraints—an implication that we will take to be the essential element of the second law.

In terms of the preceding definitions, the second law can be stated as follows: Among all the allowed states of a system with given values of energy, numbers of particles, and constraints, one, and only one, is a stable equilibrium state.

The second law cannot be derived from the laws of mechanics (classical or quantum). It asserts that a stable equilibrium state exists for each value of the energy, whereas in mechanics the only stable equilibrium state is that of minimum energy, the so-called ground state.

Some corollaries of the second law follow.

6.5.1 State Principle. By virtue of the second law, stable equilibrium states exist, and a unique stable equilibrium state corresponds to each set of values of energy, numbers of particles, and constraints. The uniqueness implies the following corollary: The value of any property of a system in a stable equilibrium state can be expressed as a function of the values of the energy, numbers of particles, and constraints only. This corollary, which is known as the *state principle,* does not apply to states that are not stable equilibrium states.

6.5.2 Reversible and Irreversible Processes. A process is *reversible* if the system and its environment can be restored to their initial states, except for differences of smaller order of magnitude than the maximum changes that occur during the process. This is the thermodynamic definition of reversibility. It is broader than the definition given in mechanics in terms of the special process of time reversal, which can be accomplished only by reversing all momenta.

Any process that does not satisfy the definition of a reversible process we shall call an *irreversible process.*

6.5.3 Impossibility of a Perpetual-Motion Machine of the Second Kind. From the definition of a stable equilibrium state, we can show that a system in such a state can receive but cannot produce work without changes in values of constraints. The plausibility of this corollary is evident when we observe that work can always be used to change from any state to a nonequilibrium state by creating relative velocity of parts or compression to a smaller volume. If a system in a stable equilibrium state were to produce work, that work could be used to cause the system itself to end in a nonequilibrium state with no net effects in the environment, and this is contrary to the definition of its initial state.

This corollary applies strictly to systems whose energy can be increased indefinitely, such as any system with particles free to move from one place to another or any system with translational degrees of freedom. Most practical systems satisfy this requirement. The generalization to systems that do not satisfy the requirement, e.g., a nuclear-spin system, is not discussed here.

A perpetual-motion machine of the second kind (PMM2) is any device that would deliver net work (i.e., raise an external weight) while undergoing a cycle and interacting with a single system in a stable equilibrium state. It is a device that permits a system to violate the corollary stated here and is, therefore, impossible.

6.5.4 Mutual Stable Equilibrium and Reservoirs. Before we discuss the next corollary, we must define the terms "mutual stable equilibrium" and "reservoir." Two systems A and B are said to be in *mutual stable equilibrium* if the combined system AB is in a stable equilibrium state. We can readily verify that, if two systems are in mutual stable equilibrium, each system must be in a stable equilibrium state. Moreover, if the two systems are brought into communication so that interactions are not prohibited, no interactions will occur unless the allowed states of at least one of the systems are altered.

A *reservoir* R is a special kind of a system that provides useful reference states for applications of the second law and satisfies the following conditions: (1) It is closed to the transfer of rest mass; i.e., its boundaries cannot be crossed by material particles. (2) Its constraints are invariant with time; e.g., its volume has the same value at all times. (3) It passes through stable equilibrium states only.

(4) In the course of finite interactions, it remains in mutual stable equilibrium with a duplicate of itself that experiences no such interactions. An example of a reservoir is a mixture of solid, liquid, and vapor water in mutual stable equilibrium while enclosed in a container.

6.5.5 Work Done Reversibly by a System in Combination with a Reservoir. A system A in combination with a reservoir R can experience a work interaction with other systems either directly or through an intermediate device X, or both, while X is undergoing a cyclic process (Fig. 6.1). During this work interaction no

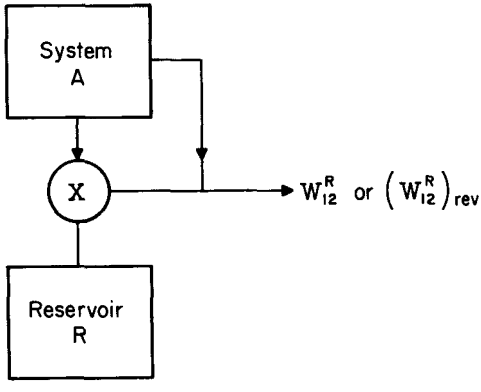


Fig. 6.1 Schematic of system A changing state from A_1 to A_2 and doing work W_{12}^R or $(W_{12}^R)_{rev}$ in combination with a cyclic device X and a reservoir R.

rest mass is exchanged between A and R while the state of A is changing from A_1 to A_2. The work done will be denoted by W_{12}^R for processes in general and by $(W_{12}^R)_{rev}$ for processes that are reversible.

By virtue of the second law, we can show that, for a given change of state from A_1 to A_2, the work $(W_{12}^R)_{rev}$ is the same for all reversible processes and for all reservoirs in mutual stable equilibrium with each other; i.e., $(W_{12}^R)_{rev}$ is fixed by the end states A_1 and A_2. Moreover, W_{12}^R cannot exceed $(W_{12}^R)_{rev}$ when both correspond to the same change of state from A_1 to A_2.

The proof of this theorem proceeds as follows. During a reversible work interaction α, the change of state from A_1 to A_2 is assumed to result in work done $(W_{12}^R)'_{rev}$ greater than $(W_{12}^R)_{rev}$. Interaction α is combined with a reversible work interaction β, which leads from A_2 to A_1 and requires work done on A equal to $(W_{12}^R)_{rev}$. The net effect of α plus β is work done $[(W_{12}^R)'_{rev} - (W_{12}^R)_{rev}]$ without any change of A. This work can be due only to a change of state of reservoir R. Such a result contradicts the corollary discussed

in Sec. 6.5.3 since R is a system in a stable equilibrium state. Since, by virtue of the second law, systems in stable equilibrium states exist, the contradiction is unacceptable. It follows that the original assumption is absurd and the theorem is proved.

A similar procedure is used to prove that W_{12}^R cannot be greater than $(W_{12}^R)_{\text{rev}}$ and that all reservoirs in mutual stable equilibrium are equivalent.

6.6 Entropy

The two laws of thermodynamics imply the existence of a property called *entropy* which is conserved in all reversible processes. Its definition is based on another property called *available work.*

6.6.1 Available Work. For a given system and reservoir, $(W_{12}^R)_{\text{rev}}$ depends on the end states of the system only, and thus a property Ω, *available work,* can be defined such that a change of value of Ω between states A_1 and A_2 is equal to the work $(W_{12}^R)_{\text{rev}}$; i.e.,

$$(\Omega_2 - \Omega_1) \equiv -(W_{12}^R)_{\text{rev}} \tag{6.3}$$

Using the second law, we can easily verify that available work is an additive or extensive property.

6.6.2 Definition of Property Entropy. Since both energy and available work are extensive properties, the difference between them is also an extensive property. *Entropy* is defined as the extensive property S whose change DS in a change of state is equal to a positive constant c_R times the difference between the corresponding changes of energy DE and available work $D\Omega$ evaluated with respect to a standard reservoir, i.e.,

$$DS \equiv c_R\,(DE - D\Omega) \tag{6.4}$$

The symbol D denotes a change, finite or infinitesimal, in magnitude of the property whose symbol it precedes.

The standard reservoir can be a mixture of solid, liquid, and vapor water in mutual stable equilibrium, and an arbitrary value, explained later, can be assigned to the constant c_R. Subject to subsequent considerations, an arbitrary value can be assigned to the entropy of an arbitrarily selected reference state of the system. Then, by means of Eq. 6.4, a value can be found for any state, equilibrium or nonequilibrium, once values for E and Ω have been determined by appropriate measurements.

6.6.3 Entropy Changes in Adiabatic Processes. If the process from state A_1 to state A_2 in system A is adiabatic (i.e., if system A experiences only work interactions) then, according to Eq. 6.1, the work done by A must equal the decrease in the energy of A. According to Eq. 6.3, for a reversible adiabatic process, the work done must also equal the decrease in available work; thus the difference between the decrease in energy and the decrease in available work must be equal to zero. It follows from Eq. 6.4 that for a reversible adiabatic process the entropy change of the system is equal to zero; i.e., the entropy is invariant. Moreover, because the corollary of Sec. 6.5.5 states that the work for any adiabatic process cannot exceed that for the reversible adiabatic process, it follows that, for an adiabatic process in general, the entropy change must be greater than or at most equal to zero. These conclusions can be written as

$$(DS)^a_{\text{rev}} = 0 \quad \text{for reversible adiabatic processes} \tag{6.5}$$

and

$$(DS)^a \geqslant 0 \quad \text{for adiabatic processes} \tag{6.6}$$

6.6.4 Principle of Increase of Entropy. When relation 6.6 is applied to any process in an isolated system (i.e., to an adiabatic process for which the work is zero), it becomes

$$DS \text{ (isolated)} \geqslant 0 \tag{6.7}$$

Thus the entropy of an isolated system may remain constant or may increase, but a decrease of entropy in an isolated system is impossible. Because any process in any system can be conceptually changed to a process in an isolated system by including all systems with which the original system interacts, the conclusion represented by relation 6.7 is of great generality. This relation is known as the *principle of increase of entropy*—although it is more strictly the principle of the impossibility of a decrease of entropy. According to this principle, whenever any process occurs in nature, the total entropy of all systems involved in the process must either increase or, if the process is reversible, remain constant.

6.7 Concepts Associated with Stable Equilibrium

6.7.1 Criterion for Stable Equilibrium. Changes of state in an isolated system can occur only until the system reaches the unique stable equilibrium state consistent with the given values of energy, numbers of particles, and constraints. It follows from relation 6.7 that the entropy of the stable equilibrium state must

be larger than that of any equilibrium or nonequilibrium state with the same values of energy, numbers of different particles, and constraints. Accordingly, a criterion for stable equilibrium is that the entropy should be at its maximum value for fixed values of energy, numbers of particles, and constraints.

6.7.2 Equations Relating Properties for Stable Equilibrium States. According to the state principle (Sec. 6.5.1), the simplest illustrations of the laws of thermodynamics are provided by applications to systems in or passing through stable equilibrium states. The properties and the relations between properties of systems in such states can be expressed by means of relatively simple mathematical forms.

The term *numbers of particles,* as used in the state principle, implies certain species of particles. The set of species selected must be a minimum set that can describe the composition of all allowed states. Such a set, which will be unique in number of members for a given system regardless of the choice of species to make up the set, is called a *set of components.*

For system A with given fixed values E of energy, $n_1, n_2, \ldots, n_n$ of the numbers of particles of components $1, 2, \ldots, n$ of the set of n components, and $\beta_1, \beta_2, \ldots, \beta_s$ of constraints $1, 2, \ldots, s$ of the set of s constraints, the uniqueness of the corresponding stable equilibrium state implies that the value of any property of A in that state must be fully and uniquely determined by the given fixed values. Moreover, for all stable equilibrium states, the values of a property F can be expressed as a function having a single value for each single set of values of $E; n_1, n_2, \ldots, n_n; \beta_1, \beta_2, \ldots, \beta_s$; namely,

$$F = F(E; n_1, n_2, \ldots, n_n; \beta_1, \beta_2, \ldots, \beta_s) \tag{6.8}$$

Relations like Eq. 6.8 for several properties can be manipulated mathematically to express the values of any property as a function of the values of any set of independent properties of stable equilibrium states.

A function of interest to subsequent considerations is that which relates the values of energy E to the values of $S; n_1, n_2, \ldots, n_n$; and $\beta_1, \beta_2, \ldots, \beta_s$; namely,

$$E = E(S; n_1, n_2, \ldots, n_n; \beta_1, \beta_2, \ldots, \beta_s) \tag{6.9}$$

6.7.3 Temperature. *The temperature* T on the Kelvin scale of a system in a stable equilibrium state is defined by the relation

$$T = \left(\frac{\partial E}{\partial S}\right)_{n,\beta} \tag{6.10}$$

where E is given by Eq. 6.9. Chapter 7 shows that, if energy can flow into either of systems A or B in mutual stable equilibrium, the temperatures T_A and T_B

must be equal. This equality is deduced from the criterion for stable equilibrium (Sec. 6.7.1) applied to the combined system AB.

6.7.4 Total Potentials. Given a system in a stable equilibrium state, the total potential μ_l of component l is defined by the relation

$$\mu_l = \left(\frac{\partial E}{\partial n_l}\right)_{S,n,\beta} \qquad \text{for } l = 1, 2, \ldots, r \tag{6.11}$$

If component l can flow into either of systems A or B in mutual stable equilibrium, we can readily show, by means of the criterion for stable equilibrium, that $(\mu_l)_A$ and $(\mu_l)_B$ must be equal.

6.7.5 Generalized Body Forces. For a given system in a stable equilibrium state, the generalized body force F_i conjugate to the constraint β_i is defined by the relation

$$F_i = \left(\frac{\partial E}{\partial \beta_i}\right)_{S,n,\beta} \qquad \text{for } i = 1, 2, \ldots, s \tag{6.12}$$

For example, if volume V is an independent constraint of the system such that

$$\beta_1 = V \tag{6.13}$$

the generalized body force F_1 conjugate to V is the negative of the pressure p:

$$F_1 = \left(\frac{\partial E}{\partial V}\right)_{S,n,\beta} = -p \tag{6.14}$$

6.7.6 Heat. We will now outline an argument to prove that certain nonwork interactions are entirely distinguishable from work in that no fraction of any one such interaction will satisfy the definition of work.

If systems A and B are in stable equilibrium states that are not in mutual stable equilibrium and each system is under constant constraints, then an interaction could occur between them. By virtue of the corollary of the second law given in Sec. 6.5.3, any such interaction must be a nonwork interaction; i.e., its effects on at least one system, say A, could not have been produced if the sole external effect was the rise of a weight. In general, however, the effects on A could be produced while weights external to A are being raised and a change of state of B takes place. It can be shown that maximum work will be delivered to the weights if the process is carried out reversibly and each system passes through stable equilibrium states. A reversible cyclic device X (Fig. 6.2) interposed between systems A and B will permit these conditions to be met. The device X may

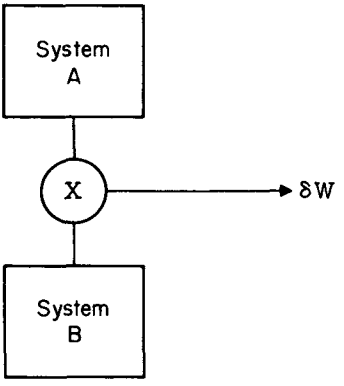


Fig. 6.2 Schematic of reversible process between systems A and B which are in stable equilibrium states but not in mutual stable equilibrium and which interact under fixed constraints via a reversible device X.

produce work, but, by virtue of the corollary of Sec. 6.5.3, the reversible interactions experienced by A with X and X with B cannot be work.

For the reversible process just described, the work δW that will be delivered to the weights is equal to the negative of the sum of the energy changes dE_A and dE_B of systems A and B. On the other hand, dE_A and dE_B satisfy the relations

$$dE_A = \left(\frac{\partial E_A}{\partial S_A}\right)_{n,\beta} dS_A = T_A \, dS_A \tag{6.15}$$

$$dE_B = \left(\frac{\partial E_B}{\partial S_B}\right)_{n,\beta} dS_B = T_B \, dS_B \tag{6.16}$$

and

$$dS_A + dS_B = 0 \tag{6.17}$$

Thus we can readily verify that δW is given by the relation

$$\delta W = -\frac{T_A - T_B}{T_A} \, dE_A \tag{6.18}$$

In the preceding relations, the symbol d denotes an infinitesimal change of the value of the property whose symbol it precedes, and δ denotes an infinitesimal amount of the quantity whose symbol it precedes when this quantity is not a

property or state function. Work is not a property of any system, and therefore δW is the appropriate way to indicate an infinitesimal amount of W.

Because dE_A is a measure of the magnitude of the interaction experienced by A with X, Eq. 6.18 expresses the work produced as a fraction of the magnitude of that interaction. As systems A and B approach mutual stable equilibrium, temperatures T_A and T_B approach equality, a necessary condition, and the fraction that represents the ratio of δW to dE_A approaches zero; i.e., the work that can be obtained from the interaction between A and B through X becomes a vanishingly small fraction of the magnitude of the interaction. Moreover, we can see from Eqs. 6.15 to 6.17 that then the interactions between A and X and between X and B are exactly equal.

In the limit, therefore, as T_A approaches T_B the cyclic device X is no longer required and the nature of the interaction approaches something completely distinguishable from work. This special kind of interaction is called *heat* and is denoted by the symbol Q. Moreover, for any process for which a system passes only through stable equilibrium states of fixed constraints and for a change of state corresponding to changes dS in entropy and dE in energy, the infinitesimal amount of heat $(\delta Q)_{\text{rev}}$ received by the system is given by the relations

$$(\delta Q)_{\text{rev}} = TdS = dE \tag{6.19}$$

If the interacting systems pass only through stable equilibrium states, the process will be reversible.

Heat is not a property of a system; thus the symbol δQ denotes an infinitesimal amount of heat and not the change in value of a property between two neighboring stable equilibrium states. Nor is heat contained within a system. Since heat is an interaction, its magnitude depends on the arbitrary selection of boundaries between systems.

The term "heat" is used here in a special sense that is quite foreign to its popular meaning. A different term might have been used were it not for the historical association of the word "heat" with thermodynamics. The word "heat" is reserved here for a very special kind of interaction between systems which, only by virtue of the second law, is entirely distinguishable from all other kinds of interactions, including the special kind called work. If a less restrictive definition had been used, some of the relations (e.g., Eq. 6.19) between heat and entropy given here would not always hold.

6.7.7 Heat and Flow of Entropy. When a heat interaction occurs between systems A and B passing through stable equilibrium states, the process is reversible. For the combined system AB, the process may be considered an adiabatic process for which the work is zero, so that, in accordance with Eqs. 6.5 and 6.7 expressed in terms of subsystems A and B,

$$dS_A + dS_B = 0$$

Because, from Eq. 6.19, dS_A is equal to $(\delta Q_A/T)_{\text{rev}}$ and different from zero, the heat interaction is associated with a flow of entropy from A to B if heat flows from A to B and, conversely, from B to A. By contrast, in a work interaction no entropy flows from A to B when work is done by A on B. In other words, although both interactions involve transfer of energy, heat is also accompanied by transfer of entropy, but work is not.

To system B we can add a part which is not in a stable equilibrium state and which may experience an irreversible process without affecting the state of the other part or the interaction between A and B. Although the process as a whole is no longer reversible, the interaction is the same as before and could appropriately be called heat. A flow of entropy from A to B is still associated with the flow of heat from A to B, but the increase of entropy in B exceeds the amount that has flowed from A; i.e., the statement

$$dS > \frac{\delta Q}{T}$$

can be written for the irreversible process and

$$dS \geqslant \frac{\delta Q}{T} \tag{6.20}$$

for processes in general.

For an adiabatic process, Eq. 6.20 reduces to Eq. 6.6. For other processes, Eq. 6.20 applies if the only nonwork interactions are heat; i.e., at least locally where the interaction occurs, the behavior on each side of the boundary is the same as if both systems were passing through stable equilibrium states with identifiable temperatures and the process was reversible. We can readily show that, when certain nonwork interactions are used in place of δQ, Eq. 6.20 will not be satisfied.

6.8 Work-Producing Systems and Devices

6.8.1 Sources of Available Work. A problem of economic and social importance in heat engineering is the production of work to be used for practical purposes such as lifting weights, driving an electric generator, propelling vehicles, or moving cutting tools. Work can be obtained from any system that is not in a stable equilibrium state as it changes toward a stable equilibrium state without changes in constraints. For example, work can be obtained by lowering a weight inside a closed system while causing, through appropriate levers, belts, and

pulleys, a weight to rise outside. The production of work need not end until the internal weight has reached the lowest position in the gravity field that the configuration of the system permits. Similarly, an electric storage battery can produce work until it is completely discharged. The maximum work that can be produced by the system adiabatically will be obtained if the process is executed reversibly, i.e., at constant entropy. By virtue of the second law, the maximum work is only a fraction of the energy of the system. This fraction is larger the farther the system is initially from stable equilibrium, namely, the larger the difference between the maximum entropy corresponding to the initial energy and the entropy of the initial nonstable state. The fraction is equal to zero when the difference in entropy just described is equal to zero, i.e., when the system is initially in a stable equilibrium state.

On the other hand, a system in any state, including a stable equilibrium state, can produce work when combined with a system in the environment which passes through stable equilibrium states but is not in mutual stable equilibrium with the first system. The maximum work that can be obtained is the maximum possible decrease in the available work when the system in the environment acts as a reservoir or the maximum possible decrease of one of the other availability functions discussed in Chap. 7 when other conditions are imposed by the environment. The work-producing process must be completely reversible to obtain this maximum. For example, if the system with the elevated weight described in the preceding paragraph is at a higher or lower temperature than the atmosphere, further work can be obtained by interposing a heat engine between the system and the atmosphere so that, as heat flows to the engine and produces work, the system is gradually cooled or heated to the temperature of the atmosphere. If the whole process is executed reversibly, the maximum possible work will be produced.

The greatest sources of work are fuels found in nature that either are not themselves in stable equilibrium states (e.g., neutrons and uranium in a reactor) or are not in mutual stable equilibrium with the atmosphere from which they have been isolated in the course of geologic history (e.g., the fossil fuels coal and petroleum). In either case the maximum work can be obtained only by a reversible approach to stable equilibrium or mutual stable equilibrium. One case requires a reversible nuclear reaction and the other, a reversible chemical reaction. A close approximation to a reversible chemical reaction is found in lead storage batteries used in automobiles for starting and for lighting. No close approximation to a reversible nuclear reaction has yet been realized in practice, but it is possible in principle.

6.8.2 **Heat Engines.** Because the cost of work includes not only fuel costs but also capital costs, which usually increase as irreversibility in the process is reduced, the economical means of producing work may in some respects depart

radically from the reversible means. For example, in fossil-fueled central power stations, fuel and oxygen in the air enter into an irreversible chemical reaction that raises the temperature of the resulting products by about 2000°K and reduces the work available by about one quarter. Energy from these hot products is then transferred irreversibly across a large temperature difference to appear as heat added to liquid and vapor water, which is the working fluid of a heat engine. The loss in available work is again about one quarter of that of the original fuel and air. Irreversibility in the heat-engine cycle accounts for another 10% loss; thus about 40% is actually realized as work delivered to the electrical distribution system.

6.9 Statistical Thermodynamics

6.9.1 Brief Review. The development of a mechanical theory of heat has been pursued concurrently with the development of thermodynamics. The purpose of such a theory is to relate heat to changes in the motion of the atoms and molecules of matter. The history of the theory can be traced back to the Greek philosophers, Democritus (~400 B.C.) and Epicurus (~300 B.C.). The French philosopher Pierre Gassendi in the seventeenth century developed a theory in which all material phenomena are attributable to the indestructible motion of atoms. Although Hooke, Leibnitz, and Daniel Bernoulli improved upon it, the theory was not firmly established until Joule demonstrated experimentally that a quantitative relation exists between heat and work when they produce identical effects.

After Joule the theory developed rapidly into a science of major importance which became known as statistical mechanics or statistical thermodynamics. Its subject is the relation between the results of thermodynamics in the limit of stable equilibrium and the details of the structure of matter. Its development, which can be traced through Helmholtz, Clausius, Maxwell, and Boltzmann, culminated in the work of Gibbs, who in 1901 presented an exposition of classical statistical thermodynamics that excels in completeness, rigor, and generality. Subsequently, von Neumann, Landau, and others adapted Gibbs's ideas to quantum statistical thermodynamics.

Perhaps because Gibbs's contribution was not fully understood, a less general and less rigorous molecular statistics prevailed in the literature, with few exceptions, until after World War II. Molecular statistics applies only to dilute systems consisting of a great number of identical constituents (such as gas molecules, electrons, or Planck oscillators) and in which the interparticle forces are so weak that a "private energy" can be identified with each constituent. On the other hand, it does not apply to dense systems, such as liquids and solids, for which the interactions between neighboring atoms are so strong that the energy of one atom cannot be affected without affecting the energy of all other atoms;

namely, the total energy of the system cannot be divided into private energies of its atoms.

Gibbs's exposition of statistical thermodynamics applies to all systems (gaseous, liquid, or solid) having many or few degrees of freedom. However, since it is based on classical mechanics, it has two shortcomings. First, it is not applicable to systems in which quantum effects are important. Second, its statistics express a subjective lack of knowledge about the precise state of a system rather than some inherent property of the system. These shortcomings can be eliminated by the description of the state in terms of quantum-mechanical probabilities given in Sec. 6.10 as quantum thermodynamics. Some key concepts from Gibbsian statistics are discussed here.

6.9.2 Some Concepts from Gibbsian Statistics. According to classical mechanics, the state of a system having N degrees of freedom (i.e., N independent ways in which the space configuration of the system may change) is fully specified by the values of N position coordinates $q_1, q_2, \ldots, q_N$ and N momentum coordinates $p_1, p_2, \ldots, p_N$. For example, the state of a system consisting of n point particles (particles without internal structure and, therefore, without internal degrees of freedom) is specified if the position coordinates along three Cartesian axes x, y, and z and the corresponding momentum coordinates along these same axes for each of the particles are specified. For such a system, the number N of the degrees of freedom is equal to three times the number n of the particles of the system.

The state of a system having N degrees of freedom can be represented geometrically in classical mechanics by a point in $2N$-dimensional phase space having as coordinates $q_1, q_2, \ldots, q_N$ and $p_1, p_2, \ldots, p_N$.

According to Gibbs, the term *thermodynamic state of a system with N degrees of freedom* does not necessarily correspond to a unique point in phase space but rather to a set of probabilities of the system being in any one of the allowed points consistent with the constraints imposed on the system. Because any set of allowed states in classical mechanics forms a continuum, Gibbs introduced a *probability density* or *distribution P*, which he called *coefficient of probability*, such that the product $P dq_1\ dq_2 \ldots dq_N\ dp_1\ dp_2 \ldots dp_N$ is equal to the probability of the system being in states having values of coordinates between q_1 and $q_1 + dq_1$, q_2 and $q_2 + dq_2, \ldots$, and q_N and $q_N + dq_N$ and values of momenta between p_1 and $p_1 + dp_1$, p_2 and $p_2 + dp_2, \ldots$, and p_N and $p_N + dp_N$. For a given thermodynamic state, the distribution P will be a corresponding function of q's and p's. Moreover, because a system must be at some one of the allowed states, the sum of all the probabilities must be equal to unity, or, equivalently, the integral of the distribution P over all q's and p's must be always equal to unity.

Various types of thermodynamic states correspond to particular mathematical forms of the distribution P. For example, Gibbs assumed without proof that for the stable equilibrium states of a system with N degrees of freedom the coefficient of probability P corresponding to the state $q_1, q_2, \ldots, q_N, p_1, p_2, \ldots, p_N$ of energy ϵ is given by the exponential relation

$$P = \exp\left(\frac{\psi - \epsilon}{\theta}\right) \tag{6.21}$$

where ψ and θ are constants. He called this distribution of probabilities a *canonical distribution.* Gibbsian statistics are not discussed further because the quantum thermodynamics presented here is more general and, in addition, relates more directly to the concepts discussed in earlier sections.

6.10 Quantum Thermodynamics

Quantum thermodynamics is a theory deduced logically from both the laws of quantum mechanics and the laws of thermodynamics. In the theory the relations between states or changes of states and detailed microscopic structures of systems are derived without the help of analogies between statistical and thermodynamic results (such as the analogy between a statistical expression for entropy and the entropy of equilibrium states of classical thermodynamics), and without hypotheses in addition to the postulates (such as the ergodic hypothesis).

The formulations of both quantum mechanics and quantum thermodynamics recognize explicitly the premise that the state must be described in terms of probabilities. Quantum mechanics deals only with states that can be described by means of wave functions, however, and quantum thermodynamics deals both with states that can be described by wave functions and with states that cannot. All these points are clarified in the sections that follow.

6.10.1 Operator and Matrix Representation of Properties. In contrast to classical mechanics, a fundamental premise of quantum mechanics is that at a given time the most that can be said about results of measurements on a system is the probability of finding particular values of properties, such as coordinates and momenta, rather than which particular values will be observed. This premise, which reflects the principle of indeterminacy first introduced by Heisenberg, requires that the state of a system be described in terms of probabilities.

The principles of quantum mechanics can be stated by a number of mathematical formulations. The best known among these are the wave formulation of Schroedinger, which is discussed in Chap. 5, the vector formulation of Dirac, and the matrix formulation of Heisenberg. We use the

matrix formulation in this discussion since it is especially suited to the needs of quantum thermodynamics.

For a given system with unambiguous values of numbers of particles and constraints and with N degrees of freedom (an ordinary closed system with N degrees of freedom), any property that in classical mechanics can be expressed as a function of position and momentum coordinates is expressed in quantum mechanics by a linear Hermitian operator (Chap. 5). Each operator, in turn, can be represented by a matrix and a complete set of orthonormal functions. For example, the Hamiltonian function of classical mechanics, which is the energy of the system, is represented by the Hamiltonian operator $\hat{H}$ or by a Hamiltonian matrix $[H]$ and a set of orthonormal functions. Operators or matrices are used also for properties such as spin, which have no classical analog.

The method for calculating elements of matrices follows. The matrix $[G]$ for the operator $\hat{G}$ is square and has elements G_{mn} given by the relation

$$G_{mn} = \int u_m^* \hat{G} u_n \, d^N q \qquad \text{for all } m \text{ and } n \tag{6.22}$$

where $\{u_n\}$ is an appropriate complete set of orthonormal functions. Each function u_m for all m depends on N independent variables $q_1, q_2, \ldots, q_N$, and the integration in Eq. 6.22 is carried over all these variables; i.e., it involves N integrals. When the set $\{u_n\}$ is used to calculate the matrix elements, we say that $[G]$ is in the *u-representation.*

Each matrix possesses *eigenvalues,* i.e., a set of numbers that can be found by solving the characteristic equation of the matrix. The eigenvalues of the matrix of a property are the eigenvalues of the operator corresponding to that property. They are real in the mathematical sense as distinguished from complex. Each one corresponds to a possible outcome of a measurement of the property performed on the system. For example, a measurement of the energy of a system in a given condition will yield any one of the real eigenvalues of the Hamiltonian matrix of the system. The number of eigenvalues of a property may be either finite (as with spin) or infinite (as with energy for most systems).

6.10.2 State in the Language of Quantum Mechanics. According to quantum mechanics it is impossible to predict from knowledge of the condition of a system which eigenvalue a measurement will yield. Instead the most that can be said at any instant of time is the probability that a measurement will yield a given eigenvalue. For the purposes of this book, the thermodynamic state or simply the state of a system will be expressed by the set of all probabilities for the outcomes of all measurements that may be performed at a given instant of time. That is, the state of a system is represented by Table 6.1, in which $\overline{F}$, $\overline{G}$, $\overline{L}$, . . . denote observable properties; $W(F_1)$ denotes the probability that any one

Table 6.1 Probabilities for Outcome of Measurement of Properties $\overline{F}$, $\overline{G}$, and $\overline{L}$ of a System

Property	$\overline{F}$	$\overline{G}$	$\overline{L}$	
Probabilities	$W(F_1)$	$W(G_1)$	$W(L_1)$	...
	$W(F_2)$	$W(G_2)$	$W(L_2)$	...
	.	.	.	
	.	.	.	
	.	.	.	
	$W(F_m)$	$W(G_m)$	$W(L_m)$	...
	.	.	.	
	.	.	.	
	.	.	.	
Sum of probabilities	1	1	1	

measurement of property $\overline{F}$ will yield the eigenvalue $F_1, \ldots$; $W(G_1)$ denotes the probability that any one measurement of property $\overline{G}$ will yield the eigenvalue $G_1, \ldots$; etc.

It can be shown that these probabilities can be combined into a linear Hermitian operator $\hat{\rho}$ or an equivalent matrix $[\rho]$ in some representation. In other words, in the matrix formulation of quantum mechanics the state is represented by a linear Hermitian operator $\hat{\rho}$ (matrix $[\rho]$), called the *density operator* (matrix). Conversely, $\hat{\rho}$ can be used to compute the probability about the result of any measurement. For example, the probability $W(Fm)$ that a measurement of property $\overline{F}$ corresponding to matrix $[F]$ will yield the eigenvalue F_m of $[F]$ is given by the relation

$$W(F_m) = \int u_m^* \hat{\rho} u_m \, d^N q \tag{6.23}$$

where u_m is the normalized eigenfunction corresponding to the eigenvalue F_m, namely,

$$\hat{F} u_m = F_m u_m \tag{6.24}$$

In words, $W(F_m)$ is given by the mth diagonal element of the density matrix $[\rho]$ in the representation determined by the eigenfunctions of operator $\hat{F}$. Again, the probability $W(G_n)$ that a measurement of property $\overline{G}$ corresponding to matrix $[G]$ will yield the eigenvalue G_n of $[G]$ is given by the relation

$$W(G_n) = \int v_n^* \hat{\rho} v_n \, d^N q \tag{6.25}$$

where v_n is the normalized eigenfunction corresponding to the eigenvalue G_n, namely,

$$\hat{G} v_n = G_n v_n \tag{6.26}$$

In words, $W(G_n)$ is given by the nth diagonal element of the density matrix $[\rho]$ in the representation determined by the eigenfunctions of the operator $\hat{G}$.

6.10.3 Values of Properties. Because measurements of a property $\overline{F}$ of a system in a given state yield different eigenvalues F_m each with a probability $W(F_m)$ specified by the density matrix $[\rho]$ that represents the state, the value F of the property corresponding to the given state is defined as the weighted average of the various possible eigenvalues, each eigenvalue being weighted by the probability that it will be observed. In other words, F is defined as an expectation value. In symbols, the value F of property $\overline{F}$ is given by the relations

$$F = \sum_m W(F_m) F_m \tag{6.27}$$

and

$$1 = \sum_m W(F_m) \tag{6.28}$$

where each eigenvalue is represented by a different symbol, even when two or more eigenvalues have the same value, and a discrete spectrum of eigenvalues is assumed for convenience.

It can be easily verified that F is also given by the sum of the diagonal elements of the matrix that is the product of the matrices $[\rho]$ and $[F]$ expressed in the same representation, i.e., by the trace of the product of $[\rho]$ and $[F]$:

$$F = \sum_m \sum_k \rho_{mk} F_{km} = \mathrm{Tr}\left[[\rho]\,[F]\right] \tag{6.29}$$

where Tr denotes the trace of the matrix that follows.

Through use of a single representation, expressions analogous to Eq. 6.29 can be written for the values $G, L, \ldots$ of properties corresponding to matrices $[G], [L], \ldots$. Specifically,

$$G = \sum_m \sum_k \rho_{mk} G_{km} = \mathrm{Tr}\left[[\rho]\,[\mathrm{G}]\right] \tag{6.30}$$

$$L = \sum_m \sum_k \rho_{mk} L_{km} = \mathrm{Tr}\left[[\rho]\,[\mathrm{L}]\right] \tag{6.31}$$

We see from Eqs. 6.29 to 6.31 that, if the state $[\rho]$ is given, the values of all properties and all probabilities about results of measurements can be computed. Conversely, for given values $F, G, L, \ldots$ of a complete set of independent properties of a system, the corresponding complete set of linear equations (6.29 to 6.31) can be solved for the elements of $[\rho]$ and, therefore, for the state.

6.10.4 State in Quantum Mechanics and in Quantum Thermodynamics. Like other matrices representing linear Hermitian operators, a state $[\rho]$ has real eigenvalues. When $y_1, y_2, \ldots, y_m, \ldots$ denote the eigenvalues of $[\rho]$, it can be shown that each y is positive, smaller than unity, and is such that

$$\sum_m y_m = 1 \tag{6.32}$$

Different combinations of eigenvalues $y_1, y_2, \ldots, y_m \ldots$ correspond to different types of states (Sec. 6.4). One combination consists of all the eigenvalues being equal to zero, except for one that is equal to unity. It is called a *pure state.* A pure state has at least one property with a value identically equal to an eigenvalue of the matrix corresponding to the property and can be fully and equivalently represented by means of a wave function (Chap. 5). As such it describes the state postulated in conventional quantum mechanics.

Any other combination of eigenvalues $y_1, y_2, \ldots, y_m, \ldots$ consists of more than one eigenvalue being different from zero. It is called a *mixed* or *nonpure state.* A nonpure state cannot be expressed by means of a wave function or a linear superposition of wave functions. As such, "mixed" or "nonpure" describes states other than those postulated in conventional quantum mechanics.

It will be seen below that the entropy of a pure state is equal to zero whereas the entropy of a nonpure state is different from zero. Since, in general, the entropy of a system is not equal to zero, the state of the system cannot be pure (cannot be described by a wave function) and, therefore, the density matrix representation is more general than the wave function representation described in Chap. 5. The distinction between the two types of states is important because, although all the energy of a system in a pure state can be converted to work without other effects, only a fraction of the energy of a system in a nonpure state can be converted to work without other effects.

6.10.5 Reversible Equation of Motion. Using the equation of motion, which is provided by quantum mechanics, we can find the rate of change of the density matrix $[\rho]$ with time for a pure state and for each of the many processes that are reversible and adiabatic. The equation of motion is the Schroedinger equation for change in state written in matrix form. An identical form for the

equation of motion can be postulated for nonpure states and for some, but not all, reversible adiabatic processes. It is given by the matrix relation:

$$\frac{\partial[\rho]}{\partial t} = \frac{i}{\hbar}\{[\rho][H] - [H][\rho]\} \tag{6.33}$$

where $[H]$ is the Hamiltonian matrix in the same representation as that used for the state $[\rho]$. In other words, Eq. 6.33 applies for both pure and nonpure states. For pure states it is the matrix equivalent of the Schroedinger equation.

Conditions for equilibrium are found from the equation of motion. If the eigenvalues of the density matrix $[\rho]$ for an equilibrium state are denoted by x_1, $x_2, \ldots, x_m, \ldots$, and the eigenvalues of the Hamiltonian matrix $[H]$ by E_1, $E_2, \ldots, E_m, \ldots$, it can be shown that, in the absence of irreversibilities, a necessary and sufficient condition for equilibrium is that the first eigenvalue x_1 be equal to the probability $W(E_1)$ that a measurement of the energy will yield the first eigenvalue E_1, that the second eigenvalue x_2 be equal to $W(E_2)$, etc. An equivalent way of specifying this is to say that an equilibrium state is described by a diagonal matrix with diagonal elements $x_1, x_2, \ldots, x_m, \ldots$ in the representation defined by the eigenfunctions of the Hamiltonian operator $\hat{H}$, the so-called energy representation.

For equilibrium states the expectation value of the energy E can be expressed as a sum over terms each of which is the product of x_m and E_m, namely,

$$E = \sum_m x_m E_m \tag{6.34}$$

6.10.6 General Expression for Entropy. According to the first and second laws of thermodynamics, entropy is a property or state function of a system. Moreover, it must be invariant in any reversible adiabatic process. From these two conditions and the requirement that it is an extensive property, the value S of entropy can be shown to be proportional to the sum of the diagonal elements (the trace) of a matrix that is the product of the matrices $[\rho]$ and $\ln[\rho]$. When $y_1, y_2, \ldots, y_m, \ldots$ denote the eigenvalues of the density matrix $[\rho]$ (corresponding to an arbitrary state, not necessarily equilibrium or stable equilibrium), the value of S is given by the relation

$$S = -k\sum_m y_m \ln y_m \tag{6.35}$$

where k is a universal constant that proves to be the Boltzmann constant.

We see from Eq. 6.35 that the entropy of a pure state is equal to zero since for such a state all eigenvalues of $[\rho]$ are equal to zero except one, which is equal to unity. Because of its zero entropy, all the energy of a pure state is available for conversion to work without any other effects.

6.10.7 Stable Equilibrium States. In the language of quantum thermodynamics, the criterion for stable equilibrium (Sec. 6.7.1) implies that the density matrix for a stable equilibrium state must be such that the entropy (Eq. 6.35) is at its maximum for given (expectation) values of energy, numbers of particles, and constraints. When this criterion is applied to a system whose numbers of particles and constraints are known without uncertainty (the probability of the value for the number of particles of a given kind which is found by a measurement is unity and the probability of the value for a given constraint which is found by a measurement is unity), it yields the eigenvalues $x_1^0, x_2^0, \ldots, x_m^0 \ldots$ for the stable equilibrium states as functions of the energy eigenvalues $E_1, E_2, \ldots, E_m, \ldots$ of the Hamiltonian matrix of the system and the temperature T of the stable equilibrium state in question. The various mathematical functions are

$$x_m^0 = \frac{\exp(-E_m/kT)}{\sum_m \exp(-E_m/kT)} = \frac{\exp(-E_m/kT)}{Q_c} \quad \text{for } m = 1, 2, \ldots \tag{6.36}$$

$$Q_c = \sum_m \exp\left(-\frac{E_m}{kT}\right) \tag{6.37}$$

$$T \equiv \left(\frac{\partial E}{\partial S}\right)_\beta \tag{6.38}$$

$$E = \sum_m x_m^0 E_m = \frac{\sum_m E_m \exp(-E_m/kT)}{\sum_m \exp(-E_m/kT)} \tag{6.39}$$

and

$$S = -k \sum_m x_m^0 \ln x_m^0 = k \ln\left[\sum_m \exp\left(-\frac{E_m}{kT}\right)\right] + \frac{E}{T} \tag{6.40}$$

The values x_m^0 are called the *canonical distribution,* and the quantity Q_c (Eq. 6.37) is called the *canonical partition function.*

Because the eigenvalues $x_1, x_2, \ldots, x_m, \ldots$ of the density matrix for any equilibrium state and, therefore, the eigenvalues $x_1^0, x_2^0, \ldots, x_m^0, \ldots$ of the density matrix for any stable equilibrium state represent probabilities that measurements of energy will yield eigenvalues $E_1, E_2, \ldots, E_m, \ldots$, respectively, Eq. 6.36 can be thought of as analogous to the canonical distribution postulated by Gibbs, provided that the two constants θ and ψ in Eq. 6.21 are expressed in terms of the temperature T and the energy eigenvalues $E_1, E_2, \ldots, E_m, \ldots$. It can be shown that θ and ψ are given by the relations

$$\theta = kT \tag{6.41}$$

and

$$\psi = kT \ln \left[\sum_m \exp\left(-\frac{E_m}{kT}\right) \right] \tag{6.42}$$

We can readily verify from Eqs. 6.40 and 6.42 that the function ψ is equal to $E - TS$, namely, the Helmholtz free energy of the system. Moreover, because the energy eigenvalues $E_1, E_2, \ldots, E_m, \ldots$ are, in general, functions of the values of the constraints, the Helmholtz free energy ψ is a function of temperature and the constraints and, therefore, as discussed in Chap. 7, is a characteristic function. This means that the problem of evaluating the properties of any ordinary closed system in a stable equilibrium state reduces to the problem of evaluating the energy eigenvalues as functions of the constraints. The evaluation of the energy eigenvalues is a difficult mathematical task that has been carried out for only a very few systems. The difficulty is numerical rather than conceptual, however.

6.11 Generalizations to All Systems

Up to this point we have limited the discussion of quantum thermodynamics to ordinary closed systems, namely, systems with unambiguous (dispersion-free) numbers of particles and constraints.

In many systems the numbers of particles are uncertain, either because the system is open to the transfer of rest mass or because particles can be created and annihilated within the system. The terminology of Gibbs suggests that such systems be called *grand systems.* When the system is simply open to rest mass, it is also called an ordinary open system.

By definition, uncertainty in the numbers of particles means that measurements of these numbers for a system in a given state yield different particle-number eigenvalues. This uncertainty is analogous to that associated in general with the energy. It follows that an expectation value rather than a dispersion-free value (i.e., rather than a unique value observed as a result of every measurement) is assigned to the number of particles of each component of a state of a grand system.

In other systems encountered in quantum field theory, the constraints are uncertain because they are quantized.

Quantum thermodynamics can be readily extended both to grand systems and to systems with uncertain numbers of particles and constraints. The extensions require the use of supermatrices instead of matrices and are not given here except for some results applicable to stable equilibrium states of grand systems only.

A stable equilibrium state of a grand system is fully described by the set of eigenvalues $x_1^0, x_2^0, \ldots, x_m^0, \ldots$ of the state matrix that maximizes the entropy of the system at given expectation values of energy and numbers of particles of the components, and at given values of the constraints. The mth eigenvalue x_m^0 is equal to the probability that a suitable measurement will yield the mth energy eigenvalue E_m, and the mth number-of-particle eigenvalues $n_{1m}, n_{2m}, \ldots, n_{nm}$ for the set of n components of the system. It can be expressed as a function of the energy eigenvalues $E_1, E_2, \ldots, E_m, \ldots$ of the system, the eigenvalues $n_{11}, n_{12}, \ldots, n_{1m}, \ldots$ of the number of particles of component 1, the eigenvalues $n_{21}, n_{22}, \ldots, n_{2m}, \ldots$ of the number of particles of component 2, ..., the total potentials $\mu_1, \mu_2, \ldots, \mu_n$ of the n components $1, 2, \ldots, n$, and the temperature T of the stable equilibrium state in question and is given by the relation

$$x_m^0 = \frac{\exp\left[(n_{1m}\mu_1 + n_{2m}\mu_2 + \ldots + n_{nm}\mu_n - E_m)/kT\right]}{\sum_m \exp\left[(n_{1m}\mu_1 + n_{2m}\mu_2 + \ldots + n_{nm}\mu_n - E_m)/kT\right]} \tag{6.43}$$

The set of values x_m^0 is called the *grand canonical distribution.* The sum in the denominator of Eq. 6.43 is called the *grand partition function* Q_g; i.e.,

$$Q_g = \sum_m \exp\left(\frac{n_{1m}\mu_1 + n_{2m}\mu_2 + \ldots + n_{nm}\mu_n - E_m}{kT}\right) \tag{6.44}$$

It appears in many relations between properties for grand systems in stable equilibrium states. For example, for a grand system with volume V as the only independent constraint, the equation of state is given by the relation

$$\frac{pV}{kT} = \ln Q_g \tag{6.45}$$

Equation 6.45 applies to grand systems in gaseous, liquid, or solid phases.

7

Thermodynamics of Stable Equilibrium States

7.1 General Remarks

The concepts and postulates presented in Chaps. 5 and 6 are sufficient for analyses of nonrelativistic physical phenomena. Mathematically, the simplest illustrations of these concepts are provided by applications to systems passing through stable equilibrium states. By virtue of the second law, the properties and the relations between properties of systems in such states can be expressed by means of relatively simple mathematical forms.

In this chapter the properties of systems passing through stable equilibrium states are discussed, and some of the concepts defined in Chaps. 5 and 6 are presented in greater detail. Ordinary closed systems are considered first, then ordinary open systems.

7.2 Ordinary Closed Systems

7.2.1 Equations of Stable Equilibrium States. A system is called an ordinary closed system if over a period of time its interactions with the environment do not involve exchange of matter. For example, a fixed amount of gas in a container with an adjustable piston is an ordinary closed system. The qualifier "ordinary" is used to denote that relativistic effects and quantization of constraints are not considered.

Since the amount of matter is a fixed parameter, it follows that, for an ordinary closed system A, all equations of stable equilibrium states (Eq. 6.8) need not be written as explicit functions of the numbers of particles of A. For example, the value E of the energy of A passing through stable equilibrium states can be regarded as a function of the values S of the entropy and $\beta_1,\beta_2,\ldots,\beta_s$ of the s constraints only. Thus

$$E = E(S;\beta_1,\beta_2,\ldots,\beta_s) \tag{7.1}$$

Of course, the exact form of any particular equation of stable equilibrium states depends on the amount of matter involved and the allowed states of the system.

The relative functional simplicity of relations of the form of Eq. 7.1 is the major reason for our advanced theoretical understanding of systems passing through stable equilibrium states. In fact, the advantages of such relations can be extended to systems passing through *quasistable* states as discussed in Sec. 7.2.2.

7.2.2 **Quasistable States.** If an ordinary closed system, A, with fixed values of constraints is in a nonstable state A_1, equilibrium or nonequilibrium, it may be possible to specify new values of constraints so that state A_1 corresponds to a stable equilibrium state B_1 of the system subjected to the new values of the constraints. When this can be done, state A_1 is called a *quasistable state* and state B_1 the *corresponding stable equilibrium state.*

The following example illustrates the meaning of quasistable and corresponding stable equilibrium states. If an amount of gas fills half the volume of a container A, as shown schematically in Fig. 7.1a, then it is in a nonequilibrium state A_1, because as time goes on the gas fills the entire container. On the other

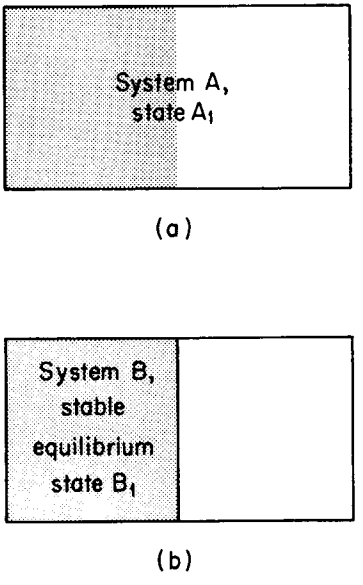


Fig. 7.1 Schematic of nonequilibrium state A_1 (a) corresponding to stable equilibrium state B_1 (b).

hand, it may happen that, with respect to half the volume of the container (system B in Fig. 7.1b), the gas is in stable equilibrium state B_1. In other words, if a partition were introduced at the midplane of container A, without any effects on the gas whatsoever, A_1 would be identical with stable equilibrium state B_1 of B. When this is possible, we say that A_1 is a quasistable state and B_1 is the corresponding stable equilibrium state, because A_1 can be made to coincide with B_1 by a mere change of the volume of A.

The relations between properties of systems passing through quasistable states are the same as those that can be written between properties of the corresponding stable equilibrium states. For example, for a given system A passing through quasistable states, the energy is given by a relation of the form of Eq. 7.1 except that the values of the constraints are those associated with the corresponding stable equilibrium states and not those of A.

7.2.3 A Condition for Mutual Stable Equilibrium. If system C consists of two ordinary closed systems, A and B, in mutual stable equilibrium (Fig. 7.2), then each of systems A, B, and C is in a stable equilibrium state, and, though interactions between A and B without changes of values of constraints are not prohibited, A does not affect the state of B and vice versa.

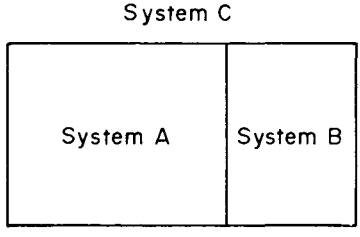


Fig. 7.2 Schematic of a system C in a stable equilibrium state consisting of two systems, A and B, in mutual stable equilibrium.

According to the criterion for stable equilibrium (Sec. 6.7.1), the entropy of C is at a maximum for the value of its energy. One of the consequences of this requirement will be seen from a possible variation in which the energy of A is increased by a small amount (dE_A), that of B is decreased by the same amount, and both A and B pass through stable equilibrium states without changes of values of their constraints.

For A and B passing through stable equilibrium states, the energies E_A and E_B and the entropies S_A and S_B satisfy relations of the forms

$$S_A = S_A(E_A) \tag{7.2}$$

$$S_B = S_B(E_B) \tag{7.3}$$

where the dependence of the entropy of each system on the values of the appropriate constraints is omitted since here these values are kept fixed. The energy E_C and entropy S_C of C are given by the relations

$$E_C = E_A + E_B \tag{7.4}$$

and

$$S_C = S_A + S_B \tag{7.5}$$

Since S_C is at a maximum for the value E_C, the variation of state described is such that

$$dE_C = dE_A + dE_B = 0 \tag{7.6}$$

and

$$dS_C = dS_A + dS_B \leqslant 0 \tag{7.7}$$

By virtue of Eqs. 7.2 and 7.3, the entropy changes dS_A and dS_B are given by the relations

$$dS_A = \left(\frac{\partial S_A}{\partial E_A}\right)_\beta dE_A \qquad dS_B = \left(\frac{\partial S_B}{\partial E_B}\right)_\beta dE_B \tag{7.8}$$

where the subscript β indicates that the values of all constraints are kept fixed while differentiating with respect to energy and the functions in Eqs. 7.2 and 7.3 are assumed to admit continuous derivatives. Substituting Eqs. 7.6 and 7.8 into relation 7.7 yields

$$\left[\left(\frac{\partial S_A}{\partial E_A}\right)_\beta - \left(\frac{\partial S_B}{\partial E_B}\right)_\beta\right] dE_A \leqslant 0 \tag{7.9}$$

If energy can be transferred either from A to B or from B to A (i.e., if dE_A can be either negative or positive), relation 7.9 is satisfied if and only if

$$\left(\frac{\partial S_A}{\partial E_A}\right)_\beta = \left(\frac{\partial S_B}{\partial E_B}\right)_\beta \tag{7.10}$$

Equation 7.10 is a necessary condition for C to be in a stable equilibrium state, namely, for A and B to be in mutual stable equilibrium. This condition can be stated in terms of an equality of temperatures, as discussed in Sec. 7.2.4.

The quantity $(\partial S/\partial E)_\beta$ of a system can be thought of as a potential that governs the tendency of energy to pass from one system to another. It can be shown to be a "capturing tendency" in that if

$$\left(\frac{\partial S_A}{\partial E_A}\right)_\beta > \left(\frac{\partial S_B}{\partial E_B}\right)_\beta \tag{7.11}$$

then dE_A may be greater than zero (energy may pass from system B to system A) in an actual process but not less than zero (because its being less than zero would be contrary to the principle of increase of entropy). Conversely, the reciprocal quantity $(\partial E/\partial S)_\beta$ can be thought of as an "escaping tendency."

7.2.4 Thermodynamic Temperature Scales. The most general definition of the temperature of a system in a stable equilibrium state is given by the relation

$$T = f\left[\left(\frac{\partial S}{\partial E}\right)_\beta\right] \tag{7.12}$$

where f denotes any arbitrarily selected monotonic function of the quantity $(\partial S/\partial E)_\beta$. It is clear that, since T is defined in terms of properties of a stable equilibrium state, it is itself a property of a stable equilibrium state. In contrast to empirical definitions of temperature scales, all temperature scales corresponding to the definition represented by Eq. 7.12 are called *thermodynamic temperature scales.*

From Eqs. 7.10 and 7.12, it follows that the necessary condition for two ordinary closed systems A and B to be in mutual stable equilibrium is equivalent to requiring that the temperatures T_A and T_B be equal or that A and B be in *temperature equality*.

7.2.5 The Kelvin Temperature Scale. The most widely used thermodynamic temperature scale is the *Kelvin scale* defined in Sec. 6.7.3. For ordinary closed systems, T in the Kelvin scale is given by the relation

$$T \equiv \left(\frac{\partial S}{\partial E}\right)_\beta^{-1} \equiv \left(\frac{\partial E}{\partial S}\right)_\beta \tag{7.13}$$

In this scale, T is measured in degrees Kelvin (°K).

The temperature scale is completely defined once the numerical value of the constant c_R in Eq. 6.4 is specified. Because the work produced by the combination of systems A, X, and R (Fig. 6.1) used in the definition of entropy change (Sec. 6.6.2) is the decrease in energy of the combination AR and because for a reversible process the entropy of that combination is conserved, we can readily show that the constant c_R is equal to the partial derivative $(\partial S_R/\partial E_R)_\beta$ of the entropy of the reservoir R or the inverse of the temperature T_R of the reservoir in degrees Kelvin. By international agreement, the value of T for solid,

liquid, and vapor water in mutual stable equilibrium is 273.16°K. One of the reasons for selecting this five-digit figure is that it causes the temperature interval between the freezing and boiling points of water under a pressure of 1 atm to be almost exactly 100 degrees.

When the number 273.15 is subtracted from the value of any Kelvin temperature (T), the result is the Celsius temperature t, or t (degrees Celsius) $\equiv T - 273.15$.

7.2.6 Analytical Descriptions of Work and Heat Interactions. As a result of interactions, a system can be passing through stable equilibrium states or quasistable states. In either of these two possibilities, the energy of A is given by an expression of the form

$$E = E(S; \beta_1, \beta_2, \ldots, \beta_s) \tag{7.14}$$

where $\beta_1, \beta_2, \ldots, \beta_s$ are the values of the constraints of each stable equilibrium or corresponding stable equilibrium state.

An energy change DE of A may come about because of interactions that involve both variations in the values of the constraints of A and other phenomena. This energy change can be expressed as a Taylor series expansion of E (Eq. 7.14) around the initial value of E. Thus

$$DE = \left(\frac{\partial E}{\partial S}\right)_{\beta} DS + \left(\frac{\partial E}{\partial \beta_1}\right)_{S,\beta} D\beta_1 + \left(\frac{\partial E}{\partial \beta_2}\right)_{S,\beta} D\beta_2 \ldots + \text{higher order terms} \tag{7.15}$$

where all partial derivatives are evaluated at the initial stable equilibrium state and subscripts denote the values that must be kept fixed during differentiation (except the one appearing in the differential). Using the Kelvin temperature scale (Eq. 7.13), we can replace the coefficient $(\partial E/\partial S)_{\beta}$ with the temperature T of the initial stable equilibrium state.

In terms of differentials Eq. 7.15 reduces to

$$dE = T\,dS + \sum_{i}^{s} \left(\frac{\partial E}{\partial \beta_i}\right)_{S,\beta} d\beta_i \tag{7.16}$$

For each such differential the values of T and $(\partial E/\partial \beta_i)_{S,\beta}$ are computed at the appropriate initial stable equilibrium state.

If the interactions resulting in the differential energy exchange dE are reversible, then at least three cases can be distinguished. In the first, A experiences a reversible adiabatic interaction by virtue of changes of constraints (namely, a reversible work-only interaction). For such an interaction the entropy is invariant,

$(dS)_{\text{rev,ad}} = 0$

and therefore Eq. 7.16 becomes

$$dE = \sum_{i}^{s} \left(\frac{\partial E}{\partial \beta_i}\right)_{S,\beta} d\beta_i \qquad \text{for a reversible, adiabatic interaction} \tag{7.17}$$

In addition, the work done $(\delta W)_{\text{rev,ad}}$ is equal to the negative of dE, and therefore

$$(\delta W)_{\text{rev,ad}} = -dE = -\sum_{i}^{s} \left(\frac{\partial E}{\partial \beta_i}\right)_{S,\beta} d\beta_i \tag{7.18}$$

The symbol δ is used in front of W because work done is not a property of a system, and, therefore, δW is not a perfect differential.

In the second case, A experiences a heat interaction; in other words, the energy change dE is brought about while A passes through stable equilibrium states and

$$d\beta_i = 0 \qquad \text{for all } i \tag{7.19}$$

The interaction is reversible and involves transfer of both energy and entropy. If the energy transferred to A is denoted by $(\delta Q)_{\text{rev}}$, it follows from Eqs. 7.16 and 7.19 that

$$(\delta Q)_{\text{rev}} = dE = T\,dS \tag{7.20}$$

The notation δQ is used because Q is not a property of a system. If dE is positive, it is said that A received an amount of heat $(\delta Q)_{\text{rev}}$, and, if dE is negative, it is said that A delivered an amount of heat $(\delta Q)_{\text{rev}}$.

Finally, in the third case A experiences a reversible interaction involving changes of both constraints and entropy while passing through stable equilibrium states. The differential energy dE is given by Eq. 7.16. It is brought about by reversible work and heat interactions. The work is given by Eq. 7.18, and the heat is given by Eq. 7.20.

As stated in Chap. 6, work and heat are modes of energy transfer across a boundary of a system and not entities stored in the bulk of a system. Energy can be stored and recovered, but work and heat cannot. Work and heat are not properties of a system.

Given two interacting systems A and B, it follows from the preceding discussion that, if throughout the period of interactions A passes through stable equilibrium states without changes of constraints, then A exchanges heat with B.

Accordingly it follows from the definition of reservoirs (Sec. 6.5.4) that reservoirs deliver or receive heat during interactions.

7.2.7 Reversible Work and Heat in the Language of Quantum Mechanics. Expressions for $(\delta W)_{\text{rev}}$ and $(\delta Q)_{\text{rev}}$ can also be found in terms of the quantum-mechanical concepts of Chap. 6. For a stable equilibrium state, the energy E of a system is given by the relation (Sec. 6.10.7)

$$E = \sum_m x_m^0 E_m \tag{7.21}$$

where x_m^0 is the canonical distribution and E_m the mth energy eigenvalue corresponding to the stable equilibrium state under consideration. For a neighboring stable equilibrium state, the energy is $E + dE$, the canonical distribution is $x_m^0 + dx_m$, and the mth energy eigenvalue is $E_m + dE_m$. It follows that the differential energy dE is given by the relation

$$dE = \sum x_m^0 \, dE_m + \sum_m E_m \, dx_m \tag{7.22}$$

Changes of eigenvalues come about because of changes in the values β_1, $\beta_2, \ldots, \beta_s$ of the constraints. The differential change dE_m of the mth eigenvalue is given by the relation

$$dE_m = \sum_{i=1}^{s} \left(\frac{\partial E_m}{\partial \beta_i}\right)_\beta d\beta_i \tag{7.23}$$

and therefore

$$\sum_m x_m^0 \, dE_m = \sum_m x_m^0 \sum_{i=1}^{s} \left(\frac{\partial E_m}{\partial \beta_i}\right)_\beta d\beta_i \tag{7.24}$$

The order of the summations in the right-hand side of Eq. 7.24 can be interchanged. Moreover, x_m^0 and the summation over the energy eigenstates can be placed after the differentiation sign since x_m^0 corresponds to the initial stable equilibrium state. Thus Eq. 7.24 becomes

$$\sum_m x_m^0 \, dE_m = \sum_{i=1}^{s} \left(\frac{\partial(\Sigma_m x_m^0 E_m)}{\partial \beta_i}\right)_{S,\beta} d\beta_i = \sum_{i=1}^{s} \left(\frac{\partial E}{\partial \beta_i}\right)_{S,\beta} d\beta_i \tag{7.25}$$

where use has been made of Eq. 7.21. Comparing Eqs. 7.18 and 7.25 yields

$$(\delta W)_{\text{rev}} = -\sum_m x_m^0 \, dE_m \tag{7.26}$$

Furthermore, substituting Eq. 7.26 into Eq. 7.22 and comparing the result with Eqs. 7.16 and 7.20 yield

$$(\delta Q)_{\text{rev}} = \sum_m E_m \, dx_m \tag{7.27}$$

7.2.8 Work Plus Heat in a General Interaction. If system A receives heat δQ from a reservoir R, does work δW on system B (Fig. 7.3) by virtue of changes of values $d\beta_1, d\beta_2, \ldots, d\beta_s$ of the constraints of A, and if, before and after these energy exchanges, A is in stable equilibrium states, then Eq. 7.16 and conservation of energy imply that the differential energy change dE is given by the relations

$$dE = T\,dS + \sum_{i=1}^{s} \left(\frac{\partial E}{\partial \beta_i}\right)_{S,\beta} d\beta_i = \delta Q - \delta W \tag{7.28}$$

Without further specifications of the interactions, it is not possible to relate δQ and δW to the changes of properties of A in a manner analogous to that used for reversible interactions. For given δQ and δW, the best that can be done is to establish either upper or lower bounds for the changes of properties. If, during the interactions, A is not passing through stable equilibrium states, then irreversible processes occur; the entropy change dS of A is different from the entropy transferred from the reservoir; and, therefore, δQ is not uniquely related to dS. Specifically, heat δQ is supplied from reservoir R at temperature $T_R = T$, the entropy transferred is $\delta Q/T$ but the entropy change dS must be such that

$$dS \geqslant \frac{\delta Q}{T} \tag{7.29}$$

Moreover, substituting Eq. 7.29 into Eq. 7.28 yields

$$\delta W \leqslant -\sum_{i=1}^{s} \left(\frac{\partial E}{\partial \beta_i}\right)_{S,\beta} d\beta_i \tag{7.30}$$

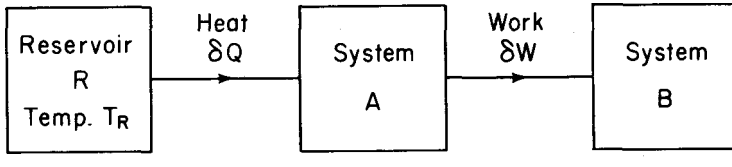


Fig. 7.3 Schematic of work-plus-heat interactions of A with B and R.

7.2.9 Explicit Forms of Equations of Stable Equilibrium States. The distribution (Eq. 6.36)

$$x_m^0 = \frac{\exp(-E_m/kT)}{Q_c} \tag{7.31}$$

can be used to derive explicit forms of equations of stable equilibrium states. For example, the entropy of stable equilibrium states is given by the relation

$$S = -k \sum_m x_m^0 \ln x_m^0 = k \ln Q_c + \frac{\Sigma x_m^0 E_m}{T} = k \ln Q_c + \frac{E}{T} \tag{7.32}$$

Solving Eq. 7.32 for the partition function Q_c yields

$$Q_c = \exp\left(-\frac{E - TS}{kT}\right) = \exp\left(-\frac{\Psi}{kT}\right) \tag{7.33}$$

where Ψ is the *Helmholtz free energy* given by the relation

$$\Psi = E - TS = -kT \ln Q_c \tag{7.34}$$

Other equations can be found in an analogous manner.

The partition function Q_c and the Helmholtz free energy Ψ are useful properties of stable equilibrium states for many applications. For example, it can be readily verified that

$$E = -T\left(\frac{\partial(\Psi/T)}{\partial \ln T}\right)_\beta \tag{7.35}$$

$$S = -\left(\frac{\partial \Psi}{\partial T}\right)_\beta \tag{7.36}$$

and

$$x_m^0 = \left(\frac{\partial \Psi}{\partial E_m}\right)_{T,\beta} \tag{7.37}$$

7.2.10 Simple Systems. Up to this point our discussion has been quite general, and no specifics have been given regarding constraints. This is appropriate since the discussion is applicable to all systems regardless of number and type of constraints. It is convenient, however, to illustrate the principles of quantum

thermodynamics by applying them to a *simple system,* i.e., any system that is subject to only one adjustable constraint β_1. For example, any system not influenced by capillarity, distortion of solid phases, and external force fields (electric, magnetic, or gravitational) is a simple system with volume V as the only adjustable constraint.

For a simple system with volume V, the energy E is called the *internal energy* and is denoted by U; thus the energy equation of stable equilibrium states becomes

$$U = U(S,V) \tag{7.38}$$

or in differential form

$$dU = T\,dS - p\,dV \tag{7.39}$$

where

$$T = \left(\frac{\partial U}{\partial S}\right)_V$$

and the pressure p is given by the relation

$$p = -\left(\frac{\partial U}{\partial V}\right)_S \tag{7.40}$$

Relation 7.38 is a *fundamental equation* and the corresponding function a *characteristic function.* A fundamental equation is a relation from which values of properties not explicit in it can be obtained by differentiation. For example, the *enthalpy, H,* the *Helmholtz free energy,* Ψ, and the *Gibbs free energy, Z,* are given by the relations

$$H = U - V\left(\frac{\partial U}{\partial V}\right)_S \tag{7.41}$$

$$\Psi = U - S\left(\frac{\partial U}{\partial S}\right)_V \tag{7.42}$$

$$Z = U - V\left(\frac{\partial U}{\partial V}\right)_S - S\left(\frac{\partial U}{\partial S}\right)_V \tag{7.43}$$

Other fundamental equations of the simple system are

$$H = H(S,p) \qquad \Psi = \Psi(T,V) \qquad Z = Z(T,p) \tag{7.44}$$

7.3 Applications

7.3.1 Efficiency of a Thermionic Converter—Carnot Efficiency. A thermionic converter receives heat at a rate Q_i from a reservoir at temperature T_E (heat source), rejects heat at a rate Q_o to a reservoir at temperature T_C (heat sink), and delivers electrical work to the environment at a rate W_T. These rates are such that

$$W_T = Q_i - Q_o \tag{7.45}$$

Strictly speaking, a thermionic converter is not an ordinary closed system of the type discussed in the preceding sections. Electrons leave from one terminal (the collector terminal), flow through an external electrical load, and return to the converter through the other terminal (the emitter terminal). In so doing, however, they deliver work to the load and leave no other effects outside the converter. For this reason the process experienced by the converter is thermodynamically equivalent to the process of a closed system subject to heat and work interactions.*

An upper limit for the efficiency η of conversion of heat into electrical work, namely,

$$\eta = \frac{W_T}{Q_i} \tag{7.46}$$

can be derived as follows. When the converter is in steady state, the only changes of entropy are those of the heat source and the heat sink. The rates of increase of these two entropies are $-Q_i/T_E$ and Q_0/T_C. According to the principle of increase of entropy, the sum of the two rates must be nonnegative, and therefore

$$-\frac{Q_i}{T_E} + \frac{Q_o}{T_C} \geqslant 0 \tag{7.47}$$

Using Eqs. 7.45 and 7.46 in Eq. 7.47 yields

$$\eta \leqslant \eta_c \tag{7.48}$$

where

*The equivalence holds only if the converter terminals are at the same temperature.

$$\eta_c = 1 - \frac{T_C}{T_E} \tag{7.49}$$

It is seen from Eq. 7.48 that η is smaller than η_c, which represents the highest attainable efficiency of any *heat engine,* (i.e., any device converting heat into work and operating between two reservoirs at temperatures T_E and T_C). The quantity η_c is also the efficiency of any reversible heat engine operating under the same conditions, because for a reversible process relation 7.48 holds with the equality sign. The efficiency η_c is called the *reversible efficiency* or the *Carnot efficiency* in honor of S. Carnot, who, before the discovery of the first and second laws, made the following statement concerning the efficiency of engines working between reservoirs at specified levels of temperature: The efficiency of a reversible engine has a fixed value that is the same for all reversible engines and is greater than that of any irreversible engine.

The efficiency η of a converter is much smaller than the Carnot efficiency because of many unavoidable irreversibilities (these are discussed later).

7.3.2 Maximum Useful Work. If a simple system A of volume V is surrounded by an atmosphere at constant temperature T_0 and pressure p_0, then the combination can do work on other systems, provided that A and the atmosphere are not in mutual stable equilibrium.

By virtue of the second law, the maximum work can be produced by the combination during reversible interactions. This maximum can be evaluated by considering the infinitesimal reversible energy exchanges shown schematically in Fig. 7.4. Heat is supplied from the atmosphere to A by means of a reversible heat engine B receiving the small heat quantity δQ_0 at T_0 for each cycle in B, and work $(\delta W)_{max}$ is delivered by A and B to other systems.

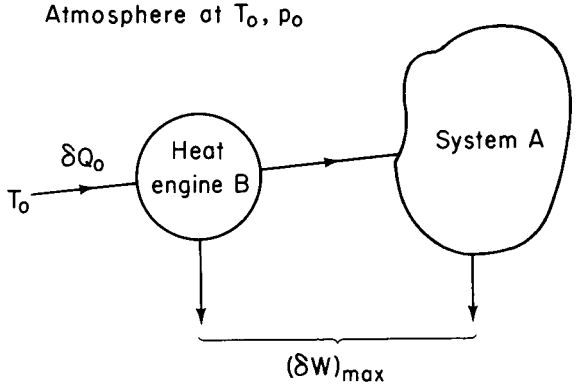


Fig. 7.4 Schematic of work done by A and the atmosphere in a nonequilibrium state.

If, for a specified change of its state, dE, dV, and dS denote, respectively, the energy, volume, and entropy changes of A, the following relations hold. By virtue of conservation of energy, $(\delta W)_{max}$ is given by the relation

$$(\delta W)_{max} = -dE + \delta Q_0 \tag{7.50}$$

Moreover, since the interactions are reversible and B undergoes a cycle, the entropy dS must be equal to the entropy $-dS_0$ supplied by the atmosphere, and therefore

$$dS = -dS_0 = \frac{\delta Q_0}{T_0} \tag{7.51}$$

Combining Eqs. 7.50 and 7.51 yields

$$(\delta W)_{max} = -dE + T_0\, dS = -d(E - T_0 S) \tag{7.52}$$

All the work $(\delta W)_{max}$ is not useful. While the volume of A changes by dV, the atmosphere receives an amount of work at least equal to $p_0\, dV$. It follows that the useful part $(\delta W_u)_{max}$ that can be obtained from the reversible interactions described is given by the relation

$$(\delta W_u)_{max} = (\delta W)_{max} - p_0\, dV = -d(E - T_0 S + p_0 V) = -d\Phi \tag{7.53}$$

where

$$\Phi = E - T_0 S + p_0 V \tag{7.54}$$

The quantity Φ is a property of the combination of A and the atmosphere. It applies to all states of A, equilibrium and nonequilibrium, and, therefore, E, S, and V are not always related by an equation of stable equilibrium states.

Equation 7.53 shows that positive useful work can be obtained from A and the atmosphere only by means of changes that correspond to a decrease in Φ. Moreover, since the useful work δW_u may be in any degree less than $(\delta W_u)_{max}$, δW_u can be zero for certain changes in spite of the decrease in Φ. For example, this may occur during irreversible processes. In general, for a change from state 1 to state 2, the useful work W_u satisfies the relation

$$W_u \leqslant \Phi_1 - \Phi_2 \tag{7.55}$$

For a given initial state 1 the *maximum useful work* $(W_u)_{max}$ that can be obtained from A and the atmosphere is given by the relation

$$(W_u)_{max} = \Phi_1 - \Phi_{min} \tag{7.56}$$

where Φ_{min} is the minimum value of Φ, which occurs when A is in mutual stable equilibrium with the atmosphere. Necessary but not sufficient requirements for this equilibrium are that the temperature and pressure of A are T_0 and p_0, respectively.

7.3.3 Equations of Stable Equilibrium States for One Structureless Particle. For a simple system A consisting of one structureless particle of mass m confined in a field-free box of volume V, the equations of stable equilibrium states are important because they yield information regarding the behavior of monatomic gases at very low densities.

Without loss of generality, we can assume that the volume V is cubical. Then the energy eigenvalues (as stated in Sec. 5.6) are given by the relation

$$\epsilon_n = \frac{h^2}{8mV^{2/3}}(n_1^2 + n_2^2 + n_3^2) \tag{7.57}$$

where n_1, n_2, and n_3 are positive integers. Substituting Eq. 7.57 into the expression for the partition function Q_c (Eq. 6.37) yields

$$\begin{aligned} Q_c &= \sum_{n_1}\sum_{n_2}\sum_{n_3} \exp\left[-\frac{h^2}{8mV^{2/3}kT}(n_1^2 + n_2^2 + n_3^2)\right] \\ &= \left[\sum_{n} \exp\left(-\frac{h^2 n^2}{8mV^{2/3}kT}\right)\right]^3 \end{aligned} \tag{7.58}$$

where the second relation results from the observation that the summations over n_1, n_2, and n_3 are independent of each other.

For practical values of m, V, and T, the value of $h^2/8mV^{2/3}kT$ is much smaller than unity. For this reason the summation in Eq. 7.58 can be replaced by an integral so that

$$Q_c = \left[\int_0^\infty \exp\left(-\frac{h^2 x^2}{8mV^{2/3}kT}\right)dx\right]^3 = \left(\frac{2\pi mkT}{h^2}\right)^{3/2} V \tag{7.59}$$

In the derivation of Eq. 7.59, the energy eigenstates are assumed to be nondegenerate. If each energy eigenstate is g-fold degenerate, then the partition function Q_c is given by the relation

$$Q_c = g\left(\frac{2\pi mkT}{h^2}\right)^{3/2} V \tag{7.60}$$

Substituting Eq. 7.60 into Eq. 7.34 for the Helmholtz free energy Ψ yields

$$\Psi = -kT \ln V - \frac{3}{2} kT \ln T - kT \ln \frac{g(2\pi mk)^{3/2}}{h^3} \tag{7.61}$$

The Helmholtz free energy can be used to compute other properties of the particle. For example, we can readily verify that

$$p = -\left(\frac{\partial \Psi}{\partial V}\right)_T \tag{7.62}$$

Thus, substituting Eq. 7.61 into Eq. 7.62 yields

$$pV = kT \tag{7.63}$$

A simple system for which the product of pressure p and volume V is proportional to temperature T is defined as a *semiperfect gas.* The constant of proportionality per particle equals the ratio of the semiperfect gas constant divided by Avogadro's number, namely, the Boltzmann constant k. Equation 7.63 shows that the simple system under consideration behaves like a semiperfect gas and that the constant k (Eq. 6.35) is indeed Boltzmann's constant.

Another example of the usage of Ψ is provided by the calculation of the energy U and entropy S of the particle. Substituting Eq. 7.61 into Eqs. 7.35 and 7.36 yields

$$U = \frac{3}{2} kT \tag{7.64}$$

and

$$S = \frac{3}{2} k \ln T + k \ln V + k \left[\frac{3}{2} + \ln \frac{g(2\pi mk)^{3/2}}{h^3}\right] \tag{7.65}$$

Together Eqs. 7.64 and 7.63 imply that the system of one structureless particle in a box is a *perfect gas* (a perfect gas is defined later).

7.3.4 The Maxwellian Velocity Distribution. For the simple system A discussed in Sec. 7.3.3, the question can be raised, "For A in a stable equilibrium state, what is the probability $f(\mathbf{v}_n)$ that a measurement will yield a velocity $\mathbf{v}_n$?"

The velocity $\mathbf{v}_n$ has components v_{nx}, v_{ny}, and v_{nz} such that

$$E_n = \frac{m}{2}(v_{nx}^2 + v_{ny}^2 + v_{nz}^2) = \frac{m}{2}|\mathbf{v}_n|^2 \tag{7.66}$$

It follows that $f(\mathbf{v}_n)$ must be proportional to the probability of finding the particle with energy E_n.

The probability of finding the particle with energy E_n is given by the canonical distribution x_n^0 (Eq. 6.36). Consequently $f(\mathbf{v}_n)$ must be proportional to $\exp(-E_n/kT)$ so that

$$f(\mathbf{v}_n) = c_2 \exp\left(-\frac{E_n}{kT}\right) = c_2 \exp\left[-\frac{m}{2kT}(v_{nx}^2 + v_{ny}^2 + v_{nz}^2)\right] \tag{7.67}$$

where c_2 is defined by the requirement that the sum of all probabilities $f(\mathbf{v}_n)$ must be equal to unity. For practical values m, V, and T, $\mathbf{v}_n$ can be thought of as continuous. Then it is readily verified that

$$f(\mathbf{v}) = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left[-\frac{m}{2kT}(v_x^2 + v_y^2 + v_z^2)\right] \tag{7.68}$$

The quantity $f(\mathbf{v})$ is called the *Maxwellian velocity distribution* or the *Maxwellian distribution.* It applies also to perfect monatomic gases with any number of particles, but then $f(\mathbf{v})\,d\mathbf{v}$ is the expectation value of the number of particles of the gas with velocities in the range $\mathbf{v}$ and $\mathbf{v} + d\mathbf{v}$.

7.3.5 The One-Dimensional Harmonic Oscillator. The properties of the one-dimensional harmonic oscillator are useful in many applications including molecular and lattice vibrations.

The ith energy eigenvalue ε_i of such an oscillator is given by the relation

$$\varepsilon_i = \left(i + \frac{1}{2}\right)h\nu \tag{7.69}$$

where i is a nonnegative integer and ν is the frequency of oscillation.

The energy eigenvalues can be used to compute all the properties of the stable equilibrium states of the harmonic oscillator. For example, the canonical distribution x_i^0 is given by the relation

$$x_i^0 = \frac{\exp(-\varepsilon_i/kT)}{q_s} = \frac{\exp\left[-\left(i + \frac{1}{2}\right)h\nu/kT\right]}{q_s} \tag{7.70}$$

where q_s is the canonical partition function given by the sum

$$q_s = \sum_i \exp\left(-\frac{\varepsilon_i}{kT}\right) = \frac{\exp(-h\nu/2kT)}{1-\exp(-h\nu/kT)} \tag{7.71}$$

Again the values of the energy ε and the entropy s of the one-dimensional harmonic oscillator are given by the relations

$$\varepsilon = \sum x_i^0 \varepsilon_i = \frac{1}{2}h\nu + \frac{h\nu}{\exp(h\nu/kT)-1} \tag{7.72}$$

and

$$s = k \ln q_s + \frac{\varepsilon}{T} = \frac{h\nu/T}{\exp(h\nu/kT)-1} - k \ln\left[1-\exp\left(-\frac{h\nu}{kT}\right)\right] \tag{7.73}$$

At very low and very high temperatures, the partition function q_s, energy ε, and entropy s of the one-dimensional harmonic oscillator assume the following limiting values.

For $kT \ll h\nu$: $\quad q_s \approx \exp\left(-\frac{h\nu}{2kT}\right) \quad \varepsilon \approx \frac{1}{2}h\nu \quad s \approx 0$

For $kT \gg h\nu$: $\quad q_s \approx \frac{kT}{h\nu}\exp\left(-\frac{h\nu}{2kT}\right) \quad \varepsilon \approx kT \quad s \approx -k \ln \frac{h\nu}{kT}$

These limiting results show that at very low temperatures the harmonic oscillator is in its ground state.

Again, in the limit of zero frequency, the partition function q_s, distribution x_n^0, energy ε, and entropy s assume the following values.

For $\nu = 0$: $\quad q_s = \infty \quad x_n^0 = 0 \quad \varepsilon = kT \quad s = \infty$

It is noteworthy that in the limit of zero frequency the harmonic oscillator in a stable equilibrium state behaves as a reservoir of infinite entropy.

7.4 Ordinary Open Systems

7.4.1 General Remarks. A system is called an ordinary open system if over a period of time its interactions with the environment include exchanges of both energy and matter.

Many practical problems require the consideration of ordinary open systems. For example, for an electron emitter adjacent to a gaseous phase, one of the quantities of interest is the electron current flowing from the emitter to the gaseous phase. If the emitter is regarded as a system, it must be an open system because otherwise the electron component would not be allowed to cross the interface between the solid and the gaseous phases.

Another example is liquid H_2O immersed in H_2O vapor. The interface between liquid and vapor is crossed by H_2O molecules, and therefore both the liquid and the vapor can be regarded as open systems. These two systems have the same *composition;* in other words, both consist of the same single component, H_2O molecules.

A third example is an amount of liquid air immersed in gaseous air. The liquid is an open system because air molecules condense and evaporate from its surface. The same qualification applies to the gaseous phase. The components of these two systems are the same (say, oxygen and nitrogen molecules), but the composition of one is different from that of the other (the liquid is richer in oxygen than the gas).

The analysis of stable equilibrium states of open systems proceeds in a manner more or less similar to that discussed for ordinary closed systems. An important difference is introduced, however, by the openness of the boundaries of the system to the transfer of matter. For a given amount of matter, the number of energy eigenstates is much larger than that of a closed system because of the many possible eigenvalues of the numbers of particles of distinguishable components. Each set of eigenvalues defines a complete set of energy eigenfunctions. Since many sets of such numbers are possible, many complete sets of energy eigenfunctions are necessary for the description of open systems.

Because of the many sets of possible eigenvalues of numbers of particles of distinguishable components, the expectation values of these numbers are independent variables rather than fixed parameters of an open system. Accordingly an equation of stable equilibrium states for property F is given by a relation of the form (Eq. 6.8)

$$F = F(E; \beta_1, \beta_2, \ldots, \beta_s; n_1, n_2, \ldots, n_r) \tag{7.74}$$

where $n_1, n_2, \ldots, n_r$ are the expectation values of the numbers of particles of the r distinguishable components. In particular, the energy E of an open system passing through stable equilibrium states is given by the relation

$$E = E(S, \beta_1, \beta_2, \ldots, \beta_s; n_1, n_2, \ldots, n_r) \tag{7.75}$$

Such relations can be written also for quasistable states (Sec. 7.2.2).

Another consequence of the many eigenvalues of the numbers of particles of distinguishable components is that the values $n_1, n_2, \ldots, n_r$ vary continuously rather than discontinuously. This result is similar to the continuous variation of the energy E of a closed system although the system may be admitting discrete energy eigenvalues only.

For analyses of stable equilibrium states, it is often convenient to number all the characteristics of energy eigenstates consecutively. Thus the energy eigenstates are denoted as $G_1, G_2, \ldots, G_n, \ldots$; the energy eigenvalues as $E_1, E_2, \ldots, E_n, \ldots$; the numbers of particles as $(n_{11}, n_{21}, \ldots, n_{r1}), (n_{12}, n_{22}, \ldots, n_{r2}), \ldots, (n_{1n}, n_{2n}, \ldots, n_{rn}), \ldots$; and the distribution as $x_1^0, x_2^0, \ldots, x_n^0, \ldots$. In this notation a different symbol is used for each characteristic of each energy eigenstate. For example, different symbols are used for the energies of a g-fold degenerate energy eigenstate even though these energies are identical. Again, different symbols are used for two eigenvalues corresponding to different numbers of particles, even though these two eigenvalues may be equal.

The Kelvin temperature scale is used for open systems. This temperature is given by Eq. 6.10, repeated here for convenience:

$$T = \left(\frac{\partial E}{\partial S}\right)_{\beta,n} \tag{7.76}$$

Finally, the total potential μ_l of component l and the generalized body force F_i conjugate to the constraint β_i are defined by Eqs. 6.11 and 6.12, respectively. These definitions are

$$\mu_l = \left(\frac{\partial E}{\partial n_l}\right)_{S,\beta,n} \qquad \text{for } l = 1, 2, \ldots, r \tag{7.77}$$

and

$$F_i = \left(\frac{\partial E}{\partial \beta_i}\right)_{S,\beta,n} \qquad \text{for } i = 1, 2, \ldots, s \tag{7.78}$$

The properties of stable equilibrium states T, μ_l, and F_i are very helpful in the subsequent discussions. They are called *thermodynamic potentials.*

7.4.2 The Generalized Gibbs Equation. By virtue of Eqs. 7.75 to 7.78, the differential energy change dE of an open system passing through stable equilibrium states is given by the relation

$$dE = \left(\frac{\partial E}{\partial S}\right)_{\beta,n} dS + \sum_{l}^{r} \left(\frac{\partial E}{\partial n_l}\right)_{S,\beta,n} dn_l + \sum_{i}^{s} \left(\frac{\partial E}{\partial \beta_i}\right)_{S,\beta,n} d\beta_i$$

$$= T\,dS + \sum_{l}^{r} \mu_l\,dn_l + \sum_{i}^{s} F_i\,d\beta_i \tag{7.79}$$

This relation, called the *generalized Gibbs equation*, has many applications.

7.4.3 Conditions for Mutual Stable Equilibrium. When two open systems A and B are in mutual stable equilibrium, the combination is in a stable equilibrium state, and neither energy nor matter crosses any common interface between the two systems even though such crossings are not prohibited.

For A and B to be in mutual stable equilibrium, their properties must satisfy certain conditions, which can be established by means of the criterion for stable equilibrium and by considering changes of states for which both A and B pass through stable equilibrium states.

According to the criterion for stable equilibrium, the entropy of the combined system A plus B must be larger than the entropy of any other state of A plus B of the same energy. For all possible changes for which

$$dE_A + dE_B = 0 \tag{7.80}$$

the entropy change $dS_A + dS_B$ of the combined system must satisfy the relation

$$dS_A + dS_B < 0 \tag{7.81}$$

One possible change consists in an energy exchange between A and B without change in any constraints or in any values of numbers of particles. Such a change leads to considerations similar to those discussed in Sec. 7.2.3 and to the conclusion that A and B must be in temperature equality, namely,

$$T_A = T_B \tag{7.82}$$

For fixed values of all constraints, another possible change consists in altering the value $(n_l)_A$ of A to $(n_l)_A + dn_l$ and the value of $(n_l)_B$ of B to $(n_l)_B - dn_l$ while keeping all other values of numbers of particles fixed. The entropies dS_A and dS_B for this change can be found from the generalized Gibbs equation (7.79). Substituting into Eq. 7.81 from Eq. 7.79 yields

$$\frac{dE_A}{T_A} - \frac{(\mu_l)_A}{T_A}\,dn_l + \frac{dE_B}{T_B} + \frac{(\mu_l)_B}{T_B}\,dn_l \leqslant 0 \tag{7.83}$$

Combining Eqs. 7.80, 7.82, and 7.83 yields

$$[(\mu_l)_A - (\mu_l)_B]\, dn_l > 0 \tag{7.84}$$

If dn_l can assume both positive and negative values, i.e., if component l is present in both A and B, condition 7.84 is satisfied if and only if

$$(\mu_l)_A = (\mu_l)_B \tag{7.85}$$

We see from Eq. 7.85 that for mutual stable equilibrium the total potentials of component l present in both A and B must be equal. Moreover, since no conditions were imposed on l, the conclusion is applicable to all components provided these components are present in both A and B.

For fixed constraints and entropies, still another possible change must be considered if some components ($F, G, \ldots$) can combine chemically to form other components ($X, Y, \ldots$) in accordance with the stoichiometric statement of the form

$$fF + gG + \ldots = xX + yY + \ldots \tag{7.86}$$

where $f, g, \ldots$ and $x, y, \ldots$ are the numbers characterizing the stoichiometry of the chemical reaction. It can be shown that, if the interacting components are present in any system in stable equilibrium, the total potentials of these components satisfy the relation[1]

$$f\mu_F + g\mu_G + \ldots = x\mu_X + y\mu_Y + \ldots \tag{7.87}$$

which is called the *equation of chemical equilibrium.*

Depending on the nature of the open systems A and B and on their independent constraints, more conditions than the three mentioned may be required for mutual stable equilibrium. The conditions can be established by considering the consequences of possible changes of different constraints.

In general, conditions for mutual stable equilibrium are very helpful in relating properties of one system to those of another. Illustrative applications are given later.

7.5 One Atom as an Ordinary Open System

An atom or a molecule can be regarded as an open system with electrons as its only component because it can exchange energy and electrons while reacting chemically with other atoms and molecules.

An atom A of atomic number Z contains Z units of positive electric charge in addition to its electrons. It has all the energy eigenvalues of the neutral atom and

of each of the multiply (positively) charged ionized atoms. In addition it has energy eigenvalues corresponding to a simply charged negative ion, namely, an ion with $Z + 1$ electrons.

The eigenvalues n_i of the number of electrons in atom A are such that

$$n_i = i = 0, 1, 2, \ldots, Z, Z + 1$$

That is, the atom can be fully ionized ($i = 0$), partially ionized ($i < Z$), neutral ($i = Z$), or singly negatively charged ($i = Z + 1$).

For each n_i a number g_{i0} of energy eigenvalues will have the minimum numerical value E_{i0}, where g_{i0} may be unity or greater than unity.* Each of the g_{i0} identical eigenvalues E_{i0} represents the energy of the so-called ground state of A with exactly n_i electrons. Moreover, an infinite number of energy eigenvalues E_{ij} ($j = 1, 2, \ldots$) each with a degeneracy g_{ij} correspond to the excited states for each n_i.

The ground-state energies decrease as the number of electrons increases from 0 to $Z + 1$. In symbols

$$E_{00} > E_{10} > \ldots > E_{Z-1,0} > E_{Z,0} > E_{Z+1,0}$$

If the arbitrary zero-energy reference level is set at the energy $E_{Z,0}$ of the neutral atom, the relative values of the ground state energies can be written as

$$I_Z > I_{Z-1} > \ldots > I_1 > 0 > -A_1$$

where I_i and A_i are the ith ionization energy and electron affinity of the atom, respectively. Finally, each of the energies E_{ij} is greater than the energy E_{i0} of the corresponding ground state.

For a stable equilibrium state at temperature T and total electron potential μ, the expectation value n of the number of electrons of the atom is given in terms of this notation by the relation

$$n = \frac{\sum_{i=1}^{Z+1} \sum_j n_i g_{ij} \exp\left[(n_i\mu - E_{ij})/kT\right]}{\sum_{i=0}^{Z+1} \sum_j g_{ij} \exp\left[(n_i\mu - E_{ij})/kT\right]} \tag{7.88}$$

where use has been made of the grand canonical distribution (Eq. 6.43). It is evident from Eq. 7.88 that the number of electrons n may assume any value

*By virtue of the third law of thermodynamics, it can be shown that the ground state of any system must be nondegenerate.

between 0 and $Z + 1$ even though the numbers n_i assume integral values only. In other words, n is a continuous variable, but n_i is not. Other properties can be computed in an analogous manner.

For given values of n and T, Eq. 7.88 can be solved for the values of μ. Although the solution is numerically tedious, some general results can be readily established.

1. For all values of T, for a fully ionized state (i.e., for the minimum value $n = 0$), μ is minus infinity, and, for a singly charged negative-ion state (i.e., for the maximum value $n = Z + 1$), μ is plus infinity. Such extreme values of μ are obtained always when the value of the number of particles of the grand system is either minimum or maximum. They are consistent with the interpretation of the potential μ as an escaping tendency (Sec. 7.7.5).

2. For values of n in the range between Z and $Z + 1$, the potential μ is positive. For all other values the potential μ is negative.

3. For a neutral-atom state (i.e., for $n = Z$) in the limit as T approaches zero, μ is equal to one-half the difference between the eigenvalues of the ground states of the singly charged negative ion and the singly charged positive ion. It can also be written in the form

$$\mu = -\frac{I_1 + A_1}{2} \tag{7.89}$$

Equation 7.89 is a special case of the more general result discussed in item 4.

4. For a positive-ion state with an integral number σ of electrons (i.e., $n = \sigma$) in the limit of small temperatures ($T \approx 0°\text{K}$), Eq. 7.88 can be approximated by the relation

$$\sum_{i=0}^{\sigma-1} (\sigma - n_i) g_{i0} \exp\left(\frac{n_i\mu - E_{i0}}{kT}\right) \approx \sum_{i=\sigma+1}^{Z+1} (n_i - \sigma) g_{i0} \exp\left(\frac{n_i\mu - E_{i0}}{kT}\right) \tag{7.90}$$

In addition to the small temperature, the approximation is based on the facts that E_{ij} is greater than E_{i0} and that μ is negative. For different ranges of negative values of μ, the exponents $n_i\mu - E_{i0}$ on either side of Eq. 7.90 can be ordered. For a given range of values of μ, if the largest exponents are $l\mu - E_l$ and $r\mu - E_r$ on the left- and right-hand side, respectively, where $l \leqslant \sigma - 1$ and $r > \sigma + 1$, then Eq. 7.90 can be further approximated by the simple relation

$$g_{l0}(\sigma - l) \exp\left(\frac{l\mu - E_{l0}}{kT}\right) = g_{r0}(r - \sigma) \exp\left(\frac{r\mu - E_{r0}}{kT}\right) \tag{7.91}$$

For Eq. 7.91 to be satisfied, the value of μ must be given by the relation

$$\mu = \frac{(E_{r0} - E_{l0}) + kT \ln \{[g_{r0}(\sigma - l)]/[g_{l0}(r - \sigma)]\}}{r - l}$$

$$= \frac{(I_{z-r} - I_{z-l}) + kT \ln \{[g_{r0}(\sigma - l)]/[g_{l0}(r - \sigma)]\}}{r - l} \quad \text{for } n = \sigma \text{ and } T \approx 0 \qquad (7.92)$$

This value of μ is acceptable if it is within the range of values assumed for the ordering of the exponents, namely, the range resulting in approximate Eq. 7.91. Otherwise, another range of μ and different values of l and r must be considered.

For electrons the total potential μ is called the *electrochemical potential.* Qualitatively, the electrochemical potential is a measure of the tendency of electrons to escape from the open system. The negative of μ represents a power to attract electrons. For this reason the quantity $-\mu$ for electrons has been identified with the chemical concept of *electronegativity*.[2] Electronegativity, which is useful for surface studies, will be discussed again in Chap. 9.

7.6 Homogeneous Systems with Volume as an Independent Constraint

7.6.1 General Remarks. Most ordinary open systems of interest to thermionic conversion have two common features. First, they can be approximated either by *homogeneous systems* or by *heterogeneous systems consisting of homogeneous parts.* Second, they can be described by two independent constraints, one being volume and the other being an applied field, usually an electric field.

For these reasons the subsequent derivations and applications are restricted to homogeneous systems with volume plus an applied field as the only independent constraints. The discussion begins with the definitions of *extensive* and *intensive properties* and of homogeneous and heterogeneous systems.

Extensions of the results to two-dimensional systems are given at the end of the chapter. Such systems are of interest in analyzing composite thermionic emitter surfaces.

7.6.2 Extensive and Intensive Properties. Properties of a system can be classified as *extensive* or *intensive.*

If system A can be subdivided into independent parts, then an *extensive property* of A is a property whose value equals the sum of the values of the corresponding properties of the independent parts. An *intensive property* of A is a property whose value at a spatial point in a part of A approaches a limit independent of the size of the part as the size is reduced until it can no longer be identified as a system. An intensive property may either assume the same value in all independent parts or vary from part to part.

Examples of extensive properties are energy, volume, number of particles, and entropy. Examples of intensive properties are temperature, pressure, and density of particles.

7.6.3 Homogeneous and Heterogeneous Systems. A given system A with uniform intensive properties is defined as a *homogeneous system.* For example, a gas in a stable equilibrium state and in the absence of force fields is a homogeneous system because the values of temperature, pressure, particle density per unit volume, etc., are the same for any part of the gas as for the whole.

A given system A made of independent homogeneous parts which are open to the transfer of matter and which differ in intensive properties is defined as a *heterogeneous system.* The homogeneous parts of a heterogeneous system are called phases. A simple example of a heterogeneous system is one consisting of liquid and vapor water in a container. The density of H_2O molecules in the liquid is different from that in the vapor. Another example is a small amount of liquid air immersed in gaseous air, both confined in a large container. The densities of particles and the composition of each of the two phases of this system are different.

For the two examples of heterogeneous systems just cited, the change from liquid to gaseous phase is gradual; in other words, there is a small part of the system which cannot be characterized either as a liquid phase or as a gaseous phase. For many practical problems, however, it is expedient to disregard this gradual change and to approximate the systems by two phases separated by a surface of discontinuity in density or composition, or both. Thus each of the two heterogeneous systems can be regarded as consisting of two phases, one liquid and one gaseous. Moreover, it is helpful to assume that the discontinuities do not introduce capillary effects. The approximation and the assumption are justified when the volumes of the phases enclosed by the surface of discontinuity are much larger than the volume of the small part with gradually changing properties.

The approximation of a heterogeneous system by two or more phases, each having volume as an independent constraint, is not restricted to gaseous–liquid phases. It may also involve solid phases. When solids are present, the approximation requires that they not be subject to shear.

A third example of a heterogeneous system is a column of plasma between two parallel electrodes. Here the density of the charged particles of the plasma varies with position between the electrodes. Such a heterogeneous system can be represented by a number of phases, each of which consists of a thin slice of the plasma at a specific distance from one of the electrodes. At each distance the intensive properties of the slice are the same across it. Unlike the previous examples, here the phases are not separated by surfaces of discontinuity.

Apart from the fact that it can be represented by a number of homogeneous phases, the third example suggests the method that might be used for the approximate analysis of any heterogeneous system with a continuously varying

intensive property. Of course, for the approximation in terms of homogeneous phases to be applicable, each phase must be a statistically independent system.

From these remarks it follows that knowledge of equations of stable equilibrium states of homogeneous systems plus the conditions for mutual stable equilibrium is sufficient for the study of stable equilibrium states of heterogeneous systems consisting of homogeneous phases.

7.6.4 The Gibbs Equation for Systems with Two Constraints. In many applications volume V is an independent constraint of an open system. For example, this is true for three-dimensional and isotropic systems such as a gas or a solid not subject to shear stresses.

If V and an applied field β are the only two independent constraints of a system A, then Eqs. 7.75 and 7.79 become

$$E = E(S;V;\beta;n_1,n_2,\ldots,n_r) \tag{7.93}$$

and

$$dE = T\,dS - p\,dV + \left(\frac{\partial E}{\partial \beta}\right)_{S,V,n} d\beta + \sum_{l=1}^{r} \mu_l\,dn_l \tag{7.94}$$

where

$$T = \left(\frac{\partial E}{\partial S}\right)_{V,\beta,n} \qquad p = -\left(\frac{\partial E}{\partial V}\right)_{S,\beta,n} \qquad \mu_l = \left(\frac{\partial E}{\partial n_l}\right)_{S,V,\beta,n} \tag{7.95}$$

The definition of the quantity p, called the pressure, is analogous to that for ordinary closed systems. If the system were not three dimensional and isotropic, however, neither volume nor pressure could be defined. Because Eq. 7.94 is applicable only to three-dimensional and isotropic systems, it is distinguished from Eq. 7.79, which is applicable to systems in general, and is referred to simply as the *Gibbs equation*.

When the independent constraint β is an applied field, the generalized body force and its conjugate generalized displacement can be expressed in a number of equivalent ways. For example, the displacement $d\beta$ can be an infinitesimal change $d\mathbf{r}$ in the position of the system in the field. For an electrostatic field the generalized force conjugate to $d\mathbf{r}$ is given by the relation

$$\left(\frac{\partial E}{\partial \beta}\right)_{S,V,n} = \left(\frac{\partial E}{\partial \mathbf{r}}\right)_{S,V,n} = -\left(\sum_{l=1}^{r} q_l n_l\right)\mathbf{E} \tag{7.96}$$

where q_l is the number of units of positive electric charge per particle of component l, **E** is the intensity of the applied electric field, and the sum in the second part of Eq. 7.96 is the net positive electric charge of the open system. Implicit in this expression for the force is the assumption that **E** does not depend on the degree of charging of the system; in other words, **E** is an applied field.

For fixed values of S, V, and $n_1, n_2, \ldots, n_l$, the differential energy change $dE|_{S,V,n}$ due only to a displacement in the field is given by the relation

$$dE|_{S,V,n} = -\left(\sum_l^r q_l n_l\right) \mathbf{E} \cdot d\mathbf{r} \tag{7.97}$$

Alternative definitions of the generalized force and its conjugate displacement can be determined by recognizing that the field **E** can be derived from an electrostatic potential ψ. Specifically

$$\mathbf{E} = -\operatorname{grad} \psi \tag{7.98}$$

Therefore Eq. 7.97 can be written in the form

$$dE|_{S,V,n} = \left(\sum_l^r q_l n_l\right) \operatorname{grad} \psi \cdot d\mathbf{r} = \left(\sum_l^r q_l n_l\right) d\psi \tag{7.99}$$

where $d\psi$ is the change in ψ corresponding to $d\mathbf{r}$. Equation 7.99 indicates that the energy E can be thought of as a function of ψ and that the generalized force is equal to the net positive electric charge, whereas its conjugate displacement equals $d\psi$. Accordingly the Gibbs equation and E are given by the relations

$$dE = T\,dS - p\,dV + \left(\sum_l^r q_l n_l\right) d\psi + \sum_l^r \mu_l\,dn_l \tag{7.100}$$

and

$$E = E(S;V;\psi;n_1,n_2,\ldots,n_r) \tag{7.101}$$

Relations analogous to Eqs. 7.97 and 7.99 can be derived for other types of applied fields, such as a gravitational field, and for several types of fields applied simultaneously. An important feature of electrostatic fields is that the net positive electric charge is independent of ψ at constant n_l. When the generalized force does not depend on the constraint ψ at constant S, V, and all values of n_l, the energy of the system can be written as a sum of two terms, one representing the energy in the absence of the force field and the other representing the energy due to the force field. Moreover, when the force field does not depend on any of the properties S and V and the constraint ψ, the energy E is a sum of these two

terms, and the temperature and pressure of the system are independent of the value of the constraint ψ. These points are clarified in Sec. 7.6.5.

7.6.5 Internal Energy—Chemical Potential. Integrating Eq. 7.99 over the range 0 to ψ for given fixed values of S, V, and $n_1, n_2, \ldots, n_r$ yields the relation

$$E = \left(\sum_l^r q_l n_l\right)\psi + U(S;V;n_1,n_2,\ldots,n_r) \tag{7.102}$$

where U is a function of all the independent properties except ψ. Equation 7.102 holds for all values of S, V, ψ, and $n_1, n_2, \ldots, n_r$. In particular, for $\psi = 0$ the equation yields

$$E = U(S;V;n_1,n_2,\ldots,n_r) \tag{7.103}$$

It is seen from Eq. 7.103 that U is the energy of the open system in the absence of the applied field. For this reason U is defined as the *internal energy*, and the definition is an extension of that given for simple systems (Sec. 7.2.10). Substituting Eq. 7.102 into the first two relations in Eq. 7.95 yields

$$T = \left(\frac{\partial E}{\partial S}\right)_{V,\psi,n} = \left(\frac{\partial\left(\sum_l^r q_l n_l\right)\psi}{\partial S}\right)_{V,\psi,n} + \left(\frac{\partial U}{\partial S}\right)_{V,n} = \left(\frac{\partial U}{\partial S}\right)_{V,n} \tag{7.104}$$

and

$$p = -\left(\frac{\partial E}{\partial V}\right)_{S,\psi,n} = -\left(\frac{\partial\left(\sum_l^r q_l n_l\right)\psi}{\partial V}\right)_{S,\psi,n} - \left(\frac{\partial U}{\partial V}\right)_{S,n} = -\left(\frac{\partial U}{\partial V}\right)_{S,n} \tag{7.105}$$

Equations 7.104 and 7.105 indicate that the values of the temperature T and the pressure p are not affected by the presence of the applied field. Moreover, it is easy to verify that either the temperature or the pressure, or both, would depend on the value of the applied field if the generalized force were a function of either S or V, or both.

Finally, substituting Eq. 7.102 into the third relation in Eq. 7.95 for the total potential μ_l yields

$$\mu_l = \left(\frac{\partial E}{\partial n_l}\right)_{S,V,\psi,n} = q_l\psi + \mu_{cl} \qquad \text{for all } l \tag{7.106}$$

where

$$\mu_{cl} = \left(\frac{\partial U}{\partial n_l}\right)_{S,V,n} \qquad \text{for all } l \tag{7.107}$$

The quantity μ_{cl}, called the *chemical potential of component l*, is independent of the value of the applied field. The relation between total and chemical potentials (Eq. 7.106) is very useful for applications of interest to thermionic conversion. It is valid only when the energy can be written as a sum of terms of the form of Eq. 7.102.

7.6.6 The T–p Equations. Up to this point equations of stable equilibrium states have been written with S, V, β, and n_1, $n_2, \ldots, n_r$ as independent properties. Alternative forms can be obtained by replacing S and V by T and p. Thus the energy E and its differential dE are given by the relations

$$E = E(T;p;\beta;n_1, n_2, \ldots, n_r) \tag{7.108}$$

and

$$dE = \left(\frac{\partial E}{\partial T}\right)_{p,\beta,n} dT + \left(\frac{\partial E}{\partial p}\right)_{T,\beta,n} dp + \left(\frac{\partial E}{\partial \beta}\right)_{T,p,n} d\beta + \sum_l^r e_l \, dn_l \tag{7.109}$$

where

$$e_l = \left(\frac{\partial E}{\partial n_l}\right)_{T,p,\beta,n} \qquad \text{for all } l \tag{7.110}$$

Equation 7.108 is called the *T–p equation for the energy.* A similar equation can be written for any other extensive property F. When this is done, the differential dF is given by the relation

$$dF = \left(\frac{\partial F}{\partial T}\right)_{p,\beta,n} dT + \left(\frac{\partial F}{\partial p}\right)_{T,\beta,n} dp + \left(\frac{\partial F}{\partial \beta}\right)_{T,p,n} d\beta + \sum_l^r f_l \, dn_l \tag{7.111}$$

where

$$f_l = \left(\frac{\partial F}{\partial n_l}\right)_{T,p,\beta,n} \qquad \text{for all } l \tag{7.112}$$

The quantities e_l and f_l for all l are of special interest. Each is called a *partial property.* For example, e_l is a *partial energy of component l.* Partial properties, which are discussed in Sec. 7.6.7, are useful for analyzing homogeneous systems.

7.6.7 Partial Properties. Many important relations regarding homogeneous systems are conveniently expressed in terms of partial properties. These properties are intensive; they depend only on the relative composition and not on the absolute values of the numbers of particles of the various components; and their magnitudes can be related to tabulated quantities established by means of measurements on or calculations for one-component systems.

If two identical homogeneous systems A and B are characterized by the same sets of values T, p, β, and $n_1, n_2, \ldots, n_r$ of their independent properties, then the value of any property of A equals that of B, and therefore the value of the partial energy e_l of A equals that of B.

In the homogeneous system C consisting of A and B together, the stable equilibrium state is characterized by the set T, p, β, $2n_1, 2n_2, \ldots, 2n_r$ since T and p are intensive properties and the applied field is the same for both A and B. In addition, the energy of A is twice that of either A or B. Since both the energy and the number of particles of component l have been increased by the same factor, it follows that the partial energy e_l of C has the same value as that of A or B. Hence e_l is intensive. Moreover, since the value of e_l is the same for the two sets of independent properties (T, p, β, and $n_1, n_2, \ldots, n_r$ and T, p, β, and $2n_1, 2n_2, \ldots, 2n_r$), it follows that, for given values of T, p, and β, the value of e_l depends only on the composition and not on the exact magnitudes of $n_1, n_2, \ldots, n_r$. The same arguments and conclusions are applicable to any other partial property f_l.

Partial properties are very useful for computing extensive properties. For example, if, in a multicomponent homogeneous system in a stable equilibrium state with $n_1, n_2, \ldots, n_r$ moles or numbers of particles of its components, the amounts $dn_1, dn_2, \ldots, dn_r$ are added under conditions of constant temperature, pressure, and value of β, then according to Eq. 7.109 the change of energy $dE|_{T,p,\beta}$ is given by the relation

$$dE|_{T,p,\beta} = \sum_{l}^{r} e_l \, dn_l \tag{7.113}$$

Equation 7.113 can be used to compute the energy E. For given values T, p, and β, the numbers $n_1, n_2, \ldots, n_r$ can be thought of as having been built up in steps. In each step the differential amounts $dn_1, dn_2, \ldots, dn_r$ are added in such a way that they do not alter the composition of the homogeneous system. In other words, the numbers of particles $dn_1, dn_2, \ldots, dn_r$ at each step are added so as to bear the same relation to each other as do the numbers of particles $n_1, n_2, \ldots, n_r$ of the components already present in the system. In symbols this means that

$$\frac{dn_1}{n_1} = \frac{dn_2}{n_2} = \ldots = \frac{dn_r}{n_r} \tag{7.114}$$

In each step the energy $dE|_{T,p,\beta}$ is given by Eq. 7.113. Because of the fixed composition, all the partial energies e_l retain the same values at all steps. Consequently the total energy E for building up the multicomponent system is given by the sum of all the dE's, namely, by the relation

$$E = \sum_{l}^{r} e_l n_l \tag{7.115}$$

Equation 7.115 is obtained for given values of T, p, and β and for a given composition. On the other hand, since the dependence of e_l on these values is implicit, the equation applies to any set of values T, p, and β and to any composition. More generally, given any extensive property F of a homogeneous multicomponent system and the corresponding partial properties f_l (Eq. 7.112), using the preceding arguments results in the relation

$$F = \sum_{l}^{r} f_l n_l \tag{7.116}$$

Apart from their usefulness for calculating values of extensive properties, Eqs. 7.115 and 7.116 yield other general results. For example, if E is regarded as a function of the two sets of quantities $e_1, e_2, \ldots, e_r$ and $n_1, n_2, \ldots, n_r$, then for any arbitrary differential changes in these quantities between stable equilibrium states, the corresponding energy change dE is given by the relation

$$dE = \sum_{l}^{r} e_l \, dn_l + \sum_{l}^{r} n_l \, de_l \tag{7.117}$$

If the arbitrary changes are introduced under conditions of constant temperature, pressure, and value of β, then comparing Eqs. 7.113 and 7.117 yields

$$n_1 \, de_1 + n_2 \, de_2 + \ldots + n_r \, de_r = 0 \qquad \text{for fixed } T, p, \text{ and } \beta \tag{7.118}$$

The analogous relation for any extensive property F is given by the expression

$$n_1 \, df_1 + n_2 \, df_2 + \ldots + n_r \, df_r = 0 \qquad \text{for fixed } T, p, \text{ and } \beta \tag{7.119}$$

Useful relations can also be established between intensive properties, such as total potentials, and partial properties. These relations are discussed later. Presently it suffices to indicate that all results regarding partial properties and their relations to other properties are applicable as well to one-component homogeneous systems. For such systems, however, one talks about *specific* rather than partial properties. For example. the extensive properties E, S, and V are given by the relations

$$E = en \tag{7.120}$$

$$S = sn \tag{7.121}$$

and

$$V = vn \tag{7.122}$$

where e = *specific energy*
s = *specific entropy*
v = *specific volume*
n = number of either particles or moles of the one-component system

The word "specific" refers to the value of a property either per particle or per mole. Values of specific properties can be measured or computed and tabulated once and for all. Steam tables are an example of such tabulated values.

7.6.8 Partial Gibbs Free Energy. An important equation of stable equilibrium states is that for the Gibbs free energy Z, given by the relation

$$Z = E + pV - TS \tag{7.123}$$

This energy can be expressed as a function of the independent variables T, p, β, and $n_1, n_2, \ldots, n_r$ so that

$$Z = Z(T;p;\beta;n_1,n_2, \ldots, n_r) \tag{7.124}$$

According to the preceding section, Z must also be given by a relation of the form

$$Z = \sum_l^r \zeta_l n_l \tag{7.125}$$

where ζ_l is the *partial Gibbs free energy of component l* and is defined by the relation

$$\zeta_l = \left(\frac{\partial Z}{\partial n_l}\right)_{T,p,\beta,n} \tag{7.126}$$

The partial Gibbs free energy ζ_l equals the total potential μ_l. Indeed, differentiating Eq. 7.123 and using the Gibbs equation (Eq. 7.94) yield

$$dZ = -S\,dT + V\,dp + \left(\frac{\partial E}{\partial \beta}\right)_{S,V,n} d\beta + \sum_l^r \mu_l\,dn_l \tag{7.127}$$

On the other hand, differentiating Eq. 7.124 and using Eq. 7.126 yield

$$dZ = \left(\frac{\partial Z}{\partial T}\right)_{p,\beta,n} dT + \left(\frac{\partial Z}{\partial p}\right)_{T,\beta,n} dp + \left(\frac{\partial Z}{\partial \beta}\right)_{T,p,n} d\beta + \sum_{l}^{r} \zeta_l \, dn_l \tag{7.128}$$

When Eqs. 7.127 and 7.128 are compared, the following relations emerge:

$$-S = \left(\frac{\partial Z}{\partial T}\right)_{p,\beta,n} \tag{7.129}$$

$$V = \left(\frac{\partial Z}{\partial p}\right)_{T,\beta,n} \tag{7.130}$$

$$\left(\frac{\partial E}{\partial \beta}\right)_{S,V,n} = \left(\frac{\partial Z}{\partial \beta}\right)_{T,p,n} \tag{7.131}$$

and

$$\mu_l = \zeta_l \qquad \text{for all } l \tag{7.132}$$

By virtue of the equality between ζ_l and μ_l, the quantities Z and E are given by the relations

$$Z = \sum_{l} \mu_l n_l \tag{7.133}$$

and

$$E = TS - pV + \sum_{l} \mu_l n_l \tag{7.134}$$

Equation 7.134 is helpful in establishing the relation between the partition function Q_g of an open system and the readily observable properties p, V, and T. This relation is discussed in Sec. 7.6.9.

7.6.9 The Partition Function. The distribution x_m^0 for an open system is given by the relation (Eq. 6.43)

$$x_m^0 = \frac{\exp\left[-(E_m - \sum_{l=1}^{r} \mu_l n_{lm})/kT\right]}{Q_g} \tag{7.135}$$

where the partition function Q_g is given by the relation (Eq. 6.44)

$$Q_g = \sum_m \exp\left(-\frac{E_m - \sum_{l=1}^{r} \mu_l n_{lm}}{kT}\right) \tag{7.136}$$

Substituting the distribution x_m^0 into the expression for the entropy S (Eq. 6.35) yields

$$S = -k \sum_m x_m^0 \ln x_m^0 = k \ln Q_g + \frac{E}{T} - \frac{\sum_{l=1}^{r} \mu_l n_l}{T} \tag{7.137}$$

or

$$E = TS - kT \ln Q_g + \sum_{l=1}^{r} \mu_l n_l \tag{7.138}$$

Comparing Eqs. 7.134 and 7.138 results in the important relation

$$\ln Q_g = \frac{pV}{kT} \tag{7.139}$$

which applies to gaseous, liquid, and solid phases.

7.6.10 The Gibbs–Duhem Equation. Equations 7.127 and 7.133 indicate that dZ can be written in the two forms

$$dZ = -S\,dT + V\,dp + \left(\sum_{l=1}^{r} q_l n_l\right) d\psi + \sum_{l=1}^{r} \mu_l\,dn_l \tag{7.140}$$

and

$$dZ = \sum_{l=1} \mu_l\,dn_l + \sum_{l=1}^{r} n_l\,d\mu_l \tag{7.141}$$

In Eq. 7.140 use has been made of Eq. 7.99. Comparing Eqs. 7.140 and 7.141 yields

$$S\,dT - V\,dp + \sum_{l=1}^{r} n_l\,d(\mu_l - q_l\psi) = 0 \tag{7.142a}$$

or

$$S\,dT - V\,dp + \sum_{l=1}^{r} n_l\,d\mu_{cl} = 0 \tag{7.142b}$$

since the difference $\mu_l - q_l\psi$ equals the chemical potential μ_{cl} of component l. Relation 7.142b is known as the *Gibbs–Duhem equation.* It is useful in studies of mixtures of gases.

7.7 On Total Potentials

7.7.1 General Remarks. Total potentials represent a set of important properties of stable equilibrium states. Their values must be considered in problems involving questions of mutual stable equilibrium, chemical equilibrium, calculation of partial and total Gibbs free energies, and electronegativity. We will see later that the total potential of electrons in a metal is related to the work function of the surface of the metal. For these and many other uses, it is helpful to discuss total potentials in some detail and to describe their relations to other properties.

7.7.2 Total Potentials and Partial Properties. The total potential μ_l equals the partial Gibbs free energy ζ_l (Eq. 7.132). It can also be expressed in terms of other partial properties. For example, substituting Eq. 7.123 into Eq. 7.126 yields

$$\mu_l = \zeta_l = \left(\frac{\partial(E + pV - TS)}{\partial n_l}\right)_{T,p,\beta,n} = e_l + pv_l - Ts_l \quad \text{for all } l \tag{7.143}$$

It is seen from Eq. 7.143 that the total potential μ_l of component l differs from the partial energy e_l of the same component by two terms: (1) the product of the total pressure p and the partial volume v_l of component l and (2) the product of the temperature T and the partial entropy s_l of component l.

The combination $E + pV$ appears frequently in thermodynamic analyses. It is the enthalpy H of a system,

$$H \equiv E + pV \tag{7.144}$$

The total potential can be expressed in terms of the enthalpy as follows:

$$\mu_l = \zeta_l = \left(\frac{\partial(H - TS)}{\partial n_l}\right)_{T,p,\beta,n} = h_l - Ts_l \quad \text{for all } l \tag{7.145}$$

where

$$h_l = \left(\frac{\partial H}{\partial n_l}\right)_{T,p,\beta,n} \tag{7.146}$$

The quantity h_l is called the *partial enthalpy of component l*.

Finally, since the order of differentiation is immaterial, Eqs. 7.129, 7.130, and 7.131 can be partially differentiated with respect to n_l and the results equated to the partial derivative of Eq. 7.132 with respect to T, p, and β, respectively. Thus the following relations are established:

$$\left(\frac{\partial \mu_l}{\partial T}\right)_{p,\beta,n} = -\left(\frac{\partial S}{\partial n_l}\right)_{T,p,\beta,n} = -s_l \tag{7.147}$$

$$\left(\frac{\partial \mu_l}{\partial p}\right)_{T,\beta,n} = \left(\frac{\partial V}{\partial n_l}\right)_{T,p,\beta,n} = v_l \tag{7.148}$$

and

$$\left(\frac{\partial \mu_l}{\partial \beta}\right)_{T,p,n} = q_l \quad \text{for } \beta \text{ and applied electrostatic field } \psi \tag{7.149}$$

Combining Eqs. 7.145 and 7.147 yields

$$\left[\frac{\partial(\mu_l/T)}{\partial T}\right]_{p,\beta,n} = -\frac{h_l}{T^2} \quad \text{for all } l \tag{7.150}$$

7.7.3 **Total Potentials in Terms of Specific Properties.** The total potential μ_l of a multicomponent system A can also be expressed in terms of specific properties of a one-component system B consisting of the component l only. System A can be in mutual stable equilibrium with system B via a semipermeable membrane as shown schematically in Fig. 7.5. The semipermeable membrane is a device that permits only component l to flow freely in either direction between A and B. No other component can flow through the membrane. For mutual stable equilibrium the total potentials of component l in A and B must be equal, i.e.,

$$\mu_l = \mu_{ll} \tag{7.151}$$

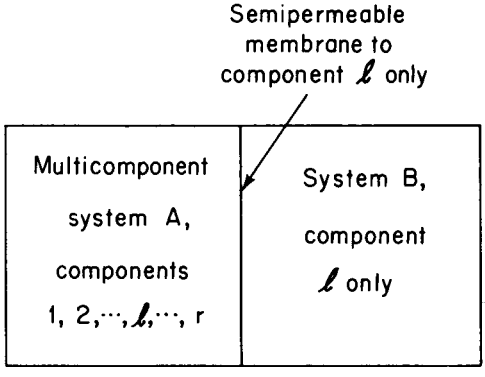


Fig. 7.5 Schematic of multi- and one-component systems in mutual stable equilibrium.

where μ_{II} is the total potential of component l in the one-component system B. The quantity μ_{II} is also equal to the specific Gibbs free energy ζ_{II} of B, and therefore

$$\mu_l = \zeta_{II} \tag{7.152}$$

On the other hand, ζ_{II} satisfies an equation of the form of Eq. 7.143, and consequently

$$\mu_{II} = \zeta_{II} = e_{II} + p_{II}v_{II} - Ts_{II} = h_{II} - Ts_{II} \tag{7.153}$$

where e_{II} = the specific energy
v_{II} = the specific volume
s_{II} = the specific entropy
p_{II} = the pressure
h_{II} = the specific enthalpy of the one-component system B
T = the common temperature of A and B

These relations are useful because specific properties are easier to compute or measure than are partial properties.

7.7.4 Total Potentials in Terms of Work. An interesting interpretation of μ_l is given by the following statement. If an infinitesimal amount of a component l is taken from a state of zero energy and entropy and introduced adiabatically into an open system A held at fixed values of all constraints and of all n_i's except n_l, the total potential μ_l of component l in A is the negative of the reversible work of the interaction divided by the amount of component l introduced. The reversible work is the maximum work delivered (or the minimum work received) by all things involved in the interaction. The proof of this statement is straightforward.[3]

7.7.5 Qualitative Interpretation. Section 7.4.3 states that, for mutual stable equilibrium with respect to transfer of a component between A and B, the value of the total potential of the component in A must be equal to that in B. This equality of values suggests that the total potential $(\mu_l)_A$ should be interpreted as a measure of either the tendency to escape from or the tendency to be attracted (captured) by A. Actually, $(\mu_l)_A$ is a measure of the tendency to escape. In a possible interaction between A and B such that the energy of the two systems combined remains fixed and only dn_l particles are transferred from A to B, it can be shown with the help of Eq. 7.94 that

$$dE_A + dE_B = T(dS_A + dS_B) + [(\mu_l)_A - (\mu_l)_B]\, dn_l = 0 \tag{7.154}$$

By virtue of the principle of increase of entropy,

$$dS_A + dS_B \geqslant 0 \tag{7.155}$$

and, therefore, Eq. 7.154 becomes

$$[(\mu_l)_A - (\mu_l)_B]\, dn_l = -T(dS_A + dS_B) \leqslant 0 \tag{7.156}$$

It is seen from relation 7.156 that if

$$(\mu_l)_A > (\mu_l)_B \tag{7.157}$$

then dn_l must be negative, namely, component l has a tendency to move from the system with the higher total potential, $(\mu_l)_A$, to the system with the lower total potential, $(\mu_l)_B$. Consequently μ_l may be thought of as an escaping tendency.

7.8 One-Particle Approximations

7.8.1 General Remarks. Up to this point our discussion has not been concerned either with the calculation of energy eigenstates of a system or with the explicit analytical manipulations of equations of stable equilibrium states. Given an amount of matter and its constraints, we have been assuming that the exact energy eigenstates can be computed and that explicit equations of stable equilibrium states in any form can be established.

Except for a very limited number of systems, however, both the calculation of exact energy eigenstates and the explicit manipulation of equations of stable equilibrium states are very difficult mathematical tasks. Results are found only after drastic approximations are introduced. For example, as stated in Chap. 5, energy eigenstates of a given system are found only after the Born–Oppenheimer, the one-electron, and other approximations are introduced.

This section outlines some of the approximate methods used in quantum-thermodynamic analyses of many-particle homogeneous systems and indicates the ranges of validity of the results. The most widely used approximations are the *one-particle approximation* for the energy eigenstates and various *gas approximations* for equations of stable equilibrium states.

7.8.2 The One-Particle Approximation for One-Component Open Systems. We will consider an open system A which consists of identical particles that interact with either very short- or very long-range forces, or both and which is in a stable equilibrium state such that the expectation value n of the number of particles is

much smaller than the maximum eigenvalue n_{max} of the possible eigenvalues n_m. Such an open system will be called a one-component *dilute gas.*

Approximate equations of stable equilibrium states for dilute gases can be computed in this way. For a given eigenvalue n_i of the number of particles, suppose that the following approximations are introduced into the calculation of the potential energy of the Hamiltonian operator: (1) short-range interparticle forces are neglected and (2) the long-range force exerted on each particle is represented by a function of the coordinates of the particular particle and not of the relative distance of that particle and others. It is evident that the approximations for both short- and long-range forces are better the smaller the value of n_i. Nevertheless, they are used for all values of n_i of a dilute gas.

As a result of approximations 1 and 2, the Hamiltonian operator corresponding to n_i can be expressed as a sum of Hamiltonian operators each associated with one particle; i.e., the Hamiltonian operator can be separated into additive parts, each of which depends on the coordinates of one particle only. Each of the separated Hamiltonian operators defines a one-particle energy eigenvalue problem. The one-particle energy eigenstates can be combined to approximate the energy eigenstates of the dilute gas and then to compute equations of stable equilibrium states. When this is done the analysis is said to be approached by means of the method of the *one-particle approximation.*

The one-particle approximation is appropriate for dilute gases in stable equilibrium states because the properties of such gases depend primarily on the energy eigenstates corresponding to small eigenvalues n_i. For example, the value n is given by the relation

$$n = \sum_i x_i^0 n_i \tag{7.158}$$

where x_i^0 is the grand canonical distribution. Since n is much smaller than n_{max} and all x_i^0 are positive, it follows that the dependence of n on eigenvalues $n_i > n$ must be weak.

In terms of the one-particle approximation, the relation between an energy eigenvalue E_i of the dilute gas and the one-particle energy eigenvalues ϵ_j can be stated as follows. In the energy eigenstate G_i of energy E_i and number of particles n_i, the gas can be thought of as made up of ν_{i1} particles with one-particle energy ϵ_1, ν_{i2} particles with one-particle energy ϵ_2, . . ., so that E_i is given by the relation

$$E_i = \nu_{i1}\epsilon_1 + \nu_{i2}\epsilon_2 + \ldots + \nu_{ij}\epsilon_j + \ldots = \sum_j \nu_{ij}\epsilon_j \quad \text{for all } i \tag{7.159}$$

where

$$\sum_j \nu_{ij} = n_i \quad \text{for all } i \tag{7.160}$$

If the stable equilibrium state of the dilute gas is characterized by a temperature T and a total potential μ, the probability x_i^0 of finding the energy E_i as a result of a measurement is given by the relation

$$x_i^0 = \frac{\exp\left[(\mu n_i - E_i)/kT\right]}{Q_g} = \frac{\exp\left\{\sum_j \nu_{ij}\left[(\mu - \epsilon_j)/kT\right]\right\}}{Q_g} \tag{7.161}$$

where the grand partition function Q_g is given by the expression

$$Q_g = \sum_i \exp\left[\sum_j \nu_{ij} \frac{\mu - \epsilon_j}{kT}\right] \tag{7.162}$$

The *expectation value* ν_j of the number of particles in the one-particle energy eigenstate of energy ϵ_j, called the *one-particle distribution function,* is given by the relation

$$\nu_j = \sum_i x_i^0 \nu_{ij} \tag{7.163}$$

The expectation value n (Eq. 7.158) can be written in the form

$$n = \sum_i x_i^0 n_i = \sum_i x_i^0 \sum_j \nu_{ij} = \sum_j \sum_i x_i^0 \nu_{ij} = \sum_j \nu_j \tag{7.164}$$

where use has been made of Eqs. 7.160 and 7.163.

As discussed in Chap. 5, depending on the nature of a collection of identical particles of a system, the acceptable energy eigenfunctions can be either antisymmetric or symmetric. Antisymmetric energy eigenfunctions are acceptable for a collection of particles called fermions, and symmetric energy eigenfunctions are acceptable for a collection of particles called bosons. The distinction between fermions and bosons can also be stated in terms of the admissible values of the occupation numbers ν_{ij}.

1. *Fermions.* The one-particle energy eigenstates can be occupied by at most one fermion, i.e., the admissible values of ν_{ij} are

$$\nu_{ij} = 0 \text{ or } 1 \qquad \text{for all } i \text{ and } j \tag{7.165}$$

For example, a spin-orbital of an atom cannot be occupied by more than one electron.

2. *Bosons.* The one-particle energy eigenstates can be occupied by any number of bosons, i.e., the admissible values of ν_{ij} are

$$\nu_{ij} = 0, 1, 2, \ldots \qquad \text{for all } i \text{ and } j \tag{7.166}$$

For example, many photons may occupy the same one-particle energy eigenstate.

Explicit use of the admissible values of ν_{ij} results in simplified forms for Q_g (Eq. 7.162) and ν_j (Eq. 7.163), as discussed in Secs. 7.8.3 and 7.8.4.

7.8.3 The Partition Function Q_g. If D_j is given by the relation

$$D_j \equiv \exp\left(\frac{\mu - \epsilon_j}{kT}\right) \quad \text{for all } j \tag{7.167}$$

and $b_1, b_2, \ldots$ are the admissible values for any of the numbers ν_{ij}, then it can easily be verified that Eq. 7.162 can be written in the forms

$$\begin{aligned} Q_g &= \sum_i D_1^{\nu_{i1}} D_2^{\nu_{i2}} \ldots D_j^{\nu_{ij}} \ldots \\ &= (D_1^{b_1} + D_1^{b_2} + \ldots)(D_2^{b_1} + D_2^{b_2} + \ldots) \ldots (D_j^{b_1} + D_j^{b_2} + \ldots) \ldots \end{aligned} \tag{7.168}$$

If the dilute gas is a Fermi–Dirac gas (i.e., if it consists of a collection of fermions), the admissible values of ν_{ij} are (Eq. 7.165) $b_1 = 0$, $b_2 = 1$, and $b_i = 0$ for $i \geq 3$. Accordingly Eq. 7.168 becomes

$$\begin{aligned} Q_g &= (1 + D_1)(1 + D_2) \ldots (1 + D_j) \ldots \\ &= \prod_j \left[1 + \exp\left(\frac{\mu - \epsilon_j}{kT}\right)\right] \end{aligned} \tag{7.169}$$

or

$$\ln Q_g = \sum_j \ln \left[1 + \exp\left(\frac{\mu - \epsilon_j}{kT}\right)\right] \tag{7.170}$$

On the other hand, if the dilute gas is a Bose–Einstein gas (i.e., if it consists of a collection of bosons), the admissible values of ν_{ij} are (Eq. 7.166) $b_1 = 0$, $b_2 = 1$, and $b_3 = 2, \ldots$. Accordingly Eq. 7.168 becomes

$$\begin{aligned} Q_g &= (1 + D_1 + D_1^2 + \ldots)(1 + D_2 + D_2^2 + \ldots) \ldots (1 + D_j + D_j^2 \ldots) \ldots \\ &= [(1 - D_1)(1 - D_2) \ldots (1 - D_j) \ldots]^{-1} \\ &= \prod_j \left[1 - \exp\left(\frac{\mu - \epsilon_j}{kT}\right)\right]^{-1} \end{aligned} \tag{7.171}$$

or

$$\ln Q_g = \sum_j \ln\left[1 - \exp\left(\frac{\mu - \epsilon_j}{kT}\right)\right]^{-1} \tag{7.172}$$

7.8.4 The Expectation Value or One-Particle Distribution ν_j. In terms of the notation introduced in the preceding section, we can readily verify that the one-particle distribution function ν_j is given by the relations

$$\nu_j = \sum_i x_i^0 \nu_{ij} = \frac{b_1 D_j^{b_1} + b_2 D_j^{b_2} + \ldots}{D_j^{b_1} + D_j^{b_2} + \ldots} \quad \text{for all } j \tag{7.173}$$

For a Fermi–Dirac gas, Eq. 7.173 becomes

$$\nu_j = \frac{D_j}{1 + D_j} = \frac{1}{\exp\left[(\epsilon_j - \mu)/kT\right] + 1} \quad \text{for all } j \tag{7.174}$$

This one-particle distribution function is called the *Fermi–Dirac factor.*

For a Bose–Einstein gas, Eq. 7.173 becomes

$$\nu_j = \frac{D_j + 2D_j^2 + 3D_j^3 + \ldots}{1 + D_j + D_j^2 + D_j^3 + \ldots} = \frac{D_j/(1 - D_j)^2}{1/(1 - D_j)}$$

$$= \frac{1}{\exp\left[(\epsilon_j - \mu)/kT\right] - 1} \quad \text{for all } j \tag{7.175}$$

This one-particle distribution function is called the *Bose–Einstein factor.*

7.8.5 The p–V–T–β–μ Relation for a Dilute Gas. For a homogeneous, one-component dilute gas with volume V and an applied field β as two independent constraints, a p–V–T–β–μ relation is found by substituting the grand partition function Q_g (Eq. 7.170 or 7.172) into Eq. 7.139. Thus

$$\frac{pV}{kT} = \ln Q_g = \sum_j \ln\left[1 \pm \exp\left(\frac{\mu - \epsilon_j}{kT}\right)\right]^{\pm 1} \tag{7.176}$$

where the plus sign in $\pm$ applies for Fermi–Dirac gases and the minus sign applies for Bose–Einstein gases. The one-particle energy eigenvalues ϵ_j are functions of both V and the applied field β. For this reason Eq. 7.176 is called the *p–V–T–β–μ relation for the dilute gas.* This relation is not a fundamental equation.

7.8.6 A Fundamental Equation for a One-Component Dilute Gas. For the gas discussed in the preceding section, a fundamental equation is given by the Gibbs free energy Z expressed as a function of n, p,T, and β, i.e.,

$$Z = n\mu = Z(n,p,T,\beta) \tag{7.177}$$

It can be derived as follows:

Substituting the distribution ν_j (Eq. 7.174 or 7.175) into Eq. 7.164 yields

$$n = \sum_j \left[\exp\left(\frac{\epsilon_j - \mu}{kT}\right) \pm 1\right]^{-1} \tag{7.178}$$

where the plus sign in $\pm$ holds for Fermi–Dirac gases and the minus sign holds for Bose–Einstein gases. It is clear that Eq. 7.178 is of the form

$$n = n(V,T,\beta,\mu) \tag{7.179}$$

since the one-particle energy eigenvalues ϵ_j are functions of V and β only. On the other hand, Eq. 7.176 is of the form

$$p = p(V,T,\beta,\mu) \tag{7.180}$$

In principle, the volume V can be eliminated from Eqs. 7.179 and 7.180 to yield a relation of the form of Eq. 7.177.

Another fundamental equation is readily derivable from Eq. 7.177 by partial differentiation with respect to n. The differentiation yields the total potential μ as a function of p, T, β, and, in general, n. This potential is important for the analysis of many processes encountered in thermionic converters.

Unfortunately, in spite of the one-particle approximation, the computational procedure for evaluating $\mu(p,T,\beta,n)$ is extremely difficult to carry out. The primary source of the difficulty is the complexity of the calculation of the one-particle energy eigenvalues ϵ_j. For an arbitrary one-particle potential energy, eigenvalues ϵ_j can be found only after tedious computer-aided analyses. Moreover, even after these eigenvalues are computed, a general analytical expression for μ cannot be found because Eqs. 7.179 and 7.180 are transcendental.

The difficulties can be bypassed by introducing further approximations into the analysis. These approximations are (1) the *Boltzmann approximations*, (2) the *constant-potential approximation*, (3) the *point-particle approximation*, and (4) the *continuous eigenvalue-spectrum approximation*. When approximations 2, 3, and 4 apply, the dilute gas is called a *Sommerfeld gas*. When approximations 1, 2, and 4 apply, the dilute gas is called a *semiperfect gas*.

Finally, when all four approximations apply, the dilute gas is called a *perfect gas.* These approximations and the ranges of their validity are discussed in more detail in the sections that follow.

7.8.7 The Boltzmann Approximations. The analysis of dilute gases is simplified when restricted to stable equilibrium states for which

$$\frac{\mu - \epsilon_0}{kT} \ll 0 \tag{7.181}$$

where ϵ_0 is the smallest one-particle energy eigenvalue.

For such states, Eqs. 7.176 and 7.178 can be written in the approximate forms

$$\frac{pV}{kT} = \exp\left(\frac{\mu}{kT}\right) \sum_j \exp\left(-\frac{\epsilon_j}{kT}\right) = q \exp\left(\frac{\mu}{kT}\right) \tag{7.182}$$

and

$$n = \exp\left(\frac{\mu}{kT}\right) \sum_j \exp\left(-\frac{\epsilon_j}{kT}\right) = q \exp\left(\frac{\mu}{kT}\right) \tag{7.183a}$$

or

$$\mu = kT \ln \frac{n}{q} \tag{7.183b}$$

where

$$q = \sum_j \exp\left(-\frac{\epsilon_j}{kT}\right) \tag{7.184}$$

The approximate forms 7.182 and 7.183 are known as the *Boltzmann approximations.* The quantity q defined by Eq. 7.184 is known as the *one-particle partition function.*

Eliminating the one-particle partition function from Eqs. 7.182 and 7.183a yields the p–V–T–n relation

$$pV = nkT \tag{7.185}$$

Equation 7.185 is applicable to all stable equilibrium states for which condition 7.181 is satisfied. It is independent of the nature of the particles of the dilute gas and the values of the one-particle energy eigenvalues ϵ_j.

7.8.8 The Constant-Potential Approximation. In the *constant-potential approximation,* the one-particle energy eigenvalues are found as sums of two parts. One part is computed as though the externally applied and the long-range forces on each particle did not exist. The other part is assumed to be equal to the average potential energy per particle due to the externally applied and the long-range forces. Thus the energy of the overall system equals that corresponding to the first part of the one-particle energy eigenvalues plus a potential energy equal to the expectation value n of the number of particles times the average potential energy per particle.

This method of evaluating the one-particle energy eigenvalues is justified when the potential energy per particle as a function of position differs only slightly from the average potential energy per particle.

By virtue of this approximation, hereafter the total potential μ encountered in the general relations for dilute gases is replaced by the chemical potential μ_c (Sec. 7.6.5), and the one-particle energy eigenvalues are replaced by the values computed in the absence of externally applied and long-range forces.

In the absence of external and long-range forces, the one-particle Hamiltonian operator $\hat{H}_{\text{part}}$ of a dilute gas can be separated into two additive parts $\hat{H}_{\text{in}}$ and $\hat{H}_{\text{tr}}$. The first part, $\hat{H}_{\text{in}}$, corresponds to the internal degrees of freedom associated with the internal structure of the particle, and $\hat{H}_{\text{tr}}$ corresponds to the translational degrees of freedom associated with the motion of the center of mass of the particle. Thus

$$\hat{H}_{\text{part}} = \hat{H}_{\text{in}} + \hat{H}_{\text{tr}} \tag{7.186}$$

An example of such a separation is given in Sec. 5.7.1.

Equation 7.186 implies that each one-particle energy eigenvalue ϵ_j can be expressed as a sum of an energy eigenvalue ε_s of the operator $\hat{H}_{\text{in}}$ and an energy eigenvalue ϵ'_r of the operator $\hat{H}_{\text{tr}}$. Thus

$$\epsilon_j = \varepsilon_s + \epsilon'_r \tag{7.187}$$

The operator $\hat{H}_{\text{tr}}$ can be further subdivided into three independent parts, $\hat{H}_x$, $\hat{H}_y$, and $\hat{H}_z$, associated with the x, y, and z Cartesian coordinates of the three-dimensional gas. Accordingly each one-particle energy eigenvalue ϵ_j can be written as

$$\epsilon_j = \varepsilon_s + \epsilon'_x + \epsilon'_y + \epsilon'_z \tag{7.188}$$

where ϵ'_x, ϵ'_y, and ϵ'_z are energy eigenvalues of $\hat{H}_x$, $\hat{H}_y$, and $\hat{H}_z$, respectively. For a dilute gas in a rectilinear container of dimensions l_x, l_y, and l_z, the energies ϵ'_x, ϵ'_y, and ϵ'_z are given by the relations

$$\epsilon'_x = \frac{h^2}{8ml_x^2} k_x^2 \qquad \epsilon'_y = \frac{h^2}{8ml_y^2} k_y^2 \qquad \epsilon'_z = \frac{h^2}{8ml_z^2} k_z^2 \tag{7.189}$$

where h is Planck's constant, m is the mass of the particle, and k_x, k_y, and k_z are integers (see Sec. 5.6).

By virtue of Eqs. 7.188 and 7.189 and in the absence of fields, relations 7.176 and 7.178 become

$$\frac{pV}{kT} = \sum_{s,k_x,k_y,k_z} \ln\left\{1 \pm \exp\left(\frac{\mu_c}{kT}\right)\exp\left(-\frac{\varepsilon_s}{kT}\right) \times \exp\left[-\frac{h^2}{8mkT}\left(\frac{k_x^2}{l_x^2} + \frac{k_y^2}{l_y^2} + \frac{k_z^2}{l_z^2}\right)\right]\right\}^{\pm 1} \tag{7.190}$$

and

$$n = \sum_{s,k_x,k_y,k_z} \left\{\exp\left(-\frac{\mu_c}{kT}\right)\exp\left(\frac{\varepsilon_s}{kT}\right) \times \exp\left[\frac{h^2}{8mkT}\left(\frac{k_x^2}{l_x^2} + \frac{k_y^2}{l_y^2} + \frac{k_z^2}{l_z^2}\right)\right] \pm 1\right\}^{-1} \tag{7.191}$$

In general, the multiple summations in Eqs. 7.190 and 7.191 cannot be carried out. To proceed, we need further approximations.

7.8.9 The Point-Particle and Continuous Eigenvalue-Spectrum Approximations. The *point-particle approximation* consists in replacing the energy spectrum of the internal structure of the particle by a g-fold degenerate value $\bar{\varepsilon}_o$. Thus the summation over s in Eqs. 7.190 and 7.191 is replaced by the degeneracy factor g of $\bar{\varepsilon}_o$.

The quantity $\bar{\varepsilon}_o$ can be regarded as the energy of the particle when its kinetic energy equals zero. For a one-component gas, $\bar{\varepsilon}_o$ may be assigned any arbitrary value, including zero, since the zero-energy reference level of any system is arbitrary. On the other hand, for a multicomponent system, such as a gas consisting of electrons, ions, and neutral atoms, the values of $\bar{\varepsilon}_o$ for the various components are interrelated, and not all of them can be assigned arbitrary magnitudes.

The *continuous eigenvalue-spectrum approximation* consists in replacing the summations over k in Eqs. 7.190 and 7.191 by integrals. This approximation is justified when

$$\frac{h^2}{8mkTl^2} \ll 1 \tag{7.192}$$

where l stands for l_x, l_y, or l_z. It is applicable when T is high or when the dimensions of the system are large, or both.

7.8.10 The Sommerfeld Gas. The *Sommerfeld gas* is defined as a dilute gas for which the constant-potential, the point-particle, and the continuous eigenvalue-spectrum approximations are applicable. Then Eqs. 7.190 and 7.191 become

$$\frac{pV}{kT} = g \int_0^\infty \int_0^\infty \int_0^\infty \ln \left\{ 1 \pm \exp\left(\frac{\mu_c - \bar{\varepsilon}_o}{kT} \right) \times \exp\left[-\frac{h^2}{8mkT} \left(\frac{x^2}{l_x^2} + \frac{y^2}{l_y^2} + \frac{z^2}{l_z^2} \right) \right] \right\}^{\pm 1} dx\, dy\, dz \tag{7.193}$$

and

$$n = g \int_0^\infty \int_0^\infty \int_0^\infty \left\{ \exp\left(\frac{\mu_c - \bar{\varepsilon}_o}{kT} \right) \times \exp\left[\frac{h^2}{8mkT} \left(\frac{x^2}{l_x^2} + \frac{y^2}{l_y^2} + \frac{z^2}{l_z^2} \right) \right] \pm 1 \right\}^{-1} dx\, dy\, dz \tag{7.194}$$

Moreover, if we introduce the variable Y, given by the relation

$$Y^2 \equiv \frac{h^2}{8kTm} \left(\frac{x^2}{l_x^2} + \frac{y^2}{l_y^2} + \frac{z^2}{l_z^2} \right)$$

Eqs. 7.193 and 7.194 can be written as

$$\frac{pV}{kT} = \frac{\pi g V}{2} \left(\frac{8mkT}{h^2} \right)^{3/2} \int_0^\infty \ln \left[1 \pm \exp\left(\frac{\mu_c - \bar{\varepsilon}_o}{kT} - Y^2 \right) \right]^{\pm 1} Y^2\, dY \tag{7.195}$$

and

$$n = \frac{\pi g V}{2} \left(\frac{8mkT}{h^2} \right)^{3/2} \int_0^\infty \left[\exp\left(Y^2 - \frac{\mu_c - \bar{\varepsilon}_o}{kT} \right) \pm 1 \right]^{-1} Y^2\, dY \tag{7.196}$$

where V denotes the volume $l_x l_y l_z$. The integrations in Eqs. 7.195 and 7.196 cannot be carried out without further approximations.

For fermions and relatively large values of μ_c so that

$$\frac{\mu_c - \bar{\varepsilon}_o}{kT} \gg 0 \tag{7.197}$$

the integral in Eq. 7.196 can be evaluated by means of a power series expansion of the integrand. Thus μ_c is given by the approximate relation

$$\mu_c - \bar{\varepsilon}_o = \Delta\epsilon^* - \frac{\pi kT}{12\Delta\epsilon^*} \tag{7.198}$$

where

$$\Delta\epsilon^* = \frac{h^2}{8m}\left(\frac{6n}{\pi g V}\right)^{2/3} \tag{7.199}$$

When the volume V is a unique function of temperature, such as the volume of a freely expanding piece of metal, $\Delta\epsilon^*$ is given by the approximate relation

$$\Delta\epsilon^* = \Delta\epsilon_0^* - \frac{2}{3}\Delta\epsilon_0^* \int_0^T \beta\, dT \tag{7.200}$$

where

$$\Delta\epsilon_0^* = \frac{h^2}{8m}\left(\frac{6n}{\pi g V_0}\right)^{2/3} \tag{7.201}$$

V_0 is the volume in the limit of 0°K, and β is the volume expansion coefficient $[\beta = (1/V)(dV/dT)]$.

Using condition 7.197 and the power series expansion technique in Eq. 7.195 yields

$$p = \frac{8\pi g (2m)^{3/2}}{15h^3}(\Delta\epsilon^*)^{5/2}\left[1 - \frac{5}{24}\left(\frac{\pi kT}{\Delta\epsilon^*}\right)^2\right]\left[1 + \frac{5}{8}\left(\frac{\pi kT}{\Delta\epsilon^*}\right)^2\right] \tag{7.202}$$

Equations 7.198 and 7.202 are used to evaluate the chemical potential and the pressure of electrons in the bulk of a metal.

The integral in Eq. 7.196 can also be evaluated for relatively small values of μ_c. The results are applicable to perfect gases, which are discussed in Sec. 7.8.13.

7.8.11 The Semiperfect Gas. If the one-particle energy eigenvalues of a dilute gas are described by the constant-potential approximation, condition 7.181 becomes

$$\frac{\mu_c - \varepsilon_0}{kT} \ll 0 \tag{7.203}$$

where ε_0 denotes the smallest energy eigenvalue of the internal structure of the particle. Then Eq. 7.196 can be approximated by the relation

$$n = \sum_{s,k_x,k_y,k_z} \left\{ \exp\left(-\frac{\mu_c}{kT}\right) \exp\left(\frac{\varepsilon_s}{kT}\right) \times \exp\left[\frac{h^2}{8mkT}\left(\frac{k_x^2}{l_x^2} + \frac{k_y^2}{l_y^2} + \frac{k_z^2}{l_z^2}\right)\right] \right\}^{-1} \tag{7.204a}$$

or, equivalently,

$$\mu_c = kT \ln \frac{n}{q_s} - kT \ln \sum_{k_x,k_y,k_z} \exp\left[-\frac{h^2}{8mkT}\left(\frac{k_x^2}{l_x^2} + \frac{k_y^2}{l_y^2} + \frac{k_z^2}{l_z^2}\right)\right] \tag{7.204b}$$

where

$$q_s = \sum_{s=0} \exp\left(-\frac{\varepsilon_s}{kT}\right) \tag{7.205}$$

For reasonably high temperatures and reasonably large values of l_x, l_y, and l_z, condition 7.192 is satisfied. and the summations in Eq. 7.204b can be replaced by integrals. Thus

$$\begin{aligned}\mu_c &= kT \ln \frac{n}{q_s} - kT \ln \int_0^\infty \int_0^\infty \int_0^\infty \exp\left[-\frac{h^2}{8mkT}\left(\frac{x^2}{l_x^2} + \frac{y^2}{l_y^2} + \frac{z^2}{l_z^2}\right)\right] dx\, dy\, dz \\ &= -kT \ln \frac{V}{n} - \frac{3}{2} kT \ln T - kT \ln q_s - kT \ln \frac{(2\pi mk)^{3/2}}{h^3} \\ &= kT \ln p - \frac{5}{2} kT \ln T - kT \ln q_s - kT \ln \frac{k\,(2\pi mk)^{3/2}}{h^3}\end{aligned} \tag{7.206}$$

This is a fundamental equation and any dilute gas described by it is called a *semiperfect gas.* Moreover, we can readily verify that the one-particle distribution function for such a gas is given by the Maxwellian distribution (Eq. 7.68).

Equation 7.206 can be used for the approximate description of the stable equilibrium states of any homogeneous open system. When this is done, the system is said to be analyzed by means of the *semiperfect gas approximation.*

Substituting Eq. 7.206 into condition 7.203 yields

$$\frac{ph^3}{(2\pi m)^{3/2}(kT)^{5/2}} \ll q_s \exp\left(\frac{\varepsilon_0}{kT}\right) \tag{7.207}$$

From the preceding discussion and relation 7.207, it follows that any homogeneous system of reasonably large dimensions (Eq. 7.192) behaves like a semiperfect gas at low pressures or high temperatures, or both.

Finally, as already stated, many equations of stable equilibrium states can be derived from a fundamental equation. For example, the energy E of a one-component open system is given by the relation

$$E = H - pV = -nT^2\left[\frac{\partial(\mu/T)}{\partial T}\right]_{p,V,n} - pV \tag{7.208}$$

where use has been made of Eq. 7.150. For a semiperfect gas, E and μ can be replaced by U and μ_c, respectively, and therefore

$$U = -nT^2\left[\frac{\partial(\mu_c/T)}{\partial T}\right]_{p,V,n} - pV = \frac{3}{2}nkT + nkT\frac{\partial \ln q_s}{\partial \ln T} \tag{7.209}$$

Again the specific entropy s is given by the relation

$$\begin{aligned} s &= -\left(\frac{\partial \mu_c}{\partial T}\right)_{p,\beta,n} \\ &= -\left(\frac{\partial \mu_c}{\partial T}\right)_{p,n} \\ &= -k \ln p + \frac{5}{2}k(1 + \ln T) + k\left(\ln q_s + T\frac{\partial \ln q_s}{\partial T}\right) + k \ln \frac{k(2\pi mk)^{3/2}}{h^3} \end{aligned} \tag{7.210}$$

7.8.12 Specific Heats. For a homogeneous system A, the quantity c_θ, given by the relation

$$c_\theta = T\left(\frac{\partial s}{\partial T}\right)_{\theta,\ \text{all other properties in } s} \tag{7.211}$$

is defined as the *specific heat per particle of A under constant property* θ. For example, if the only independent constraints of A are volume and an applied

field, then the *specific heat* c_p *per particle under constant pressure* is given by the relation

$$c_p = T\left(\frac{\partial s}{\partial T}\right)_{p,n} \tag{7.212}$$

and the *specific heat* c_V *per particle under constant volume* is given by the relation

$$c_V = T\left(\frac{\partial s}{\partial T}\right)_{V,n} \tag{7.213}$$

For a semiperfect gas, substituting Eq. 7.210 into Eqs. 7.212 and 7.213 yields

$$c_p = \frac{5}{2}k + 2kT\frac{\partial \ln q_s}{\partial T} + kT^2\frac{\partial^2 \ln q_s}{\partial T^2} \tag{7.214}$$

$$c_V = \frac{3}{2}k + 2kT\frac{\partial \ln q_s}{\partial T} + kT^2\frac{\partial^2 \ln q_s}{\partial T^2} \tag{7.215}$$

and

$$c_p - c_V = k \tag{7.216}$$

Since the partition function q_s of the internal degrees of freedom of each particle depends on T, we can see from Eqs. 7.214 and 7.215 that the specific heats per particle of a semiperfect gas under both constant pressure and constant volume are functions of temperature.

7.8.13 The Perfect Gas. A semiperfect gas consisting of particles whose internal structure can be represented by a g-fold degenerate energy eigenvalue $\bar{\varepsilon}_o$ is defined as a *perfect gas.* For such a gas

$$q_s = \sum \exp\left(-\frac{\varepsilon_s}{kT}\right) = g\exp\left(-\frac{\bar{\varepsilon}_o}{kT}\right) \tag{7.217}$$

$$\begin{aligned}\mu_c &= kT\ln p - \frac{5}{2}kT\ln T - kT\ln\frac{gk\,(2\pi mk)^{3/2}}{h^3} + \bar{\varepsilon}_o \\ &= -kT\ln\frac{V}{n} - \frac{3}{2}kT\ln T - kT\ln\frac{g\,(2\pi mk)^{3/2}}{h^3} + \bar{\varepsilon}_o\end{aligned} \tag{7.218}$$

$$U = \frac{3}{2}nkT + n\bar{\varepsilon}_o \tag{7.219}$$

$$c_p = \frac{5}{2}k \qquad c_V = \frac{3}{2}k \qquad c_p - c_V = k \tag{7.220}$$

Of course, the p–V–T–n relation of a perfect gas is the same as that of a semiperfect gas or that resulting from the Boltzmann approximations, namely,

$$pV = nkT$$

In addition, the one-particle distribution function is the Maxwellian distribution (Eq. 7.68).

More generally, a perfect gas can be defined as a substance that obeys the p–V–T–n relation resulting from the Boltzmann approximations and has specific heats c_p and c_V independent of temperature and such that $c_p - c_V = k$. For such a substance the specific energy u, the specific enthalpy h, and the specific entropy s at temperatures T_0 and T and pressures p_0 and p obey the relations

$$\begin{aligned} u - u_0 &= c_V(T - T_0) \\ h - h_0 &= c_p(T - T_0) \\ s - s_0 &= c_p \ln \frac{T}{T_0} - k \ln \frac{p}{p_0} \end{aligned} \tag{7.221}$$

Other relations between properties of stable equilibrium states can be derived by suitable combinations of Eqs. 7.221. For example,

$$\begin{aligned} \mu_c = h - Ts &= c_pT - c_pT \ln T + kT \ln p + h_0 - c_pT_0 \\ &\quad + (c_p \ln T_0' - k \ln p_0' - s_0)T \end{aligned} \tag{7.222}$$

where the reference state (T_0, P_0) for the enthalpy may be different from that (T_0', P_0') for the entropy. If T_0 is zero on the Kelvin scale and a new entropy constant is defined by the relation

$$s_{00} \equiv s_0 - c_p \ln T_0' + k \ln p_0'$$

then Eq. 7.222 becomes

$$\mu_c = c_pT - c_pT \ln T + kT \ln p + h_0 - Ts_{00} \tag{7.223}$$

7.9 Mixtures of Gases

7.9.1 General Remarks. Although the number of gases of chemical elements is finite, the number of mixtures of different compositions that can be made from the various elements is unlimited. In dealing with gaseous mixtures in any general way, therefore, we must find means of relating the properties of the mixture to the properties of its components. This problem is so complicated, however, that explicit answers can be found only for certain stable equilibrium states.

This section presents some results regarding gaseous mixtures which are useful in analyzing plasmas formed in thermionic converters.

7.9.2 The Gibbs–Dalton Law. If a mixture of gases (A) consists of r different gaseous components at pressure p and temperature T, under certain conditions to be subsequently derived, the properties of A can be expressed in a simple manner in terms of the properties of r one-component systems $1, 2, \ldots, r$, each consisting of one component of A and being in mutual stable equilibrium with A through a membrane permeable to the component. Systems A and $1, 2, \ldots, r$ are shown schematically in Fig. 7.6. All are in the same applied field with a potential ψ.

If the pressures of A and the lth one-component system were changed by dp and dp_{II}, respectively, while ψ, the temperature of all parts, and the pressures of all other one-component systems are held fixed, then the pressure, the temperature, and the chemical potential of each one-component system (except the lth) remain unaltered. For mutual stable equilibrium, the total potential of each component must be the same on either side of its semipermeable membrane and, since ψ is uniform for all the systems, the chemical potential of each

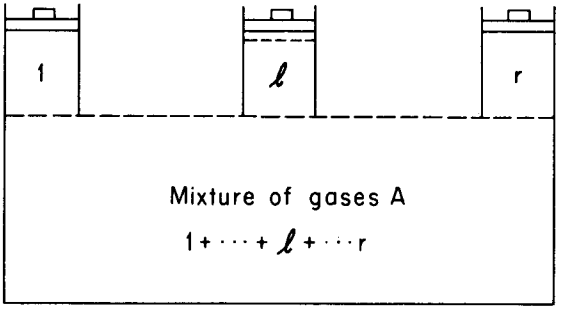


Fig. 7.6 Schematic of a mixture of gases A in mutual stable equilibrium with one-component systems $1, 2, \ldots, r$.

component must be the same on either side of its semipermeable membrane. It follows that the chemical potentials of all components in A except the lth remain unchanged.

Applying the Gibbs–Duhem equation (Eq. 7.142b) to the change between the stable equilibrium states of A yields

$$-V\,dp + n_l\,d\mu_{cl} = 0 \qquad \text{for constant } T,\ \psi, \text{ and } n \tag{7.224}$$

Applying the same equation to the lth one-component system yields

$$-V_{ll}\,dp_{ll} + n_{ll}\,d\mu_{cll} = 0 \qquad \text{for constant } T,\ \psi, \text{ and } n \tag{7.225}$$

where V is the volume of A; n_l and μ_{cl} are the number of particles and the chemical potential, respectively, of the lth component of A; and V_{ll}, n_{ll}, and μ_{cll} are the volume, the number of particles, and the chemical potential, respectively, of the one-component system consisting of component l. For mutual stable equilibrium, the changes $d\mu_{cl}$ and $d\mu_{cll}$ must be equal, namely,

$$d\mu_{cl} = d\mu_{cll} \tag{7.226}$$

Substituting Eq. 7.226 into Eqs. 7.224 and 7.225 yields

$$dp = \frac{V_{ll}/n_{ll}}{V/n_l}\,dp_{ll} = \frac{v_{ll}}{V/n_l}\,dp_{ll} \qquad \text{for all } l \tag{7.227}$$

Moreover, if we assume that for A

$$\frac{V}{n_l} = v_{ll} \qquad \text{for all } l \tag{7.228}$$

then it follows that

$$dp = dp_{ll} \qquad \text{for all } l \tag{7.229}$$

or equivalently

$$p = p_{11} + p_{22} + \ldots + p_{rr} \tag{7.230}$$

In words, if relation 7.228 is satisfied, the pressure of a mixture of different gases equals the sum of the pressures of each of the different gases existing by itself at the temperature of the mixture and with the same value of the chemical

potential as it has in the mixture. The pressures in the sum are called *partial pressures.* This conclusion is known as the *Gibbs–Dalton law for mixtures of gases.*

By virtue of Eq. 7.228, we can readily show that the partial pressure of a component of a Gibbs–Dalton mixture of gases equals the pressure the component would have if it alone occupied the volume of the mixture at the temperature of the mixture.

By virtue of Eqs. 7.148, 7.224, and 7.225, the condition for the applicability of the Gibbs–Dalton law (Eq. 7.228) can also be expressed in terms of total potentials. Specifically

$$\left(\frac{\partial \mu_l}{\partial p}\right)_{T,\psi,n} = \left(\frac{\partial \mu_{ll}}{\partial p}\right)_{T,\psi,n_{ll}} \tag{7.231}$$

Many other results follow from Eq. 7.228. For example, the density $1/v$ or n/V of A equals the sum of the densities of the one-component systems, as expressed by the relation

$$\frac{1}{v} = \sum_l \frac{1}{v_{ll}} \tag{7.232}$$

Again, it can be shown that the energy E, enthalpy H, entropy S, Helmholtz free energy Ψ, and Gibbs free energy Z are given by the relations

$$\frac{E}{V} = \sum_l \left(\frac{E}{V}\right)_{ll} \quad \frac{H}{V} = \sum_l \left(\frac{H}{V}\right)_{ll} \quad \frac{S}{V} = \sum_l \left(\frac{S}{V}\right)_{ll}$$

$$\frac{\Psi}{V} = \sum_l \left(\frac{\Psi}{V}\right)_{ll} \quad \frac{Z}{V} = \sum_l \left(\frac{Z}{V}\right)_{ll} \tag{7.233}$$

Finally, it can be shown that a mixture of perfect gases obeys the Gibbs–Dalton law and behaves like a perfect gas.[4]

7.10 Mixtures of Ionized Gases

7.10.1 The Saha Equation. We will consider a system A consisting of a mixture of an electrically neutral gaseous component X and its electrically charged products of dissociation in a state of chemical equilibrium for the reaction

$$X \rightleftharpoons X^+ + e^- \tag{7.234}$$

where X, X^+, and e^- denote the neutral component, the positive-ion component, and the negative-ion component of A, respectively. For brevity these components are referred to subsequently as molecules, ions, and electrons.

The presence of ions and electrons results in a nonuniform electric field of potential energy ψ, and therefore A is heterogeneous. At high temperatures and low pressures, however, A can be subdivided into independent parts, each of which is homogeneous and behaves like a perfect gas. In other words, in each part, each of the components (molecules, ions, and electrons) is assumed to consist of particles with an internal structure that can be represented by a single degenerate energy eigenvalue, and the short-range interparticle forces are assumed to be zero. Thus each of the components behaves like a perfect gas subject to the Gibbs–Dalton rules. At high temperatures and low pressures, using the point-particle approximation for each of the internal structures of the particles is reasonable. In addition, for the same conditions and for sufficiently small parts, neglecting interparticle forces may be justified for molecules, ions, and electrons equally well even though the ions and the electrons interact by means of long-range Coulomb forces. The reason is that the smaller the part, the better the uniformity of ψ within the part. Of course, the part must be sufficiently large for the helpful continuous eigenvalue-spectrum approximation to be applicable (Eq. 7.192).

If molecules are denoted by subscript a, ions by subscript i, and electrons by subscript e, applying the equation of chemical equilibrium to the various components of one part of A yields (Eq. 7.87)

$$\mu_a - \mu_i - \mu_e = 0 \tag{7.235}$$

where μ_a, μ_i, and μ_e denote the total potential of molecules, ions, and electrons, respectively. The total potentials are related to the chemical potentials of the components of each part by the relations

$$\mu_a = \mu_{ca} \qquad \mu_i = q_i\psi + \mu_{ci} \qquad \mu_e = q_e\psi + \mu_{ce} \tag{7.236}$$

where q_i = the positive charge per ion
q_e = the negative charge per electron
ψ = electrostatic potential of the part under consideration
μ_{ca} = chemical potential of molecules
μ_{ci} = chemical potential of ions
μ_{ce} = chemical potential of electrons

Conservation of charge requires that q_i equal $-q_e$. Therefore combining Eqs. 7.235 and 7.236 yields

$$\mu_{ca} - \mu_{ci} - \mu_{ce} = 0 \tag{7.237}$$

We see from Eqs. 7.235 and 7.237 that the equation of chemical equilibrium can be expressed in terms of either total potentials or chemical potentials.

The value of the chemical potential of a component in Gibbs–Dalton mixture A equals that of the pure component in mutual stable equilibrium with A. Accordingly μ_{cl} for the lth component is given by the relation (Eq. 7.218)

$$\mu_{cl} = kT \ln p_l - \frac{5}{2} kT \ln T - kT \ln \frac{g_l k (2\pi m_l k)^{3/2}}{h^3} + \bar{\varepsilon}_{ol} \tag{7.238}$$

where p_l and m_l are the partial pressure and mass, respectively, of the lth component in the mixture and g_l and $\bar{\varepsilon}_{ol}$ are the degeneracy and single energy eigenvalue, respectively, for the internal structure of the particles of the lth component. Substituting Eq. 7.238 evaluated for molecules, ions, and electrons into Eq. 7.237 yields

$$kT \ln \frac{p_a}{p_i p_e} + \frac{5}{2} kT \ln T + kT \ln \frac{g_i g_e}{g_a} \frac{(2\pi k)^{3/2} k}{h^3} \left(\frac{m_i m_e}{m_a}\right)^{3/2} + \bar{\varepsilon}_{oa} - \bar{\varepsilon}_{oi} - \bar{\varepsilon}_{oe} = 0 \tag{7.239}$$

In the context of the point-particle approximation, the quantity $\bar{\varepsilon}_{oi} + \bar{\varepsilon}_{oe} - \bar{\varepsilon}_{oa}$ equals the energy required to ionize a molecule at 0°K. To be acceptable, this quantity must be assigned the value of the ionization energy V_i of the molecule, and therefore

$$\bar{\varepsilon}_{oi} + \bar{\varepsilon}_{oe} - \bar{\varepsilon}_{oa} \equiv V_i \tag{7.240}$$

Substituting Eq. 7.240 into Eq. 7.239 and rearranging terms yield

$$\frac{p_i p_e}{p_a} = \frac{g_i g_e}{g_a} \frac{(2\pi k)^{3/2} k}{h^3} \left(\frac{m_i m_e}{m_a}\right)^{3/2} T^{5/2} \exp\left(-\frac{V_i}{kT}\right) \tag{7.241}$$

Equation 7.241 is known as the *Saha equation* for a mixture of perfect gases.

Now all the parts of A are in mutual stable equilibrium. Consequently the temperature T and each of the total potentials μ_a, μ_i, and μ_e must have the same values in every part of A (Sec. 7.4.3). If the subscript 0 denotes a reference part of the mixture for which $\psi = 0$, it follows that

$$\mu_a = \mu_{a0} \qquad \mu_i = \mu_{i0} \qquad \mu_e = \mu_{e0} \tag{7.242}$$

Substituting Eqs. 7.236 and 7.238 into Eq. 7.242 yields

$$p_a = p_{a0} \qquad p_i = p_{i0} \exp\left(-\frac{q_i\psi}{kT}\right) \qquad p_e = p_{e0} \exp\left(-\frac{q_e\psi}{kT}\right) \tag{7.243}$$

In words, the partial pressure p_a of the molecules is uniform, whereas the partial pressures p_i of the ions and p_e of the electrons vary exponentially with ψ throughout the mixture A.

The density of the lth component of the mixture equals p_l/kT, and therefore the charge densities ρ_i of the ions and ρ_e of the electrons are $q_i p_i/kT$ and $q_e p_e/kT$, respectively. Thus

$$\rho_i = \rho_{i0} \exp\left(-\frac{q_i\psi}{kT}\right) > 0$$

$$\rho_e = \rho_{e0} \exp\left(-\frac{q_e\psi}{kT}\right) < 0 \tag{7.244}$$

Moreover, the total charge density ρ and Poisson's equation are given by the relations

$$\rho = \rho_i + \rho_e = \rho_{i0} \exp\left(-\frac{q_i\psi}{kT}\right) + \rho_{e0} \exp\left(-\frac{q_e\psi}{kT}\right) \tag{7.245}$$

and

$$\text{div grad } \psi = -\frac{\rho_{i0}}{\varepsilon_0}\left[\exp\left(-\frac{q_i\psi}{kT}\right) + \frac{\rho_{e0}}{\rho_{i0}} \exp\left(-\frac{q_e\psi}{kT}\right)\right] \tag{7.246}$$

where ε_0 is the dielectric constant of vacuum. For given boundary conditions, Eq. 7.246 can be integrated to yield the electrostatic potential distribution in the mixture A in a stable equilibrium state.

7.10.2 Mutual Stable Equilibrium of a Mixture of Ionized Gases with a Solid. If the mixture A of ionized gases discussed in the preceding section is in mutual stable equilibrium with a solid made up of the same components as A and the total potentials of molecules, ions, and electrons are denoted, respectively, by μ_a^g, μ_i^g, and μ_e^g in the mixture and by μ_a^s, μ_i^s, and μ_e^s in the solid, then it is necessary that

$$\mu_a^g = \mu_a^s \qquad \mu_i^g = \mu_i^s \qquad \mu_e^g = \mu_e^s \tag{7.247}$$

Using Eqs. 7.236 and 7.247 in Eq. 7.238 for the different components yields

$$p_a = \frac{g_a k (2\pi m_a k)^{3/2}}{h^3} T^{5/2} \exp\left(\frac{\mu_a^s - \bar{\varepsilon}_{oa}^g}{kT}\right) \tag{7.248}$$

and

$$p_l = \frac{g_l k (2\pi m_l k)^{3/2}}{h^3} T^{5/2} \exp\left(\frac{\mu_l^s - q_l\psi - \bar{\varepsilon}_{ol}^g}{kT}\right) \quad \text{for } l = i \text{ or } e \tag{7.249}$$

It is noteworthy that the coefficients of the exponentials in Eqs. 7.248 and 7.249 depend on properties of the particles of the components of the gaseous mixture and not on properties of the solid.

7.10.3 Random-Flux Densities. In a mixture of perfect gases, each component behaves like a perfect gas, and its one-particle distribution function is given by the Maxwellian distribution. The random flux density Γ_{rl} of the lth component is defined by the relation

$$\Gamma_{rl} = \frac{1}{v_l}\left(\frac{m_l}{2\pi kT}\right)^{3/2} \int_{-\infty}^{\infty} dV_x \int_{-\infty}^{\infty} dV_y \int_0^{\infty} dV_z V_z \times \exp\left[-\frac{m_l}{2kT}(V_x^2 + V_y^2 + V_z^2)\right] \tag{7.250}$$

where v_l is the specific volume of the lth component. If the integrations indicated in the right-hand side of Eq. 7.250 are carried out and use is made of the perfect-gas relation $v_l = kT/p_l$, we find that Γ_{rl} is given by the relation

$$\Gamma_{rl} = \frac{p_l}{(2\pi m_l kT)^{1/2}} \tag{7.251}$$

where p_l is partial pressure of the lth component.

For electrons, substituting p_e from Eq. 7.249 into Eq. 7.251 and multiplying the result by the electron charge e yield

$$\begin{aligned} J_{re} &= e\Gamma_{re} \\ &= \frac{2\pi m_e k^2 e g_e}{h^3} T^2 \exp\left(\frac{\mu_e^s + e\psi - \bar{\varepsilon}_{oe}^g}{kT}\right) \\ &= AT^2 \exp\left(\frac{\mu_e^s + e\psi - \bar{\varepsilon}_{oe}^g}{kT}\right) \end{aligned} \tag{7.252}$$

where

$$A = \frac{2\pi m_e k^2 e g_e}{h^3} = 120 \text{ amp/°K}^2\text{-cm}^2$$

The quantity J_{re} is called the *electron random current density.* Subsequently Eq. 7.252 is used for calculating the current emitted from a thermionic emitter.

7.11 Surfaces

7.11.1 General Remarks. Any system whose particles can be regarded as having only two translational degrees of freedom will be called a *surface system* or simply a *surface.* For example, a monatomic layer of cesium particles on a tungsten substrate can be regarded as a surface system.

The analysis of stable equilibrium states of homogeneous surfaces is based on the same methods as those employed in the preceding sections for analyzing three-dimensional homogeneous systems. For surfaces, however, the results differ from those derived for three-dimensional systems because area A rather than volume V is an independent constraint.

Some results applicable to surfaces in stable equilibrium states are given here without elaborate proofs. Unless otherwise stated, all symbols have the same meanings as they did in the discussions of three-dimensional systems.

7.11.2 Basic Relations. If a homogeneous open-surface system consists of r distinguishable components and the area A and applied field β are the only independent constraints, the Gibbs equation (Eq. 7.94) becomes

$$dE = T\,dS - \tau\,dA + \left(\frac{\partial E}{\partial \beta}\right)_{S,A,n} d\beta + \sum_{l}^{r} \mu_l\,dn_l \tag{7.253}$$

where

$$T = \left(\frac{\partial E}{\partial S}\right)_{A,\beta,n} \qquad \tau = -\left(\frac{\partial E}{\partial A}\right)_{S,\beta,n} \qquad \mu_l = \left(\frac{\partial E}{\partial n_l}\right)_{S,A,\beta,n} \tag{7.254}$$

The Kelvin temperature T and the total potential μ_l of the lth component of the surface are defined in a manner analogous to that used for other systems. The quantity τ, called the *surface spreading force*, has dimensions of force per unit length. It is analogous to but differs from the pressure of a three-dimensional system because here area rather than volume is the independent constraint.

The conditions for mutual stable equilibrium between a surface and another system are the same as those between any two systems. They require equality of temperatures and of total potentials of each component present in both the surface and the other system. In addition, the equation of chemical equilibrium (Eq. 7.87) applies to a surface. Finally, for a homogeneous open surface, relation 7.139 becomes

$$\frac{\tau A}{kT} = \ln Q_g \tag{7.255}$$

where Q_g is the grand partition function for the surface.

7.11.3 Dilute Surfaces. A two-dimensional system for which the one-particle approximation (Sec. 7.8.2) holds is defined as a *dilute surface.* All relations derived for dilute gases can be readily adapted to dilute surfaces.

For example, for a one-component dilute surface for which the constant-potential approximation (Sec. 7.8.8) is applicable, Eqs. 7.190 and 7.191 become

$$\frac{\tau A}{kT} = \sum_{s,k_x,k_y} \ln \left\{ 1 \pm \exp\left(\frac{\mu_c}{kT}\right) \exp\left(-\frac{\varepsilon_s}{kT}\right) \times \exp\left[-\frac{h^2}{8mkT}\left(\frac{k_x^2}{l_x^2} + \frac{k_y^2}{l_y^2}\right)\right]\right\}^{\pm 1} \tag{7.256}$$

and

$$n = \sum_{s,k_x,k_y} \left\{ \exp\left(-\frac{\mu_c}{kT}\right) \exp\left(\frac{\varepsilon_s}{kT}\right) \exp\left[\frac{h^2}{8mkT}\left(\frac{k_x^2}{l_x^2} + \frac{k_y^2}{l_y^2}\right)\right] \pm 1 \right\}^{-1} \tag{7.257}$$

Again, when condition 7.203 and the continuous eigenvalue-spectrum approximation (Sec. 7.8.9) are applicable, a dilute surface is defined as a *semiperfect surface.* Then we can readily verify that the following relations hold.

$$\tau A = nkT \tag{7.258}$$

$$\mu_c = kT \ln \tau - 2kT \ln T - kT \ln q_s - kT \ln \frac{2\pi mk^2}{h^2} \tag{7.259}$$

$$U = nkT + nkT \frac{\partial \ln q_s}{\partial \ln T} \tag{7.260}$$

In the study of adsorbed layers regarded as surfaces, it is helpful to express Eqs. 7.258 and 7.259 in terms of the coverage θ rather than the area A and the

expectation value n of the number of adsorbed particles. The coverage θ of a surface system is defined by the relation

$$\theta = \frac{n/A}{\sigma_f} \tag{7.261}$$

where σ_f is the maximum number of particles (the number of adsorption sites) which can be accommodated per unit area. Substituting Eq. 7.261 into Eqs. 7.258 and 7.259 yields

$$\frac{\tau}{\theta} = \sigma_f kT \tag{7.262}$$

and

$$\mu_c = kT \ln \sigma_f \theta - kT \ln T - kT \ln q_s - kT \ln \frac{2\pi mk}{h^2} \tag{7.263}$$

It is noteworthy that the approximations involved in defining a semiperfect surface are valid only for very small values of the coverage θ, usually less than 0.1. Accordingly Eq. 7.263 for μ_c as a function of θ and T is applicable only for θ smaller than about 0.1. For larger values of θ, not only are the approximations introduced in the derivation of the relations for semiperfect surfaces not valid but also the surface cannot be regarded as dilute.

7.11.4 Chemical Potential of Nondilute Surfaces. The relation between the chemical potential μ_c and the coverage θ and temperature T of a surface is of prime importance to the study of thermionic emission phenomena. For values of θ smaller than about 0.1, this relation is given by Eq. 7.263.

A general technique for evaluating the chemical potential μ_c is based on the knowledge of the τ–A–T–β–n or, equivalently, of the τ–θ–T–β relation for the surface. For a coverage θ, the chemical potential μ_c can be expressed in the form

$$\mu_c(\theta,T) = \mu_c(\theta^*,T) + \int_{\theta^*}^{\theta} \left(\frac{\partial \mu_c}{\partial \theta}\right)_T d\theta \tag{7.264}$$

where θ^* is an arbitrary value of the coverage. The derivative $(\partial \mu_c/\partial \theta)_T$ can be written as

$$\left(\frac{\partial \mu_c}{\partial \theta}\right)_T = \left(\frac{\partial \mu_c}{\partial \tau}\right)_T \left(\frac{\partial \tau}{\partial \theta}\right)_{T,\beta} = \frac{1}{\sigma_f \theta} \left(\frac{\partial \tau}{\partial \theta}\right)_{T,\beta} \tag{7.265}$$

The relation between $(\partial\mu_c/\partial\tau)_T$ and $\sigma_f\theta$ in Eq. 7.265 is obtained by combining Eq. 7.261 and the Gibbs–Duhem equation for a one-component surface, i.e.,

$$S\,dT - A\,d\tau + n\,d\mu_c = 0 \tag{7.266}$$

Substituting Eq. 7.265 into Eq. 7.264 yields

$$\mu_c(\theta,T) = \mu_c(\theta^*,T) + \int_{\theta^*}^{\theta} \frac{1}{\sigma_f\theta}\left(\frac{\partial\tau}{\partial\theta}\right)_{T,\beta} d\theta \tag{7.267}$$

The value of θ^* can be taken so small that $\mu_c(\theta^*,T)$ is given by Eq. 7.263. Then Eq. 7.267 becomes

$$\mu_c(\theta,T) = kT \ln \frac{\sigma_f\theta^* h^2}{2\pi mkTq_s} + \int_{\theta^*}^{\theta} \frac{1}{\sigma_f\theta}\left(\frac{\partial\tau}{\partial\theta}\right)_{T,\beta} d\theta \tag{7.268}$$

Equation 7.268 yields the chemical potential of any one-component surface, dilute or nondilute, if the τ–θ–T–β relation is known. This equation is used in Chap. 9 for calculating emission properties of cesiated emitters.

7.12 Lattice Vibrations

7.12.1 General Remarks. In calculating energy eigenstates of crystals, we treat electrons and ion cores (nuclei) independently of each other. The electron system is discussed in Sec. 5.10. Here we present some approximate methods for evaluating energy eigenstates and equations of stable equilibrium states of crystal ion cores.

To a first approximation, ion cores can be regarded as a system that undergoes small oscillations about a position of equilibrium. The oscillations are analyzed in terms of a set of modes of vibration each of which is considered as an independent, one-dimensional harmonic oscillator.

The enumeration of the fundamental frequencies of the harmonic oscillators is a very difficult problem. Only approximate answers have been given. Einstein, who developed one of the earliest answers, proposed that a crystal with n ion cores should be described by $3n$ independent but identical oscillators, each with a fundamental frequency ν_E. This is known as Einstein's theory of lattice vibrations. It was improved by Debye who suggested another approximation, namely, that the $3n$ fundamental frequencies could be taken as the $3n$ lowest frequencies of a continuum with the same elastic constants as the crystal. This approximation is known as Debye's theory of lattice vibrations.

7.12.2 Partition Function. For a harmonic oscillator with fundamental frequency ν_m and energy eigenvalues ϵ_{mi}, where

$$\epsilon_{mi} = jh\nu_m + \frac{\epsilon_{i0}}{3} \qquad \text{for } j = 0, 1, 2, \ldots$$

and ϵ_{i0} denotes the minimum energy per ion core, the partition function q_{sm} is given by the relation

$$q_{sm} = \frac{\exp(-\epsilon_{i0}/3kT)}{1 - \exp(-h\nu_m/kT)} \tag{7.269}$$

For a set of $3n$ independent oscillators, the canonical partition function Q is given by the product of the partition functions q_{sm} for each oscillator and therefore

$$\ln Q = -\frac{n\epsilon_{i0}}{kT} - \sum_{m=1}^{3n} \ln\left[1 - \exp\left(-\frac{h\nu_m}{kT}\right)\right] \tag{7.270}$$

If the oscillators are distributed among the several oscillator frequencies in such a manner that the function describing the distribution is continuous, and, if $g(\nu)\,d\nu$ denotes the number of oscillators between ν and $\nu + d\nu$, then the summation in Eq. 7.270 can be replaced by an integral so that

$$\ln Q = -\frac{n\epsilon_{i0}}{kT} - \int_0^\infty g(\nu) \ln\left[1 - \exp(-h\nu/kT)\right] d\nu \tag{7.271}$$

Since the number of oscillators is equal to $3n$, the distribution $g(\nu)$ must satisfy the relation

$$\int_0^\infty g(\nu)\,d\nu = 3n \tag{7.272}$$

In the Einstein theory, $g(\nu)$ is assumed to be a delta function centered at the Einstein frequency ν_E. Thus, Eq. 7.271 becomes

$$\ln Q_E = -\frac{n\epsilon_{i0}}{kT} - 3n \ln\left[1 - \exp\left(-\frac{h\nu_E}{kT}\right)\right] \tag{7.273}$$

In the Debye theory, an expression for $g(\nu)$ is obtained from considerations of standing waves in an elastic continuum having the same elastic properties and density as the crystal. Since an infinite number of normal modes of vibration are possible in a continuum, Debye suggested the following method for selecting the

$3n$ modes necessary for a lattice of n ion cores. Modes with wavelengths smaller than approximately twice the internuclear spacing are unrealistic. For this reason, only those $3n$ normal modes having the greatest realizable wavelengths, (i.e., the lowest realizable frequencies) should be considered.

In a rectangular parallelepiped with edge lengths l_x, l_y, and l_z and containing an elastic continuum, three modes of vibration (one longitudinal with propagation speed c_l and two orthogonal transverse with propagation speed c_t) are possible for each allowed wavelength. For present purposes, the boundary of the system can be assumed to be rigid so that each standing wave must have a node at the boundary. Then the displacements of the standing waves are proportional to

$$\sin\left(\frac{\pi k_x x}{l_x}\right)\sin\left(\frac{\pi k_y y}{l_y}\right)\sin\left(\frac{\pi k_z z}{l_z}\right)$$

where k_x, k_y, and k_z are integers. These standing waves are formed by superimposing traveling waves with wavelength λ related to the integers k_x, k_y, and k_z by the relation

$$\frac{1}{\lambda^2}=\left(\frac{k_x}{2l_x}\right)^2+\left(\frac{k_y}{2l_y}\right)^2+\left(\frac{k_z}{2l_z}\right)^2$$

It follows that the allowed frequencies are

$$\nu_l=\frac{c_l}{\lambda}=c_l\left[\left(\frac{k_x}{2l_x}\right)^2+\left(\frac{k_y}{2l_y}\right)^2+\left(\frac{k_z}{2l_z}\right)^2\right]^{1/2}$$

and

$$\nu_t=\frac{c_t}{\lambda}=c_t\left[\left(\frac{k_x}{2l_x}\right)^2+\left(\frac{k_y}{2l_y}\right)^2+\left(\frac{k_z}{2l_z}\right)^2\right]^{1/2}$$

It can be shown that the number $g_l(\nu)\,d\nu$ of longitudinal modes with frequencies between ν and $\nu + d\nu$ is given by the relation

$$g_l(\nu)\,d\nu=\frac{4\pi V\nu^2}{c_l^3}d\nu$$

where $V = l_x l_y l_z$. Similarly, the number $g_t(\nu)\,d\nu$ of transverse modes with frequencies between ν and $\nu + d\nu$ is given by the relation

$$g_t(\nu)\,d\nu=\frac{8\pi V\nu^2}{c_t^3}\,d\nu$$

and therefore

$$g(\nu)\,d\nu = 4\pi V\left(\frac{1}{c_l^3}+\frac{2}{c_t^3}\right)\nu^2\,d\nu = \frac{12\pi V}{c_a^3}\nu^2\,d\nu \tag{7.274}$$

where c_a denotes an average velocity of propagation defined by the relation

$$\frac{3}{c_a^3}=\frac{1}{c_l^3}+\frac{2}{c_t^3}$$

In the Debye approximation, the value of $g(\nu)$ is given by Eq. 7.274 for ν between zero and a frequency ν_D and is equal to zero for ν greater than ν_D. The Debye frequency ν_D, therefore, is determined from the relation (Eq. 7.272)

$$\int_0^{\nu_D}\frac{12\pi V}{c_a^3}\nu^2\,d\nu = \frac{4\pi V}{c_a^3}\nu_D^3 = 3n \tag{7.275a}$$

or

$$\nu_D = \left(\frac{3c_a^3}{4\pi}\frac{n}{V}\right)^{1/3} \tag{7.275b}$$

Use of Eq. 7.275 in Eq. 7.274 yields

$$g(\nu)\,d\nu = \begin{cases} 9n\,\dfrac{\nu^2}{\nu_D^3}\,d\nu & \text{for } 0<\nu\leqslant\nu_D \\ 0 & \text{for } \nu_D<\nu \end{cases} \tag{7.276}$$

It is convenient to define the *Debye temperature* θ_D by the relation

$$\theta_D = \frac{h\nu_D}{k} = \frac{hc_a}{k}\left(\frac{3n}{4\pi V}\right)^{1/3} \tag{7.277}$$

In terms of θ_D, the partition function in the Debye approximation becomes (Eq. 7.271)

$$\ln Q = -\frac{n\epsilon_{i0}}{kT} - 9n\frac{T^3}{\theta_D^3}\int_0^{\theta_D/T}\xi^2\ln\left[1-\exp(-\xi)\right]d\xi \tag{7.278}$$

In some analyses another term is added to the right-hand side of Eq. 7.278 to account for the effects of the internal structure of the ion core. This term is relatively small and will not be considered here.

7.12.3 Gibbs Free Energy. The partition function estimates given in Sec. 7.12.2 can be used to compute contributions of the ion cores to the properties of a crystal. We will compute only the Gibbs free energy per particle. This energy is given by the relation

$$\zeta_i = \frac{-kT \ln Q + p_i V}{n} \tag{7.279}$$

Use of the Debye approximation for Q (Eq. 7.278) yields

$$\zeta_i = \epsilon_{i0} + 9k \frac{T^4}{\theta_D^3} \int_0^{\theta_D/T} \xi^2 \ln [1 - \exp(-\xi)]\, d\xi + p_i \nu_i \tag{7.280}$$

where ν_i denotes the specific volume of the ion cores.

Equation 7.280 can be further simplified by considering different ranges of temperature. For example, for T greater than the Debye temperature θ_D, it can be shown that

$$\zeta_i = \epsilon_{i0} - 3kT \left(\ln \frac{T}{\theta_D} + \frac{1}{3} + \frac{3}{8} \frac{\theta_D}{T} - \frac{1}{40} \frac{\theta_D^2}{T^2} \right) + p_i \nu_i$$

$$\approx \epsilon_{i0} - 3kT \left(\ln \frac{T}{\theta_D} + \frac{1}{3} + \frac{3}{8} \frac{\theta_D}{T} \right) \quad \text{for } T \gg \theta_D \tag{7.281}$$

In the second part of Eq. 7.281 the term $p_i \nu_i$ is omitted because it is usually much smaller than the other terms.

References

1. G. N. Hatsopoulos and J. H. Keenan, *Principles of General Thermodynamics*, p. 327, John Wiley & Sons, Inc., New York, 1965.
2. E. P. Gyftopoulos and G. N. Hatsopoulos, Quantum-Thermodynamic Definition of Electronegativity, *Proc. Nat. Acad. Sci., U. S. A.*, **60**(3): 786-793 (1968).
3. Ref. 1, p. 474.
4. Ref. 1, p. 288.

8

Thermodynamics of Steady States

8.1 General Remarks

From the practical standpoint the most important and interesting applications of the concepts and principles discussed in Chaps. 5 and 6 involve interactions between systems some or all of which are passing through nonequilibrium states. For example, energy conversion, chemical reactions, and flow of energy and matter are processes occurring in systems in nonequilibrium states. These processes result from interactions between various phases that are not in mutual stable equilibrium.

Unfortunately two major difficulties are involved in analyzing a collection of systems passing through nonequilibrium states. The first and foremost difficulty is the lack of knowledge of the exact equation of motion. In such systems entropy may be generated; i.e., the processes may be irreversible. Although they result in drastic modifications of the usefulness of the energy contained in a system, irreversible processes are difficult to describe as a function of time.†

The second difficulty is numerical and one that would persist even if the general equation of motion were known. In general, the number of properties required to characterize nonequilibrium states is much larger than that for stable equilibrium states. Accordingly the relations between properties of nonequilibrium states are numerically much more complicated than those for stable equilibrium states and, therefore, much more difficult to analyze.

Because of these two difficulties, the theoretical understanding of nonequilibrium states is limited, and much more work needs to be done. Thus far the applications of the principles of Chaps. 5 and 6 have been restricted primarily to a special class of states having fixed time rates of change of extensive properties which can be determined by the values of a relatively small number of properties. This number of properties, however, is larger than that required to characterize a stable equilibrium state. The resulting formalism has become known as *thermodynamics of steady states* or *rate processes.*

This chapter summarizes the concepts and procedures employed in thermodynamics of steady states, first as an extension of ideas applicable to stable equilibrium states and then from a more general point of view.

†The Boltzmann equation discussed in Sec. 8.8 is an equation of motion that describes irreversible processes. Its range of validity is very limited, however.

8.2 Steady-Rate Processes

8.2.1 Steady States. If system A is interacting with other systems I, II, III, . . ., and, by virtue of these interactions, the state of A is invariant with time, then A is said to be in a *steady state*. For example, if a copper bar A is interposed between two constant-temperature systems I and II (Fig. 8.1), then after an initial transient the condition of the bar is fixed by the temperatures of I and II and remains unaltered at all times. In other words, the bar is in a steady state.

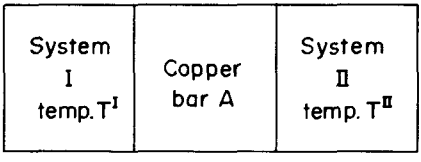


Fig. 8.1 Schematic of an ordinary closed system A in a steady state.

From this definition it follows that the values of all properties of A in a steady state are independent of time. The same statement applies to equilibrium states. The two types of states are different, however. In a steady state the time independence is the result of interactions of A with two or more systems, but in an equilibrium state the time independence is achieved in the absence of any interactions whatsoever with other systems in the environment. Thus a steady state is a nonequilibrium state since the condition of A would change with time if the interactions were suddenly interrupted.

All systems I, II, III, . . . interacting with A in a steady state cannot be in mutual stable equilibrium. If they were, A would reach a state in mutual stable equilibrium with them, and therefore A would be in a stable equilibrium state.

8.2.2 Fluxes and Flux Densities. In general, interactions of A with other systems in the environment involve transfer of matter, transfer of momentum, transfer of energy, etc. In a steady state the rate of transfer of an extensive, conserved property to A is fixed with time and equals the rate of transfer of the same property from A to other systems. For example, the rate of energy transferred from I to the copper bar (Fig. 8.1) must be time independent and must equal the rate of energy transferred from the bar to II. Otherwise the bar is not in a steady state.

The fixed rates of transfer of extensive properties to and from a system in a steady state also illustrate the difference between equilibrium and steady states. For a system in either an equilibrium or a stable equilibrium state, all rates of transfer are equal to zero, but this is not true for a system in a steady state.

The rate at which an extensive property i is transferred to or from a system in a steady state is called a *flux of i.* Sometimes it is convenient to express a flux as an integral of a flux per unit area over a given surface area. The flux per unit area of a property i is called the *flux density of i.* For example, the energy flux can be expressed as an integral of the energy flux density over the surface area over which energy is transferred either to the system or from that system to others.

Fluxes (flux densities) require scalar, vector, or tensor quantities for their representation, depending on the nature of the flux (flux density) and that of the system and its steady state. This point will be clarified in subsequent sections.

Because of the fixed fluxes associated with the interactions between A in a steady state and I, II, and III, . . ., the collection of systems A, I, II, III, . . . is said to experience a *steady-rate process.*

Finally, since the fluxes do not affect the steady state of A, they can be regarded as resulting from interactions between I, II, III, . . . not in mutual stable equilibrium via A in a steady state. To be sure, A plays an important role in establishing the steady-rate process. For example, in the process shown in Fig. 8.1, the copper bar is essential in maintaining the constant-temperature difference between systems I and II and fixes the value of the energy flux through the bar. For the same temperature difference but a bar of a different metal, the value of the energy flux would be different.

8.2.3 Fluxes and Thermodynamic Potentials. If systems A, I, II, III, . . . are experiencing a steady-rate process, the question arises, "How are the values of the fluxes related to the states of A, I, II, III, . . . ?" Thus far it has not been possible to give a general answer to this question.

An approximate answer can be given when each of systems I, II, III, . . . (but not A) can be regarded as passing through stable equilibrium states in spite of the nonzero fluxes associated with these systems. The approximation is partly due to the characterization of the states of I, II, III, . . . as stable when in fact *they are not* since the fluxes are not equal to zero.†

For a steady-rate process in which systems I, II, III, . . . can be regarded as passing through stable equilibrium states, we will show in the following paragraphs that the values of the fluxes must depend at least on the values of the thermodynamic potentials of I, II, III, ...

For simplicity we will consider the steady-rate process shown in Fig. 8.2, where A, I, and II are simple, one-component, open systems and I and II are in

†In this approximation, systems I, II, III, . . . are replaced by quasi reservoirs, i.e., reservoirs with intensive properties independent of the values of the fluxes.

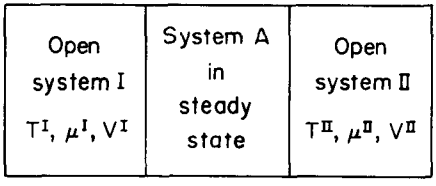


Fig. 8.2 Schematic of a simple one-component system A in a steady state.

stable equilibrium states defined by temperatures T^{I} and T^{II}, total potentials μ^{I} and μ^{II}, and volumes V^{I} and V^{II}.

If I and II were in mutual stable equilibrium, A would be in mutual stable equilibrium with them, and therefore all fluxes would be equal to zero. It follows that each flux must be a function of at least those quantities necessary to determine whether I and II are in mutual stable equilibrium.

In Sec. 7.4.3, it is shown that the quantities entering the conditions for mutual stable equilibrium are the thermodynamic potentials T and μ. Consequently each flux associated with the steady-rate process shown in Fig. 8.2 must be at least a function of the quantities T^{I}, μ^{I}, T^{II}, and μ^{II}. This conclusion can be readily generalized to other steady-rate processes involving systems in stable equilibrium states characterized by a sufficient number of thermodynamic potentials other than T and μ.

8.2.4 Alternate Form of Thermodynamic Potentials. The conditions for mutual stable equilibrium and the thermodynamic potentials appearing in these conditions can be written in a variety of forms. For example, as discussed in Sec. 7.2.4, different temperature scales can be defined. Accordingly the temperature equality can be written in different mathematical forms. For the purposes of the subsequent discussion of the Onsager relations, a set of thermodynamic potentials different from that given in Sec. 7.4.3 is derived.

If systems I and II are in mutual stable equilibrium, then, according to the criterion of stable equilibrium (Sec. 6.7.1), any change of state of the combined system I and II at fixed values of total energy, total number of particles of each distinguishable component, and constraints must correspond to a nonincrease of entropy. In symbols

$$dS^{I} + dS^{II} \leqslant 0 \tag{8.1}$$

where dS^{I} and dS^{II} are the entropy changes of I and II, respectively, under the conditions specified. For changes such that both I and II pass through stable equilibrium states, the entropy changes dS^{I} and dS^{II} can be computed by means

of either the generalized Gibbs equation (Eq. 7.79) or the Gibbs equation (Eq. 7.94), depending on the nature of I and II.

For the systems in Fig. 8.2 and for constant volume processes, the Gibbs equation yields

$$dS^{\mathrm{I}} = \tau^{\mathrm{I}}\, dE^{\mathrm{I}} - \mu^{\mathrm{I}}\tau^{\mathrm{I}}\, dn^{\mathrm{I}} \tag{8.2}$$

and

$$dS^{\mathrm{II}} = \tau^{\mathrm{II}}\, dE^{\mathrm{II}} - \mu^{\mathrm{II}}\tau^{\mathrm{II}}\, dn^{\mathrm{II}} \tag{8.3}$$

where τ denotes T^{-1} and where the changes of energy, dE^{I} and dE^{II}, and of number of particles, dn^{I} and dn^{II}, must satisfy the conditions

$$dE^{\mathrm{I}} = -dE^{\mathrm{II}} \qquad dn^{\mathrm{I}} = -dn^{\mathrm{II}} \tag{8.4}$$

and dE^{I} and dn^{I} can be either positive or negative. Substituting Eqs. 8.2 to 8.4 into Eq. 8.1 yields that I and II are in mutual stable equilibrium if and only if

$$\tau^{\mathrm{I}} = \tau^{\mathrm{II}} \qquad \mu^{\mathrm{I}}\tau^{\mathrm{I}} = \mu^{\mathrm{II}}\tau^{\mathrm{II}} \tag{8.5}$$

Volume V and the thermodynamic potentials τ and $\mu\tau$ represent another set of quantities sufficient to characterize a stable equilibrium state of a one-component open system with volume as an independent constraint. This conclusion and the discussion in the preceding section imply that each flux associated with the steady-rate process shown in Fig. 8.2 must be at least a function of the quantities τ^{I}, $\mu^{\mathrm{I}}\tau^{\mathrm{I}}$, τ^{II}, and $\mu^{\mathrm{II}}\tau^{\mathrm{II}}$. An analogous statement is applicable to systems with more distinguishable components and more constraints than those of the systems in Fig. 8.2. It will be seen later that the thermodynamic potentials in the form appearing in Eq. 8.5 are useful for defining the Onsager reciprocal relations.

8.2.5 Entropy Generation. If a collection of systems A, I, II, III, ... is experiencing a steady-rate process, the process may or may not be reversible depending on whether the entropy of all the systems involved in the process is invariant or increasing. In symbols

$$\sum_i dS^i = 0 \qquad \text{for a reversible steady-rate process} \tag{8.6a}$$

and

$$\sum_i dS^i > 0 \qquad \text{for an irreversible steady-rate process} \tag{8.6b}$$

where dS^i is the differential entropy change of the ith system and the summation is over all systems except A. System A is not included, because it is in a steady state, and therefore dS^A is equal to zero.

Conditions 8.6a and 8.6b can also be expressed per unit time so that

$$\sum_i \frac{dS^i}{dt} = 0 \qquad \text{for a reversible steady-rate process} \tag{8.7a}$$

and

$$\sum_i \frac{dS^i}{dt} > 0 \qquad \text{for an irreversible steady-rate process} \tag{8.7b}$$

The quantities $\Sigma_i\, dS^i$ and $\Sigma_i\, (dS^i/dt)$ are called the *entropy generation* and the *rate of entropy generation,* respectively, of the steady-rate process.

When I, II, III, . . . are restricted to pass through stable equilibrium states only, the entropy generation and the rate of entropy generation can be computed by means of the generalized Gibbs equation for each of the systems I, II, III, ... Thus it is possible to analyze whether or not such steady-rate processes are reversible. The method of analysis is given later.

8.3 Boundary-Value Processes

8.3.1 Definition. The general conclusion regarding the dependence of fluxes on thermodynamic potentials (Sec. 8.2.4) has not been explored in terms of practical mathematical relations, except for the special types of steady-rate processes discussed in this section.

If system A is in a steady state such that (1) each of the finite systems I, II, III, . . . interacting with A passes through stable equilibrium states only and (2) at any time the thermodynamic potentials associated with the stable equilibrium of I, II, III, . . . fully and uniquely define the steady state of A and the values of the fluxes, then the process will be called a *boundary-value process.* Boundary-value processes are the only ones that have been adequately analyzed thus far. They involve the restricted class of nonequilibrium states mentioned in Sec. 8.1 which are determined by a relatively small number of properties. These properties are the thermodynamic potentials of systems I, II, III, ...

This definition implies that any property of A and any flux associated with A can be expressed as a function of the thermodynamic potentials of I, II, III, . . . only. For example, if A in Fig. 8.2 experiences a boundary-value process, the flux I_i of property i of A is given by a relation of the form

$$\mathrm{I}_i = \mathrm{I}_i(\tau^{\mathrm{I}}, \mu^{\mathrm{I}}\tau^{\mathrm{I}}, \tau^{\mathrm{II}}, \mu^{\mathrm{II}}\tau^{\mathrm{II}}) \qquad \text{for all } i \tag{8.8}$$

This relation can also be written in the entirely equivalent form

$$I_i = I_i(\tau, \mu\tau, \Delta\tau, \Delta\mu\tau) \qquad \text{for all } i \tag{8.9}$$

where

$$\tau = \tau^{I} \qquad \mu\tau = \mu^{I}\tau^{I} \qquad \Delta\tau = \tau^{II} - \tau^{I} \qquad \Delta\mu\tau = \mu^{II}\tau^{II} - \mu^{I}\tau^{I} \tag{8.10}$$

The explicit dependence of I_i on the thermodynamic potentials is determined by the nature of A and its steady state (this is discussed later). Here it suffices to indicate only that

$$I_i(\tau, \mu\tau, 0, 0) \equiv 0 \qquad \text{for all } i \tag{8.11}$$

since all fluxes are equal to zero when systems I and II are in mutual stable equilibrium, namely, when

$$\Delta\tau = 0 \qquad \text{and} \qquad \Delta\mu\tau = 0 \tag{8.12}$$

Relations analogous to Eqs. 8.9 and 8.11 can be written for boundary-value processes involving a number of thermodynamic potentials larger than that considered in Fig. 8.2.

8.3.2 Linear and Nonlinear Processes. For a given boundary-value process, if the expressions for the fluxes are continuous functions of their arguments, each flux can be written in the form of a power series of the thermodynamic potential differences involved in the process. For example, for the process shown in Fig. 8.2, the flux I_i to or from system A can be expressed as a power series of the form

$$I_i = L_{i\tau}\,\Delta\tau - L_{i\mu\tau}\,\Delta\mu\tau + \left\{ \begin{array}{c} \text{higher order terms of} \\ \Delta\tau,\ \Delta\mu\tau \end{array} \right\} \tag{8.13}$$

where the values of the coefficients $L_{i\tau}$ and $L_{i\mu\tau}$, as well as those of the coefficients of the higher order terms, depend on the values of τ and $\mu\tau$ only. In addition, no term independent of $\Delta\tau$ and $\Delta\mu\tau$ appears in the expansion, because of Eq. 8.11.

When all fluxes of a boundary-value process depend only on the first power of each thermodynamic potential difference, the process is called a *linear process.* When all or some of the fluxes depend on powers of thermodynamic potential differences higher than the first, the boundary-value process is called a *nonlinear*

process. For small magnitudes of thermodynamic potential differences, linear processes can be regarded as first-order approximations to nonlinear processes.

The presentation that follows is restricted to linear processes both because they are better understood theoretically and because they can be used to represent many transport phenomena of interest to thermionic conversion.

8.4 One-Dimensional Rate Processes

8.4.1 Continuous Rate Processes. We will consider systems A, 1, and 2, shown in Fig. 8.3a, which are experiencing a boundary-value process without chemical reactions and such that A is separable spatially into laminas 1, 2, . . ., j, . . . by means of faces perpendicular to the x direction spaced a distance dx apart (Fig. 8.3b), and the laminas satisfy the following conditions: (1) any one lamina j can be regarded as an independent system partaking in a boundary-value process; (2) the boundary-value process for j is determined by two sets of thermodynamic potentials, one set for each face of the lamina; (3) the thermodynamic potentials at the interface between laminas j and $j + 1$ have the

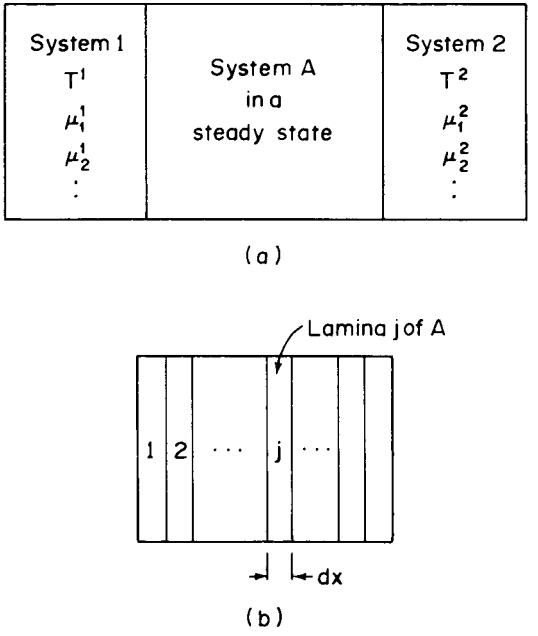


Fig. 8.3 System A in a steady state (a) which can be divided into independent systems 1, 2, . . ., j, . . . (b).

same values for both laminas; and (4) any thermodynamic potential can be regarded as a continuous function of the position x of the face.

When conditions 1 to 4 are satisfied, each lamina can be analyzed separately and the results for all laminas combined to yield the characteristics of the overall process. The overall process will be called a *continuous one-dimensional rate process.*

Specifically, if the thermodynamic potentials at the two faces of lamina j are τ, $\mu_1\tau$, $\mu_2\tau$, . . . and $\tau + d\tau$, $\mu_1\tau + d(\mu_1\tau)$, $\mu_2\tau + d(\mu_2\tau)$, . . ., this lamina can be regarded as part of the elementary process shown in Fig. 8.4, i.e., a lamina in a steady state by virtue of interactions with systems I and II, each of which is passing through stable equilibrium states. Systems I and II are not the same as systems 1 and 2 in Fig. 8.3. The condition of the lamina as part of the elementary process in Fig. 8.4 is entirely equivalent to that of the lamina as part of the process in Fig. 8.3.

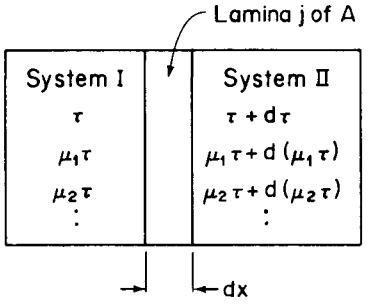


Fig. 8.4 Elementary process for lamina j entirely equivalent to that of the lamina as part of the process in Fig. 8.3. The values of τ and all $\mu\tau$ vary with position x of the lamina.

8.4.2 Entropy Generation in an Elementary Process. For the process shown in Fig. 8.4 and for each face of the lamina having a unit cross-sectional area, the rate of entropy generation, dJ_s, can be computed by means of the Gibbs equations for the two systems that pass through stable equilibrium states under conditions of constant volume (Sec. 8.2.5). Differentiating the two equations with respect to time and adding the results yield the relation

$$d\dot{S} = dJ_s$$
$$= \tau\dot{E}^{\mathrm{I}} - \sum_l \mu_l\tau\dot{n}_l^{\mathrm{I}} + [\tau + d\tau]\dot{E}^{\mathrm{II}} - \sum_l [\mu_l\tau + d(\mu_l\tau)]\dot{n}_l^{\mathrm{II}} \qquad (8.14)$$

where the dot above a symbol denotes the rate of change with time.

Now, when the energy-flux density and lth-component particle-flux density from system I to the lamina are denoted by J_ε and J_l, respectively, then in steady state

$$J_\varepsilon = -\dot{E}^{\mathrm{I}} = \dot{E}^{\mathrm{II}} \quad \text{and} \quad J_l = -\dot{n}_l^{\mathrm{I}} = \dot{n}_l^{\mathrm{II}} \quad \text{for } l = 1, 2, \ldots \tag{8.15}$$

Substituting Eqs. 8.15 into Eq. 8.14 yields

$$dJ_s = J_\varepsilon \, d\tau + \sum_l J_l \, d(-\mu_l \tau) \tag{8.16}$$

Equation 8.16 gives the differential rate of entropy generation in terms of fluxes and thermodynamic potential differences. It can also be written in the form of a differential equation with respect to x, which upon integration yields the rate of entropy generation in system A shown in Fig. 8.3. The differential equation is presented in Sec. 8.4.3.

8.4.3 Conjugate Fluxes and Forces. In general, we see from Eq. 8.16 that dJ_s is different from zero when the differentials of the thermodynamic potentials are not all equal to zero, namely, when the lamina is not in a stable equilibrium state. In addition to their relation to dJ_s, these differentials, called *forces,* are the causes for all the fluxes through the lamina.

The relation between the rate of entropy generation and the forces can be stated in the form of a differential equation. Differentiating both sides of Eq. 8.16 by the thickness dx of the lamina yields

$$\frac{dJ_s}{dx} = J_\varepsilon \frac{d\tau}{dx} + \sum_l J_l \frac{d(-\mu_l \tau)}{dx} \tag{8.17a}$$

or

$$J_s' = J_\varepsilon \tau' + \sum_l J_l (-\mu_l \tau)' \tag{8.17b}$$

where a prime denotes a derivative with respect to x. The derivatives τ' and $(-\mu_l \tau)'$ for all l are also called forces. Integrating Eq. 8.17b over the range of x defined by system A would yield the rate of entropy generation ΔJ_s in the process shown in Fig. 8.3.

A pair of flux and force appearing in either Eq. 8.16 or Eq. 8.17b is called *conjugates* or a *pair of conjugate flux and force.* For example, J_ε and $d\tau$ are conjugates, and J_l and $(-\mu_l \tau)'$ are also conjugates.

8.4.4 Flux Densities as Functions of a State and Forces. For the elementary boundary-value process shown in Fig. 8.4, the flux density J_m of extensive conserved property m can be expressed in the form

$$J_m = J_m[\tau, \mu_1\tau, \mu_2\tau, \ldots; \tau', (\mu_1\tau)', (\mu_2\tau)', \ldots] \quad \text{for all } m \tag{8.18}$$

This form is obtained from the general expression for J_m of a boundary-value process (Sec. 8.3.1) after substituting $\tau' dx$ for $\Delta\tau$ and $(\mu_l\tau)' dx$ for $\Delta(\mu_l\tau)$ for all l. In the form of Eq. 8.18, J_m is said to be a *function of the state* τ, $\mu_1\tau$, $\mu_2\tau$, . . . *and the forces* τ', $(\mu_1\tau)'$, $(\mu_2\tau)'$, ...

The function J_m can be expanded into a Maclaurin power series of the forces. For linear processes this series is of the form

$$J_m = \left(\frac{\partial J_m}{\partial \tau'}\right)_{\eta_k'=0} \tau' + \sum_l \left(\frac{\partial J_m}{\partial (\mu_l\tau)'}\right)_{\eta_k'=0} (\mu_l\tau)' \quad \text{for all } m \tag{8.19}$$

where the subscript $\eta_k' = 0$ denotes that in the evaluation of a partial derivative all values of the forces are taken equal to zero. The term $J_m[\tau, \mu_1\tau, \mu_2\tau, \ldots; \tau' = 0; (\mu_1\tau)' = 0, \ldots]$ does not appear in the Maclaurin expansion, because when all forces are equal to zero all fluxes are equal to zero.

The number of terms, G, in the right-hand side of Eq. 8.19 equals the number of thermodynamic potentials necessary to specify the stable equilibrium states of either system I or system II in Fig. 8.4. Moreover, the value of each of the derivatives $(\partial J_m/\partial\tau')_{\eta_k'=0}$ and $[\partial J_m/\partial(\mu_l\tau)']_{\eta_k'=0}$ depends only on the values of the thermodynamic potentials τ, $\mu_1\tau$, and $\mu_2\tau$, which define the stable equilibrium state of I, or simply depends only on the state.

The number of forces in the right-hand side of Eq. 8.17b for the rate of entropy generation is G; i.e., the number is the same as that in the right-hand side of Eq. 8.19 for any flux density. Since each force is associated with a conjugate flux, the number of flux densities in the right-hand side of Eq. 8.17b is also equal to G. Accordingly, if each of the fluxes in the right-hand side of Eq. 8.17b is expressed in the form of Eq. 8.19, a set of G linear equations results. This set of equations, which involves a square matrix of coefficients of the forces, is known as the *set of phenomenological equations of a linear process.*

Each coefficient of the set of phenomenological equations is represented by the symbol L with appropriate subscripts, as indicated by the restatement of Eq. 8.19 in the specific form

$$J_m = L_{m\varepsilon}\tau' + L_{m1}(-\mu_1\tau)' + L_{m2}(-\mu_2\tau)' + \ldots \tag{8.20}$$

Comparing Eqs. 8.19 and 8.20 indicates that each coefficient L is determined by a partial derivative of J_m with respect to one of the forces. From the discussion of these derivatives, it follows that each coefficient is independent of the value of any force and depends only on the values of τ, $\mu_1\tau$, $\mu_2\tau$, ...

In a general coefficient, L_{mn}, subscript m denotes the property m in the flux equation in which the coefficient appears and subscript n denotes the force that is multiplied by the coefficient. When the subscripts are identical, the coefficient multiplies the force conjugate to the flux in question.

Each coefficient L_{mn} is called a *conductivity* and can be expressed in terms of more-familiar conductivities such as the thermal conductivity of Fourier's law of heat conduction or the electric conductivity of Ohm's law of current flow, etc. Examples of such relations are given later.

8.4.5 Onsager Reciprocal Relations. For the set of phenomenological equations of the form of Eq. 8.20, it can be shown that the square matrix of the coefficients L_{mn} is symmetrical. In other words

$$L_{mn} = L_{nm} \qquad \text{for all } m \text{ and } n \tag{8.21}$$

The equalities represented by Eq. 8.21 are known as the *Onsager reciprocal relations.*

These reciprocal relations were originally derived by Onsager in 1931 for two-dimensional heat conduction in an anisotropic medium.[1] In 1945 Casimir[2] showed them to be of general validity for linear processes. Subsequently De Groot[3] and Prigogine[4] applied the theory to a wide variety of phenomena.

8.5 Applications

8.5.1 General Remarks. The phenomenological equations can be used to analyze and correlate a variety of processes involving flows of energy and matter. These processes are often referred to as generalized diffusion phenomena and include, among other things, thermoelectric effects and transport of charged particles through ionized gases.

The following sections treat in detail a number of processes describable by the phenomenological equations. The purposes of these treatments are manyfold: first, to provide some familiarity with the use of the phenomenological equations; second, to illustrate the simplifications introduced into the analysis by the Onsager reciprocal relations; third, to indicate how the various coefficients L_{mn} can be measured experimentally; and, finally, to show relations that exist between the coefficients L_{mn} and more familiar quantities such as electrical and thermal conductivities and diffusion coefficients.

8.5.2 Flow of a Substance Through a Porous Barrier. We will consider a substance consisting of identical particles that flow through a porous barrier A (Fig. 8.5) so that a linear phenomenological equation holds for each flux. For

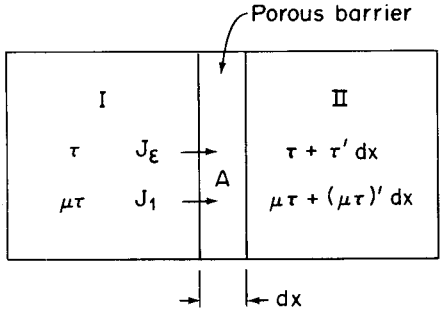


Fig. 8.5 Flow of a substance through a porous barrier.

example, the substance could be a gas like helium and the barrier could be a solid material like a stretched rubber membrane. The substance on either side of the membrane is regarded to be approximately in a stable equilibrium state.

The rate of entropy generation J_s' per unit length and per unit cross-sectional area of A is given by the relation

$$J_s' = J_\varepsilon \tau' + J_1 (-\mu_1 \tau)' \tag{8.22}$$

where J_ε denotes the energy-flux density and J_1 the particle-flux density through A. The flux densities J_ε and J_1 are given by the phenomenological equations

$$J_\varepsilon = L_{\varepsilon\varepsilon}\tau' + L_{\varepsilon 1} (-\mu_1 \tau)' \tag{8.23}$$

and

$$J_1 = L_{1\varepsilon}\tau' + L_{11} (-\mu_1 \tau)' \tag{8.24}$$

Equations 8.23 and 8.24 can be substituted into Eq. 8.22 to establish the conditions under which the flow of the substance is irreversible. Indeed, the substitution yields that J_s' is given by the expression

$$J_s' = L_{\varepsilon\varepsilon} (\tau')^2 - 2L_{\varepsilon 1}\tau' (\mu_1 \tau)' + L_{11} [(\mu_1 \tau)']^2 \tag{8.25}$$

where use is made of the Onsager reciprocal relation

$$L_{\varepsilon 1} = L_{1\varepsilon} \tag{8.26}$$

By virtue of the principle of increase of entropy (Sec. 6.6.4), J_s' must be nonnegative for all values of τ' and $(\mu_1 \tau)'$. For this to be true, the necessary and

sufficient conditions that must be satisfied by the conductivities L_{mn} are given by the relations

$$L_{\varepsilon\varepsilon} \geqslant 0 \qquad L_{11} \geqslant 0 \tag{8.27a}$$

$$L_{\varepsilon\varepsilon}L_{11} - L_{\varepsilon 1}^2 \geqslant 0 \tag{8.27b}$$

When condition 8.27b is satisfied with the inequality sign, the process is irreversible because J_S' is positive. This procedure, for the derivation of the conditions satisfied by the conductivities L_{mn} for a process to be irreversible, can be readily extended to any linear process involving many components and many fluxes.

Several relations between the quantities appearing in Eqs. 8.23 and 8.24 can be deduced from the requirement of the Onsager reciprocal relation (Eq. 8.26). Among many other uses, these relations suggest procedures for experimental measurement of the conductivities.

For example, the values of the forces can be adjusted so that the particle-flux density J_1 is equal to zero. Equation 8.24 shows that this happens when the ratio of the forces satisfies the relation

$$\left[-\frac{(-\mu_1\tau)'}{\tau'}\right]_{J_1=0} = \left[\frac{\Delta(\mu_1\tau)}{\Delta\tau}\right]_{J_1=0} = \frac{L_{1\varepsilon}}{L_{11}} \tag{8.28}$$

where Δ denotes small differences. Since each L in the ratio $L_{1\varepsilon}/L_{11}$ is fixed by the thermodynamic potentials τ and $\mu_1\tau$ and is independent of the values of the fluxes, we see from Eq. 8.28 that the ratio $\Delta(\mu_1\tau)/\Delta\tau$ for zero particle flux is also independent of the values of the fluxes.

On the other hand, the particle flow may occur under isothermal conditions such that $\tau' = 0$. For example, isothermal flow occurs when systems A, I, and II in Fig. 8.5 are in contact with an ordinary closed system III at temperature T equal to $1/\tau$ (Fig. 8.6). Under this condition Eqs. 8.23 and 8.24 yield

$$\left(\frac{J_\varepsilon}{J_1}\right)_{\tau'=0} = \frac{L_{\varepsilon 1}}{L_{11}} \tag{8.29}$$

For the reasons stated, we see from Eq. 8.29 that the ratio $(J_\varepsilon/J_1)_{\tau'=0}$ depends on the values of the thermodynamic potentials τ and $\mu_1\tau$ only.

By virtue of the Onsager reciprocal relation, the ratios $L_{1\varepsilon}/L_{11}$ and $L_{\varepsilon 1}/L_{11}$ are identical, and therefore Eqs. 8.28 and 8.29 yield

$$\left[\frac{\Delta(\mu_1\tau)}{\Delta\tau}\right]_{J_1=0} = \left(\frac{J_\varepsilon}{J_1}\right)_{\tau'=0} \tag{8.30}$$

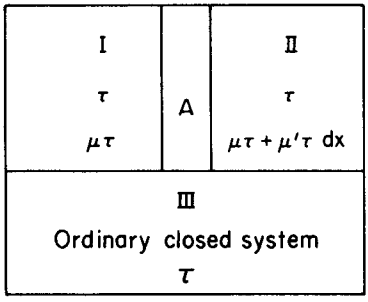


Fig. 8.6 Isothermal flow of a substance through barrier A.

This result can be restated in somewhat more familiar terms. To this end, using Eq. 7.145 yields

$$\frac{d(\mu_1 \tau)}{d\tau} = \frac{d}{d\tau}(h_1 \tau - s_1) = h_1 + \tau \frac{dh_1}{d\tau} - \frac{ds_1}{d\tau} \tag{8.31}$$

where h_1 and s_1 are the specific enthalpy and specific entropy, respectively, per particle of the substance. In addition, for any one-component simple system

$$T\,ds_1 = dh_1 - v_1\,dp \tag{8.32a}$$

or

$$\tau \frac{dh_1}{d\tau} - \frac{ds_1}{d\tau} = v_1 \tau \frac{dp}{d\tau} \tag{8.32b}$$

where v_1 is the specific volume and p is the pressure. Combining Eqs. 8.31 and 8.32b yields

$$\frac{d(\mu_1 \tau)}{d\tau} = h_1 + v_1 \tau \frac{dp}{d\tau} \tag{8.33a}$$

or, in terms of small differences of thermodynamic potentials,

$$\frac{\Delta(\mu_1 \tau)}{\Delta\tau} = h_1 + v_1 \tau \frac{\Delta p}{\Delta\tau} \tag{8.33b}$$

It follows from Eqs. 8.30 and 8.33b that

$$\left(\frac{J_\varepsilon}{J_1}\right)_{\tau'=0} = h_1 + \nu_1 \tau \left(\frac{\Delta p}{\Delta \tau}\right)_{J_1=0} \tag{8.34a}$$

or

$$\left(\frac{\Delta p}{\Delta \tau}\right)_{J_1=0} = \frac{1}{\nu_1 \tau}\left[\left(\frac{J_\varepsilon}{J_1}\right)_{\tau'=0} - h_1\right] \tag{8.34b}$$

Equation 8.34b gives the ratio of pressure difference over temperature difference of systems I and II (Fig. 8.5) for which the particle flux is stopped. Before we consider an interesting application of this result to thermionic converters, given later, it is instructive to discuss the quantities appearing in Eq. 8.34b.

The ratio $(J_\varepsilon/J_1)_{\tau'=0}$, usually denoted by λ_1, i.e.,

$$\lambda_1 \equiv \left(\frac{J_\varepsilon}{J_1}\right)_{\tau'=0} \tag{8.35}$$

is called the *isothermal energy of transport per particle.* It represents the energy-flux density by virtue of flow of particles and without any additional mode of energy transfer such as heat conduction. The value of λ_1 is not equal to the specific energy e_1 of the substance in the stable equilibrium state characterized by τ and $\mu_1 \tau$, because in a flow process the fraction of fast particles crossing a unit area per unit time is greater than the fraction existing in the stable equilibrium state.

The value λ_1 depends on the nature of both the substance and the barrier through which it flows. For example, for flow of a liquid or a gas through a barrier with pores that are large compared with the mean free path of the particles, λ_1 is given by the relation[5]

$$\lambda_1 \equiv \left(\frac{J_\varepsilon}{J_1}\right)_{\tau'=0} = h_1 \tag{8.36}$$

In general, this case corresponds to bulk flow through a pipe. For bulk flow of a monatomic perfect gas, the specific enthalpy h_1 (and, therefore, λ_1) equals $(\frac{5}{2})kT$, whereas the specific energy e_1 equals $(\frac{3}{2})kT$ (Sec. 7.8.13). For other cases, λ_1 also differs from h_1.

In general, the difference q_1^*, given by the relation

$$q_1^* \equiv \left(\frac{J_\varepsilon}{J_1}\right)_{\tau'=0} - h_1 \equiv \lambda_1 - h_1 \tag{8.37}$$

is defined as the *heat of transport per particle.* It can be shown[6] that q_1^* is equal to the heat passing from system II to system I via system III in Fig. 8.6 when one particle is transferred from I to II via A.

The value of the heat of transport can be computed for a few special cases only. For example, for bulk flow q_1^* equals zero. Again, for a perfect gas flowing through pores that are small compared with the mean free path of the gas molecules, q_1^* is found[7] to be equal to $-kT/2$.

Conditions of flow through small pores are equivalent to those prevailing in a low-pressure thermionic converter in the following sense. By analogy with the process shown in Fig. 8.5, the cesium-vapor environment of a thermionic converter can be regarded as system I, the interelectrode spacing as a small pore of A, and one of the electrodes as system II. In steady state the flux of cesium from the cesium vapor to the electrode via the interelectrode space is equal to zero. When the cesium-vapor pressure is so low that the mean free path of cesium atoms is much larger than the interelectrode spacing,

$$q_1^* \equiv \left(\frac{J_\varepsilon}{J_1}\right)_{\tau'=0} - h_1 = -\frac{kT}{2}$$

Substituting this value into Eq. 8.34b yields

$$\left(\frac{\Delta p}{\Delta \tau}\right)_{J_1=0} = -\frac{p}{2\tau} \tag{8.38}$$

or

$$\frac{p^{\mathrm{I}}}{p^{\mathrm{II}}} = \left(\frac{T^{\mathrm{I}}}{T^{\mathrm{II}}}\right)^{1/2} \tag{8.39}$$

where use has been made of the perfect gas relation

$$pv_1 = kT$$

Equation 8.39, known as the *Knudsen effect,* means that, when the cesium vapor is at temperature T^{I} and pressure p^{I} and the electrode is at temperature T^{II}, the pressure at the electrode is p^{II} different from p^{I}. The pressure p^{II} is given by Eq. 8.39, provided the conditions for the validity of this equation are satisfied.

Finally, a relation between the conductivities L_{mn} and other, more familiar conductivities can be derived as follows. Eliminating $(\mu_1 \tau)'$ from Eqs. 8.23 and 8.24 yields that J_ε is given by the relation

$$J_\varepsilon = \frac{L_{\varepsilon 1}}{L_{11}} J_1 + \left(L_{\varepsilon\varepsilon} - \frac{L_{\varepsilon 1}^2}{L_{11}}\right)\tau' = \lambda_1 J_1 + J_k \tag{8.40}$$

where λ_1 is the isothermal energy of transport (Eqs. 8.29 and 8.35) and J_k is defined by the identity

$$J_k \equiv \left(L_{\varepsilon\varepsilon} - \frac{L_{\varepsilon 1}^2}{L_{11}}\right)\tau' \equiv (J_\varepsilon)_{J_1 = 0} \tag{8.41}$$

The quantity J_k is called the *Fourier heat-flux density.*

For J_1 equal to zero, the energy flux through the barrier is in the form of heat. This heat can be computed by means of the *Fourier equation for heat conduction* given by the relation

$$(J_\varepsilon)_{J_1 = 0} = -k_A \frac{dT}{dx} = k_A \frac{\tau'}{\tau^2} \tag{8.42}$$

where k_A denotes the *thermal conductivity* of the barrier. Combining Eqs. 8.41 and 8.42 yields

$$(J_\varepsilon)_{J_1 = 0} = J_k = \left(L_{\varepsilon\varepsilon} - \frac{L_{\varepsilon 1}^2}{L_{11}}\right)\tau' = k_A \frac{\tau'}{\tau^2} \tag{8.43a}$$

or

$$L_{\varepsilon\varepsilon} - \frac{L_{\varepsilon 1}^2}{L_{11}} = k_A T^2 \tag{8.43b}$$

Equation 8.43b represents one of the relations between the coefficients L of the phenomenological equations and more-familiar conductivities such as k_A.

8.6 Thermoelectric Effects

8.6.1 General Remarks. The term "thermoelectric effects" refers to the coupled flows of energy and electrons through solids resulting from applied thermodynamic potential differences. The study of these effects provides another illustration of the concepts of steady-rate processes. What follows is a discussion of some concepts and relations associated with an elementary thermoelectric process useful to thermionic conversion.

8.6.2 One-Dimensional Electron Flow. We will consider electrons flowing only in the x direction through a conductor of unit cross-sectional area under

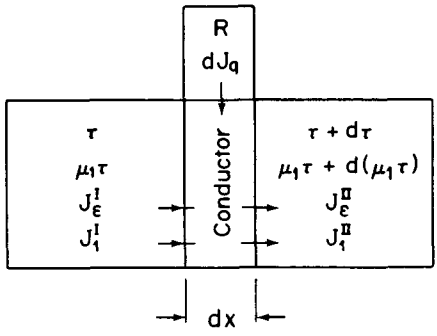


Fig. 8.7 Elementary thermoelectric process.

conditions that can be approximated by a continuous one-dimensional rate process of the following type. When the conductor is subdivided into elementary laminas of length dx, each lamina experiences a boundary-value process by virtue of interactions with systems I, II, and R, as shown in Fig. 8.7. It is seen from this figure that systems I and II are substituted for adjacent pieces of conductor and R is used to supply only a heat flux dJ_q. Each of systems I and II is passing approximately through stable equilibrium states.

For the lamina in Fig. 8.7, the energy-flux density J_ε and electron-flux density J_1 are defined by the relations

$$J_\varepsilon \equiv J_\varepsilon^{\mathrm{I}} \equiv -\dot{E}^{\mathrm{I}} \tag{8.44}$$

and

$$J_1 \equiv J_1^{\mathrm{I}} \equiv -\dot{n}_1^{\mathrm{I}} \tag{8.45}$$

where E^{I} is the energy and n_1^{I} the number of electrons of system I and the dot denotes a time derivative. Similarly, the energy- and electron-flux densities for system II are defined by the relations

$$J_\varepsilon^{\mathrm{II}} \equiv \dot{E}^{\mathrm{II}} \qquad J_1^{\mathrm{II}} \equiv \dot{n}_1^{\mathrm{II}} \tag{8.46}$$

Since the lamina is in a steady state, the preceding flux densities must satisfy the relations

$$dJ_q = J_\varepsilon^{\mathrm{II}} - J_\varepsilon^{\mathrm{I}} = dJ_\varepsilon \tag{8.47}$$

and

$$J_1^{II} - J_1^{I} = dJ_1 = 0 \tag{8.48}$$

In words, Eq. 8.47 is a statement of conservation of energy, and Eq. 8.48 is a statement of conservation of particles (electrons) as applied to the process shown in Fig. 8.7.

Finally, Eq. 8.17b, combined with the fact that dJ_q is negligible compared with J_ε (Eq. 8.47), yields that the rate of entropy generation J_s' per unit length is given by the relation

$$J_s' = J_\varepsilon \tau' - J_1 (\mu_1 \tau)' \tag{8.49}$$

where μ_1 is the total potential of the electrons. It is noteworthy that J_s' in Eq. 8.49 is independent of the heat-flux density dJ_q. The reason is that, for all practical purposes, this heat flux is supplied to the lamina under isothermal conditions, and therefore it does not involve any entropy generation.

8.6.3 Phenomenological Equations for a Homogeneous Conductor. For the process in Fig. 8.7, conjugate fluxes and forces are defined by Eq. 8.49. Accordingly the phenomenological equations for J_ε and J_1 can be expressed in the forms

$$J_\varepsilon = L_{\varepsilon\varepsilon} \tau' - L_{\varepsilon 1} (\mu_1 \tau)' \tag{8.50}$$

and

$$J_1 = L_{\varepsilon 1} \tau' - L_{11} (\mu_1 \tau)' \tag{8.51}$$

where use has been made of the Onsager reciprocal relation $L_{1\varepsilon} = L_{\varepsilon 1}$.

Eliminating $(\mu_1 \tau)'$ from Eqs. 8.50 and 8.51 yields that J_ε is given by the relation

$$J_\varepsilon = \frac{L_{\varepsilon 1}}{L_{11}} J_1 + \left(L_{\varepsilon\varepsilon} - \frac{L_{\varepsilon 1}^2}{L_{11}} \right) \tau' = \lambda_1 J_1 + J_k \tag{8.52}$$

where λ_1 is the isothermal energy of transport and J_k is the Fourier heat-flux density. Relation 8.52 is used in the energy-balance equations of a thermionic converter.

Differentiating Eq. 8.52 with respect to x and using Eqs. 8.47 and 8.48 yield

$$J_\varepsilon' = J_q' = \lambda_1' J_1 + J_k' \tag{8.53}$$

Since the coefficients L do not depend on the values of the forces, the derivative λ_1' of λ_1 with respect to x can be written in the form

$$\lambda_1' = \left(\frac{\partial \lambda_1}{\partial \tau}\right)_{\mu_1} \tau' + \left(\frac{\partial \lambda_1}{\partial \mu_1}\right)_{\tau} \mu_1' \tag{8.54}$$

and therefore

$$J_q' = \left(\frac{\partial \lambda_1}{\partial \tau}\right)_{\mu_1} J_1 \tau' + \left(\frac{\partial \lambda_1}{\partial \mu_1}\right)_{\tau} J_1 \mu_1' + J_k' \tag{8.55}$$

Finally, the quantity λ_1 can be used to express the electron-flux density J_1 in the form

$$J_1 = L_{11}\left[\frac{L_{\varepsilon 1}}{L_{11}} \tau' - (\mu_1 \tau)'\right] = L_{11}\,[(\lambda_1 - \mu_1)\tau' - \tau \mu_1']\tag{8.56}$$

8.6.4 Special Case 1: Heat Conduction for $J_1 = 0$. For J_1 equal to zero, the energy flow through the conductor is in the form of heat. This heat can be computed by means of the Fourier equation for heat conduction and is given by the relation

$$(J_\varepsilon)_{J_1 = 0} = -k_M \frac{dT}{dx} = k_M \frac{\tau'}{\tau^2} \tag{8.57}$$

where k_M denotes the thermal conductivity of the conductor. On the other hand

$$(J_\varepsilon)_{J_1 = 0} = J_k = \left(L_{\varepsilon\varepsilon} - \frac{L_{\varepsilon 1}^2}{L_{11}}\right)\tau' \tag{8.58}$$

From Eqs. 8.57 and 8.58 it follows that

$$L_{\varepsilon\varepsilon} - \frac{L_{\varepsilon 1}^2}{L_{11}} = k_M T^2 \tag{8.59}$$

For a homogeneous conductor of finite length and fixed k_M, we see from Eqs. 8.55 and 8.57 that, when J_1 and J_q' are equal to zero, the temperature distribution across the conductor is linear. On the other hand, when J_1 is equal to zero but J_q' is independent of x, the temperature distribution is parabolic.

8.6.5 Special Case 2: Isothermal Electron Flow for $\tau' = 0$. For isothermal electron flow, systems I, II, and R in Fig. 8.7 must be at the same temperature $1/\tau$.

Now the electric potential difference $d\psi$ between two points of a conductor, isothermal or nonisothermal, is defined by the relation

$$d\psi \equiv -\frac{d\mu_1}{e} \tag{8.60}$$

where e is the charge of the electron and $d\mu_1$ is the difference between the total potentials of the two points. Moreover, the electric-current density J_e resulting from an electron-flux density J_1 is given by the relation

$$J_e = -eJ_1 \tag{8.61}$$

For an isothermal conductor the electric-potential difference $d\psi$ and the electron current density J_e are related by *Ohm's equation*, namely,

$$\frac{d\psi}{dx} = \psi' = -rJ_e \tag{8.62}$$

where r is the *electric resistivity* of the conductor. By virtue of Eqs. 8.60 and 8.61, Ohm's equation can be expressed in the form

$$\mu_1' = -re^2 \, (J_1)_{\tau'=0} \tag{8.63}$$

On the other hand, for $\tau' = 0$, Eq. 8.51 yields

$$(J_1)_{\tau'=0} = -L_{11}\tau\mu_1' \tag{8.64}$$

Combining Eqs. 8.63 and 8.64 results in the relation

$$L_{11} = \frac{T}{re^2} \tag{8.65}$$

Maintaining an electric conductor at constant and uniform temperature requires a continuous exchange of heat between the conductor and other systems since heat is generated as current flows through a resistor. The rate of heat exchange $(J_q')_{\tau'=0}$ per unit length (the Joulean heat) is given by the relation

$$(J_q')_{\tau'=0} = -rJ_e^2 \tag{8.66}$$

where the minus sign indicates that heat flows away from the conductor regardless of the direction of the electric-current flow. The same heat rate can also be computed by using Eq. 8.55, which, under isothermal conditions, yields

$$(J'_q)_{\tau'=0} = \left(\frac{\partial \lambda_1}{\partial \mu_1}\right) \mu'_1 (J_1)_{\tau'=0} \tag{8.67}$$

Comparing Eqs. 8.66 and 8.67 and using Eq. 8.63 result in the relation

$$\left(\frac{\partial \lambda_1}{\partial \mu_1}\right)_\tau = 1 \tag{8.68}$$

After integration Eq. 8.68 yields

$$\lambda_1 = \mu_1 + \frac{s_1(\tau)}{\tau} \tag{8.69}$$

where $s_1(\tau)$ is a function of temperature only. We can easily verify that $s_1(\tau)$ is the *isothermal entropy of transport per electron.*

8.6.6 Electric-Conduction Equation. Substituting Eqs. 8.65 and 8.69 into Eq. 8.56 yields

$$J_1 = \frac{1}{re^2}\left[\frac{s_1(\tau)}{\tau^2}\tau' - \mu'_1\right] \tag{8.70}$$

or, by virtue of Eqs. 8.60 and 8.61,

$$\psi' = -\frac{\mu'_1}{e} = -\frac{s_1(\tau)}{e\tau^2}\tau' - rJ_e \tag{8.71}$$

Equation 8.70 gives the electron-flux density in terms of the gradients of temperature and total potential. Equation 8.71 gives the electric-potential gradient in terms of the temperature gradient and the electron current density. In a thermionic converter this equation is used to relate the electric-potential difference across the load to the difference between the μ's of the electrodes.

The preceding relations can be further rearranged to define and discuss such thermoelectric effects and coefficients as the Thomson effect and the Thomson coefficient, the Peltier effect and the Peltier coefficient, the Seebeck effect and the Seebeck coefficient, the Kelvin relations between various coefficients, and, finally, the performance characteristics of a thermoelectric generator. These rearrangements and discussions are beyond the scope of this book, however, and for their study the reader is referred to the literature.[8]

8.7 Chemical Reactions

8.7.1 General Remarks. The rate processes discussed up to this point did not involve chemical reactions. Accordingly the conditions for mutual stable equilibrium did not include the equation of chemical equilibrium (Eq. 7.87), and the phenomenological equations did not include forces arising from total potential differences that do not satisfy the equation of chemical equilibrium.

Chemical reactions provide a mechanism by which sets of components in different interacting systems can differ. For example, H_2 and O_2, which are each actual components of two systems, can combine upon entering a third system to form H_2O, which is a component of the third system but not of the first and the second.

This section presents the procedures used in analyzing rate processes involving chemical reactions. Because of its importance in thermionic conversion, the only example discussed is that of a mixture of ionized gases consisting of electrons, positive ions, and neutral particles capable of reacting with each other in accordance with an ionization equation. The results can be generalized to other processes involving one or more chemical reactions other than ionization.

8.7.2 Three-Component Three-Temperature Ionized Gases. We will consider a gaseous system A of unit cross-sectional area which is in a steady state and which consists of a molecular species X and its products of dissociation defined by the chemical reaction

$$X \rightleftharpoons X^+ + e^- \tag{8.72}$$

where X, X^+, and e^- denote molecular species, singly charged positive ionic species, and electron species, respectively, which for brevity will be called molecules, ions, and electrons.

The phenomena experienced by A can be subdivided into *spatial* and *local*. The spatial phenomena involve the transfer of energy and particles from one position to another. The local phenomena involve the transfer of energy from one species to another and the transformation of two species into the third. The local phenomena occur at each elementary volume.

In general, the analysis of the spatial and local phenomena is very complicated except when A can be regarded as a system involved in a continuous rate process and each species of A can be regarded as a one-component system in a steady state by virtue of interactions with the other species and with systems in the environment. The following discussion is applicable to a gaseous system A that is separable into three independent systems A_e, A_i, and A_a consisting of electrons, ions, and molecules, respectively. Each of the three one-component systems

occupies the same Cartesian space and is in a steady state defined by interactions with the other two and with systems in the environment passing approximately through stable equilibrium states. For convenience all spatial phenomena will be assumed to depend on the Cartesian coordinate x only.

In the described gaseous system A, a lamina dA of length dx can be represented by the systems and interactions shown in Fig. 8.8. We see from this

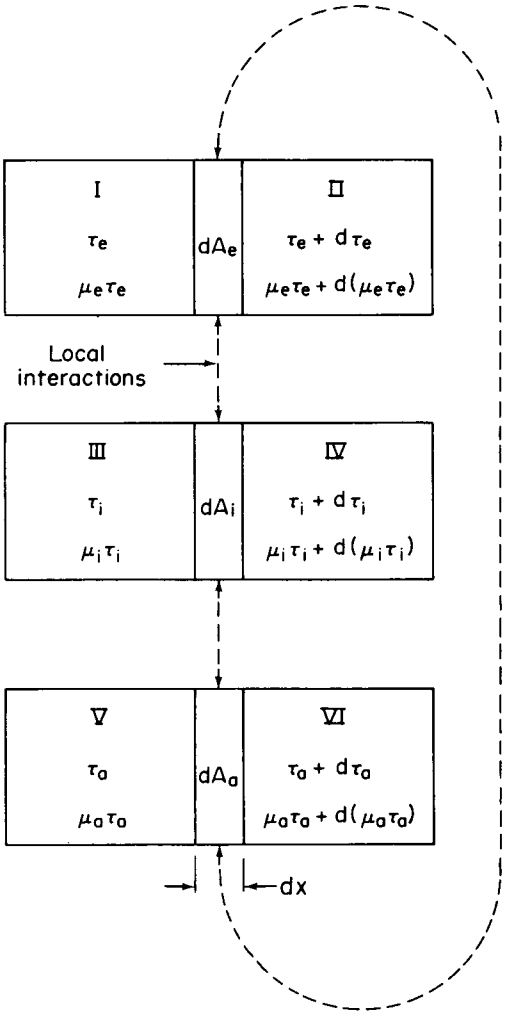


Fig. 8.8 Elementary process in a three-component three-temperature mixture of electrons (e), ions (i), and molecules (a).

figure that lamina dA_e of the electron system is maintained in a steady state by interactions with systems I and II passing through stable equilibrium states and by interactions with systems dA_i and dA_a. The interactions with I and II contribute to the spatial phenomena in dA_e. The interactions with dA_i and dA_a contribute to the local phenomena in dA_e. Because of these local phenomena, the fluxes at the interface between I and dA_e are not equal to the fluxes between dA_e and II. Similar remarks can be made about systems dA_i and dA_a and their interactions with others. Finally, the temperatures of systems I, III, and V in Fig. 8.8 are assumed to be different from each other. For this reason and because of the presence of three components, A will be called a *three-component three-temperature ionized gas*.

The conjugate fluxes and forces involved in the elementary process shown in Fig. 8.8 can be determined from the expression for the rate of entropy generation dJ_s. This expression is discussed in Sec. 8.7.3.

8.7.3 Rate of Entropy Generation. Because of the complexity of the process under consideration and in order to avoid multiple subscripts, some quantities are represented by symbols different from those used previously. All new symbols are defined as needed.

In terms of the thermodynamic potentials shown in Fig. 8.8, the rate of entropy generation dJ_s is given by the relation

$$\begin{aligned} dJ_s = {} & \tau_e \dot{E}^{\mathrm{I}} - \mu_e \tau_e \dot{n}_e^{\mathrm{I}} + (\tau_e + d\tau_e)\dot{E}^{\mathrm{II}} - [\mu_e \tau_e + d(\mu_e \tau_e)]\, \dot{n}_e^{\mathrm{II}} \\ & + \tau_i \dot{E}^{\mathrm{III}} - \mu_i \tau_i \dot{n}_i^{\mathrm{III}} + (\tau_i + d\tau_i)\dot{E}^{\mathrm{IV}} - [\mu_i \tau_i + d(\mu_i \tau_i)]\, \dot{n}_i^{\mathrm{IV}} \\ & + \tau_a \dot{E}^{\mathrm{V}} - \mu_a \tau_a \dot{n}_a^{\mathrm{V}} + (\tau_a + d\tau_a)\dot{E}^{\mathrm{VI}} - [\mu_a \tau_a + d(\mu_a \tau_a)]\, \dot{n}_a^{\mathrm{VI}} \end{aligned} \tag{8.73}$$

where E^{I} to E^{VI} are the energies of systems I to VI, and n_e^{I} and n_e^{II} are the numbers of electrons of systems I and II, etc.

Since the laminas dA_e, dA_i, and dA_a are individually and collectively in steady states and the process is continuous, it follows that

$$u_e = -\dot{E}^{\mathrm{I}} \quad \Gamma_e = -\dot{n}_e^{\mathrm{I}} \quad \dot{E}^{\mathrm{II}} = u_e + \frac{\partial u_e}{\partial x}\, dx \quad \dot{n}_e^{\mathrm{II}} = \Gamma_e + \frac{\partial \Gamma_e}{\partial x}\, dx \tag{8.74}$$

$$u_i = -\dot{E}^{\mathrm{III}} \quad \Gamma_i = -\dot{n}_i^{\mathrm{III}} \quad \dot{E}^{\mathrm{IV}} = u_i + \frac{\partial u_i}{\partial x}\, dx \quad \dot{n}_i^{\mathrm{IV}} = \Gamma_i + \frac{\partial \Gamma_i}{\partial x}\, dx \tag{8.75}$$

$$u_a = -\dot{E}^{\mathrm{V}} \quad \Gamma_a = -\dot{n}_a^{\mathrm{V}} \quad \dot{E}^{\mathrm{VI}} = u_a + \frac{\partial u_a}{\partial x}\, dx \quad \dot{n}_a^{\mathrm{VI}} = \Gamma_a + \frac{\partial \Gamma_a}{\partial x}\, dx \tag{8.76}$$

$$\frac{\partial u_e}{\partial x} + \frac{\partial u_i}{\partial x} + \frac{\partial u_a}{\partial x} = u_e' + u_i' + u_a' = 0 \tag{8.77}$$

$$\frac{\partial \Gamma_e}{\partial x} = \frac{\partial \Gamma_i}{\partial x} = -\frac{\partial \Gamma_a}{\partial x} = \dot{\xi}' \tag{8.78a}$$

or

$$\Gamma_e' = \Gamma_i' = -\Gamma_a' = \dot{\xi}' \tag{8.78b}$$

where u_e = electron-energy-flux density from I to dA_e
u_i = ion-energy-flux density from III to dA_i
u_a = molecule-energy-flux density from V to dA_a
Γ_e = electron-flux density from I to dA_e
Γ_i = ion-flux density from III to dA_i
Γ_a = molecule-flux density from V to dA_a

Equation 8.77 is a statement of conservation of energy in the overall system dA. Equations 8.78a and 8.78b are statements of conservation of charged and neutral particles. The quantity $\dot{\xi}'$ represents the *ionization rate per unit length* (the chemical reaction rate per unit length).

Substituting Eqs. 8.74 to 8.78 into Eq. 8.73 and neglecting second-order effects yield, after some rearranging, the relation

$$dJ_s = u_e\, d\tau_e + u_i\, d\tau_i + u_a\, d\tau_a - \Gamma_e\, d(\mu_e\tau_e) - \Gamma_i\, d(\mu_i\tau_i) - \Gamma_a\, d(\mu_a\tau_a) + (u_i'\, dx)\,\Delta\tau_{ie} + (u_a'\, dx)\,\Delta\tau_{ae} - (\dot{\xi}'\, dx)\,\Delta(\mu\tau)_R \tag{8.79}$$

or

$$J_s' = u_e\tau_e' + u_i\tau_i' + u_a\tau_a' - \Gamma_e\,(\mu_e\tau_e)' - \Gamma_i\,(\mu_i\tau_i)' - \Gamma_a\,(\mu_a\tau_a)' + u_i'\,\Delta\tau_{ie} + u_a'\,\Delta\tau_{ae} - \dot{\xi}'\,\Delta(\mu\tau)_R \tag{8.80}$$

where

$$\Delta\tau_{ie} = \tau_i - \tau_e \qquad \Delta\tau_{ae} = \tau_a - \tau_e \tag{8.81}$$

$$\Delta(\mu\tau)_R = \mu_e\tau_e + \mu_i\tau_i - \mu_a\tau_a \tag{8.82}$$

It is clear from Eq. 8.80 that nine conjugate pairs of fluxes and forces are associated with the elementary process shown in Fig. 8.8. They are listed in

Table 8.1 Conjugate Pairs of Fluxes and Forces Associated with Elementary Processes Shown in Fig. 8.8.

Index	Forces	Fluxes
1	$\tau_e' \propto$ electron temperature gradient	u_e = energy-flux density in electron system
2	$\tau_i' \propto$ ion temperature gradient	u_i = energy-flux density in ion system
3	$\tau_a' \propto$ molecule temperature gradient	u_a = energy-flux density in molecule system
4	$-(\mu_e\tau_e)' \propto$ electron total potential gradient	Γ_e = electron-flux density
5	$-(\mu_i\tau_i)' \propto$ ion total potential gradient	Γ_i = ion-flux density
6	$-(\mu_a\tau_a)' \propto$ molecule total potential gradient	Γ_a = molecule-flux density
7	$\Delta\tau_{ie} \propto$ ion-electron temperature difference	$\frac{\partial u_i}{\partial x} = u_i'$, energy-flux density per unit length of ion system
8	$\Delta\tau_{ae} \propto$ molecule–electron temperature difference	$\frac{\partial u_a}{\partial x} = u_a'$, energy-flux density per unit length of molecular system
9	$-\Delta(\mu\tau)_R \propto$ total potential difference	$\dot{\xi}'$ = ionization rate per unit length

Table 8.1, where each pair is tagged with a number from 1 to 9. The conditions necessary for all nine forces to be equal to zero coincide with the conditions necessary for all systems in Fig. 8.8 to be in mutual stable equilibrium, namely,

$\tau_e = \tau_i = \tau_a = \tau$ independent of x

μ_e, μ_i, and μ_a independent of x

and

$\mu_e + \mu_i - \mu_a = 0$

8.7.4 Uncoupling of Spatial and Local Phenomena. If the forces are relatively small, each of the flux densities in Table 8.1 can be expressed by a phenomenological equation of the form

$$(\text{Flux density})_i = \sum_j L_{ij}\,(\text{force})_j \qquad \text{for } i, j = 1, 2, \ldots, 9 \tag{8.83}$$

where the subscripts $i, j = 1, 2, \ldots, 9$ are chosen as indicated in Table 8.1 and the coefficients L_{ij} satisfy the Onsager reciprocal relations and are defined operationally by the expression

$$L_{ij} = \left(\frac{\partial(\text{flux density})_i}{\partial(\text{force})_j}\right)_{\text{all forces} = 0} \tag{8.84}$$

Many systems of practical interest involving chemical reactions can be approximated by laminas for which the forces of types 1 to 6 are small. The same statement, however, may not apply to forces of types 7 to 9, and therefore the phenomenological equations may be grossly approximate for the effects of these forces. Nonetheless, if we assume that, for the ionized gases under consideration, all nine forces are relatively small, then nine phenomenological equations can be written for the flux densities listed in Table 8.1.

These phenomenological equations require 81 coefficients L_{ij}. By virtue of the Onsager reciprocal relations, only 45 of the coefficients are independent. Moreover, 18 of the 45 coefficients are equal to zero for the reason discussed in the following paragraph.

In general, fluxes and forces do not have the same tensorial character; e.g., some are scalars and some are vectors. Accordingly under symmetry operations, such as rotations and reflections, the Cartesian components of these quantities transform in different ways. Consequently an effect of symmetry properties of the material of a system may be that the components of the fluxes do not depend on all components of the forces. This is known as the *Curie symmetry principle.*[9]

By virtue of the Curie symmetry principle, in the three-component three-temperature ionized gases under consideration, fluxes 1 to 6 are not coupled to forces 7 to 9. Thus the number of coefficients L_{ij} is reduced by 18 or, alternatively, the six phenomenological equations for the spatial flux densities u_e, u_i, u_a, Γ_e, Γ_i, and Γ_a are not coupled to the three phenomenological equations for the local flux densities u_i', u_a', and $\dot{\xi}'$. These two sets of equations are discussed separately.

8.7.5 Spatial Phenomenological Equations. The spatial phenomenological equations consist of six linear equations expressing each of the flux densities u_e, u_i, u_a, Γ_e, Γ_i, and Γ_a in terms of the six forces τ_e', τ_i', τ_a', $(\mu_e\tau_e)'$, $(\mu_i\tau_i)'$, and $(\mu_a\tau_a)'$. They are

$$u_e = L_{11}\tau_e' + L_{12}\tau_i' + L_{13}\tau_a' - L_{14}\,(\mu_e\tau_e)' - L_{15}\,(\mu_i\tau_i)' - L_{16}\,(\mu_a\tau_a)' \tag{8.85}$$

$$u_i = L_{21}\tau_e' + L_{22}\tau_i' + L_{23}\tau_a' - L_{24}(\mu_e\tau_e)' - L_{25}(\mu_i\tau_i)' - L_{26}(\mu_a\tau_a)' \quad (8.86)$$

$$u_a = L_{31}\tau_e' + L_{32}\tau_i' + L_{33}\tau_a' - L_{34}(\mu_e\tau_e)' - L_{35}(\mu_i\tau_i)' - L_{36}(\mu_a\tau_a)' \quad (8.87)$$

$$\Gamma_e = L_{41}\tau_e' + L_{42}\tau_i' + L_{43}\tau_a' - L_{44}(\mu_e\tau_e)' - L_{45}(\mu_i\tau_i)' - L_{46}(\mu_a\tau_a)' \quad (8.88)$$

$$\Gamma_i = L_{51}\tau_e' + L_{52}\tau_i' + L_{53}\tau_a' - L_{54}(\mu_e\tau_e)' - L_{55}(\mu_i\tau_i)' - L_{56}(\mu_a\tau_a)' \quad (8.89)$$

$$\Gamma_a = L_{61}\tau_e' + L_{62}\tau_i' + L_{63}\tau_a' - L_{64}(\mu_e\tau_e)' - L_{65}(\mu_i\tau_i)' - L_{66}(\mu_a\tau_a)' \quad (8.90)$$

Solving Eqs. 8.88 to 8.90 for $(\mu_e\tau_e)'$, $(\mu_i\tau_i)'$, and $(\mu_a\tau_a)'$ and substituting the results into Eqs. 8.85 to 8.87 yield

$$u_e = k_{ee}\tau_e' + k_{ei}\tau_i' + k_{ea}\tau_a' + \lambda_{ee}\Gamma_e + \lambda_{ei}\Gamma_i + \lambda_{ea}\Gamma_a \quad (8.91)$$

$$u_i = k_{ie}\tau_e' + k_{ii}\tau_i' + k_{ia}\tau_a' + \lambda_{ie}\Gamma_e + \lambda_{ii}\Gamma_i + \lambda_{ia}\Gamma_a \quad (8.92)$$

$$u_a = k_{ae}\tau_e' + k_{ai}\tau_i' + k_{aa}\tau_a' + \lambda_{ae}\Gamma_e + \lambda_{ai}\Gamma_i + \lambda_{aa}\Gamma_a \quad (8.93)$$

where the quantities $k_{\alpha\beta}$ and $\lambda_{\alpha\beta}$ for $\alpha,\beta = e$, i, and a are transport coefficients that are combinations of the L_{ij}'s of Eqs. 8.85 to 8.90. By virtue of the Onsager reciprocal relations, we can verify that

$$k_{\alpha\beta} = k_{\beta\alpha} \qquad \text{for } \alpha,\beta = e,\ i, \text{ and } a$$

Each of Eqs. 8.91 to 8.93 appears as a sum of three conduction-like terms plus three convection-like terms. Accordingly each coefficient $k_{\alpha\beta}$ is interpreted as a thermal conductivity and each $\lambda_{\alpha\beta}$ as an isothermal energy of transport per particle. For example, k_{ee} represents the energy flux density entering dA_e and due only to a unit force τ_e'. This flux arises because of energy transfer during elastic collisions between the particles of dA_e and those of systems I and II in Fig. 8.8. On the other hand, k_{ei} represents the energy flux density entering dA_e and due only to a unit force τ_i'. This flux arises because of energy transfer during elastic collisions between the particles of dA_e and those of systems III, dA_i, and IV in Fig. 8.8. Similar statements are applicable to the other thermal conductivities.

Again, the isothermal energy of transport λ_{ee} represents the energy transferred by an electron from I to II in Fig. 8.8 as a result of a unit force $\tau_e(\mu_e)'$ only. On the other hand, λ_{ei} represents the energy transferred by an electron from I to II as a result of a unit force $\tau_i(\mu_i)' = \tau_e(\mu_i)'$ only.

If no particle diffusion were present and only bulk flow existed, $\lambda_{\alpha\alpha}$ for $\alpha = e$, i, and a would be equal to the specific enthalpy h_α, and $\lambda_{\alpha\beta}$ for $\alpha \neq \beta$ would be

equal to zero. This interpretation of the $\lambda_{\alpha\beta}$ follows directly from Eqs. 8.91 to 8.93 and suggests that heats of transport $q^*_{\alpha\beta}$ can be defined by the relations

$$q^*_{ee} = \lambda_{ee} - h_e \qquad q^*_{ei} = \lambda_{ei} \qquad q^*_{ea} = \lambda_{ea} \tag{8.94}$$

$$q^*_{ie} = \lambda_{ie} \qquad q^*_{ii} = \lambda_{ii} - h_i \qquad q^*_{ia} = \lambda_{ia} \tag{8.95}$$

$$q^*_{ae} = \lambda_{ae} \qquad q^*_{ai} = \lambda_{ai} \qquad q^*_{aa} = \lambda_{aa} - h_a \tag{8.96}$$

The quantities $q^*_{\alpha\beta}$ for all α and β are measures of the departure of the flows of particles from bulk flows. They are identically equal to zero when bulk-flow conditions prevail.

8.7.6 **Local Phenomenological Equations.** In general, the forces $\Delta\tau_{ie}$, $\Delta\tau_{ae}$, and $\Delta(\mu\tau)_R$ are relatively large, and the relations between the fluxes u'_i, u'_a, and $\dot{\xi}'$ and the forces are nonlinear. Nevertheless, if the forces $\Delta\tau_{ie}$, $\Delta\tau_{ae}$, and $\Delta(\mu\tau)_R$ are assumed to be relatively small, then the local phenomenological equations are given by the relations

$$u'_i = L_{77}\,\Delta\tau_{ie} + L_{78}\,\Delta\tau_{ae} - L_{79}\,\Delta(\mu\tau)_R \tag{8.97}$$

$$u'_a = L_{87}\,\Delta\tau_{ie} + L_{88}\,\Delta\tau_{ae} - L_{89}\,\Delta(\mu\tau)_R \tag{8.98}$$

$$\dot{\xi}' = L_{97}\,\Delta\tau_{ie} + L_{98}\,\Delta\tau_{ae} - L_{99}\,\Delta(\mu\tau)_R \tag{8.99}$$

These relations can be combined to yield the expressions

$$u'_i = k_{11}\,\Delta\tau_{ie} + k_{12}\,\Delta\tau_{ae} + (h_i + q^*_i)\dot{\xi}' \tag{8.100}$$

$$u'_a = k_{21}\,\Delta\tau_{ie} + k_{22}\,\Delta\tau_{ae} + (h_a + q^*_a)\dot{\xi}' \tag{8.101}$$

$$\frac{\dot{\xi}}{L_{99}} = (h_i + q^*_i)\,\Delta\tau_{ie} + (h_a + q^*_a)\,\Delta\tau_{ae} - \Delta(\mu\tau)_R \tag{8.102}$$

where definitions for all coefficients are analogous to those for the spatial phenomenological equations and where $k_{12} = k_{21}$. Here, however, energy transfer is associated with inelastic rather than elastic collisions, and, moreover, these collisions occur between dissimilar particles, say, electrons and ions.

8.7.7 **Approximate Calculation of Chemical Potentials.** To evaluate the chemical potential part of each total potential μ, we can assume that each of the

one-component systems in Fig. 8.8 is describable by the perfect-gas approximation. Thus the chemical potential can be computed by means of the perfect-gas relations and can be expressed in terms of partial properties. For example, using Eq. 7.238 for the electron system yields

$$\mu_e\tau_e = \mu_{ce}\tau_e + \psi\tau_e$$

$$= k \ln p_e - \frac{5}{2} k \ln T_e - k \ln \frac{g_e k (2\pi m_e k)^{3/2}}{h^3} + \bar{\varepsilon}_{oe}\tau_e + \psi\tau_e \tag{8.103}$$

where ψ is the electron motive of the applied electric field (the negative of the product of the charge of the electron and the electrostatic potential), and the other quantities are as defined in Sec. 7.10.1. Differentiating Eq. 8.103 with respect to x yields

$$(\mu_e\tau_e)' = k\frac{p_e'}{p_e} - \frac{5}{2} k \frac{T_e'}{T_e} + \bar{\varepsilon}_{oe}\tau_e' + eE_x\tau_e + \psi\tau_e'$$

$$= k\frac{n_e'}{n_e} - kT_e\tau_e' + eE_x\tau_e + \left(\frac{5}{2}kT_e + \bar{\varepsilon}_{oe} + \psi\right)\tau_e' \tag{8.104a}$$

where E_x is the applied field, e is the charge of the electron, n_e is the electron density per unit volume, and use is made of the perfect-gas relation

$$p_e v_e = \frac{p_e}{n_e} = kT_e$$

The quantity $(5/2)kT_e + \bar{\varepsilon}_{oe} + \psi$ is the total enthalpy h_e per electron in the presence of the applied field, and therefore

$$(\mu_e\tau_e)' = k\frac{n_e'}{n_e} - kT_e\tau_e' + eE_x\tau_e' + h_e\tau_e' \tag{8.104b}$$

Similar expressions can be written for $(\mu_i\tau_i)'$ and $(\mu_a\tau_a)'$.

Regarding the force $\Delta(\mu\tau)_R$, repeated use of Eq. 7.238 yields

$$\Delta(\mu\tau)_R = -k \ln \left[\frac{n_a}{n_e n_i}\left(\frac{T_e T_i}{T_a}\right)^{3/2} C \exp\left(-\frac{V_i}{kT_e}\right)\right]$$

$$+ (\bar{\varepsilon}_{oi} - \psi)\,\Delta\tau_{ie} - \bar{\varepsilon}_{oa}\,\Delta\tau_{ae} \tag{8.105}$$

where V_i is the ionization energy of the molecule and C is a constant given by the relation

$$C = \frac{g_e g_i}{g_a}\left(\frac{m_e m_i}{m_a}\right)^{3/2}\frac{(2\pi k)^{3/2}}{h^3}$$

Using Eq. 8.104b and the analogous expressions for $(\mu_i\tau_i)'$ and $(\mu_a\tau_a)'$ in Eqs. 8.91 to 8.93 yields that

$$u_e = k_{ee}\tau_e' + k_{ei}\tau_i' + k_{ea}\tau_a' + (h_e + q_{ee}^*)\Gamma_e + q_{ei}^*\Gamma_i + q_{ea}^*\Gamma_a \tag{8.106}$$

$$u_i = k_{ie}\tau_e' + k_{ii}\tau_i' + k_{ia}\tau_a' + q_{ie}^*\Gamma_e + (h_i + q_{ii}^*)\Gamma_i + q_{ia}^*\Gamma_a \tag{8.107}$$

$$u_a = k_{ae}\tau_e' + k_{ai}\tau_i' + k_{aa}\tau_a' + q_{ae}^*\Gamma_e + q_{ai}^*\Gamma_i + (h_a + q_{aa}^*)\Gamma_a \tag{8.108}$$

Similarly, if the forces $\Delta\tau_{ie}$, $\Delta\tau_{ae}$, and $\Delta(\mu\tau)_R$ are assumed to be relatively small, using Eq. 8.105 in the local phenomenological equations yields

$$u_i' = k_{11}\ \Delta\tau_{ie} + k_{12}\ \Delta\tau_{ae} + (h_i + q_i^*)\dot{\xi}' \tag{8.109}$$

$$u_a' = k_{21}\ \Delta\tau_{ie} + k_{22}\ \Delta\tau_{ae} + (h_a + q_a^*)\dot{\xi}' \tag{8.110}$$

$$\frac{\xi'}{L_{99}} = \left(\frac{5}{2}kT_i + q_i^*\right)\Delta\tau_{ie} + \left(\frac{5}{2}kT_a + q_a^*\right)\Delta\tau_{ae} + k\ln\left[\frac{n_a}{n_e n_i}\left(\frac{T_e T_i}{T_a}\right)^{3/2} C\exp\left(-\frac{V_i}{kT_e}\right)\right] \tag{8.111}$$

For the subsequent discussion it is helpful to combine Eqs. 8.109, 8.110, and 8.77 and find that u_e', u_i', and u_a' are given by the relations

$$u_e' = (k_{11} + k_{21})\,\Delta\tau_{ei} + (k_{12} + k_{22})\,\Delta\tau_{ea} - (h_i + h_a + q_i^* + q_a^*)\dot{\xi}' \tag{8.112}$$

$$u_i' = (k_{11} + k_{12})\,\Delta\tau_{ie} - k_{12}\,\Delta\tau_{ia} + (h_i + q_i^*)\dot{\xi}' \tag{8.113}$$

and

$$u_a' = -k_{21}\,\Delta\tau_{ae} + (k_{21} + k_{22})\,\Delta\tau_{ae} + (h_a + q_a^*)\dot{\xi}' \tag{8.114}$$

where

$$\Delta\tau_{\alpha\beta} = \tau_\alpha - \tau_\beta \qquad \text{for } \alpha, \beta = e, i, \text{ and } a \tag{8.115}$$

Since each u_α' is positive when energy flows into system α and since energy flows from a system of high temperature to one of lower temperature, it follows from Eqs. 8.112 to 8.115 that the conductivities k_{ij} must satisfy the relations

$$k_{11} + k_{12} > 0 \qquad k_{22} + k_{12} > 0 \qquad k_{12} = k_{21} < 0 \tag{8.116}$$

and therefore

$$k_{11} > 0 \qquad k_{22} > 0 \tag{8.117}$$

Finally, in spite of the perfect-gas approximation for the chemical potentials, the preceding phenomenological equations for three-component three-temperature ionized gases are still quite involved and have not been studied. Instead, various further approximate versions of these equations have been investigated. In one of these versions, the following approximations are introduced.

$$k_{\alpha\beta} = q^*_{\alpha\beta} = 0 \qquad \text{for } \alpha \neq \beta \text{ and } \alpha,\beta = e, i, \text{ and } a \tag{8.118}$$

$$L_{mn} = 0 \qquad \text{for } m \neq n \text{ and } m,n = 4, 5, \text{ and } 6 \tag{8.119}$$

and

$$-k_{12} = -k_{21} \gg 0 \qquad k_{12} + k_{11} = k_{12} + k_{22} = 0 \tag{8.120}$$

The justifications for these approximations are as follows: The ions and molecules have approximately the same mass, but the mass of the electrons is much smaller. Because of the difference in mass, energy exchanges between electrons and the heavier components are negligible, and the flux of electrons is not affected by the states of the ions and molecules. The analytical equivalent of these remarks is reflected in Eqs. 8.118 and 8.119. Finally, because ions and molecules have approximately equal masses, the energy exchange between them is very effective and the temperature difference between them is negligible. This observation, together with the one regarding the small electron mass, leads to the approximation represented by Eqs. 8.120.

Using approximations 8.118 to 8.120 clearly leads to substantial simplifications of the phenomenological equations. Further simplifications may be introduced depending on the importance of each of the components of the mixture to the application under consideration. For example, under certain conditions prevailing in a thermionic converter, the energetics of the heavy particles are not important. Thus the only quantities of interest are the electron-energy and electron-flux densities and the ion-particle-flux density. By virtue of the approximations represented by Eqs. 8.118 to 8.120, these densities are given by the relations

$$u_e = k_{ee}\tau'_e + (h_e + q^*_{ee})\Gamma_e \tag{8.121}$$

$$\Gamma_e = L_{44}\left[(kT_e + q^*_{ee})\tau'_e - k\frac{n'_e}{n_e} - eE_x\tau_e\right] \tag{8.122}$$

$$\Gamma_i = L_{55}\left[(kT_i + q^*_{ii})\tau'_i - k\frac{n'_i}{n_i} + eE_x\tau_i\right] \tag{8.123}$$

$$u'_e = (h_i + h_a + q^*_i + q^*_a)\dot{\xi}' = (h_e + q^*_e)\dot{\xi}' \tag{8.124}$$

$$\frac{\dot{\xi}'}{L_{99}} = (kT_i + q^*_i + q^*_a)\,\Delta\tau_{ie} + k\ln\left[\frac{n_a}{n_e n_i}T^{3/2}C\exp\left(-\frac{V_i}{kT_e}\right)\right] \tag{8.125}$$

It is important to emphasize again that Eqs. 8.124 and 8.125 are grossly approximate since, in problems of interest to thermionic converters, the forces $\Delta\tau_{ie}$ and $\Delta(\mu\tau)_R$ are relatively large.

8.8 Kinetic Theory of Semiperfect Gases

8.8.1 General Remarks. Up to this point the phenomenological equations for various steady-rate processes have been discussed without reference to the explicit dependence of the conductivities on the values of the thermodynamic potentials. If the phenomenological equations are to be useful in analyzing a given steady-rate process, however, it is clear that the explicit values of the conductivities must be known.

This section presents some theoretical approximate methods for calculating the values of the conductivities. The methods are applicable to systems in steady states near stable equilibrium which can be regarded as semiperfect gases.

All statements about steady-rate processes made in the preceding sections are based on the criterion for stable equilibrium. Although this criterion yields valuable information regarding possible relations between fluxes and forces associated with a system A, it provides no information about the interdependence of these quantities and the properties of A. For example, it does not provide the means for calculating the explicit values of the conductivities L as functions of the properties of A.

In principle the analysis of nonequilibrium states, steady or others, must be approached by means of the concepts and postulates discussed in Chap. 6, which imply the following program. As a result of interactions with other systems, system A can be passing through nonequilibrium states characterized by a density operator $\hat{\rho}$ (Sec. 6.10.2). This operator is an explicit function of time and other variables. If the complete equation of motion for $\hat{\rho}$ were known, then, given the state $\hat{\rho}$ at one instant of time, we can determine the state of A at all times.

Now, by suitable mathematical manipulations of the density operator $\hat{\rho}$ and operators corresponding to properties, the values of properties and fluxes of A can be computed, and thus the analysis of the nonequilibrium states of A would be completed. Approximations can be introduced in the analysis so that, for example, for a continuous-rate process the relations between fluxes and forces become linear. These relations, however, would be complete in that, given the forces, we would know the values of fluxes and vice versa.

Unfortunately, thus far this program has not been and cannot be carried out rigorously because the complete equation of motion for $\hat{\rho}$ is not known. Except for some reversible adiabatic processes (Sec. 6.10.5), the equation of motion for other reversible adiabatic processes and for irreversible processes is not given by Eq. 6.33.

The preceding fundamental difficulty notwithstanding, Boltzmann postulated an equation for analyzing nonequilibrium states of semiperfect gases which has found a broad range of applications in a variety of fields. The theory based on Boltzmann's equation, called the *kinetic theory of semiperfect gases*, is used for calculating the conductivities of the phenomenological equations for systems which experience continuous-rate processes and which can be regarded as semiperfect gases.

The general definition of semiperfect gases and the Boltzmann equation are discussed in Secs. 8.8.2 and 8.8.3.

8.8.2 **General Definition of Semiperfect Gases.** We will consider an open system A consisting of identical particles and passing through nonequilibrium states by virtue of interactions with other systems. The states of A are such that it can be regarded as a dilute gas for which the one-particle approximation is valid (Sec. 7.8.2).

In addition to the one-particle approximation, we will assume that A satisfies the following conditions.

1. System A is subdivisible into small, homogeneous, open parts or phases. Around position $\mathbf{r}$ the volume of the phase is $d^3\mathbf{r}$. For the subdivision to be possible, the potential energy within each phase must be uniform and the spectrum of eigenfunctions of each phase must be continuous. In other words, the volume $d^3\mathbf{r}$ around position $\mathbf{r}$ must be sufficiently small so that the phase is approximately homogeneous, and yet $d^3\mathbf{r}$ must be sufficiently large so that the continuous-eigenvalue-spectrum approximation is applicable (Sec. 7.8.9). Because of the small $d^3\mathbf{r}$, the values of the properties of all phases collectively can be regarded as continuous functions of position $\mathbf{r}$.

2. Each phase is passing through nonequilibrium states by virtue of interactions with neighboring phases.

3. The states of each phase are such that the probability of finding more than one particle in any volume $d^3\mathbf{r}$ is negligibly small and need not be considered in

the analysis. Accordingly the states of each open phase can be expressed in terms of the eigenstates corresponding to zero and one particle only.

It can be verified that condition 3 for nonequilibrium states is equivalent to the requirement for the validity of the Boltzmann approximations (Sec. 7.8.7) for the corresponding stable equilibrium states. Here corresponding nonequilibrium and stable equilibrium states have the same expectation values of energy and number of particles but different values of entropy.

When the one-particle approximation and conditions 1 to 3 are valid, the system will be defined as a semiperfect gas because in the limit of stable equilibrium states the definition is identical to that given in Sec. 7.8.11.

8.8.3 The Boltzmann Equation. We will consider an arbitrary, open, homogeneous, one-component phase of volume $d^3\mathbf{r}$ around position $\mathbf{r}$ in a state $[\rho(\mathbf{r},t)]$ in the momentum representation. The argument $\mathbf{r}$ is used in $[\rho(\mathbf{r},t)]$ to indicate its association with the independent phase around $\mathbf{r}$ and not a variation of the elements of $[\rho(\mathbf{r},t)]$ within the phase. Such a variation would be meaningless.

The $(\alpha m;\alpha m)$th diagonal element $\rho_{\alpha m;\alpha m}$ of $[\rho(\mathbf{r},t)]$ is equal to the probability that, upon measurement, the phase is found to have a total momentum eigenvalue $\mathbf{M}_{\alpha m}(\mathbf{r})$ corresponding to an eigenvalue $n_\alpha(\mathbf{r})$ of the number of particles. For discrete spectra the total momentum eigenvalue $\mathbf{M}_{\alpha m}(\mathbf{r})$ is given by the relation

$$\mathbf{M}_{\alpha m}(\mathbf{r}) = \sum_j \nu_{\alpha m}(\mathbf{r},\mathbf{p}_j)\,\mathbf{p}_j \tag{8.126}$$

where $\mathbf{p}_j$ is the momentum eigenvalue of a single particle of the phase, $\nu_{\alpha m}(\mathbf{r},\mathbf{p}_j)$ is the number of particles with momentum $\mathbf{p}_j$, and

$$\sum_j \nu_{\alpha m}(\mathbf{r},\mathbf{p}_j) = n_\alpha(\mathbf{r}) \qquad \text{for all } \alpha \text{ and } m \tag{8.127}$$

In addition, the expectation value $\nu(\mathbf{r},\mathbf{p}_j,t)$ of the number of particles in the phase which have momentum $\mathbf{p}_j$ is given by the relation

$$\nu(\mathbf{r},\mathbf{p}_j,t) = \sum_{\alpha,m} \rho_{\alpha m;\alpha m}\nu_{\alpha m} \tag{8.128}$$

and the expectation value $n(\mathbf{r},t)$ of the total number of particles in the phase is given by the relation

$$\begin{aligned} n(\mathbf{r},t) &= \sum_{\alpha,m} \rho_{\alpha m;\alpha m} n_\alpha = \sum_{j,\alpha,m} \rho_{\alpha m;\alpha m}\,\nu_{\alpha m}(\mathbf{r},\mathbf{p}_j) \\ &= \sum_j \nu(\mathbf{r},\mathbf{p}_j,t) \end{aligned} \tag{8.129}$$

The definition of $\nu(\mathbf{r},\mathbf{p}_j,t)$ is analogous to that of the one-particle distribution function associated with stable equilibrium states (Sec. 7.8.2). It is noteworthy, however, that, although a given state $[\rho(\mathbf{r},t)]$ corresponds to one expectation value $\nu(\mathbf{r},\mathbf{p}_j,t)$, a given one-particle distribution function $\nu(\mathbf{r},\mathbf{p}_j,t)$ for all $\mathbf{p}_j$ corresponds, in general, to a very large number of states $[\rho(\mathbf{r},t)]$. One important difference between the various states corresponding to a given $\nu(\mathbf{r},\mathbf{p}_j,t)$ is the value of the entropy. This value may range from zero for a phase in a pure state to a maximum consistent with $\nu(\mathbf{r},\mathbf{p}_j,t)$. In particular, if $\nu(\mathbf{r},\mathbf{p}_j,t)$ is a Maxwellian distribution, the maximum is that of the stable equilibrium state corresponding to the values of energy, number of particles, and constraints of the phase.†

For a phase in states that are adequately represented in terms of the eigenstates corresponding to zero and one particles only, the preceding relations are greatly simplified. For example, the probability of finding a particle in the phase with a momentum $\mathbf{p}_j$ is given by the diagonal element $\rho_{1j;1j}$ of $[\rho(\mathbf{r},t)]$ corresponding to the eigenstate of the phase with one particle of momentum $\mathbf{p}_j$ ($n_1 = 1$). Again, from Eq. 8.128 it follows that the expectation value $\nu(\mathbf{r},\mathbf{p}_j,t)$ of the number of particles with momentum $\mathbf{p}_j$ is a number smaller than unity and equal to $\rho_{1j;1j}$, namely,

$$\nu(\mathbf{r},\mathbf{p}_j,t) = \rho_{1j;1j} \tag{8.130}$$

We see from Eq. 8.130 that $\nu(\mathbf{r},\mathbf{p}_j,t)$ uniquely defines $[\rho(\mathbf{r},t)]$ and, therefore, the state.

The requirement for the validity of Eq. 8.130 is satisfied by all phases of a semiperfect gas. Accordingly for a semiperfect gas the expectation value of the number of particles of a phase of volume $d^3\mathbf{r}$ around position $\mathbf{r}$ with a precise momentum is equal to the probability of finding a particle with that momentum in the phase. The discussion that follows applies to semiperfect gases only.

In addition to continuity of values of all properties with respect to $\mathbf{r}$, the momentum eigenstates of a semiperfect gas can be regarded as continuous. Thus the expectation value of the number of particles of a phase can be expressed in the form $f(\mathbf{r},\mathbf{p},t)\, d^3\mathbf{r}\, d\mathbf{p}$ or, equivalently, in the form $f(\mathbf{r},\mathbf{v},t)\, d^3\mathbf{r}\, d\mathbf{v}$, where $\mathbf{v}$ is the velocity corresponding to $\mathbf{p}$. The two spaces $\mathbf{r}$ and $\mathbf{v}$ together are called the *μ-space.*

At time t the expectation value $f(\mathbf{r},\mathbf{v},t)\, d^3\mathbf{r}\, d\mathbf{v}$ represents the probability of finding a particle in $d^3\mathbf{r}$ around $\mathbf{r}$ with velocity between $\mathbf{v}$ and $\mathbf{v} + d\mathbf{v}$.

†A little appreciated corollary of this statement is that in a many-particle system a Maxwellian distribution is necessary but *not sufficient* to characterize a stable equilibrium state. The insufficiency is equally true for the Fermi–Dirac and the Bose–Einstein factors (Sec. 7.8.4).

Alternatively, the quantity $f(\mathbf{r},\mathbf{v},t)$ is a continuous function of both $\mathbf{r}$ and $\mathbf{v}$ and represents the expected number of particles at time t per unit volume $d^3\mathbf{r}\, d^3\mathbf{v}$ in μ-space.

Many properties of semiperfect gases can be computed with the aid of $f(\mathbf{r},\mathbf{v},t)$. For example, the number of particles n per unit volume and the momentum $\mathbf{M}$ per unit volume are given by the relations

$$n = \int d^3\mathbf{v} f(\mathbf{r},\mathbf{v},t) \tag{8.131}$$

and

$$\mathbf{M} = m \int d^3\mathbf{v}\, \mathbf{v}\, f(\mathbf{r},\mathbf{v},t) \tag{8.132}$$

where m is the mass of the particle.

The change of $f(\mathbf{r},\mathbf{v},t)$ in time is described by an integro-differential equation postulated by Boltzmann. The Boltzmann equation includes explicitly effects due to both long-range and short-range forces between particles of neighboring phases and implicitly effects due to irreversible processes. The implicit dependence on the latter is discussed in Sec. 8.8.4.

The Boltzmann equation is given by the relation

$$\frac{\partial f}{\partial t} = -\mathbf{v}\cdot\nabla_r f - \frac{\mathbf{F}}{m}\cdot\nabla_v f + \int d^3\mathbf{v}_1 \int d^2\Omega\, g\, \sigma(g,\theta)\,(f'f_1' - ff_1) \tag{8.133}$$

where $\nabla_r f$ and $\nabla_v f$ = gradients of f in $\mathbf{r}$ and $\mathbf{v}$ spaces, respectively

$\mathbf{F}$ = long-range force

$g \equiv |\mathbf{v} - \mathbf{v}_1| \equiv |\mathbf{v}_1' - \mathbf{v}'| \equiv g'$

$f' \equiv f(\mathbf{r},\mathbf{v}',t)$

$f_1 \equiv f(\mathbf{r},\mathbf{v}_1,t)$

$\sigma(g,\theta)$ = differential elastic scattering cross section

θ = polar scattering angle in the center of mass coordinates

$d^2\Omega$ = an element of solid angle

Unprimed quantities refer to values prior to a collision, and primed quantities refer to values following a collision. The force $\mathbf{F}$ is independent of $\mathbf{v}$. The equation is applicable, however, for an applied magnetic field $\mathbf{B}$ for which the force on the particle is proportional to $\mathbf{v} \times \mathbf{B}$.

In Eq. 8.133 the first two terms in the right-hand side describe the change of f per unit time due to streaming of particles from a phase around $\mathbf{r}$ into neighboring phases. The streaming results from the nonuniformity of f from phase to phase and the long-range force $\mathbf{F}$. On the other hand, the integral term

or *collision integral* describes the change of f per unit time due to binary collisions associated with the short-range forces and occurring between the particles of neighboring phases. As discussed later, the form of the collision integral is justified only when all phases can be regarded as statistically independent systems at all times.

The form of the collision integral can be derived as follows: In the phase $d^3\mathbf{r}$ around $\mathbf{r}$, $f\,d\mathbf{v}$ is the expected number of particles per unit volume having velocities between $\mathbf{v}$ and $\mathbf{v} + d\mathbf{v}$. One of these particles can experience a *direct collision* with a particle of another phase having velocity $\mathbf{v}_1$. After the collision the velocities of the two particles are $\mathbf{v}'$ and $\mathbf{v}_1'$, and therefore a direct collision reduces the number of particles $f\,d\mathbf{v}$ of the phase under consideration. On the other hand, particles of two phases characterized by velocities $\mathbf{v}'$ and $\mathbf{v}_1'$ may collide in a *restituting collision* as a result of which their velocities become $\mathbf{v}$ and $\mathbf{v}_1$. It follows that a restituting collision increases the number of particles $f\,d\mathbf{v}$ of the phase under consideration.

For computing the expectation value of all collisions per unit volume and per unit time (i.e., the collision integral), we must consider the relative motion between particles of two phases corresponding to velocities $\mathbf{v}$ and $\mathbf{v}_1$. If the particle having velocity $\mathbf{v}$ is surrounded by an action sphere of radius equal to the range of the interparticle forces (Fig. 8.9), then all particles having a velocity

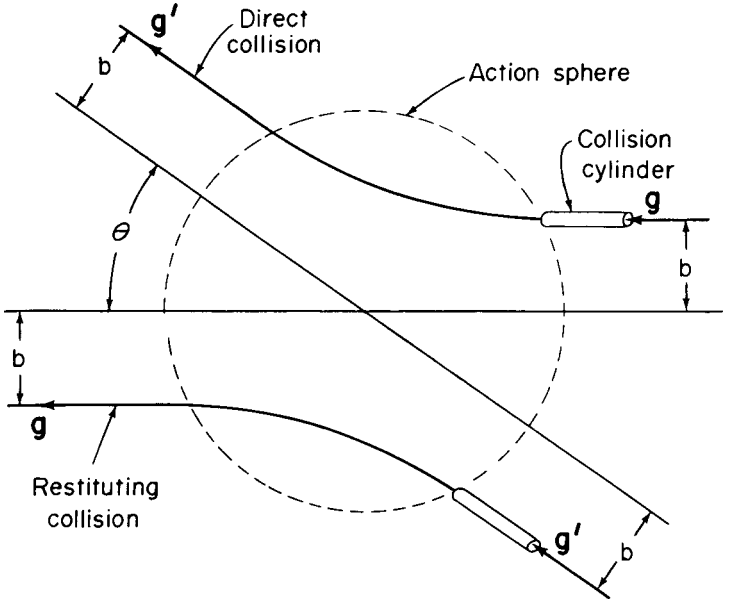


Fig. 8.9 Schematic of direct and restituting elastic collisions in the center of mass frame.

between $\mathbf{v}_1$ and $\mathbf{v}_1 + d\mathbf{v}_1$ which enter the action sphere with an *impact parameter* between b and $b + db$ and which lie in the small cylinder $gb\,db\,d\zeta\,dt$, where ζ is the polar angle, will in the time interval dt collide with a particle having velocity $\mathbf{v}$ and be deflected by an angle θ. The relation between b, g, and θ, which is determined by conservation of energy and conservation of momentum for the elastic collisions under consideration, is given by the expression[10]

$$\theta = \pi - 2\int_0^{\eta_0} \frac{d\eta}{[1 - \eta^2 - (2/mg^2)\,\phi\,(b/\eta)]^{1/2}} \tag{8.134}$$

where $\phi(\mathbf{r})$ is the potential energy for the short-range force between the colliding particles and η_0 is the smallest root of the quantity in the denominator of the integrand. Since the probability of finding a particle with velocity between $\mathbf{v}$ and $\mathbf{v} + d\mathbf{v}$ is $f\,d^3\mathbf{r}\,d\mathbf{v}$ and the probability of finding a particle with velocity between $\mathbf{v}_1$ and $\mathbf{v}_1 + d\mathbf{v}_1$ is $f_1\,d^3\mathbf{r}\,d\mathbf{v}$,† it follows that the number dC_d of direct collisions per unit volume and per unit time is given by the relation

$$dC_d = gb\,db\,d\zeta\,ff_1\,d\mathbf{v}\,d\mathbf{v}_1 = g\,\sigma(g,\theta)\,d^2\Omega\,ff_1\,d\mathbf{v}\,d\mathbf{v}_1 \tag{8.135}$$

where

$$\sigma(g,\theta)\,d^2\Omega = b\,db\,d\zeta \tag{8.136}$$

and

$$d^2\Omega = \sin\theta\,d\theta\,d\zeta \tag{8.137}$$

Equation 8.135 is valid when the two probabilities $f\,d^3\mathbf{r}\,d\mathbf{v}$ and $f_1\,d^3\mathbf{r}\,d\mathbf{v}_1$ are statistically independent. The differential elastic collision cross section $\sigma(g,\theta)$ is uniquely determined by the function $\phi(\mathbf{r})$ and Eq. 8.134.

On the basis of similar considerations, the number dC_r of restituting collisions per unit volume and per unit time between particles of velocities $\mathbf{v}'$ and $\mathbf{v}_1'$ is given by the relation

$$dC_r = g\,\sigma(g,\theta)\,d^2\Omega\,f'f_1'\,d\mathbf{v}'\,d\mathbf{v}_1' \tag{8.138}$$

†Strictly speaking, f_i should be evaluated for a value of $\mathbf{r}$ different from that used in f since the two f's correspond to neighboring phases. The difference in $\mathbf{r}$'s, however, is negligibly small.

In this relation the pair of velocities $\mathbf{v}'$ and $\mathbf{v}_1'$ is related to the pair $\mathbf{v}$ and $\mathbf{v}_1$ by virtue of conservations of energy and momentum. In particular, it can be shown that[11]

$$d\mathbf{v}'\, d\mathbf{v}_1' = d\mathbf{v}\, d\mathbf{v}_1$$

Thus the expectation value C of all binary elastic scattering collisions per unit volume of μ-space and per unit time is given by the relation

$$C = \int (dC_r - dC_d) = \int d^3\mathbf{v}_1 \int\int d^2\Omega\, g\, \sigma(g,\theta)\,(f'f_1' - ff_1) \tag{8.139}$$

namely, the collision integral in Eq. 7.136.

When the one-component semiperfect gas is in a steady state, f does not depend explicitly on time, and it satisfies the *steady-state Boltzmann equation*, which is given by the relation

$$\mathbf{v}\cdot\nabla_r f + \frac{\mathbf{F}}{m}\cdot\nabla_v f = \int d^3\mathbf{v}_1 \int d^2\Omega\, g\, \sigma(g,\theta)\,(f'f_1' - ff_1) \tag{8.140}$$

A set of steady-state Boltzmann equations for a mixture of ionized gases is given in Sec. 8.8.5.

Finally, fluxes of various properties of a phase $d^3\mathbf{r}$ around position $\mathbf{r}$ can be computed by means of $f(\mathbf{r},\mathbf{v},t)$. For example, the kinetic-energy-flux density $\mathbf{w}$ is given by the relation

$$\mathbf{w} = \int d^3\mathbf{v}\left(\frac{mv^2}{2}\right)\mathbf{v}\, f(\mathbf{r},\mathbf{v},t) \tag{8.141}$$

8.8.4 **Discussion of the Boltzmann Equation.** Conditions 1 to 3 (given in Sec. 8.8.3), which must be satisfied for a system to be regarded as a semiperfect gas, are helpful both for interpreting and for delineating the range of validity of the Boltzmann equation. The reasons are as follows:

1. Condition 1. For a given system A, the momentum eigenvalues $\mathbf{p}_j$ and the expectation value of the number of particles with momentum $\mathbf{p}_j$ are independent of the Cartesian coordinates of A. If a distribution $f(\mathbf{r},\mathbf{v},t)$ is to be defined as a function of $\mathbf{r}$, A must be subdivisible into phases as stated in condition 1.

2. Condition 2. Both the streaming and the collision terms in the Boltzmann equation (Eq. 8.133) are due to interactions of a phase with other phases as stated in condition 2. In addition, when the gradient $\nabla_r f$ and the long-range force $\mathbf{F}$ are equal to zero and for all initial conditions $f(\mathbf{r},\mathbf{v},t = 0)$, it can be shown that the solution of the Boltzmann equation tends to a time-independent

Maxwellian distribution as time goes on.[12] Since the system is a semiperfect gas, the Maxwellian distribution corresponds to a stable equilibrium state. In the absence of interactions with other systems, the transition from a nonequilibrium to a stable equilibrium state can be accomplished only by means of irreversible processes. Hence each phase must be also subject to such processes.

3. Condition 3. The distribution function f is uniquely related to the state only when each phase may have at most one particle; hence the need for condition 3. This condition is also consistent with the use of the distribution function f for calculating the values of properties and fluxes. Finally, this condition is consistent with the inclusion of only binary collisions in the Boltzmann equation.

Apart from conditions 1 to 3, one more condition is implicit in the form of the collision integral (Eq. 8.139). Although the form of the collision integral appears intuitively understandable, it must be regarded as a postulate of a nonmechanical nature, i.e., a statement that cannot be derived from the laws of classical or quantum mechanics. This important point is discussed elsewhere.[13] For present purposes it suffices to indicate the following.

As already stated, in the absence of interactions with other systems, the solution of the Boltzmann equation tends to the Maxwellian distribution as time approaches infinity. The tendency to the Maxwellian distribution is a consequence of the inclusion of binary collisions and the form of the collision integral in the Boltzmann equation. Since such a time dependence of f corresponds to an irreversible process, it appears that binary interparticle collisions are the source of irreversibility. This appearance, however, should not be mistaken either as a proof of the second law or as an explanation of irreversibility.

A possible explanation of the irreversibility described by the Boltzmann equation can be as follows: Interactions between phases of a semiperfect gas are due to short- and long-range forces. In general, these forces cause correlations between phases; i.e., they make the phases statistically interdependent. Each phase is also subject to irreversible processes. One of the effects of the irreversible processes may be the destruction of the correlations between neighboring phases. Thus the probability of having a particle with velocity $\mathbf{v}$ in a phase around $\mathbf{r}$ and another particle with velocity $\mathbf{v}_1$ in a second phase around $\mathbf{r}_1$ is proportional to $f(\mathbf{r},\mathbf{v},t)f(\mathbf{r}_1,\mathbf{v}_1,t)$. This effect is implicit in the form of the collision integral. Moreover, the destruction of correlations always causes an increase of entropy. In the absence of irreversible processes the collision integral would have a form different from that given by Eq. 8.139.

8.8.5 Steady-State Boltzmann Equations for a Mixture of Ionized Gases. We will consider a system A that consists of a mixture of monatomic neutral particles and their products of dissociation defined by the ionization reaction

8.72. Both the electron and ion species are singly charged and A is in a steady state.

If each component of A can be regarded as a semiperfect gas in a steady state, then a one-particle distribution function $f_\alpha(\mathbf{r}_\alpha,\mathbf{v}_\alpha)$ can be associated with each component of the mixture, namely, $f_e(\mathbf{r}_e,\mathbf{v}_e)$ for electrons ($\alpha = e$), $f_i(\mathbf{r}_i,\mathbf{v}_i)$ for ions ($\alpha = i$), and $f_a(\mathbf{r}_a,\mathbf{v}_a)$ for neutral particles ($\alpha = a$). According to the theory of semiperfect gases, each one-particle distribution function satisfies a steady-state Boltzmann equation given by the relation

$$\mathbf{v}_\alpha \cdot \nabla_r f_\alpha + \frac{\mathbf{F}_\alpha}{m_\alpha} \cdot \nabla_v f_\alpha = \sum_\beta C_{\alpha\beta} + C_\alpha^{\text{in}} \qquad \text{for } \alpha,\beta = e,i, \text{ and } a \tag{8.142}$$

where $F_\alpha = e_\alpha \mathbf{E}$

e_α = electric charge of species α

m_α = mass of species α

$\mathbf{v}_\alpha$ = velocity of species α

$\mathbf{E}$ = electric field

$C_{\alpha\beta}$ = collision integral for binary elastic collisions between species α and β

C_α^{in} = number of particles of species α generated per unit volume of μ-space and per unit time by virtue of inelastic collisions.

The collision integral $C_{\alpha\beta}$ is given by the Boltzmann form

$$C_{\alpha\beta} = \int d^3\mathbf{v}_\beta \int d^2\Omega\, g_{\alpha\beta}\, \sigma_{\alpha\beta}\,(g_{\alpha\beta},\theta)\,(f'_\alpha f'_\beta - f_\alpha f_\beta) \tag{8.143}$$

where

$$g_{\alpha\beta} \equiv |\mathbf{v}_\alpha - \mathbf{v}_\beta| \equiv |\mathbf{v}'_\beta - \mathbf{v}'_\alpha| \equiv g'_{\alpha\beta}$$

$\sigma_{\alpha\beta}(g_{\alpha\beta},\theta)$ is the differential elastic scattering cross section between species α and β, and other quantities have definitions similar to those given in Sec. 8.8.3. For collisions between particles of the same species, Eq. 8.143 is identical to Eq. 8.139 with appropriate adjustments of subscripts. For elastic collisions between dissimilar particles, the form of $C_{\alpha\beta}$ is derived in a manner equivalent to that used for deriving Eq. 8.139. In general, Eq. 8.142 is analogous to Eq. 8.140 except for the inclusion of effects of elastic and inelastic collisions between dissimilar particles.

The integro-differential equations (8.142) represent a set of coupled equations for $f_\alpha, \alpha = e,\ i,$ and a. For continuous boundary-value processes only, the solutions of this set are related to the forces resulting from the thermodynamic potentials at the boundaries of the mixture. For linear processes the solutions can be manipulated mathematically and reduced to a set of nine phenomenologi-

cal equations. The conductivities L in these equations are explicitly known functions of the thermodynamic potentials. The necessary mathematical manipulations are brought forth via the hydrodynamic equations discussed in Sec. 8.8.6.

8.8.6 Hydrodynamic Equations. Multiplying Eq. 8.142 by either a scalar or a vector function of $\mathbf{v}_\alpha$ and integrating the result over the entire α-velocity space yield a relation called a *hydrodynamic equation*. Hydrodynamic equations are helpful in deriving phenomenological equations.

For example, multiplying Eq. 8.142 by unity and by $m_\alpha \mathbf{v}_\alpha$ and $m_\alpha \mathbf{v}_\alpha^2/2$ for all α and integrating each result over all $\mathbf{v}_\alpha$ yield

$$\nabla_r \cdot \mathbf{\Gamma}_\alpha = \dot{\xi}' \qquad \text{for } \alpha = e \text{ and } i \text{ and } \nabla_r \mathbf{\Gamma}_a = -\dot{\xi}' \tag{8.144}$$

$$\nabla_r p_\alpha + \nabla_r \cdot \bar{\bar{\pi}}_\alpha - e_\alpha n_\alpha \mathbf{E} - \sum_{\beta \neq a} \mathbf{R}_{\alpha\beta} - \mathbf{R}_\alpha^{\text{in}} = 0 \tag{8.145}$$

$$\nabla_r \cdot \mathbf{w}_\alpha = e_\alpha \mathbf{\Gamma}_\alpha \cdot \mathbf{E} - \sum_{\beta \neq \alpha} Q_{\alpha\beta} - Q_\alpha^{\text{in}} \tag{8.146}$$

where e is the charge of the electron, $e_e = -e$ and $e_i = +e$,

$$n_\alpha \equiv \int d^3\mathbf{v}_\alpha \, f_\alpha \tag{8.147}$$

$$p_\alpha \equiv \int d^3\mathbf{v}_\alpha \, m_\alpha v_\alpha^2 f_\alpha \tag{8.148}$$

$$\pi_{\alpha jk} \equiv \int d^3\mathbf{v}_\alpha \, m_\alpha \left(v_{\alpha j} v_{\alpha k} - \frac{v_\alpha^2}{3} \delta_{jk} \right) f_\alpha \tag{8.149}$$

$$\mathbf{w}_\alpha \equiv \int d^3\mathbf{v}_\alpha \frac{m_\alpha v_\alpha^2}{2} \mathbf{v}_\alpha f_\alpha \tag{8.150}$$

$$\mathbf{\Gamma}_\alpha \equiv \int d^3\mathbf{v}_\alpha \, \mathbf{v}_\alpha f_\alpha \tag{8.151}$$

$$\mathbf{R}_{\alpha\beta} \equiv \int d^3\mathbf{v}_\alpha \, m_\alpha \mathbf{v}_\alpha C_{\alpha\beta} \tag{8.152}$$

$$Q_{\alpha\beta} \equiv -\int d^3\mathbf{v}_\alpha \frac{m_\alpha \mathbf{v}_\alpha^2}{2} C_{\alpha\beta} \tag{8.153}$$

$$\dot{\xi}' \equiv \int d^3\mathbf{v}_e \, C_e^{\text{in}} \equiv \int d^3\mathbf{v}_i \, C_i^{\text{in}} \equiv -\overline{\int} d^3\mathbf{v}_a \, C_a^{\text{in}} \tag{8.154}$$

$$\mathbf{R}_\alpha^{\text{in}} \equiv \int d^3\mathbf{v}_\alpha \, m_\alpha \mathbf{v}_\alpha \, C_\alpha^{\text{in}} \tag{8.155}$$

$$Q_\alpha^{\text{in}} \equiv -\int d^3\mathbf{v}_\alpha \frac{m_\alpha v_\alpha^2}{2} C_\alpha^{\text{in}} \tag{8.156}$$

where $v_{\alpha k}$ = component of $\mathbf{v}_\alpha$ along the kth Cartesian axis
δ_{jk} = the Kronecker delta
p_α = force per unit area
$\pi_{\alpha jk}$ = αjkth component of the viscosity tensor $\overline{\overline{\pi}}$
$\mathbf{w}_\alpha$ = kinetic-energy-flux density for species α
Γ_α = particle-flux density for species α
$\mathbf{R}_{\alpha\beta}, Q_{\alpha\beta}$ = flux densities of momentum and energy transfer, respectively, between species α and β due to elastic collisions
$\xi', \mathbf{R}_\alpha^{\text{in}}, Q_\alpha^{\text{in}}$ = rates of generation of particles, momentum, and energy of species α, respectively, due to inelastic collisions

Identities 8.154 are statements of conservation of mass for a system in a steady state. On the other hand, by virtue of conservations of momentum and energy, we can verify that

$$\sum_\beta \int d^3\mathbf{v}_\alpha C_{\alpha\beta} = 0 \tag{8.157}$$

$$\sum_\alpha \mathbf{R}_{\alpha\beta} = 0 \tag{8.158}$$

and

$$\sum_\alpha Q_{\alpha\beta} = 0 \tag{8.159}$$

Equation 8.157 has already been used in the derivation of Eq. 8.144. It is clear that additional hydrodynamic equations can be established by taking higher order moments of the Boltzmann equation for species α with respect to $\mathbf{v}_\alpha$.

Together with identities 8.147 to 8.156, hydrodynamic equations 8.144 to 8.146 provide the means for calculating the conductivities L of the phenomenological equations. This point can be clarified as follows:

In principle, we see from Eqs. 8.144 to 8.156 that all quantities entering into these equations can be evaluated if the one-particle distribution functions f_α and the collision laws for elastic and inelastic encounters are known. The distribution functions are solutions of the Boltzmann equations 8.142 and are functions of the forces resulting from the thermodynamic potentials. For linear processes, Eqs. 8.144, 8.150, and 8.151 can be easily expressed in the forms of the phenomenological equations discussed in Sec. 8.7.5.

Exact calculation of one-particle distribution functions from the coupled Eqs. 8.142 is a very difficult mathematical task. The difficulty is due both to the nonlinear nature of the collision integrals and to the complicated dependence of the collision cross sections on the relative velocity of the colliding particles. Even if the collision integrals are linearized, the problem is still mathematically complicated.

Because of this difficulty a variety of approximations are introduced into the integro-differential equations. The discussion of these approximations is beyond the scope of this book. They are discussed elsewhere.[14] In Sec. 8.8.7 the results of a special class of approximations are given without proofs.

8.8.7 Solutions of Linearized Boltzmann Equations. For the mixture of ionized gases discussed in Sec. 8.8.5, we will assume that the phase of volume $d^3\mathbf{r}$ around position $\mathbf{r}$ is in a steady state near the corresponding stable equilibrium state. The term "near" means that the difference of entropies of the stable equilibrium and steady states is only a small fraction of the entropy of the stable equilibrium state. For these steady states, solutions of Eqs. 8.142 have been found under the following restrictions.[15]

Restriction 1. For electrons and ions, the one-particle distribution functions are written in the form

$$f_\alpha = f_\alpha^0 \, [1 + \phi_\alpha(\mathbf{r}_\alpha, \mathbf{v}_\alpha)] \qquad \text{for } \alpha = e \text{ and } i \tag{8.160}$$

where

$$f_\alpha^0 \equiv n_\alpha \left(\frac{m_\alpha}{2\pi k T_\alpha}\right)^{3/2} \exp\left(-\frac{m_\alpha v_\alpha^2}{2kT_\alpha}\right) \tag{8.161}$$

$$n_\alpha \equiv \int d^3\mathbf{v}_\alpha \, f_\alpha^0 \tag{8.162}$$

T_α is the temperature of the stable equilibrium state corresponding to the steady state of the alpha component of the elementary phase under consideration and $\phi_\alpha(\mathbf{r}_\alpha, \mathbf{v}_\alpha)$ is a slowly varying scalar function of $\mathbf{r}_\alpha$ and $\mathbf{v}_\alpha$, which is a measure of the departure of f_α from the stable equilibrium state one-particle distribution function f_α^0, namely, the Maxwellian distribution, and is such that

$$|\phi_\alpha(\mathbf{r}_\alpha, \mathbf{v}_\alpha)| \ll 1 \tag{8.163}$$

The particle density n_α and the temperature T_α are functions of the position $\mathbf{r}$ of the phase. The restrictions imposed on ϕ_α are consistent with the assumption

that the steady states under consideration are near the corresponding stable equilibrium states.

On the other hand, for neutral particles, the one-particle distribution function f_a is assumed to be Maxwellian, namely,

$$f_a \approx f_a^0 = n_a \left(\frac{m_a}{2\pi k T_a}\right)^{3/2} \exp\left(-\frac{m_a v_a^2}{2kT_a}\right) \tag{8.164}$$

where n_a and T_a are functions of position only. This assumption is reasonable for mixtures in which the neutral-particle-flux density is almost equal to zero and in which the neutral–electron and neutral–ion collision frequencies are much smaller than the neutral–neutral collision frequency, i.e., when f_a is essentially unperturbed by the motions of the charged particles. By virtue of Eq. 8.164, the Boltzmann equation for the neutral particles and the hydrodynamic equations resulting therefrom are not needed in the analysis.

Regarding electrons and ions, substituting Eq. 8.160 into Eq. 8.142 yields a set of two nonlinear coupled equations for the ϕ_α's. By virtue of the restrictions assumed for ϕ_α, however, these equations can be linearized, and therefore the problem of their solution is simplified. For the linearized Boltzmann equations, it can be shown that the solutions $\phi_\alpha(\mathbf{r}_\alpha, \mathbf{v}_\alpha)$ are of the form[16]

$$\phi_\alpha(\mathbf{r}_\alpha, \mathbf{v}_\alpha) = \omega_\alpha(\mathbf{r}_\alpha, v_\alpha) \cdot \frac{\mathbf{v}_\alpha}{v_\alpha} \qquad \text{for } \alpha = e \text{ and } i \tag{8.165}$$

where the vector ω_α is a linear function of the thermodynamic potentials and a function that depends only on the magnitude v_α of $\mathbf{v}_\alpha$.

By virtue of Eqs. 8.162 and 8.165, it is noteworthy that ϕ_α satisfies the relations

$$\int d^3\mathbf{v}_\alpha \, \phi_\alpha f_\alpha^0 = 0 \tag{8.166}$$

$$\int d^3\mathbf{v}_\alpha \, m_\alpha v_\alpha^2 \phi_\alpha f_\alpha^0 = 0 \tag{8.167}$$

Accordingly

$$p_\alpha = n_\alpha k T_\alpha = \int d^3\mathbf{v}_\alpha \frac{m_a v_\alpha^2}{3} f_\alpha^0 \tag{8.168}$$

In other words, the force per unit area p_α is equal to the pressure of the corresponding stable equilibrium state. Moreover, Eqs. 8.166 and 8.167 are consistent with the remark made earlier that T_α is the temperature of the corresponding stable equilibrium state.

Restriction 2. The mass of the electrons is much smaller than that of either the ions or the neutral particles. In symbols

$$\frac{m_e}{m_\alpha} \ll 1 \qquad \text{for } \alpha = i \text{ and } a$$

By virtue of this inequality, the elastic collision integrals C_{ei}, C_{ea}, and C_{ie} can be approximated by the relations

$$C_{e\beta} \approx -\nu_{e\beta}(v_e) f_e^0 \phi_e \qquad \text{for } \beta = i \text{ and } a \tag{8.169}$$

and

$$C_{ie} \approx \frac{m_e}{p_i} f_i^0 \mathbf{v}_i \cdot \int d^3\mathbf{v}_e \, \nu_{ei}(v_e) \mathbf{v}_e f_e \tag{8.170}$$

where $\nu_{e\beta}$ is the collision frequency given by the relation

$$\nu_{e\beta} \equiv n_\beta v_e \int d^2\Omega \, (1 - \cos\theta) \, \sigma_{e\beta}(v_e, \theta) \qquad \text{for } \beta = i \text{ and } a \tag{8.171}$$

The effect of approximations 8.169 and 8.170 is that the electron and ion Boltzmann equations are decoupled in velocity space. Another effect is that the electron–heavy particle energy exchanges reduce to zero, namely,

$$Q_{ei} = Q_{ie} = Q_{ea} = 0 \tag{8.172}$$

Finally, the electron–ion momentum exchange $\mathbf{R}_{\alpha\beta}$ for $\alpha,\beta = e$ and i can be neglected,† namely,

$$\mathbf{R}_{\alpha\beta} \approx 0 \qquad \text{for } \alpha,\beta = e \text{ and } i \tag{8.173}$$

Restriction 3. The ion and neutral-particle temperatures T_i and T_a, respectively, are assumed to be equal for all elementary phases. Because of this assumption it is not necessary to include in the analysis the hydrodynamic equation for the ion kinetic energy (Eq. 8.146 for $\alpha = i$).

†In Ref. 15 the term $\mathbf{R}_{ie}$ is retained in the Boltzmann equation for the ion-distribution function. This is not consistent either with the fact that m_e is much smaller than m_i or with the Onsager reciprocal relations. For these reasons, here $\mathbf{R}_{ie}$ is assumed to be approximately equal to zero.

Restriction 4. Finally, it is assumed that the inelastic collision terms C_α^{in} can be neglected from Eqs. 8.142 for $\alpha = e$ and i when these equations are solved for the one-particle distribution functions. It can be shown[15] that consistent with this approximation is the neglect of the terms $\mathbf{R}_\alpha^{\text{in}}$. Another effect of the approximation is that the spatial and the local phenomena are decoupled, namely, the Curie principle is satisfied.

Using restrictions 1 to 4 in the Boltzmann equations for electrons and ions and a special technique for the approximate solution of these equations yields the one-particle distribution functions[15] f_e and f_i. Substituting these distribution functions into the relevant hydrodynamic equations and related identities results in the following set of expressions.

$$\mathbf{w}_e = \left(\frac{5}{2}kT_e + k_e^T kT_e\right)\Gamma_e - k_e \nabla_r T_e \tag{8.174}$$

$$\Gamma_e = M_e[\nabla_r p_e + en_e\mathbf{E} + k_e^T n_e k \nabla_r T_e] \tag{8.175}$$

$$\Gamma_i = -M_i[\nabla_r p_i - en_i\mathbf{E} + k_i^T n_i k \nabla_r T_i] \tag{8.176}$$

$$\nabla_r \cdot \mathbf{w}_e = e\Gamma_e \cdot \mathbf{E} - Q_e^{\text{in}} \tag{8.177}$$

$$\nabla_r \cdot \Gamma_e = \dot{\xi}' \qquad \nabla_r \cdot \Gamma_i = \dot{\xi}' \tag{8.178}$$

where all the coefficients k_e^T, k_e, and M_e are explicitly known functions of the temperature T_e and the collision cross sections and where the coefficients k_i^T and M_i are explicitly known functions of the temperature T_i and the collision cross sections. These functions result from a series of somewhat involved calculations[15] and will not be given here.

The quantities M_α and k_α^T are called the *mobility* and *thermal diffusion ratio*, respectively, of species α, and k_e is called the *electron thermal conductivity*.

The approximations introduced for the derivation of Eqs. 8.174 to 8.178 are equivalent to those for Eqs. 8.121 to 8.125. For mixtures of ionized gases in which the spatial phenomena depend only on the Cartesian coordinate x, the former set of equations can be expressed in the form of the latter, and thus the values of the coefficients of the phenomenological equations can be determined.

For example, if the electron motive is ψ and the energy for the internal structure of the electrons is $\bar{\varepsilon}_{oe}$, it can be readily verified that

$$\begin{aligned} u_e &= w_e + (\bar{\varepsilon}_{oe} + \psi)\Gamma_e \\ &= \left(\frac{5}{2}kT_e + k_e^T kT_e + \bar{\varepsilon}_{oe} + \psi\right)\Gamma_e + k_e T_e^2 \tau_e' \end{aligned} \tag{8.179}$$

$$\Gamma_e = M_e n_e T_e \left[(kT_e + k_e^T kT_e)\tau_e' - k\frac{n_e'}{n_e} - eE_x\tau_e \right] \tag{8.180}$$

$$\Gamma_i = M_i n_i T_i \left[(kT_i + k_i^T kT)\tau_i' - k\frac{n_i'}{n_i} + eE_x\tau_i \right] \tag{8.181}$$

Comparing Eqs. 8.179 to 8.181 with Eqs. 8.121 to 8.123 yields

$$\begin{aligned} k_{ee} &= k_e T_e^2 \qquad q_{ee}^* = k_e^T kT_e \qquad q_{ii}^* = k_i^T kT_i \\ L_{44} &= M_e n_e T_e \qquad L_{55} = M_i n_i T_i \end{aligned} \tag{8.182}$$

Apart from L_{44} and L_{55}, the other coefficients L of the spatial phenomenological equations can be determined by means of their relations to k_{ee}, q_{ee}^*, and q_{ii}^*.

References

1. L. Onsager, Reciprocal Relations in Irreversible Processes. I, *Phys. Rev.*, 37: 405-426 (1931); also Reciprocal Relations in Irreversible Processes. II, *Phys. Rev.*, 38: 2265-2279 (1931).
2. H. B. G. Casimir, On Onsager's Principle of Microscopic Reversibility, *Rev. Mod. Phys.*, 17: 343-350 (1945).
3. S. R. DeGroot, *Thermodynamics of Irreversible Processes*, Interscience Publishers, New York, 1950.
4. I. Prigogine, *Introduction to Thermodynamics of Irreversible Processes*, Charles C. Thomas, Publisher, Springfield, Ill., 1955.
5. G. N. Hatsopoulos and J. H. Keenan, *Principles of General Thermodynamics*, p. 641, John Wiley & Sons, Inc., New York, 1965.
6. Ref. 5, p. 642.
7. Ref. 5, p. 643.
8. Ref. 5, Chap. 51, pp. 675-694.
9. S. R. DeGroot and P. Mazur, *Non-Equilibrium Thermodynamics*, Interscience Publishers, New York, 1962.
10. D. J. Rose and M. Clark, *Plasmas and Controlled Fusion*, p. 25, The MIT Press, Cambridge, Mass., 1961.
11. S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases*, p. 65, Cambridge University Press, New York, 1961.
12. Ref. 11, Chap. 4, pp. 69-88.
13. G. E. Uhlenbeck and G. W. Ford, *Lectures in Statistical Mechanics*, Chap. VII, pp. 118-138, American Mathematical Society, Providence, R. I., 1963.
14. Ref. 11, Chap. 7, p. 107.
15. D. R. Wilkins and E. P. Gyftopoulos, Transport Phenomena in Low-Energy Plasmas, *J. Appl. Phys.*, 37: 3533-3540 (1966).
16. Ref. 11, p. 119.

9

Emission Phenomena

9.1 General Remarks

Emission phenomena are essential and inherent to the operation of thermionic converters. This fact is amply supported by the theoretical and experimental results presented in Vol. I, Chaps. 2 to 4. For example, the performance limits of any converter are primarily determined by the thermionic electron emission capabilities of the electrode surfaces (Chap. 2). Again, the performance of a cesium diode depends strongly on the interrelation between the values of the cesium pressure and of the thermionic particle emission rates from the cesiated emitter and collector (Chap. 4).

Emission phenomena, i.e., the processes by which electrons, ions, neutral particles, and photons are emitted (desorbed) from or collected (adsorbed) at surfaces of solids, occur spontaneously when a solid is kept at a temperature different from 0°K and is immersed in vacuum or in a vapor. Moreover, when the particles are electrically charged, the emission phenomena are also called *thermionic emission phenomena.*

The definition of the surface of a solid is somewhat ambiguous and arbitrary. In this book we will use the following definition. For a given solid with a finite number of particles, we will assume that the centers of the nuclei of the atoms of the solid occupy fixed positions.† Thus a geometric surface can be determined which contains all the centers of the nuclei of the outermost atomic layer or the outermost lattice points of the solid. This geometric surface will be called the *surface of the solid* or simply the *surface.*

Each surface has two distinguishing features: (1) the chemical nature of the atoms associated with the lattice points on the surface, which we will call the *composition of the surface,* and (2) the geometric pattern of the lattice points on the surface, which we will call the *structure of the surface.* It is found experimentally that the structure depends on both the composition and the procedures used for preparing the surface.

Emission phenomena, and particularly thermionic emission phenomena, have been investigated experimentally with many surface specimens. The results indicate that the phenomena depend very strongly on the type of surface, i.e., the composition and structure of the surface.

By and large, special procedures must be used to achieve and maintain a surface with a specified composition and structure. Although the composition

†Positions are defined by expectation values.

and structure of a surface are usually difficult to identify and describe precisely, it is convenient for the systematic discussion of both experimental and theoretical results to classify surfaces into four types: (1) *pure uniform surfaces;* (2) *pure nonuniform surfaces;* (3) *composite surfaces;* and (4) *complex surfaces* (oxide cathodes).

The theory of emission phenomena is not yet developed to the point where exact quantitative predictions can be made about the emission properties of a surface of a given type. The problem is very difficult, and only approximate results have been derived to date. The difficulties can be appreciated through the following discussion.

If an isothermal solid is in a steady state and is interacting with an adjacent vapor and with other systems (Fig. 9.1), then the state is determined by the values of a certain number of independent properties, e.g., the temperature of the solid and the degree of coverage of the surface by vapor molecules. The values of some properties are functions of the state of the vapor; e.g., the degree of coverage depends on the pressure of the vapor. The theory of emission phenomena from the surface might be developed along four steps.

1. The properties that characterize a particular emission process are identified; e.g., it may be established that thermionic electron emission depends on the value of the work function.

2. The relations between emission rates and the relevant properties are derived; e.g., the relation between electron flux density and work function is determined.

3. The values of various properties are expressed as functions of the independent properties that determine the steady state of the solid; e.g., the value of the work function is expressed as a function of the temperature of the solid, the degree of coverage of the surface, etc.

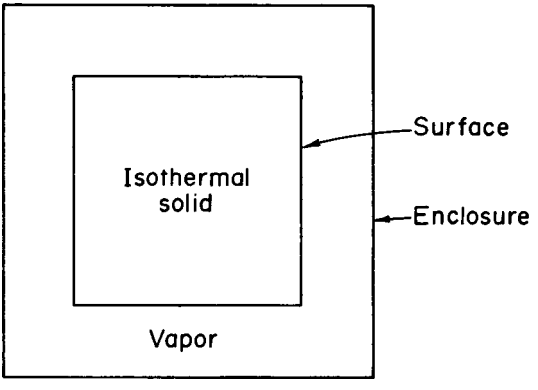


Fig. 9.1 Schematic of an isothermal solid immersed in a vapor.

4. The results of the first three steps are combined with the conservation laws to obtain relations that involve only independent properties of both the solid and the vapor; e.g., the ion emission rate is expressed as a function of the temperature of the solid and the pressure of the vapor.

Unfortunately, the first three steps of this procedure cannot be carried out rigorously at the present time. Because the interparticle forces are very strong, quantum-thermodynamic analysis of a solid is a very difficult mathematical task. For this reason, the theory is developed on the basis of a number of conceptual and numerical approximations.

Specifically, the analysis begins with consideration of the solid and the adjacent medium in states that are in mutual stable equilibrium. For these states steps 1 and 2 can be carried out almost exactly. The implementation of step 3, however, still presents serious if not insurmountable computational difficulties. It cannot be pursued without the introduction of both conceptual simplifications and numerical approximations. To proceed with step 3, we must postulate a tractable model for the behavior of the various particles in the system. Different investigators use different models, and each model provides the means for the complete analysis of the states in mutual stable equilibrium.

Next, we assume that, when the solid is in a steady state, the properties characterizing the emission phenomena and their relations to the independent properties are the same as those found under conditions of mutual stable equilibrium, except for some corrections introduced to account for processes that may be effective only in steady state. Using this important simplifying assumption, we can complete the first three steps of the analysis of the steady state of the solid. These results are used in step 4 to find the desired interrelations.

The details of this way of thinking about emission phenomena are presented in subsequent sections. These general remarks are made to forewarn the reader about the difficulties of the problem and to justify the organization of the chapter.

Despite the experimental and theoretical difficulties, substantial progress has been made during the past few decades. Although exact results are missing, a continuous interplay between measurement and analysis has resulted in a partial but reasonable understanding of emission phenomena. This chapter summarizes this understanding, with particular emphasis on results obtained from measurements and analyses of thermionic emissions of electrons, ions, and neutral particles from electronic conductors.

For the problem of photon emission and absorption and its relation to thermal radiation in thermionic converters, the reader is referred to the literature. (A. J. Chapman, *Heat Transfer*, Chap. 11, Macmillan Company, New York, 1974; R. J. McCandless and P. R. Hill, Radiation Heat Transfer Calculations, in *Proceedings*

of Thermionic Conversion Specialists Conference, Palo Alto, Calif., Oct. 30–Nov. 1, 1967, pp. 370-375.)

9.2 Electron Emission from Pure Uniform Surfaces

9.2.1 General Remarks. A pure uniform surface of a single crystal substance, also called a *monocrystalline surface*, is defined as the surface obtained by cleaving the crystal along a specific crystallographic orientation. An example is the (1 0 0)-surface of the tungsten crystal; it is a plane on which the centers of the nuclei form a square matrix. The structure of most monocrystalline surfaces is relatively easy to describe.

Emission characteristics of pure uniform surfaces provide an essential starting point for discussing the characteristics of all other types of surfaces. For this reason, thermionic emission of electrons from pure uniform surfaces is considered in detail here and in Sec. 9.3. Both experimental and theoretical results are presented, and an attempt is made to interpret the former in the light of the latter. Although many of the theoretical results are applicable to pure uniform surfaces of other elements, the emphasis of this presentation is on refractory metals.

Initially, emission of ions and neutral particles is neglected. Appendix 9.A shows that, for pure refractory metal surfaces and practical thermionic emitter temperatures, the rates of emission of ions and neutral particles are much smaller than those of electrons. It follows that metal ions and neutral atoms play a minor role in emission measurements performed on pure refractory metal surfaces and, therefore, can be neglected in a first-order analysis of these measurements.

When heavy-particle emissions are neglected, a finite piece of a metal can be regarded as a one-component open system consisting of electrons moving about the positively charged nuclei of the atoms of the metal. The electrons are held together by force fields, which include both the forces between electrons and the forces due to the positive charges of the nuclei. In general, the values of the potential function of the force fields depend on both the state of the metal and the states of other systems with which the metal may be interacting.

An experimental setup for the study of thermionic electron emission from a pure uniform surface can be represented by the two-electrode-plus-battery system shown in Fig. 9.2. In this figure the electrons, i.e., the single component of both electrodes, are denoted by dots and the nuclei by crosses. When the electrodes are kept at temperatures different from 0°K, the electrons of both electrodes are dispersed throughout both the lattices of the nuclei and the interelectrode space.

In Fig. 9.2 the pure uniform surface under study is that of metal electrode 1, also called the emitter. This surface passes through the centers of the positively

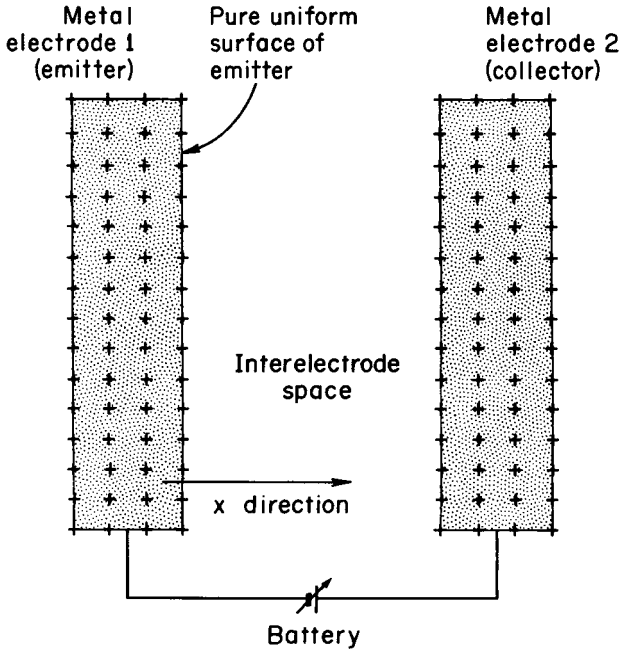


Fig. 9.2 Schematic of experimental setup for measuring thermionic electron emission from a pure uniform surface. Interelectrode space is evacuated except for electrons originating from the electrodes.

charged nuclei of the outermost atomic layer of the emitter. Metal electrode 2, also called the collector, and the battery are included in the figure to prevent ambiguity about sources and sinks of electrons. The collector surface is defined in a manner analogous to that used for the emitter.

For a given emitter state and various values of the properties of the collector, the thermionic electron current from emitter to collector via the interelectrode space may be equal to or different from zero. For example, if the emitter and collector surfaces are identical, if the emitter, collector, and battery temperatures are equal, and if the voltage of the battery is zero, then the two electrodes are in mutual stable equilibrium and the thermionic electron flux from emitter to collector is equal to zero.

In this discussion we assume that the electrode surfaces are parallel and that their linear dimensions are much larger than the interelectrode spacing. Thus any spatially dependent phenomena in the interelectrode space can be regarded as one dimensional, i.e., as dependent on the Cartesian coordinate x only (Fig. 9.2). Moreover, the theoretical analysis of the thermionic electron emission problem is

developed along the lines indicated in the introductory section. First, the emitter is assumed to be in a stable equilibrium state in mutual stable equilibrium with the collector. Thus random fluxes are computed as discussed in Sec. 9.2.2. Then the results of the analysis are modified and adapted to the steady state of the emitter corresponding to the same temperature as that of the stable equilibrium state. This topic is discussed in Secs. 9.2.3 to 9.2.5.

Because steady-state measurements are performed with the aid of an externally applied electric field between emitter and collector (i.e., the voltage of the battery in Fig. 9.2 is different from zero), one of the modifications introduced in the stable equilibrium state results is the effect of such a field. One effect of this type, the Schottky effect, is discussed in Sec. 9.2.6. Finally, the methods used for correlating experimental and theoretical results on thermionic electron emission are presented in Sec. 9.2.7.

9.2.2 Electron-Random-Current Density. This section defines an electron-random-current density $e\Gamma_1$ that depends solely on properties of a pure uniform surface of a metal in a stable equilibrium state and is related to thermionic electron emission from the surface. We will show that $e\Gamma_1$ is given by the relation

$$e\Gamma_1 = AT^2 \exp\left(-\frac{\psi_1 - \mu_e}{kT}\right) \tag{9.1}$$

where $A \equiv 120$ amp/cm^2-°K^2
T = temperature of the metal
ψ_1 = motive just outside the emitter
μ_e = total potential or the electrochemical potential of the electrons

Explicit definitions of these quantities are given later, but here it suffices to indicate that the derivation of Eq. 9.1 is based on the following considerations.

For the two electrodes in Fig. 9.2 to be in a stable equilibrium state at temperature T, the composition and structure of the surface of the collector can be taken to be identical to those of the pure uniform surface of the emitter,† and emitter and collector must be at the same temperature T and electrically connected by means of an isothermal conductor. This state is shown schematically in Fig. 9.3.

For the analysis of this state, we will assume that the electron motive ψ (Vol. I, Sec. 2.2.3) throughout the system is known. The motive is the scalar quantity whose negative gradient at any position is equal to the force exerted on an electron at that position. It is a concept associated with the one-particle

†It is convenient to assume identity of the two surfaces but not essential to the subsequent discussion.

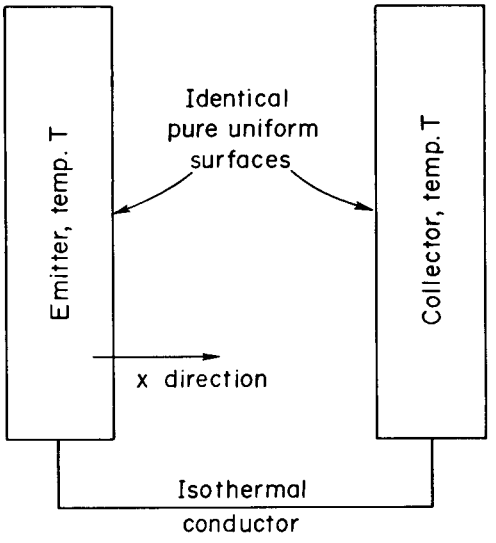


Fig. 9.3 Schematic of two-electrode system in a stable equilibrium state.

picture for the electronic structure of the two electrodes (Sec. 5.10.3). A representative plot of ψ vs. x is shown in Fig. 9.4.

In the bulk of the emitter, the motive is a periodic function of x.† Its values are determined by electrostatic, exchange, and correlation effects, and its period is determined by the Wigner–Seitz unit cell of the crystal lattice of the emitter (Sec. 5.10.2). For all practical purposes, the periodicity holds up to points in the bulk that are at a distance of about one lattice constant from the surface. These points are defined as the *points just inside,* and the surface that contains them is the *surface just inside.*

The justification for the surface just inside is as follows: In the bulk, each nucleus of the crystal lattice of the emitter is symmetrically surrounded by other nuclei, and therefore the electron charge distribution around each nucleus is symmetrical and the motive is periodic from unit cell to unit cell. But the nuclei on the pure uniform surface are not surrounded symmetrically by other nuclei, and, as a result, the force fields and the electron charge distribution are not symmetrical around the nuclei on the surface. The effects of this asymmetry do not extend very deep into the bulk, because the number of electrons in a metal is very large and rearranging the electron charge distribution over distances of the order of atomic dimensions is sufficient to shield the asymmetry of the force fields.

†In general, the motive depends on all three coordinates x, y, and z.

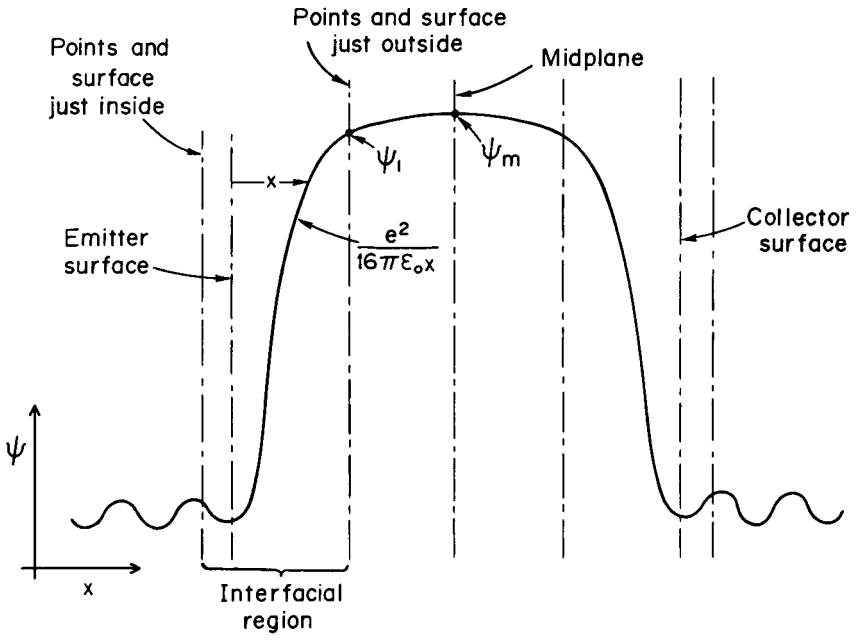


Fig. 9.4 Representative motive ψ vs. x for two identical electrodes in mutual stable equilibrium.

In the vicinity of the surface of the emitter, the motive rises abruptly. For example, a rise of about 10 to 20 ev occurs over a distance of a few hundred angstroms. We will see later that the values of the motive in this region are determined primarily by exchange and correlation effects and to a lesser degree by electrostatic effects. Moreover, here the exchange and correlation effects are substantially different from those in the bulk. Beyond the surface of the emitter, at distances greater than atomic dimensions, we find that the exchange-plus-correlation force can be approximated by the image force, $-e^2/16\pi\varepsilon_0 x^2$, and that the corresponding motive is given by the expression $-e^2/16\pi\varepsilon_0 x$, where x is the distance from the surface and the zero-energy reference level is at infinity.

The abrupt rise of the motive in the vicinity of the surface indicates that electrons are attracted to the emitter. This attraction is qualitatively understandable. For example, it is clear that electrons must be bound to the crystal lattice of the emitter because, if they were not, the emitter would not exist as a solid. Again, the binding can be explained on the basis of the densities of electrons in the bulk and in the interelectrode space. From experience we know that, far away from the surface and in the interelectrode space, the density of

electrons is much smaller than that in the bulk. The difference in densities tends to force the electrons from the emitter into the interelectrode space. In a stable equilibrium state, this tendency must be balanced by binding forces developed in the vicinity of the surface, hence the rising motive in that region.

Far away from the surface and in the interelectrode space, the motive rises gradually with distance from the emitter. For all practical purposes, the gradual rise of the motive is determined by the space-charge effect of electrons and not by exchange and correlation effects. For example, from 500 to 1000 Å from the emitter, the change in motive caused by the space-charge effect of electrons can be equal to several tenths of an electron volt, whereas that caused by exchange and correlation effects of the emitter is less than 4×10^{-3} ev. The gradual rise of the motive extends up to the midplane between the electrodes. At this plane the motive is maximum because of symmetry. Beyond the midplane, the motive varies with position, as shown in Fig. 9.4; i.e., it is the mirror image with respect to the midplane of the motive to the left of that plane.

Qualitatively, the motive in Fig. 9.4 is reasonable and is consistent with what is suggested by experience. Calculating its exact values, however, is a very difficult mathematical task that has defied the capability of presently available methods. The difficulty is primarily due to the lack of an accurate and tractable model to describe the electrons in the bulk and in the vicinity of the surface of a metal.

This difficulty notwithstanding, the qualitative features of the motive suggest separation of the system shown in Fig. 9.3 into three one-component (electron) open systems in mutual stable equilibrium. One of the open systems corresponds to the interelectrode space, and each of the other two corresponds to each of the electrodes. This separation leads to the definition of the points just outside and of electron-random-current density $e\Gamma_1$ at these points. The three open systems are defined as follows:

(a) Interelectrode System. For practical electrode temperatures and for interelectrode distances greater than about 1000 Å, in the range of the gradual change of the motive in Fig. 9.4 the density n_e of electrons is orders of magnitude smaller than the density of electrons in the bulk of a metal. For example, n_e is smaller than about 10^{15} electrons/cm^3, whereas the density of electrons in a metal is about 10^{23} electrons/cm^3. The relatively small magnitude of n_e and the absence of strong exchange and correlation effects (i.e., short-range forces) suggest that the electrons in the interelectrode space can be regarded as a heterogeneous perfect gas (Sec. 7.8.13). Thus many relations between properties of this gas can be readily derived.

If the values of the electron density n_{em} and electron motive ψ_m at the position of the maximum motive in Fig. 9.4 are known, an electron open system in a stable equilibrium state at temperature T can be defined such that: (1) the system extends from one electrode surface to the other; (2) the force field in the

system is determined only by the space-charge effect of the electrons; (3) at the midplane of the system, the value of the density of electrons and that of the motive are n_{em} and ψ_m, respectively; and (4) the system is electrically neutral by virtue of positive surface charges at its boundaries. A system that satisfies conditions 1 to 4 will be called the *interelectrode system* and is shown schematically in Fig. 9.5.

The motive ψ' of the interelectrode system is given by the relation (Eq. 10.63)

$$\xi = \sqrt{2} \arccos \exp\left(-\frac{\gamma}{2}\right) \tag{9.2}$$

where

$$\gamma = \frac{\psi_m - \psi'}{kT}$$

$$\xi = \left(\frac{e^2 n_{em}}{\varepsilon_0 kT}\right)^{1/2} (x - x_m)$$

$$= 0.14 \left(\frac{n_{em}}{T}\right)^{1/2} (x - x_m) \qquad \text{for } n_{em} \text{ in electrons/cm}^3,\ T \text{ in } ^\circ\text{K, and } x \text{ in cm}$$

where x is the distance from a boundary and $x - x_m$ (> 0) is the distance from the maximum motive. A graph of ψ' vs. x for an arbitrary ψ_m and for $n_{em} = 10^{11}$ electrons/cm^3, $T = 2000^\circ$K, and $x_m = 2 \times 10^{-3}$ cm is shown in Fig. 9.6.

Being a perfect gas, the interelectrode system can be subdivided into differential, homogeneous, open phases in mutual stable equilibrium, and, for each of these phases, the perfect-gas relations (Sec. 7.8.13) can be used to compute a variety of properties. For example, the chemical potential of each phase is given by Eq. 7.218. For a phase at position x, if the internal structure of the electrons is represented by a twofold degenerate ($g_e = 2$) energy eigenvalue $\bar{\varepsilon}_{oe} = 0$, Eq. 7.218 yields that the chemical potential μ^g_{ec} is given by the relation

$$\mu^g_{ec} = -kT \ln\left[\frac{2\,(2\pi m_e kT)^{3/2}}{n_e h^3}\right]$$

$$= -kT \ln\left[\frac{4.47 \times 10^{21}}{n_e}\left(\frac{T}{1000}\right)^{3/2}\right] \tag{9.3}$$

where n_e is the density of electrons at x in electrons/cm^3 and T is in $^\circ$K. Since n_e is smaller than about 10^{15} electrons/cm^3, Eq. 9.3 shows that μ^g_{ec} is negative throughout the interelectrode system.

The value of μ^g_{ec} defines the value of the total potential or electrochemical potential μ_e of the electron phase since μ_e equals the sum of the chemical

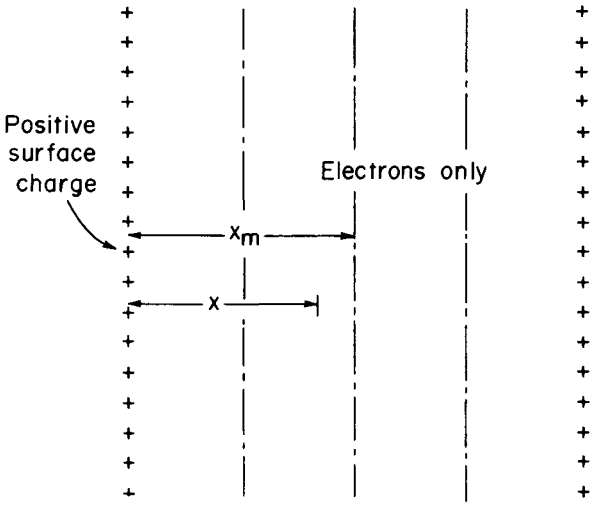


Fig. 9.5 Interelectrode system, with positive charge on each boundary surface equal to half the total electron charge between the two surfaces.

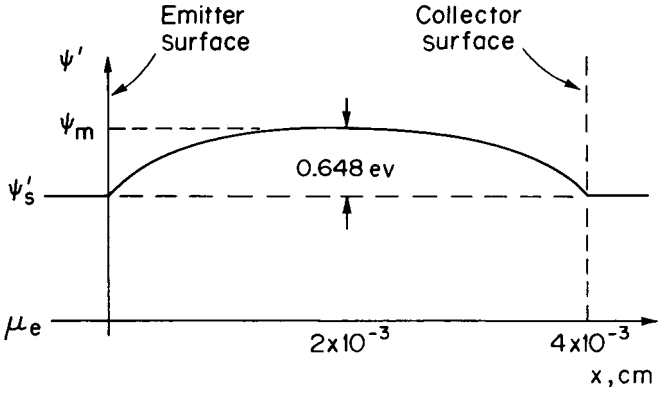


Fig. 9.6 Motive of interelectrode system for $x_m = 2 \times 10^{-3}$ cm, $T_E = 2000°K$, and $n_{em} = 10^{11}$ electrons/cm^3.

potential $\mu_{\xi c}^{\xi}$ plus the electron motive ψ' (Eq. 7.106). The value of μ_e with respect to ψ' is shown in Fig. 9.6. It is the same for all phases since all phases are in mutual stable equilibrium.

Combining Eq. 7.251 and the perfect-gas relation between pressure and temperature, we find that the electron-random-flux density Γ_{re} at position x is given by the expression

$$\Gamma_{re} = n_e \left(\frac{kT}{2\pi m_e}\right)^{1/2} \tag{9.4}$$

It is also given by an expression of the form of Eq. 7.252:

$$e\Gamma_{re} = AT^2 \exp\left(-\frac{\psi' - \mu_e}{kT}\right) \tag{9.5}$$

since here the motive ψ' is equal to the quantity $-e\psi$ of Sec. 7.10.3 and $\bar{\varepsilon}_{oe}$ equals zero.

Finally, the value of the positive charge density per unit surface area of either boundary of the interelectrode system can be found by computing $\varepsilon_0 E_0$, where E_0 is the gradient of the motive at the boundary. For all practical interelectrode systems in a stable equilibrium state, this charge density is many orders of magnitude smaller than the positive charge per unit surface area of the emitter or the collector corresponding to the positive nuclei on either of these surfaces.

(b) Electrode System. Two open systems, one for each electrode, can be defined by a virtual elimination of the interelectrode system described from the system shown in Fig. 9.3. This elimination is possible because the field in the interelectrode system is weak and is independent of the fields of the positive nuclei.

The elimination implies that each of the positive and negative charges of each electrode is reduced by an amount equal to the charge at one boundary of the interelectrode system. Each electrode with the reduced charge is electrically neutral and will be called an *electrode system.*

For all practical purposes, the charge reduction does not affect the motive except in the region between the two open electrode systems. The motive ψ'' of the two electrode systems vs. x is shown in Fig. 9.7. At each position x, its value is given by the motive ψ in Fig. 9.4 minus the difference $\psi' - \psi_s'$ of Fig. 9.6. The difference $\psi' - \psi_s'$ represents the effects of the charges that have been eliminated.

Figure 9.7 shows that, starting from the bulk of the emitter and up to certain points beyond the surface, the motive is almost identical to that in Fig. 9.4; in other words, the difference $\psi' - \psi_s'$ is negligible compared with the overall

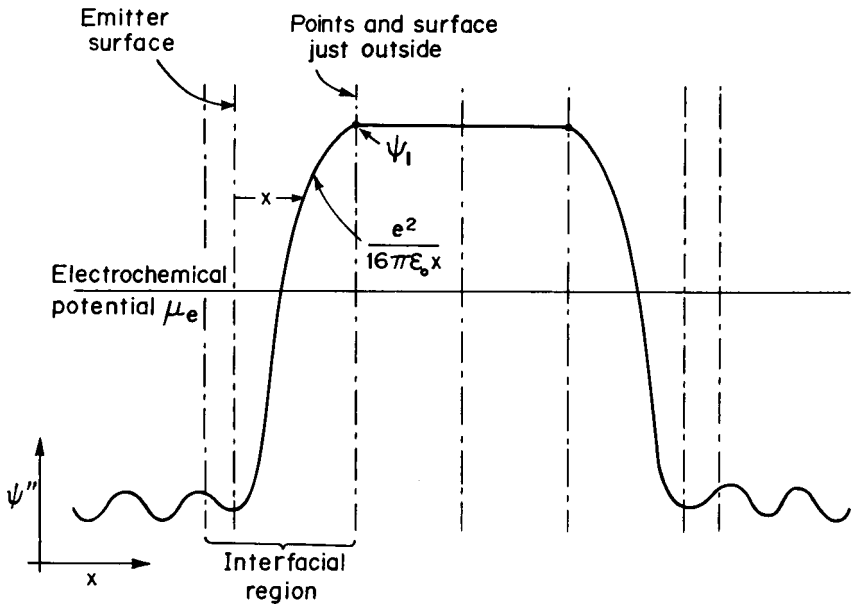


Fig. 9.7 Representative motive ψ'' vs. x for the two open electrode systems. The value of ψ_1 is almost equal to the value of ψ_s' in Fig. 9.6.

motive. The points up to which the difference $\psi' - \psi_s'$ of Fig. 9.6 is negligible compared with the overall motive in Fig. 9.4 are defined as the *points just outside*, and the surface that contains these points is the *surface just outside*. The points just outside are located at a distance of about 500 Å from the surface of the electrode, and their motive is denoted by ψ_1.

The definition of the points just outside has two implications: (1) the value of ψ_1 of the motive at these points is almost equal to ψ_s'; and (2) the values of the motive in the vicinity of the points just outside can be approximated by the expression $-e^2/16\pi\varepsilon_0 x$.

Between the points just outside the emitter and just outside the collector, the motive ψ'' is almost constant, and, beyond the points just outside the collector, it is again almost identical to that in Fig. 9.4. Included in Fig. 9.7 is the value of the electrochemical potential μ_e relative to the motive ψ'' for each of the electrode systems. It is equal to the value μ_e of the interelectrode system because of the condition for mutual stable equilibrium.

The most important result of this discussion is that, for all practical purposes, the force fields of the emitter electrode open system in a stable equilibrium state extend only up to the surface just outside. Beyond this surface the force fields

are determined by the systems with which the electrode is in mutual stable equilibrium. From this result it follows that the value of the motive ψ_1 at the surface just outside the emitter and values of properties at this surface are characteristic of the pure uniform surface under consideration.

At the motive ψ_1 in Fig. 9.7, one property of interest is the electron-random-current density $e\Gamma_1$. This density cannot be computed for the electrode system corresponding to the emitter because of lack of an accurate and tractable model for the behavior of electrons in a solid.† On the other hand, since the value ψ_1 is almost equal to the motive ψ_s' of the interelectrode system and since there are no sinks of electrons in the interelectrode space of the system in Fig. 9.3, $e\Gamma_1$ must be equal to the electron-random-current density of the interelectrode system at the points just outside, and therefore (Eq. 9.5)

$$e\Gamma_1 = AT^2 \exp\left(-\frac{\psi_1 - \mu_e}{kT}\right) \tag{9.6}$$

Equation 9.6 can be rewritten in the form

$$e\Gamma_1 = AT^2 \exp\left(-\frac{\phi}{kT}\right) \tag{9.7}$$

where

$$\begin{aligned} \phi &\equiv \psi_1 - \mu_e \\ &= -(\mu_{ec}^g)_1 \\ &= -(\mu_{ec}^s)_1 \end{aligned} \tag{9.8}$$

The quantity ϕ is called the work function of the pure uniform surface. Its value is equal to the negative of the chemical potential $(\mu_{ec}^g)_1$ of the interelectrode system or $(\mu_{ec}^s)_1$ of the emitter at the points just outside.

The value of the work function ϕ depends on the value of the temperature T because both the motive and the electrochemical potential (Eq. 9.8) are functions of temperature. Over the range from 0°K to practical thermionic emission temperatures, the total change in ϕ is only a very small fraction of the value of ϕ itself although at the present time both theoretical and experimental information about the magnitude and sign of the rate of change of ϕ with T (i.e., of the derivative $d\phi/dT$ vs. T) is incomplete.

†The Sommerfeld gas model often used to describe electrons in a metal is grossly inaccurate.

Because ϕ depends on T, Eq. 9.7 does not describe explicitly the variation of Γ_1 with T. For surfaces for which ϕ is a linear function of temperature, however, Eq. 9.7 can be easily modified to indicate explicitly the dependence of Γ_1 on T. This modification is discussed in Sec. 9.2.7.

In Sec. 9.2.3 it is shown that ϕ is a key characteristic property of the electron-emission capabilities of a pure uniform surface. Apart from temperature, its values depend strongly on the structure and the composition of the surface. These dependences are discussed in Sec. 9.3, together with some approximate methods for calculating the values of ϕ.

9.2.3 Electron-Saturation-Current Density. This section defines an electron current density characteristic of the emission capability of a pure uniform surface of an emitter in a steady state.

We will consider the two electrodes in Fig. 9.2 to be in a steady state such that: (1) Net electron current flows from the emitter to the collector via the interelectrode space and back to the emitter via the battery. (2) The emitter is isothermal at temperature T. (3) The collector is isothermal at a very low temperature T_C. (4) The voltage of the battery is set so that, at the points just outside the emitter, the force on an electron is equal to zero, whereas beyond these points the force field is accelerating for electrons originating from the emitter. This state will be called the *zero-field steady state.*

For the zero-field steady state, the value of the electron current density or any other flux through the interelectrode space cannot be computed rigorously for the reasons given in Sec. 9.1. Approximate expressions for the fluxes can be derived, however, if the following assumptions are valid.

Assumption 1. Despite the nonzero fluxes, both emitter and collector are assumed to be in stable equilibrium states. This assumption is justified when each flux density is only a very small fraction of the corresponding random flux density in the bulk of the electrode under consideration.

If the emitter is regarded as an open system, it follows from assumption 1 that the values of the electrochemical potential μ_e and of the motive up to the points just outside the emitter are the same as those discussed in Sec. 9.2.2. Similar remarks apply to the collector at temperature T_C. Beyond the points just outside the emitter, the motive is decreasing since the field is accelerating for electrons originating from the emitter. A representative plot of the motive vs. x is shown in Fig. 9.8. The figure includes the electrochemical potentials of the two electrode systems.

Assumption 2. The thermionic electron current density J_S that flows from emitter to collector via the interelectrode space is assumed to be a fraction of the random electron current density at the points just outside the emitter when the two electrodes are in mutual stable equilibrium, the fraction being $(1 - \bar{r}_1)$, where $\bar{r}_1$ is the average reflection coefficient of a barrier of height ψ_1

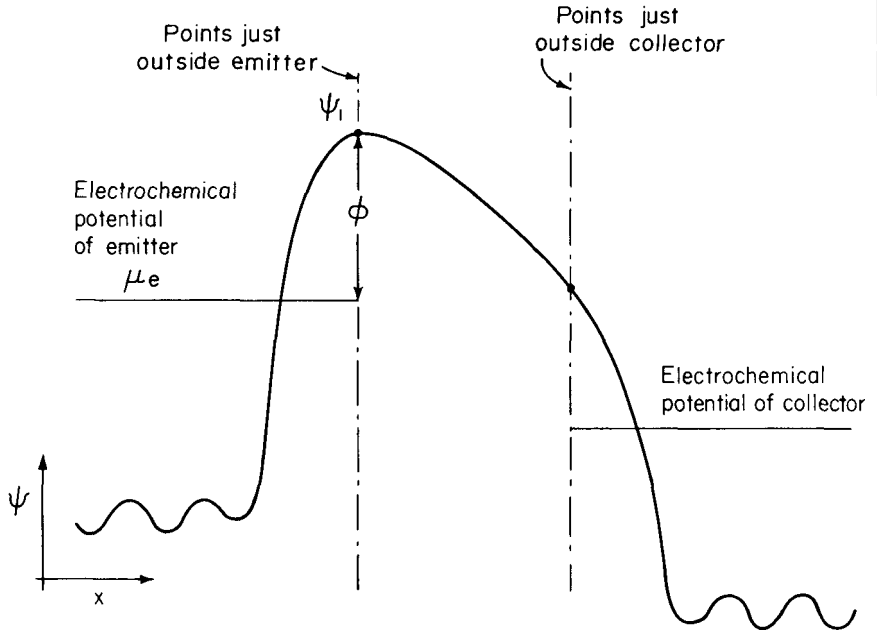


Fig. 9.8 Representative plot of motive ψ vs. x for a two-electrode system in a zero-field steady state.

(Sec. 5.5.2) penetrated by a swarm of electrons emitted thermionically from the emitter and moving in the direction from emitter to collector. Accordingly, J_S is given by the relation

$$J_S = e(1 - \bar{r}_1)\Gamma_1 = AT^2(1 - \bar{r}_1) \exp\left(-\frac{\phi}{kT}\right) \tag{9.9}$$

where Γ_1 and ϕ are defined in Sec. 9.2.2. The quantity J_S is called the electron-saturation-current density. It depends only on properties of the emitter. Equation 9.9 is called the Richardson, the Richardson–Dushmann, or the Richardson–Dushmann–Laue equation.

The factor $1 - \bar{r}_1$ is included in Eq. 9.9 to account for quantum-mechanical reflection properties of the barrier in the vicinity of the emitter surface. Approximate methods for calculating $\bar{r}_1$ are discussed in Ref. 1. It is shown that $\bar{r}_1$ is a function of temperature. In general, however, $\bar{r}_1$ for metal surfaces is much smaller than unity and can be neglected from Eq. 9.9.

Consistent with assumption 2 is the neglect of any current contribution to J_S from the collector. Since current emitted from an electrode is assumed to be

proportional to exp $(-\phi/kT)$ and the collector is kept at very low temperature, the electron current originating from the collector is negligible.

Implicit in the form of Eq. 9.9 is also the approximation that the motive between the points just outside the emitter and the collector does not affect the current flow through the interelectrode space. This is an excellent approximation for most applications. On the other hand, it proves invalid when *tunneling effects* are important.

The current density J_S given by Eq. 9.9 is compared with experimental results later. Here it suffices to remark that:

1. Equation 9.9 does not indicate the explicit dependence of J_S on T, because both ϕ and $\bar{r}_1$ are functions of the emitter temperature T. The explicit dependence of J_S on T can be put in evidence only in certain cases that are discussed in Sec. 9.2.7.

2. If quantum-mechanical reflection effects were included in its derivation, Eq. 2.4, Vol. I, Chap. 2, would be formally identical to Eq. 9.9. Because the two equations are obtained apparently from different points of view, the identity should not be mistaken as supporting evidence for either point of view. This remark can be verified by reviewing the premises of the analysis presented in Chap. 2. In Chap. 2 the derivation of J_S is based on the free electron or Sommerfeld gas approximation and the resulting Fermi factor for the one-electron distribution function. The Fermi factor corresponds to a unique state only when that state is a stable equilibrium state (Sec. 8.8.3). It follows that Eq. 2.4 represents a result which is associated with a stable equilibrium state and which can be interpreted as an electron-saturation-current density only in the light of assumptions that are essentially equivalent to assumptions 1 and 2 introduced earlier.

In Chap. 2, the Fermi factor is expressed in terms of the Fermi energy. The Fermi energy of a Sommerfeld gas corresponds to the electrochemical potential. It follows that, if reflection effects are introduced into Eq. 2.4, it becomes identical to Eq. 9.9. In general, however, the electrochemical potential of a metal differs from the Fermi energy of the Sommerfeld gas approximation for the electronic structure of the metal. The reason is that the electrochemical potential includes exchange and correlation effects and effects resulting from the periodic variation of the motive inside the metal whereas the Fermi energy of the Sommerfeld gas approximation does not.

If the Fermi energy is replaced by the electrochemical potential, the identity of Eqs. 2.4 and 9.9 would hold regardless of what model is used for the electronic structure of the metal in the derivation of Eq. 2.4. The reason is that the identity is based on the assumption that electrons in the metal are in mutual stable equilibrium with electrons outside the metal and on the fact that outside the metal the perfect-gas approximation is valid.

9.2.4 Velocity Distribution. For the stable equilibrium states discussed in Sec. 9.2.2, the one-electron distribution function for each elementary phase in the interelectrode system is represented by the Maxwellian velocity distribution $f(\mathbf{v})$, namely,

$$f(\mathbf{v}) = n_e \left(\frac{m_e}{2\pi kT}\right)^{3/2} \exp\left(-\frac{m_e v^2}{2kT}\right) \tag{9.10}$$

where n_e is the density of electrons per unit volume of the elementary phase in question.

For a pure uniform metal surface in a steady state, a practical but not necessarily rigorous approximation for the one-electron distribution function at the points just outside the emitter surface is given by the relation

$$f(\mathbf{v}) = n_e f_h(\mathbf{v}) \tag{9.11}$$

where $f_h(\mathbf{v})$ is the normalized half-Maxwellian velocity distribution given by the relation

$$f_h(\mathbf{v}) = \begin{cases} 2\left(\dfrac{m_e}{2\pi kT}\right)^{3/2} \exp\left(-\dfrac{m_e v^2}{2kT}\right) & \text{for } v_x \geq 0 \\ 0 & \text{for } v_x < 0 \end{cases} \tag{9.12}$$

9.2.5 Kinetic Energy Flux. When the velocity distribution is adequately represented by a half-Maxwellian and reflection effects are negligible, the kinetic-energy-flux density w_S and the energy-flux density u_S at the surface just outside are given by the relations

$$w_S = \int_0^\infty dv_x \int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} dv_z \frac{m_e v^2}{2} v_x f(\mathbf{v}) = 2kT \frac{J_S}{e} \tag{9.13}$$

and

$$u_S = (\psi_1 + 2kT)\frac{J_S}{e} \tag{9.14}$$

9.2.6 Schottky Effect. We will consider the system in Fig. 9.2 to be in a steady state such that the value of the voltage of the battery results in a reduction of the motive in the region between the emitter surface and the points just outside the emitter. The voltage creates an applied field of strength E which is

accelerating for electrons originating from the emitter. This state will be called the *accelerating-field steady state.*

The accelerating field increases the amount of current flowing from emitter to collector with respect to its value for the zero-field steady state. For moderate field strengths E, the increase is called the Schottky effect.

The analysis of the Schottky effect is approached in a manner analogous to that used in Sec. 9.2.3. By definition, the applied field decreases the values of the motive in the region beyond the emitter surface with respect to those corresponding to zero-field strength at the points just outside. Schematics of the two motives in the vicinity of the surface are shown in Fig. 9.9. We see from Fig. 9.9 that, for the accelerating-field steady state, the maximum motive ψ_m is at distance x_m and that ψ_m is less than ψ_1 and x_m is less than x_1.

In the range between 0 and x_m, the motive is the algebraic sum of the motive corresponding to the emitter in a stable equilibrium state plus the motive ψ_a that results from the applied field E. The motive ψ_a varies linearly with distance and is given by the relation

$$\psi_a = -eEx \qquad \text{for } 0 < x < x_m \tag{9.15}$$

In the range x greater than x_m, the overall motive is derived by means of the methods discussed in Chap. 10.

The distance x_m is determined by the points at which the slope of the motive corresponding to the stable equilibrium state is equal and opposite to the slope of ψ_a. For moderate field strengths these points are in the range in which the motive for the stable equilibrium state depends primarily on the image force, and therefore

$$\frac{e^2}{16\pi\varepsilon_0 x_m^2} = eE \tag{9.16a}$$

or

$$x_m = \left(\frac{e}{16\pi\varepsilon_0 E}\right)^{1/2} = \frac{1.9 \times 10^{-4}}{E^{1/2}} \text{ cm} \qquad \text{for } E \text{ in volts/cm} \tag{9.16b}$$

With x_m given by Eq. 9.16b, we can easily verify that the difference $\Delta\psi$ between ψ_1 and ψ_m is given by the relation

$$\Delta\psi = \psi_1 - \psi_m = e\left(\frac{eE}{16\pi\varepsilon_0}\right)^{1/2}$$

$$= 3.4 \times 10^{-4} E^{1/2} \text{ ev} \qquad \text{for } E \text{ in volts/cm} \tag{9.17}$$

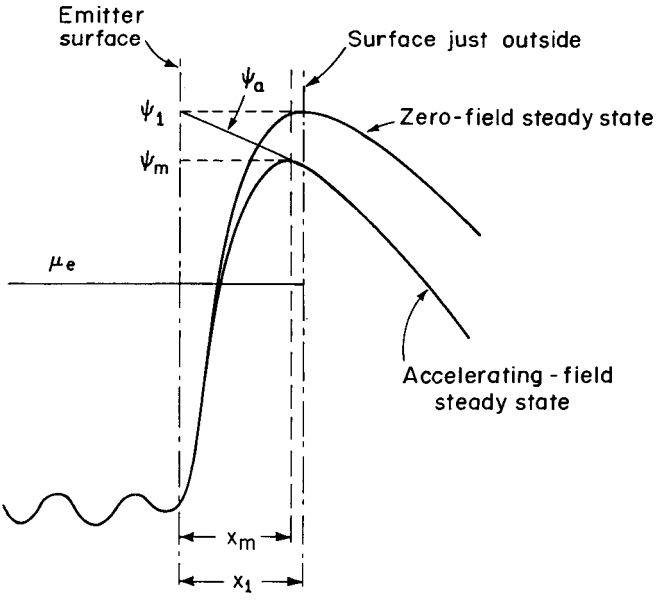


Fig. 9.9 Representative plots of motive vs. x in the vicinity of the emitter surface for zero-field and accelerating-field steady states.

The steady-state current density J is assumed to be given by Eq. 9.9 with ψ_1 replaced by ψ_m, so that

$$J = AT^2(1-\bar{r}_1)\exp\left(-\frac{\psi_m-\mu_e}{kT}\right)$$

$$= AT^2(1-\bar{r}_1)\exp\left(-\frac{\phi-\Delta\psi}{kT}\right)$$

$$= AT^2(1-\bar{r}_1)\exp\left(-\frac{\phi}{kT}\right)\exp\left[\frac{e\,(eE/16\pi\varepsilon_0)^{1/2}}{kT}\right] \tag{9.18}$$

In these relations the reflection coefficient for the barrier of height ψ_m is assumed to be equal to $\bar{r}_1$; i.e., it is the same as that of the barrier ψ_1. Moreover, the reflection effects of the motive beyond x_m are neglected. In principle, these two approximations need not be introduced.

Finally, at ψ_m the velocity distribution is assumed to be approximately half-Maxwellian, and the energy-flux density is assumed to be given by Eq. 9.14 with ψ_1 replaced by ψ_m.

The E-dependent exponential in Eq. 9.18 describes the Schottky effect on current density. This effect can be represented graphically as follows: For a given temperature T, the graph of ln J vs. $E^{1/2}$ is a straight line with a positive slope. This representation is called the *Schottky plot.* The intercept of the Schottky plot with the ln J-axis yields the electron-saturation-current density.

The effect of strong applied fields is the subject of *field emission* and is not described by Eq. 9.17. Field emission, which includes tunneling effects, will not be discussed in this book.

9.2.7 **Apparent Emission Constant.** For temperatures greater than 0°K, when ϕ is a linear function of temperature and $\bar{r}_1$ is much smaller than unity, Eq. 9.9 (or Eq. 9.18 for E equal to zero) can be written in a form that indicates the explicit dependence of the electron-saturation-current density J_S on T.

Indeed, when it is a linear function of temperature, ϕ is given by the relation

$$\phi = \phi^* + \alpha T \tag{9.19}$$

where ϕ^* is a quantity independent of T and α is the *temperature coefficient* of ϕ, namely,

$$\alpha = \frac{d\phi}{dT} = \text{constant for } T > 0 \tag{9.20}$$

For most uniform surfaces the linearity expressed by Eq. 9.19 is valid over a wide range of temperatures, except for T near 0°K where it cannot be valid (see also Sec. 9.3.10). Although Eq. 9.19 does not hold for T near zero, to a good approximation, ϕ^* can be regarded as the work function at 0°K and will be called the *apparent work function.*

In terms of the apparent work function, Eq. 9.9 can be expressed in the form

$$J_S = A^* T^2 \exp\left(-\frac{\phi^*}{kT}\right) \tag{9.21}$$

where A^* is the *apparent emission constant* given by the relation

$$A^* = A(1 - \bar{r}_1) \exp\left(-\frac{\alpha}{k}\right)$$

$$\approx A \exp\left(-\frac{\alpha}{k}\right) \quad \text{for } \bar{r}_1 \ll 1 \text{ and for all } T > 0 \tag{9.22}$$

When $\bar{r}_1$ is much smaller than unity, we see from Eq. 9.22 that A^* is independent of T, and therefore the explicit dependence of J_S on T is given by Eq. 9.21.

By virtue of Eq. 9.21, a graph of ln (J_S/T^2) vs. $1/kT$ should be a straight line with a negative slope equal to $-\phi^*$ and an intercept with the logarithmic axis equal to ln A^*. Such a graphic representation of the relation between J_S and T is called a *Richardson plot.*

When α is a positive quantity, namely, when ϕ is a linearly increasing function of T, we see from Eq. 9.22 that the value of A^* is smaller than that of A, i.e., smaller than 120 amp/cm^2-°K^2. On the other hand, when α is negative, namely, when ϕ is a linearly decreasing function of T, the value of A^* is greater than 120 amp/cm^2-°K^2.

9.2.8 Experimental Results. Since the procedures used in measuring electron thermionic emission from pure uniform surfaces are discussed in Chap. 13, we will mention procedures only briefly here. The experimental setup is somewhat similar to that shown in Fig. 9.2. The collector is kept at a very low temperature so that the back-emission current is negligible. At a given emitter temperature T, the electron current density J from emitter to collector is measured as a function of the applied field E. The results are graphically depicted in the form of a Schottky plot, i.e., as a graph of ln J vs. $E^{1/2}$. The intercept of the Schottky plot with the ln J-axis yields the saturation current density J_S at the given T. Then saturation current densities at different emitter temperatures are graphically depicted, in the form of a Richardson plot, as graphs of ln (J_S/T^2) vs. $1/kT$. The data on the Richardson plot are correlated by a relation of the form of Eq. 9.21. The end result of this procedure is the determination of values of ϕ^* and A^* for the surface under consideration. These two values provide the means for comparing theoretical and experimental data on electron emission from pure uniform surfaces. In addition, they can be used in Eqs. 9.20 and 9.21 to compute the value of ϕ if the average reflection coefficient $\bar{r}_1$ is much smaller than unity.

Other comparisons between theory and experiment are based on measurements of the velocity distribution of the emitted electrons and of the energy flux through the emitter.

9.3 Calculating the Work Function of a Pure Uniform Surface

9.3.1 General Remarks. This section presents some approximate methods for calculating the electron-saturation-current density J_S from a pure uniform surface of a refractory metal.

In the context of the one-electron picture, J_S depends primarily on the value of the work function ϕ and, to a lesser degree, on the relatively small value of

the average reflection coefficient $\bar{r}_1$. Accordingly, the calculation of J_S is equivalent to that of ϕ and $\bar{r}_1$. To be useful, the calculation of ϕ must be fairly accurate, because even small errors in ϕ result in large errors in J_S. For example, at 2000°K an error of 6% in the work function ϕ of about 3 ev would result in an error of about 200% in the saturation-current density J_S.

In general, rigorous calculations of the values of ϕ and $\bar{r}_1$ are very difficult mathematical tasks, and thus far only approximate results have been obtained by use of both fundamental theory and experience.

The starting point of most analyses is the calculation of the work function at 0°K. This work function is assumed to be approximately equal to the apparent work function ϕ^*. The problem is approached by means of the Born–Oppenheimer approximation for the electronic structure of the emitter. Two methods have been developed along this approximation. In the first, the value of $\phi(T = 0)$ is found by estimating the quantities ψ_1^* and μ_e^* that are required by the relation

$$\phi(T = 0) = \psi_1^* - \mu_e^* \approx \phi^* \tag{9.23}$$

Both one-electron and many-electron models have been used, and a number of important qualitative results have been obtained (Secs. 9.3.3 to 9.3.6). Quantitatively, however, the results leave much to be desired.

In the second method, the one-electron description of the electronic structure is interpreted in terms of monocentric spin orbitals and ϕ^* is estimated by means of empirical relations employed in physical chemistry (Sec. 9.3.9).

The next step in analyses of ϕ is estimating the temperature coefficient α (Eq. 9.20). A number of approximate models are used for this purpose. One model is based on the Debye theory of lattice vibrations, which yields estimates both of ϕ as a function of T and of α (Secs. 9.3.7 and 9.3.10).

The value of the average reflection coefficient $\bar{r}_1$ depends on the shape of the motive in the vicinity of the points just outside and the distribution of electrons in different states. Calculations of $\bar{r}_1$ have been made for approximate motive shapes and electron distributions.[1] These calculations will not be discussed here. For pure uniform surfaces, both theory and experiment suggest that $\bar{r}_1$ is small compared with unity.

First, to clarify the mathematical difficulties and to provide a frame of reference for interpreting various approximations, we outline in Sec. 9.3.2 a rigorous program for calculating ϕ. Then, we discuss the approximate methods for calculating ϕ^* and ϕ and review briefly the procedures for calculating α.

9.3.2 Rigorous Program for Calculating ϕ. The rigorous calculation of ϕ would require implementation of the following practically impossible and unnecessarily complicated program.

We will consider a finite piece of a metal inside a large container as shown schematically in Fig. 9.1 and will regard it as a two-component open system A consisting of electrons and positive nuclei. Thus its energy eigenstates are determined by many possible combinations of eigenvalues of the numbers of electrons and positive nuclei.

We assume that the metal is in a stable equilibrium state specified by the following conditions: (1) the temperature is equal to T; (2) the expectation values of the positive and negative charges are equal to each other, i.e., the metal is electrically neutral; (3) the set of energy eigenstates is restricted so that the expectation values of the positions of the positive nuclei form a regular lattice enclosed by a specified pure uniform surface, with a lattice constant equal to the experimental value of the freely expanding metal at temperature T; and (4) with respect to a common but otherwise arbitrary zero-energy reference level, all the necessary many-particle energy eigenstates can be enumerated consecutively. Thus, $G_1, G_2, \ldots, G_j, \ldots$, will denote the restricted energy eigenstates; $E_1, E_2, \ldots, E_j, \ldots$, the energy eigenvalues; $U_1, U_2, \ldots, U_j, \ldots$, the energy eigenfunctions; and $(n_{e1}, n_{i1}), (n_{e2}, n_{i2}), \ldots, (n_{ej}, n_{ij}), \ldots$, the eigenvalues of the number of electrons and positive nuclei. In this notation a different symbol is used for each characteristic of each energy eigenstate even though the values of some symbols may be identical. The subscripts e and i denote electrons and positive nuclei, respectively, and eigenfunction U_j is of the form $U_j[q_e(1), q_e(2), \ldots, q_e(n_{ej}), q_i(1), q_i(2), \ldots, q_i(n_{ij})]$, where $q_e(m)$ and $q_i(n)$ denote the observables, coordinates and spin, of the mth electron and the nth ion, respectively, of energy eigenstate G_j.

For the assumed stable equilibrium state, the value of the electrochemical potential μ_e of the electrons and the value of the motive ψ_1 at the points just outside define the value of the work function ϕ (Eq. 9.8). In principle, μ_e and ψ_1 can be computed as follows:

The state is described by the grand canonical distribution x_j^0 given by the relation

$$x_j^0 = \frac{\exp\left[(\mu_e n_{ej} + \mu_i n_{ij} - E_j)/kT\right]}{\sum_j \exp\left[(\mu_e n_{ej} + \mu_i n_{ij} - E_j)/kT\right]} \tag{9.24}$$

where μ_i is the total potential of the positive nuclei. The expectation value N of the number of electrons satisfies the relations

$$N = \sum_j x_j^0 n_{ej} = \frac{\sum_j n_{ej} \exp\left[(\mu_e n_{ej} + \mu_i n_{ij} - E_j)/kT\right]}{\sum_j \exp\left[(\mu_e n_{ej} + \mu_i n_{ij} - E_j)/kT\right]} \tag{9.25}$$

and

$$\frac{N}{Z} = \sum x_j^0 n_{ij} = \frac{\sum_j n_{ij} \exp\left[(\mu_e n_{ej} + \mu_i n_{ij} - E_j)/kT\right]}{\sum_j \exp\left[(\mu_e n_{ej} + \mu_i n_{ij} - E_j)/kT\right]} \tag{9.26}$$

where Z is the number of positive charges on each nucleus. In principle, Eqs. 9.25 and 9.26 can be solved for μ_e as a function of T, if the exact values of the many-particle energy eigenvalues E_j are known. These eigenvalues are impossible to compute, however, and therefore the function $\mu_e(T)$ cannot be accurately evaluated.

The calculation of ψ_1 requires reduction of the many-particle problem into a one-electron problem and involves the following four steps: (1) calculate the motive ψ_j' for the metal in a pure energy eigenstate G_j; (2) calculate the motive ψ for the metal in a stable equilibrium state at temperature T and electrical neutrality; (3) calculate the part of the motive ψ which is independent of the field of the positively charged nuclei; and (4) calculate ψ_1. The details of these calculations are given in the following paragraphs.

To calculate the motive ψ_j' corresponding to the energy eigenfunction U_j, we will fix the observables $q_e(1)$ of electron number 1. For the fixed position of this electron, the electron charge density at position $q_e(2)$ of electron number 2 will be proportional to the integral I_{ej} of the probability density $U_j^* U_j$ over the observables of all particles except electrons 1 and 2. The energy eigenfunction U_j is antisymmetric with respect to the observables of the electrons, and therefore the value of I_{ej} will be the same whether the integration is performed over the observables of electrons 3, 4, . . ., n_{ej} or over any other $n_{ej} - 2$ electrons. The integral I_{ej} yields the probability of finding electron number 2 at $q_e(2)$ if electron number 1 is at $q_e(1)$, and therefore it is a function of the position corresponding to $q_e(1)$. Since there are $n_{ej} - 1$ electrons besides the one held fixed, the total electron charge density at $q_e(2)$ will be proportional to $(n_{ej} - 1)I_{ej}$.

An identical procedure can be used to find that the charge density of the positive nuclei is proportional to $(n_{ij} - 1)I_{ij}$, where I_{ij} denotes the integral of $U_j^* U_j$ over all particles except two ions, one held fixed at $q_i(1)$ and the other being at $q_i(2)$. Since U_j is symmetric or antisymmetric with respect to the observables of the positive nuclei, the value of I_{ij} will be the same whether the integration is performed over the observables of positive nuclei 3, 4, . . ., n_{ij} or over any other $n_{ij} - 2$ positive nuclei. The integral I_{ij} is a function of the position corresponding to $q_i(1)$.

The two charge densities $(n_{ej} - 1)I_{ej}$ and $(n_{ij} - 1)I_{ij}$ can be used in Poisson's equation to compute the potential energy of one electron or one positive nucleus at position $q(1)$, the so-called one-electron and one-ion potential

energies. The one-electron potential energy is equal to the motive ψ_j' and depends on the position $q_e(1)$. It is noteworthy that ψ_j' includes electrostatic and all other effects since it is derived from an exact energy eigenfunction.

The method of calculation of ψ_j' can be repeated for all energy eigenstates G_1, $G_2, \ldots, G_j, \ldots$. Thus a set of motives $\psi_1', \psi_2', \ldots, \psi_j' \ldots$ is obtained.

In general, the preceding program for calculating $\psi_1', \psi_2', \ldots, \psi_j', \ldots$ is impossible to carry out both because the exact many-particle energy eigenfunctions cannot be computed and because the equation for the evaluation of each ψ_j' is complicated. Nevertheless, the idea of the program has been used successfully by Slater[2] to discuss the electronic structure of the helium atom and by Wigner[3] to discuss correlation in a metal.

When the metal is in a stable equilibrium state, the motive ψ is given by the expectation value of the motives of the various energy eigenstates. Specifically,

$$\psi = \sum_j x_j^0 \, \psi_j' = \frac{\sum_j \psi_j' \exp\left[(n_{ej}\mu_e + n_{ij}\mu_i - E_j)/kT\right]}{\sum_j \exp\left[(n_{ej}\mu_e + n_{ij}\mu_i - E_j)/kT\right]} \tag{9.27}$$

Since each ψ_j' is a function of position, Eq. 9.27 shows that the motive ψ is a function of both position and temperature.

At points far away from the surface, the motive ψ becomes practically independent of contributions other than pure electrostatic; in other words, the integrals I_{ej} and I_{ij} become independent of $q_e(1)$ and $q_i(1)$, respectively. The points nearest the surface at which this independence occurs define the points just outside and, therefore, define ψ_1 (see also Appendix 9.A). Thus the value of the work function ϕ can be computed as the difference $\psi_1 - \mu_e$. Since both ψ_1 and μ_e are functions of T, it is clear that ϕ is a function of T.

Although this approach to the calculation of ϕ cannot be implemented, it provides a means for identifying the difficulties and a frame of reference for interpreting the drastic approximations that must be used to avoid them.

One difficulty arises from the complexity of Eqs. 9.25 to 9.27 at finite temperature T. Initially this difficulty is avoided by restricting the theory to the calculation of μ_e^* and ψ_1^* in the limit of 0°K. In this limit the only energy eigenstate of interest is the ground state of the neutral metal. The second and much more severe difficulty arises from the complexity of the calculation of exact many-particle energy eigenstates. We avoid this difficulty by regarding the metal as an electron open system and by using an approximate and tractable one-electron model. A method of proceeding along these lines and the results obtained therefrom are summarized in Secs. 9.3.3 to 9.3.6.

9.3.3 Interfacial Electron Motive ψ^*—The Double Layer. In general, calculating the motive at the points just outside is part of the calculation of the

motive throughout the bulk and in the interfacial region (Fig. 9.7). The motive in the interfacial region will be called the *interfacial electron motive* or simply the *interfacial motive.*

For system A described in Sec. 9.3.2, in the limit of 0°K, we can readily verify that the state is represented solely by the ground state of the neutral metal. Let this state be G_N, its energy eigenvalue E_N, its energy eigenfunction U_N, and its number of electrons N. The number N can be taken equal to the number of valence electrons only. The electrons of the inner shells are tightly bound to the nuclei and can be regarded as parts of the positive nuclei.

In the bulk, approximate values of the motive ψ^* corresponding to G_N can be computed by assuming that: (1) The number of positive nuclei is fixed at the eigenvalue required for electrical neutrality; in other words, A is approximated by a one-component (electrons only) open system. (2) The nuclei are immobile; in other words, the Hamiltonian operator for A is represented by the Born–Oppenheimer approximation. The Born–Oppenheimer approximation describes a many-electron problem. (3) The ground state G_N can be represented by a one-electron approximation; in other words, the energy eigenfunction U_N can be expressed in the form of a Slater determinant of N one-electron spin-orbitals u_i, and the energy eigenvalue E_N can be written as a sum of the one-electron energy eigenvalues ϵ_i of the one-electron spin-orbitals.

To a first approximation, the one-electron eigenstates can be computed as solutions of a one-electron eigenvalue problem of the form of Eq. 5.188 in which the potential energy is assumed to be the sum of the self-consistent electrostatic or Hartree potential energy $V(\mathbf{r})$ plus an approximate exchange energy correction V_{ex} plus an approximate correlation energy correction V_c. Although correlation effects can never be described exactly by a one-electron approximation, they are usually included in the one-electron problem via V_c.

Now, by definition, the potential energy $V(\mathbf{r}) + V_{ex} + V_c$ is approximately equal to the motive ψ^* in the bulk. It is a periodic function of position and extends up to the points just inside the surface.

In the interfacial region, the motive can be computed by means of a procedure similar to that used for the bulk. The analysis begins with the calculation of an average one-electron electrostatic potential energy that is consistent with an assumed net charge density at each position beyond the points just inside. This potential energy, also called the *Hartree part of the interfacial motive,* differs from $V(\mathbf{r})$ in the bulk. The reason is that the electron charge distribution around lattice sites near the surface does not have the same symmetry as that inside the metal. For example, along the direction normal to the surface, this charge distribution, together with the positive nuclei at the surface, results in a net electric dipole moment, whereas in the bulk the net electric dipole moment is equal to zero. The dipole moment is often thought of as arising from a double charge layer at the surface and is called the *double layer.*

The Hartree part is grossly inadequate for the interfacial motive because it does not include exchange and correlation effects. These effects can be included by means of approximate exchange and correlation energy corrections to the Hartree part, and thus an interfacial motive is estimated. Improved estimates can be obtained by means of iterations.

Correlation effects are easily understood and are predominant for a charged particle at a position far from the surface. "Far from the surface" means at distances much greater than atomic dimensions. For such a position the surface of the metal can be replaced by a structureless, perfectly conducting plane. With respect to this plane, the electrostatic and exchange effects are practically equal to zero, and the correlation energy is that resulting from a unit charge distributed on the surface according to the rules of image fields. Thus, with respect to a zero-energy reference level at infinity, the motive ψ is given by the relation

$$\psi = -\frac{e^2}{16\pi\varepsilon_0 x} \qquad \text{for positions far from the surface}$$

This form of the motive is independent of the sign of the charge.

Analyses show that exchange and correlation effects beyond the points just inside are both much more predominant than and different from those in the bulk. Consequently the interfacial motive differs from that in the bulk because of differences in three quantities in each of the two regions: (1) the electron charge distribution and, therefore, the double layer; (2) the exchange energy; and (3) the correlation energy. Although values of these three quantities are very closely interrelated, the distinction is helpful in developing approximate computational methods for analysis of the motive.

The first quantum-mechanical analysis of the interfacial motive of a metal surface, more or less along the lines cited, was performed by Bardeen[4] in his treatment of the double-layer contribution to the work function ϕ^*. In this analysis no attempt is made to account explicitly for the dependence of the motive on the structure of the surface. Moreover, because of the complexity of the problem, many approximations are introduced and numerical results are found only for a monovalent metal with a density of valence electrons approximately equal to that of sodium. On the basis of numerical results, Bardeen reached the following conclusions:

1. The interfacial motive is due primarily to exchange and correlation effects and to a lesser degree to ordinary electrostatic effects. The same conclusions apply to the work function ϕ^*.

2. No single interfacial motive is satisfactory for electrons of all velocities, because exchange and correlation effects are velocity dependent; in other words, each electron velocity has its own motive.

3. The dipole moment of the double layer per unit surface area or, equivalently, the electrostatic potential energy V_0 is relatively small for monovalent metals.

Quantum-mechanical calculations of interfacial electron motives performed by Juretschke,[5] Loucks and Cutler,[6] Gadzuk,[7] and others are more refined than those performed by Bardeen.[4] The results, however, are quantitatively but not qualitatively different from those of Bardeen.

A question that can be raised here is: Are all three conclusions about the interfacial motive generally valid? The question cannot be answered rigorously because of lack of exact results for comparisons. Nevertheless, some theoretical considerations suggest that at least conclusions 2 and 3 should not be used without further qualifications.

Conclusion 1 is generally accepted because exchange and correlation effects are found to be important in determining both the electronic structure and the interfacial motive of a metal.

Conclusion 2 cannot be correct, however. As discussed by Slater,[8] the dependence of exchange and correlation effects on velocity or, equivalently, on the magnitude of the reduced wave vector $\mathbf{k}$ is a consequence of the approximations introduced in the calculations rather than a representation of a true physical phenomenon. In fact, if exchange effects were k-dependent, we would conclude that the density of states at the Fermi surface is equal to zero, although such a conclusion is contrary to experimental observations.

The discussion in Sec. 9.3.2 and the rigorous definition of ψ indicate clearly that the motive must be unique for all electrons. This uniqueness is also indicated by the results of a self-consistent many-electron theory of work functions and surface potential characteristics for selected metals which has been developed by Smith.[9] The interfacial motive computed by means of this theory is independent of $\mathbf{k}$ and is the same for all electrons.

Finally, conclusion 3 about the dipole moment and V_0 is hard to judge. We will see later that some approximate calculations of the work function ϕ^* of monovalent elements indicate that the double-layer contribution to the interfacial motive is negligible. The correctness of this result cannot be judged, however, because the magnitude of the error in other terms that determine the work function is not known. Many investigators believe that for multivalent elements the double-layer contribution to the interfacial motive and the work function may be larger than that for monovalent metals. In fact, it is often stated that differences in the values of work functions of surfaces of the same metal but of different crystallographic orientations are due only to differences in the strength of the double layer. This belief may not be justified, however. Exchange and correlation effects may be different for different surfaces of the same element and, hence, may be partly the cause of the differences in work functions. Moreover, from all analyses it is clear that the electrostatic effect is

strongly interrelated to the other effects. Accordingly, if the double layer changes from one crystallographic orientation to another, the other effects must change also. It appears that no general conclusion can be stated until surface exchange and correlation effects are analyzed both more accurately and by means of three-dimensional models (namely, accurate models which also account explicitly for the crystallographic orientation of the surface).

By virtue of the second law of thermodynamics, Koskinen[10] has shown that the difference in work functions of two surfaces of the same metal is due to dipole-moment effects only. The "proof" is not correct, however, because essential to the proof he presented is the assumption that a piece of metal enclosed by surfaces of different crystallographic orientations is homogeneous and this assumption is not valid. The total potentials of atoms on surfaces of different crystallographic orientations are different. The difference manifests itself either in terms of evaporation of atoms from a surface with a large total potential and condensation of atoms on a surface with a smaller total potential or in terms of migration of atoms between two such surfaces, or both.[11] These two processes may be fast or slow depending on the value of the temperature. At relatively low temperatures, for which evaporation is sufficiently slow, a piece of metal may last for a long time while bounded by surfaces of different orientations. The differences in total potentials of different surfaces are balanced by stresses set up in the metal. Thus the metal is heterogeneous, and its intensive properties are therefore not uniform. In fact, the gradients of intensive properties can be detected experimentally by measurements of two-dimensional thermoelectric effects.[12] Finally, in Sec. 9.3.7 it is shown that differences in work functions of two surfaces of the same metal depend on both the differences in total potentials of the atoms on these surfaces and the differences in dipole moments and exchange and correlation effects.

9.3.4 Electrochemical Potential μ_e^*. Values of the electrochemical potential μ_e^* for metals can be computed by several methods based on various one-electron approximations for the electronic structure of the metal. For such approximations, μ_e^* is equal to the value of the Fermi energy ϵ_F of the metal.

One method consists in using the Sommerfeld electron gas model, which is discussed in Sec. 7.8.10. In addition to μ_e^*, this model also yields values of μ_e as a function of temperature. The Sommerfeld gas model, however, neglects both exchange and correlation effects and also the periodic variation of the one-electron potential energy in the metal.

Another method consists in using a so-called *good one-electron calculation.* Approximate corrections for exchange and correlation effects are included in this calculation, and the periodic variation of the one-electron potential energy is taken into account.

Values of μ_e^* with respect to the minimum one-electron energy level of the ground state of monovalent and some multivalent metals, obtained by means of good one-electron calculations, are listed in the second column of Table 9.1. Values of μ_e^* computed by means of the Sommerfeld gas approximation for a number of free electrons per atom in the metal equal to one or to the number of valence electrons of the free atoms are given in the last two columns of the table.

Table 9.1 Values of Electrochemical Potential μ_e^* Computed by Means of Good One-Electron Calculations and Sommerfeld Gas Approximation

Metal	Good one-electron calculation (μ_e^*), ev	Valence ν_0	Sommerfeld gas approximation (μ_e^*), ev	
			Number of free electrons	
			1	ν_0
Lithium	4.05 (Ref. 13)	1	4.8	4.8
Sodium	3.23 (Ref. 13)	1	3.2	3.2
Rubidium	1.85 (Ref. 14)	1	1.85	1.85
Cesium	1.55 (Ref. 14)	1	1.65	1.65
Niobium	5.2 (Ref. 15)	5	5.4	15.8
Tantalum	7.5 (Ref. 15)	5	5.54	16.2
Tungsten	7.5 and 8.8 (Ref. 16)	6	5.95	19.6
Rhenium	9.8 (Ref. 17)	7	6.13	22.4

Comparing the results listed in Table 9.1 shows that the Sommerfeld model proves to be a good approximation for the electrons in monovalent metals such as sodium, rubidium, and cesium but is grossly inaccurate for lithium and for multivalent metals such as tungsten.

9.3.5 Wigner–Bardeen Analysis of Total Potentials. The general definition of the total potential of a component of an open system is given by the relation (Eq. 7.77)

$$\mu_l = \left(\frac{\partial E}{\partial n_l}\right)_{S,\beta,n} \tag{9.28}$$

In the limit of 0°K, the energy E of the system is solely a function of the numbers of particles of the various components.

Wigner and Bardeen[18] used Eq. 9.28 in the limit of 0°K to compute the total potentials of a two-component open system consisting of N_e electrons and N_i positive ion cores. The calculation is based on an estimate of the energy $E(N_e,N_i)$ with respect to a zero-energy reference level at infinity. The estimate is obtained by means of an extension of the method developed by Wigner and by

Seitz[19] for calculating cohesive energies of monovalent metals. Accordingly, the results of Wigner and Bardeen apply to monovalent metals only.

The calculated $E(N_e,N_i)$ is given by the relations

$$E(N_e,N_i) \approx N_e\left(E_0 + \frac{0.6e^2}{4\pi\varepsilon_0 r_s}\frac{N_e}{N_i}\right) + \epsilon_F \frac{3N_e^{5/3}}{5N_i^{5/3}} - 0.458\frac{e^2}{4\pi\varepsilon_0 r_s}\frac{N_e^{4/3}}{N_i^{1/3}}$$

(Electrostatic) (Kinetic) (Exchange)

$$-\frac{e^2}{4\pi\varepsilon_0} N_e C\left[\left(\frac{N_i}{N_e}\right)^{1/3} r_s\right] - V_D(N_e - N_i) \qquad (9.29)$$

(Correlation) (Double layer)

and

$$r_s = \left(\frac{3V}{4\pi N}\right)^{1/3} \qquad (9.30)$$

where E_0 = energy of the lowest level occupied by an electron in the field of the ion cores

V = volume of valence electrons of the metal

N = number of valence electrons of the metal

ϵ_F = Fermi energy of the Sommerfeld gas model for the electrons

$r_s\,C(r_s)$ = Wigner's *correlation energy function*

V_D = potential energy of the surface double layer when $N_e = N_i$

The quantity r_s is called the *radius of the Seitz sphere* or the *radius of the s sphere.* As indicated under each term in the righthand side of Eq. 9.29, these terms represent the electrostatic potential, kinetic, exchange, correlation, and double-layer energies. For numerical results, Eq. 9.29 requires, among other things, knowledge of the values of E_0 and V_D. This requirement is equivalent to the need to calculate the motive in the interfacial region.

The correlation function $r_s\,C(r_s)$ has been computed numerically by Wigner.[20] Different approximate analytical expressions for his numerical results have been suggested. Two of these expressions are given in the next section. As pointed out by Slater,[21] however, Wigner's arguments for the derivation of his correlation function are very involved and difficult to verify. For this reason, the results for the values $r_s\,C(r_s)$ should be used with caution.

When the metal is electrically neutral ($N_e = N_i$), the total potential or electrochemical potentials μ_e^* of the electrons and μ_i^* of the ion cores are given by the relations

$$\mu_e^* = \left.\frac{\partial E(N_e,N_i)}{\partial N_e}\right|_{N_e=N_i} \qquad (9.31)$$

and

$$\mu_i^* = \left.\frac{\partial E(N_e,N_i)}{\partial N_i}\right|_{N_e=N_i} \tag{9.32}$$

The explicit forms of Eqs. 9.31 and 9.32 and their use in calculating ϕ^* are discussed in the next section.

9.3.6 Calculating ϕ^*. By virtue of Eq. 9.23, if the energy E_0 in Eq. 9.29 is computed with respect to a zero-energy reference level at infinity, the work function ϕ^* is given by the relation

$$\phi^* = -\mu_e^* = -\left(E_0 + 1.2\frac{e^2}{4\pi\varepsilon_0 r_s}\right) - \epsilon_F + \frac{4}{3}0.458\frac{e^2}{4\pi\varepsilon_0 r_s} + \frac{e^2}{4\pi\varepsilon_0}\left[C(r_s) - \frac{r_s\,C'(r_s)}{3}\right] + V_D \tag{9.33}$$

where $C'(r_s)$ is the derivative of $C(r_s)$ with respect to r_s.

Equation 9.33 shows that the only term that depends on the structure of the surface is V_D. Therefore this equation suggests that differences in work functions of surfaces of the same element but of different crystallographic orientations are caused only by differences in the magnitudes of the double layers at these surfaces. As stated in Sec. 9.3.3, however, this conclusion is questionable (see also Sec. 9.3.7).

An approximate analytical expression for Wigner's correlation function $r_s\,C(r_s)$ is given by the relation

$$r_s\,C(r_s) = \frac{0.295r_s}{r_s + 5.1a_0} \tag{9.34}$$

where a_0 is the Bohr radius. Thus, if energies are measured in terms of Rydberg units, lengths are expressed in terms of the Bohr radius a_0, and Eq. 9.34 is substituted in Eq. 9.33, ϕ^* is given by the relation

$$\phi^* = -\left(E_0 + \frac{2.4}{r_s}\right) - \epsilon_F + \frac{1.22}{r_s} + \frac{0.78r_s + 3}{(r_s + 5.1)^2} + V_D \quad \text{Ry} \tag{9.35}$$

For the Sommerfeld gas model, the Fermi energy ϵ_F is given by the relation

$$\epsilon_F = \frac{3.68}{r_s^2} \quad \text{Ry} \tag{9.36}$$

For numerical results, Eq. 9.35 requires knowledge of the values of both E_0 and V_D. The energy E_0 can be expressed in terms of other properties of the metal as follows: According to the interpretation of total potentials in terms of work (Sec. 7.7.4), in the limit of 0°K the work required to extract an electron plus an ion from a system is equal to the negative sum of the total potentials of the electrons and the ions, i.e., $-\mu_e^* - \mu_i^*$. On the other hand, this work is also equal to the sum of the energy h_s^* required to extract an atom from the system at $T = 0$ plus the first ionization energy V_i of the atom. From the equality of these two quantities, it follows that

$$\phi^* = -\mu_e^* = h_s^* + V_i + \mu_i^*$$

$$= h_s^* + V_i - 0.4\epsilon_F - \frac{0.895}{r_s} + \frac{0.2r_s}{(r_s + 5.1)^2} + V_D \quad \text{Ry} \tag{9.37}$$

where use has been made of Eq. 9.32. We see from Eq. 9.37 that, in addition to V_D, the values of ϕ^* depend on h_s^* and V_i rather than E_0. The values of h_s^* and V_i are available from experiments.

Wigner and Bardeen[18] computed ϕ^* for the alkali metals by using Eq. 9.37 and experimental values for h_s^* and V_i, a zero value for V_D, and the Sommerfeld-gas-model Fermi energy (Eq. 9.36), except for lithium, for which ϵ_F is taken from the work of Seitz.[19] The results are in good agreement with experimental values determined from surfaces that more likely than not were either nonuniform or contaminated, or both. On the basis of this agreement between computed and experimental values of ϕ^*, Wigner and Bardeen concluded that the strength of the double layer of monovalent metal surfaces is probably very small. Although this may be correct, it is questionable whether such a conclusion can be reached by use of Eq. 9.37 since the accuracy of each term in the equation is not known.

Another equation for ϕ^* has been suggested by Bardeen.[22] Among other things, the equation involves an alternative analytical expression for the correlation terms in Eq. 9.33. It is given by the relation

$$\phi^* = -\langle V_{ps} \rangle - \frac{3.68}{r_s^2} + \frac{1.22}{r_s} + \frac{0.59r_s + 6.9}{(r_s + 7.9)^2} + V_D \quad \text{Ry} \tag{9.38}$$

where $\langle V_{ps} \rangle$ is an average pseudopotential in the bulk of the metal. For numerical results, Eq. 9.38 requires knowledge of the values of the quantity $V_D - \langle V_{ps} \rangle$.

A many-electron calculation of the work functions of a large number of metals, monovalent and multivalent, in the limit of 0°K has been performed by

Smith.[9] The results are informative but are not sufficiently accurate for the purposes of thermionic conversion.

Finally, an equation for the approximate calculation of ϕ as a function of T has been derived by Hatsopoulos and Gyftopoulos.[2 3] This equation, which is discussed in Sec. 9.3.7, provides the means for discussing the dependence of ϕ on a number of thermophysical properties of a metal and for the approximate calculation of the temperature coefficient α.

9.3.7 A Relation Between ϕ and Other Thermophysical Properties. Hatsopoulos and Gyftopoulos[2 3] derived an equation for ϕ of all metals as a function of T. The derivation is based on a combination of perfect-gas relations and results obtained from the Debye theory of lattice vibrations.

A finite piece of metal enclosed by a pure uniform surface and being in a state of mutual stable equilibrium with its own vapor at temperature T can be regarded as a two-component open system consisting of ν_0 valence electrons per atom and of positive ion cores each with ν_0 positive units of charge. Moreover, the vapor can be regarded as a heterogeneous mixture of electrons and of positive ions and neutrals of the free atoms of the metal.

At temperatures of practical interest, each component of the vapor mixture behaves like a perfect gas, and, to an excellent approximation, the densities of positive ions other than singly charged ones are negligibly small.

As discussed in Appendix 9.A, in the context of the one-particle picture for both electrons and ions, two motives are necessary for analysis of the stable equilibrium states of the metal–vapor system. One motive applies to electrons, and the other applies to singly charged positive ions. For a surface that is a copious emitter of electrons rather than ions, representative graphs of the two motives vs. x are superimposed in Fig. 9.10. In these graphs each atom in the metal is regarded as consisting of one valence electron and a singly charged positive ion core which are energetically equivalent to an atom with ν_0 valence electrons and an ion core with ν_0 positive units of charge. The equivalence is defined so that the atom total potential μ_a has the same value for either type of atom in the metal.

In Fig. 9.10 the zero-energy reference level is taken at a very large distance from the surface. At such a distance any differential phase of the vapor is electrically neutral. With respect to this reference, the figure includes the relative magnitudes of the total potentials μ_e and μ_i for electrons and singly charged ions, respectively; the maximum ion motive ψ_m; and the minimum one-particle energies $-\epsilon_{e0}$ and $-\epsilon_{i0}$ for electrons and singly charged positive ions in the metal, respectively. Moreover, the figure includes various quantities measured from different references, such as the total potentials or partial Gibbs free energies ζ^s_{e0} for electrons and ζ^s_{i0} for singly charged ions with respect to $-\epsilon_{e0}$ and $-\epsilon_{i0}$, respectively; the motive changes in the interfacial region $-(V_D + 2x_e)$

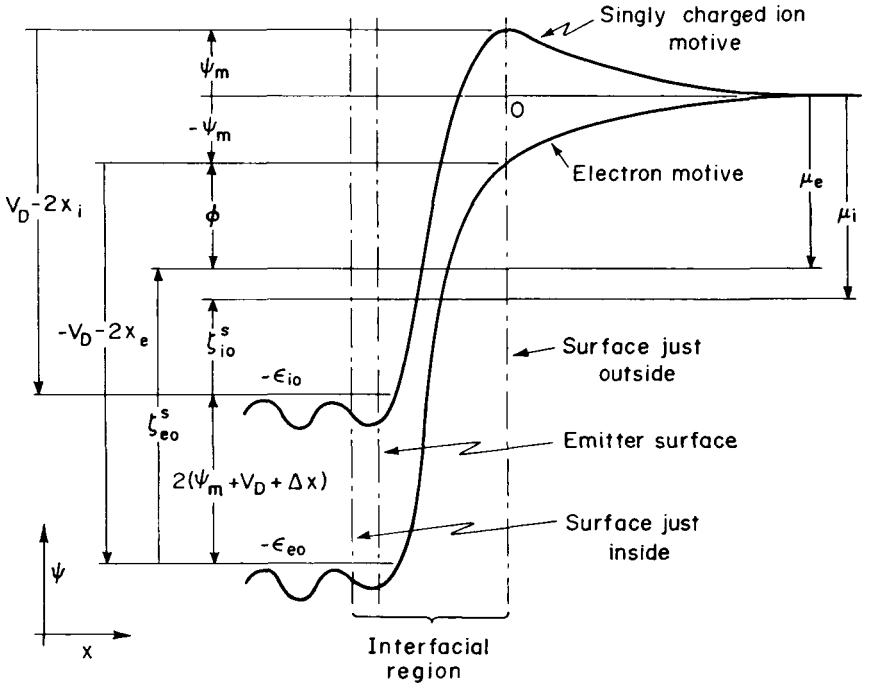


Fig. 9.10 Representative singly charged ion and electron motives vs. x for an emitter in mutual stable equilibrium with its own vapor. $\Delta x = x_e - x_i$, where $2x_i$ and $2x_e$ are exchange-plus-correlation energies for ions and electrons, respectively.

for electrons and $V_D - 2x_i$ for ions from $-\psi_m$ to $-\epsilon_{e0}$ and from ψ_m to $-\epsilon_{i0}$, respectively; and the work function ϕ.

When both the electron and the singly charged ion motives are considered, the surface just outside is determined by the points at which one of the two motives is maximum. In Fig. 9.10, the ion motive exhibits a maximum whose value is denoted by ψ_m. Moreover, since the two motives beyond the points just outside are equal and opposite, it follows that the work function ϕ is given by the relation

$$\phi = -\mu_e - \psi_m \qquad \text{for } T > 0 \tag{9.39}$$

In the limit of 0°K, ψ_m is equal to zero, and therefore ϕ is equal to $-\mu_e$.

In the interfacial region the changes of the motives from $-\psi_m$ to $-\epsilon_{e0}$ and from ψ_m to $-\epsilon_{i0}$ consist of electrostatic, exchange, and correlation energy contributions. The electrostatic or double-layer energy $-V_D$ for electrons is

equal and opposite to that for ions. Thus, if the exchange-plus-correlation energies for electrons and ions are denoted by $-2x_e$ and $-2x_i$, respectively, the changes of the motives are as shown in Fig. 9.10.

The difference $2\Delta x$, defined by the relation

$$2\Delta x = 2(x_e - x_i) \tag{9.40}$$

is probably small because, in most of the interfacial region near the points just outside, exchange-plus-correlation effects are approximately described by the image-force law, which is independent of the sign of the charge and therefore in this region exchange-plus-correlation energies for electrons are identical to those for ions.

We see from Fig. 9.10 that the difference $\epsilon_{e0} - \epsilon_{i0}$ and $-\psi_m$ satisfies the relations

$$\epsilon_{e0} - \epsilon_{i0} = 2(\psi_m + V_D + \Delta x) = \zeta^s_{e0} - \zeta^s_{i0} - (\mu_e - \mu_i) \tag{9.41}$$

and

$$-\psi_m = \frac{1}{2}(\mu_e - \mu_i) - \frac{1}{2}(\zeta^s_{e0} - \zeta^s_{i0}) + V_D + \Delta x \tag{9.42}$$

Moreover, substituting Eq. 9.42 into Eq. 9.39 yields

$$\phi = -\mu_e + \frac{1}{2}(\mu_e - \mu_i) - \frac{1}{2}(\zeta^s_{e0} - \zeta^s_{i0}) + V_D + \Delta x$$

$$= -\frac{1}{2}\mu_a - \frac{1}{2}(\zeta^s_{e0} - \zeta^s_{i0}) + V_D + \Delta x \tag{9.43}$$

where μ_a, the total potential of neutral atoms either in the metal or in the vapor, satisfies the equation of chemical equilibrium $\mu_a = \mu_e + \mu_i$.

The values of μ_a can be expressed in terms of properties of the vapor by means of the perfect-gas relations applied to an electrically neutral differential phase of the vapor (Sec. 7.10.1). The values of ζ^s_{e0} can be expressed in terms of quantities obtained from a good one-electron calculation of the ground state of the metal and in terms of the volume expansion coefficient of the metal and the temperature coefficient of the electrochemical potential of the Sommerfeld gas model. Finally, the values of ζ^s_{i0} can be computed by means of the Debye theory of lattice vibrations of the metal (Sec. 7.12). Thus it is found that ϕ is given by the relations[2 3]

$$\phi = \frac{1}{2}\left\{V_i + kT \ln \frac{(kT)^{3/2}}{p} \frac{g_a (2\pi m_a)^{3/2}}{g_i^s h^3} - \zeta_{e0}^s - 3kT\left[\ln \frac{T}{\theta_D} + \frac{1}{3} + \frac{3\theta_D}{8T}\right]\right\}$$

$$+ V_D + \Delta x$$

$$= \frac{1}{2}\left\{V_i + kT \ln \frac{(kT)^2}{\Gamma_a} \frac{2\pi m_a g_a}{g_i^s h^3} - \zeta_{e0}^s - 3kT\left[\ln \frac{T}{\theta_D} + \frac{1}{3} + \frac{3\theta_D}{8T}\right]\right\}$$

$$+ V_D + \Delta x \quad \text{for } T > \theta_D \tag{9.44}$$

where V_i = first ionization energy of the free atom of the metal
g_a = degeneracy of the free atom of the metal
m_a = mass of the free atom of the metal
p = vapor pressure at the temperature under consideration
g_i^s = degeneracy of singly charged ion cores in the metal
θ_D = Debye temperature of the metal
Γ_a = atom-random-flux density at the surface

ζ_{e0}^s is given by the relations

$$\zeta_{e0}^s = \Delta\epsilon - \frac{2}{3}\Delta\epsilon \int_0^T \beta \, dT - \frac{(\pi kT)^2}{12\,\Delta\epsilon} = \Delta\epsilon - \frac{2}{3}\bar{\beta}\,\Delta\epsilon\, T - \frac{(\pi kT)^2}{12\,\Delta\epsilon} \tag{9.45a}$$

$$\Delta\epsilon = \nu_0 \, [\epsilon_F - \epsilon(\nu_0 - 1)] \tag{9.45b}$$

where ϵ_F = Fermi energy
β = volume expansion coefficient
$\bar{\beta}$ = average volume expansion coefficient in range from zero to T
$\epsilon(\nu_0 - 1)$ = maximum one-electron energy of the ground state of the metal that corresponds to $\nu_0 - 1$ valence electrons per atom

The definitions of ϵ_F and $\epsilon(\nu_0 - 1)$ are illustrated in Fig. 9.11, which represents a good one-electron calculation for the ground state of tungsten. Equation 9.45a is based on the assumption that changes of ζ_{e0}^s with volume and temperature are described approximately by the Sommerfeld gas model (Sec. 7.8.10) although the value $\Delta\epsilon$ of ζ_{e0}^s for $T = 0$ is computed by means of a good one-electron model.

Except for V_D and Δx, the values of the quantities in the right-hand side of Eq. 9.44 are known provided that a good one-electron calculation has been performed for the metal in question, including an accurate estimate of the density of states.

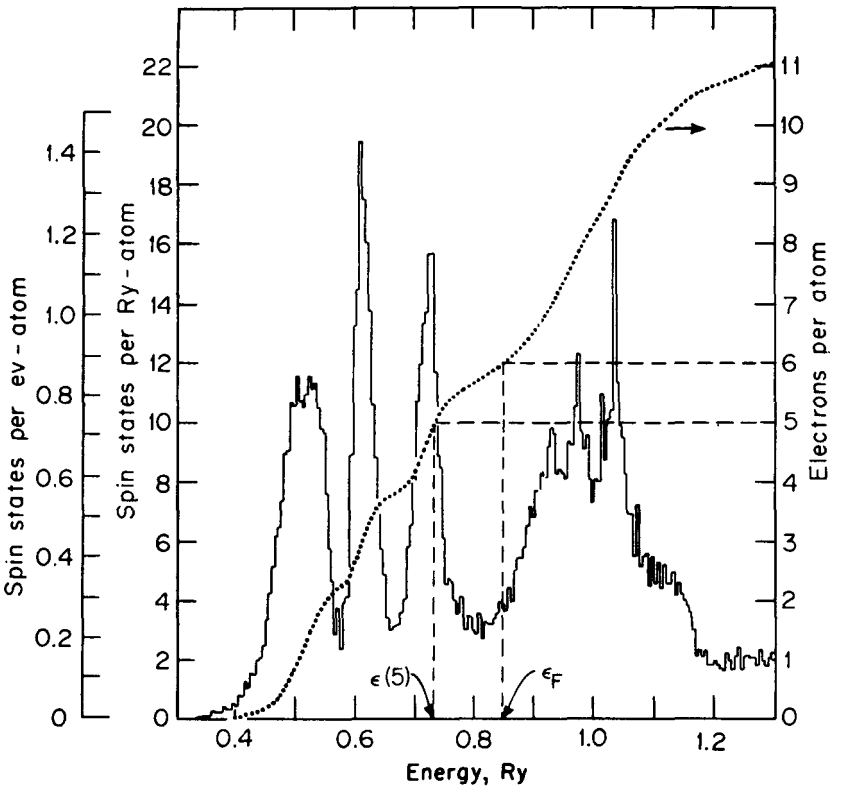


Fig. 9.11 One-electron calculation of ϵ_F, $\epsilon(5)$, and density of states of the ground state of tungsten. Number of valence electrons for tungsten is $\nu_0 = 6$. [Adapted from I. F. Mattheis, Fermi Surface in Tungsten, *Physical Review*, 139(6A): 1893-1904 (1965)].

We see from Eq. 9.44 that three terms in the right-hand side depend on the structure of the surface, i.e., the pressure p† (random flux Γ_α), the energy V_D of the double layer, and the difference Δx defined by the exchange-plus-correlation energies for electrons and ions (Eq. 9.40). Although detailed experimental data are not available, the dependence of p on the structure of the surface can be inferred from the different degrees of stability of surfaces of different crystallographic orientations.[11] Equivalently, the dependence of p on the structure of the surface may be deduced from the difference between the surface free energies of two surfaces of different structures. Thus it follows from

† Alternatively, this dependence can be expressed in terms of the total potential μ_a.

Eq. 9.44 that the change of the work function of a metal from one surface to another is not caused solely by a change of the energy of the double layer.

Another interesting feature of Eq. 9.44 is its linear dependence on $\Delta\epsilon$. Since different models may yield appreciably different values for $\Delta\epsilon$, the linear dependence of ϕ on $\Delta\epsilon$ illustrates the need for a good one-electron calculation of the electronic structure.

Values of the quantities V_i, θ_D, and $\bar{\beta}$ and average p vs. T are known for all the elements.[24] A few graphs of average p vs. T are shown in Fig. 9.12. For a metal for which a good one-electron calculation has been performed, including an accurate estimate of the density of states, the average pressure can be used to calculate a typical value of $\phi - V_D - \Delta x$ by means of Eq. 9.44.

Computed values of $\phi - V_D - \Delta x$ and ranges of experimental values of ϕ for Li, Na, K, Rb, Cs, W, and Re are listed in Table 9.2. The values of the data V_i, m_a, g_a, g_i^s, ν_0, θ_D, and $\bar{\beta}$ used in the calculations are listed in Table 9.3. The

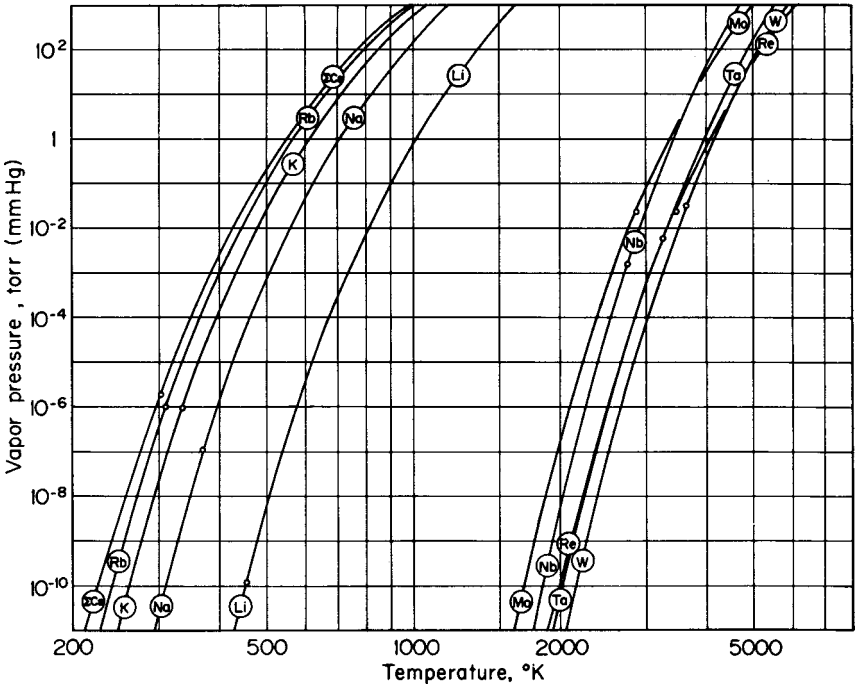


Fig. 9.12 Graphs of average vapor pressure p vs. temperature T for some elements. (Adapted from R. E. Honig, Vapor-Pressure Curves of the Elements, Radio Corporation of America, 1962.)

Table 9.2 Computed Values of $\phi - V_D - \Delta x$ for Metals for Which a Good One-Electron Calculation Is Available

Element	T, °K	p (Fig. 9.12), torr	$\Delta\epsilon$, ev	$\phi - V_D - \Delta x$ (Eq. 9.44), ev	Range of experimental ϕ (Ref. 26), ev
Lithium	1000	7×10^{1}	4.05 (Ref. 13)	1.7	2.3 to 2.5
Sodium	500	7×10^{-4}	3.23 (Ref. 13)	1.6	2.2 to 2.35
Potassium	360	7×10^{-6}	2.05†	1.7	2.15 to 2.25
Rubidium	360	6.5×10^{-5}	1.85 (Ref. 14)	1.7	2.1 to 2.2
Cesium	360	2.3×10^{-4}	1.55 (Ref. 14)	1.65	1.8 to 1.9
Tungsten	2100	3×10^{-11}	8.1 (Ref. 16)	4.3	4.3 to 5.4
Rhenium	2100	1×10^{-9}	5.1 (Ref. 17)	5.5	5.1 to 5.6

†$\Delta\epsilon$ taken equal to the Fermi energy of the Sommerfeld gas.

Table 9.3 Data[25] Used in the Calculation Summarized in Table 9.2

Element	V_i, ev	m_a, amu	g_a	g_i^s	ν_0	θ_D, °K	$\bar{\beta}$, 10^{-6}/°K
Lithium	5.39	6.94	2	1	1	370	140 (300)†
Sodium	5.14	22.99	2	1	1	158	210 (300)
Potassium	4.34	39.1	2	1	1	90	250 (300)
Rubidium	4.18	85.47	2	1	1	52	270 (250)
Cesium	3.89	132.9	2	1	1	44	290 (285)
Niobium	6.88	92.9	2	1	5	230	25.8 (1000)
Rhenium	7.88	186.2	2	1	7	430	20.7 (1000)

†Temperatures are given in parentheses.

value of β is taken equal to the tabulated value of β at the temperature indicated in parentheses in the last column of Table 9.3. Comparing the computed values of $\phi - V_D - \Delta x$ with experimental values of ϕ (last two columns in Table 9.2) suggests that the value of $V_D + \Delta x$ is of the order of a fraction of one electron volt.

In conclusion, the theories summarized up to this point are informative but do not provide the means for estimating the effect of the surface structure on either ϕ^* or ϕ. A calculation of the effect of the surface structure on ϕ^* made by Smoluchowski[27] is discussed in the next section.

9.3.8 Effect of Surface Structure on ϕ^*. An approximate quantum-mechanical analysis of the effect of surface structure on the value of the work function ϕ^*

was made by Smoluchowski.[27] Although this analysis disregards exchange and correlation effects, it provides qualitatively some insight into the processes that contribute to the differences between the work functions of surfaces of different crystallographic orientations.

Smoluchowski represented a metal by a model consisting of a number of polyhedral unit cells. These cells are arranged in the form of a crystal as shown schematically in Fig. 9.13a. Each polyhedron is filled with equal amounts of positive and negative (electron) charges. In the bulk of the model, both types of charges are uniformly distributed.

When the crystal is cleaved along a surface of given crystallographic orientation, the electron charge distribution in the surface polyhedra created by cleaving is no longer uniform. According to Smoluchowski, two factors contribute to this nonuniformity. First, because part of the crystal has been removed, the forces acting on the electrons near the surface are weaker than those in the bulk of the crystal. Thus the electrons spread beyond the boundaries of the polyhedra at the surface. The spreading of the electronic charge is shown schematically in Fig. 9.13b. Because of the spreading, the polyhedra at the surface appear as positively charged, and a double layer is formed on the surface. The effect of this double layer is an increase of the work function with respect to its value without spreading. The increase is called the *spreading effect.*

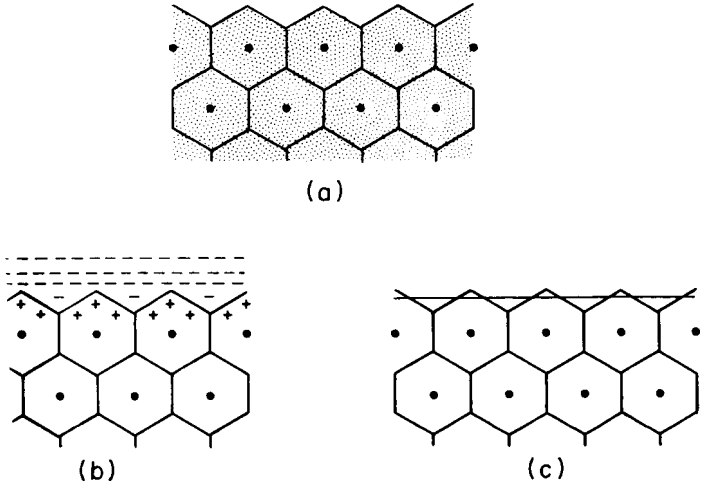


Fig. 9.13 Formation of a double layer at a surface. (a) Unperturbed distribution. (b) Redistribution of charges due to spreading. (c) Redistribution of charges due to smoothing.

The other factor that contributes to the electron charge nonuniformity at the surface is smoothing, which is explained as follows: The energy of electrons enclosed in a volume bounded by a large flat plane is lower than the energy of electrons surrounded by the geometrically intricate faces of the surface polyhedra. Because of the energy difference when a crystal is cleaved, electrons flow from the hills into the valleys formed by the surface polyhedral cells. Thus a net positive charge appears on the hills and a net negative charge appears in the valleys, as shown schematically in Fig. 9.13c. This charge redistribution results in another double layer on the surface. The effect of this double layer is a decrease of the work function. The decrease is called the *smoothing effect.*

The spreading and smoothing effects are interrelated and depend on the geometric arrangement of the surface polyhedra. Smoluchowski estimated the magnitudes of both effects and found that the spreading effect is almost independent of the surface structure whereas the smoothing effect shows considerable variation from one crystal to another of the same material and is the more complete the shorter the spacing between the peaks and valleys of the geometric surface determined by the faces of the polyhedra on the surface of the model. As a result of these findings, Smoluchowski concluded that the work function of a crystal face is higher the more densely the surface is packed with polyhedra, because then a given amount of smoothing results in a smaller decrease of the work function than it would have if the surface were more open.

The ideas of spreading and smoothing of the electron charge at a surface are intuitively understandable and reasonable. Their interpretation solely in terms of double layers, however, may not be realistic because of the close interrelation between the values of electrostatic, exchange, and correlation effects in the interfacial region.

9.3.9 An Empirical Correlation for ϕ^*. An empirical correlation for work functions ϕ^* of pure uniform surfaces of various metals is given by the expression[28]

$$\phi^* = 0.98 \frac{\nu_s + 1}{r_m} + 1.57 \quad \text{ev} \tag{9.46}$$

where ν_s is the *surface valence* and r_m is the atomic radius in angstroms. The explicit definition of ν_s is given later.

Many empirical correlations[29-33] have been suggested which relate the work function of metals to known experimental properties either of the metal itself or of the free atoms of the metal. For example, work functions have been correlated with compressibility, energy of the lattice, surface energy, atomic volume, first atomic ionization energy, etc. With the exception of Eq. 9.46,

however, none of these correlations depends on the structure of the surface, and therefore none is helpful in thermionic conversion.

The justification for Eq. 9.46 involves the following steps:

1. The metal is interpreted approximately as a collection of perturbed atoms of its element; this interpretation is analogous to the valence-bond picture used for molecules.

2. The negative of the electrochemical potential μ_e with respect to the points just outside or the work function ϕ is identified with the negative of the electrochemical potential of a surface "atom" or the electronegativity of this "atom."

3. An empirical correlation of molecular chemistry is used to express the electronegativity in terms of other properties of the surface "atom."

In the context of the one-electron approximation, each energy eigenfunction of a metal is represented by a Slater determinant of one-electron spin-orbitals. The spin-orbitals are determined by an approximate one-electron eigenvalue problem. For example, when correlation effects are neglected, the "best" one-electron spin-orbitals are determined by the Hartree–Fock one-electron equation. Because of numerical difficulties, many investigators have suggested alternative ways for expressing the Slater determinant in terms of spin-orbitals. One of these ways consists in using spin-orbitals that are localized around each nucleus of the metal. In this approximation each nucleus is associated with a number of monocentric spin-orbitals equal to the number of electrons per atom of the metal. The monocentric spin-orbitals are determined both by the field of the nucleus to which they belong and by interactions with all other charges of the metal. Thus each nucleus and its localized spin-orbitals can be regarded as a perturbed atom or simply an "atom."

Many mathematical methods have been suggested to define the localized spin-orbitals, namely, the one-electron equation satisfied by the localized spin-orbitals.[34,35] These methods will not be discussed. For present purposes it suffices to indicate that the energies of the localized spin-orbitals are not spectroscopically observable and cannot be interpreted as ionization energies of the metal. Nevertheless, the metal can be regarded as a collection of individual "atoms" which is thermodynamically equivalent to the whole metal. Moreover, since each of these "atoms" is defined by its interactions with all the other "atoms," it follows that those on the surface will be different from those in the bulk.

When the metal is in a stable equilibrium state, each "atom" can be regarded as an electron open system in mutual stable equilibrium with the other "atoms." Accordingly, the electrochemical potential of each "atom" must have the same value as the electrochemical potential μ_e of the metal.

For a pure metal enclosed by a pure uniform surface, the work function ϕ is equal to the negative of μ_e with respect to the points just outside, and therefore ϕ must be equal to the negative of the electrochemical potential of an "atom" on the surface with respect to the points just outside.

In Ref. 36 the negative of the electrochemical potential of an atom is identified with the electronegativity of the atom. It follows that the work function is equal to the electronegativity of an "atom" on the surface. Moreover, if the electronegativity of an "atom" on the surface in the limit of 0°K is denoted by $x_s(0)$, it follows that ϕ^* is given by the relation

$$\phi^* \equiv x_s(0) \tag{9.47}$$

In the limit of 0°K, Steiner and Gyftopoulos[28] proposed that the electronegativity $x_s(0)$ can be computed by a direct extension to "atoms" on a surface of a correlation that has been derived for atoms in a molecule. To this end, they observed that, in the works of Gordy[37] and of Pritchard and Skinner,[38] the electronegativity $x(0)$ of a neutral atom in a molecule is approximated by the relation

$$x(0) = 0.98\frac{\nu + 1}{r} + 1.57 \quad \text{ev} \tag{9.48}$$

where ν is the number of electrons per atom which participate in bonding and r is the effective size in angstroms of the atom in the bonded state. Then they assumed that Eq. 9.48 is also applicable to "atoms" on a surface of a metal and consequently that

$$\phi^* = x_s(0) = 0.98\frac{\nu_s + 1}{r_m} + 1.57 \quad \text{ev} \tag{9.49}$$

where ν_s is the number of electrons per "atom" on a surface which participate in bonding and r_m is the *effective size* of the "atom" in angstroms. The quantity ν_s will be called the *surface valence.*

The effective size of "atoms" on the surface is assumed to be equal to the atomic radius given by Slater.[39] The surface valence ν_s is derived from the metallic valence ν_m and fractional bond numbers of atoms in the bulk as follows:

The metallic valence ν_m is defined as the number of electrons per "atom" which participate in bonding in the bulk of a metal. The fractional bond number n is defined so that $2n$ is equal to the average number of bonding electrons shared by two interacting "atoms" at an interatomic separation R. According to

Pauling,[40] the fractional bond numbers n_1 and n_2 associated with the interatomic separations R_1 and R_2, respectively, satisfy the empirical relation:

$$R_2 - R_1 = 0.26 \ln \frac{n_1}{n_2} \tag{9.50}$$

For body-centered cubic (bcc) and hexagonal close-packed (hcp) crystal structures, fractional bond numbers between atoms farther distant than next-nearest neighbors are negligible. For face-centered cubic (fcc) crystal structures, fractional bond numbers between atoms farther distant than nearest neighbors are negligible. Thus for each crystal structure metallic valence and fractional bond numbers satisfy the relations

$$\nu_m = 8n_a + 6n_b \quad \text{for bcc} \tag{9.51a}$$

$$\nu_m = 6n_a + 6n_b \quad \text{for hcp} \tag{9.51b}$$

$$\nu_m = 12n_a \quad \text{for fcc} \tag{9.51c}$$

where n_a and n_b are the fractional bond numbers for nearest and next-nearest neighbors, respectively. Tabulated values of metallic valences and interatomic distances can be used in Eqs. 9.50 and 9.51 to compute the fractional bond numbers n_a and n_b for different metals.

Steiner and Gyftopoulos assumed that the fractional bond numbers of "atoms" on a surface are identical to those of atoms in the bulk. Thus the surface valence ν_s is given by the relations

$$\nu_s = N_a n_a + N_b n_b \quad \text{for bcc} \tag{9.52a}$$

$$\nu_s = N_a n_a + N_b n_b \quad \text{for hcp} \tag{9.52b}$$

$$\nu_s = N_a n_a \quad \text{for fcc} \tag{9.52c}$$

where N_a and N_b are the numbers of nearest and next-nearest neighbors of "atoms" on the surface, respectively. The numbers N_a and N_b are fixed by the crystallographic orientation of the surface. Values of N_a and N_b for planes of various crystallographic orientations of the bcc, hcp, and fcc structures are listed in Table 9.4.

Equation 9.49 has been used to compute work functions ϕ^* for a large number of pure uniform surfaces. The values of metallic valences, atomic radii, and interatomic distances used in the calculations are listed in Table 9.5. The results for tungsten, molybdenum, tantalum, and rhenium are listed in Tables 9.6 to 9.9, respectively. The tables also contain experimental values derived from

Table 9.4 Ligands of Surface Atoms

Plane	Number of neighbors: Nearest (N_a)	Next nearest (N_b)
	bcc	
1 1 0	6	4
1 1 2	5	3
1 0 0	4	5
1 1 1	4	3
1 1 6	4	3
	fcc	
1 1 1	9	
1 0 0	8	
1 1 0	7	
	hcp ($c/a < 1.633$)	
0 0 0 1	3	6
$1\,0\,\bar{1}\,0$	4	4

thermionic-emission data. We see from these tables that the predicted work functions are in good agreement with the experimental data.

Work functions of pure uniform surfaces of most metals have not yet been measured. Predicted values for a number of metals of interest are listed in Table 9.10.

Qualitatively, the empirical correlation represented by Eq. 9.49 suggests that the work function ϕ^* of a pure uniform surface is determined by three factors: (1) the number of ligands possessed by an "atom" on the surface; (2) the interatomic distances between that "atom" and its ligands; and (3) the metallic valence. As indicated by the results, a surface exhibits a high work function when the number of ligands and the metallic valence are large and the interatomic distances are small.

9.3.10 Temperature Coefficient α. As stated in Sec. 9.3.2, the work function ϕ depends on T because both the electrochemical potential μ_e and the motive of the points just outside are functions of T. A number of attempts have been made to compute the temperature coefficient α of ϕ (the derivative $d\phi/dT$) of pure uniform surfaces. Each calculation is based on approximate models for the electronic structure and the state of the positive nuclei of the metal. The results indicate that α is small and of the order of the Boltzmann constant of k. They

Table 9.5 Properties of Atoms in the Bulk

Element	Metallic valence[40] (ν_m)	Atomic radius[39] (r_m), Å	Interatomic distances to neighbors,[40] Å	
			Nearest (R_a)	Next nearest (R_b)
		bcc		
Tungsten	6	1.35	2.74	3.16
Molybdenum	6	1.45	2.72	3.15
Tantalum	5	1.45	2.86	3.31
Vanadium	5	1.35	2.62	3.02
Lithium	1	1.45	3.04	3.50
		fcc		
Iridium	6	1.35	2.71	3.83
Platinum	6	1.35	2.78	3.93
Nickel	6	1.35	2.49	3.52
Gold	5.56	1.35	2.88	4.07
Copper	5.56	1.35	2.56	3.62
Thorium	4	1.80	3.60	5.10
		hcp		
Rhenium	6	1.35	2.74	2.76
Osmium	6	1.30	2.68	2.74
Zirconium	4	1.55	3.18	3.23
Titanium	4	1.40	2.90	2.95

Table 9.6 Work Functions of Tungsten

Plane	Work function ϕ^*, ev			
	Theory	Experiment		
		Ref. 41	Ref. 42	Ref. 43
1 1 0	5.50	5.29	5.35	5.40
1 1 2	4.94	4.65	4.80	4.80
1 0 0	4.66	4.52	4.60	4.55
1 1 1	4.47	4.38	4.40	4.42
1 1 6	4.47	4.29	4.32	

Table 9.7 Work Functions of Molybdenum

	Work function ϕ^*, ev			
		Experiment		
Plane	Theory	Ref. 42	Ref. 44	Ref. 45
1 1 0	5.23	5.00	5.10	4.90
1 1 2	4.71	4.55		
1 0 0	4.44	4.40	4.40	4.35
1 1 1	4.27	4.10	4.15	4.18
1 1 6	4.27	4.00		

Table 9.8 Work Functions of Tantalum

	Work function ϕ^*, ev	
Plane	Theory	Experiment[42]
1 1 0	4.75	4.80
1 1 2	4.31	4.3 to 4.4
1 1 0	4.08	4.15
1 1 1	3.94	4.00
1 1 6	3.94	3.90

Table 9.9 Work Functions of Rhenium

	Work function ϕ^*, ev	
Plane	Theory	Experiment[46]
0 0 0 1	5.53	5.59
$1\,0\,\bar{1}\,0$	5.20	5.15

are obtained, however, as differences between relatively large contributions from a number of physical processes, none of which can be evaluated accurately. For this reason, at the present time neither the sign nor the magnitude of α can be stated with certainty. In addition, because of experimental difficulties, measured values of α are also uncertain. For example, reported experimental values of α for tungsten surfaces of various crystallographic orientations are listed in Table 9.11. We see from this table that, say, the values of α for the (1 1 0) surface range from -17×10^{-5} to 6.3×10^{-5} ev/°K or $-2k$ to $0.7k$.

Table 9.10 Predicted Work Functions ϕ^* (ev) for Various Metal Surfaces

Element	Plane 1 1 0	1 1 2	1 0 0	1 1 1	0 0 0 1	1 0 $\bar{1}$ 0
			bcc			
Vanadium	4.97	4.50	4.28	4.11		
Lithium	2.78	2.66	2.61	2.58		
			fcc			
Iridium	4.84		5.20	5.56		
Platinum	4.84		5.20	5.56		
Nickel	4.84		5.20	5.56†		
Gold	4.65		4.99	5.32		
Copper	4.65		4.99	5.32		
Thorium	3.38		3.57	3.75		
			hcp			
Osmium					5.59	5.34
Zirconium					4.04	3.89
Titanium					4.29	4.13

†An experimental value of 5.42 ev has been reported[4,7] for the (1 1 1) surface of nickel.

Since an excellent discussion of various calculations of α is given by Herring and Nichols,[1] it is not repeated here. Instead, an estimate of α will be obtained by means of Eqs. 9.43 to 9.45.

If a piece of metal enclosed by a pure uniform surface is in a state in mutual stable equilibrium with the vapor of the metal at temperature T, then the work function ϕ depends on T both explicitly and implicitly (Sec. 9.3.2). The explicit dependence arises because both the electron total potential and the electron motive at the points just outside are explicit functions of T (Eqs. 9.25 to 9.27). The implicit dependence arises because the energy eigenvalues of the metal are functions of the values of the constraints and these, in turn, are determined by T since the metal expands with increasing temperature.

By virtue of these remarks, ϕ can be regarded as a function $\phi(T, V)$, where the volume V of the metal is a function of temperature. Alternatively, ϕ can be regarded as a function $\phi(T,p)$, where the vapor pressure p is a function of T. As the temperature varies, the vapor pressure varies along the $p(T)$ function or the *vapor-pressure saturation line* for the surface type in question. Typical average p

Table 9.11 Published Temperature Coefficients $d\phi/dT$ of the Work Function of Tungsten[48]

Investigator†	Date	Surface	Method	Pressure conditions, torr	Temperature range, °K	$d\phi/dT \times 10^{-5}$, ev/°K
Langmuir	1936	Polycrystalline wire	Retarding field	4×10^{-4}	300 to 1000	−5.5
Reimann	1937	Polycrystalline wire	Displacement of characteristic	No details	300 to 1100	7.5
Potter	1940	Polycrystalline wire	Vibrating capacitor	7×10^{-9}	300 to 950	5.2
Kruger and Stabenow	1935	Polycrystalline wire	Cooling effect	10^{-6}	2100 to 2700	6.0
Hutson	1955	Single crystal wire	Thermionic velocity distribution	10^{-9}	700 to 2000	Inconclusive
Van Oostrom	1963	(1 1 1)	Field emission	10^{-10}	78 and 293	0.6
		(1 1 6)	Field emission	10^{-10}	78 and 293	5.7
		(1 1 0)		10^{-10}		−17.0
		(1 1 2)		10^{-10}		−14.3
Swanson et al.	1967	(1 0 0)	Velocity distribution of field-emitted electrons	10^{-10}	78 to 700	−10.9
		(1 1 1)		10^{-10}		3.5
		(1 1 6)		10^{-10}		5.0
		(3 1 0)		10^{-10}		−3.2
Fine, Madey, and Scheer	1965	(1 1 0)	Retarding-field surface ionization	10^{-10}	1800 to 2200	−16
		(1 1 0)		10^{-8}		−13
Yates and Madey	1967	(1 1 0) and (1 0 0) Polycrystalline foil	Retarding field	10^{-10}	300 to 1300	$<\|2\|$
				10^{-10}	300 to 1300	11 ± 1
Ref. 48	1969	(1 1 0)	Vibrating capacitor	10^{-10}	78 to 1300	6.3 ± 0.6
		(1 0 0)		10^{-10}	78 to 1300	9.16 ± 0.4

†See Ref. 48 for complete data on sources of information.

vs. T functions for some elements are shown in Fig. 9.12. Thus, in general, the temperature coefficient α is given by the relation

$$\alpha = \left(\frac{\partial \phi}{\partial T}\right)_n$$

$$= \left(\frac{\partial \phi}{\partial T}\right)_{V,n} + \left(\frac{\partial \phi}{\partial V}\right)_{T,n} \frac{dV}{dT}$$

$$= \left(\frac{\partial \phi}{\partial T}\right)_{p,n} + \left(\frac{\partial \phi}{\partial p}\right)_{T,n} \frac{dp}{dT} \tag{9.53}$$

depending on whether ϕ is regarded as a function either of T and V or of T and p. The derivative dV/dT is related to the expansion coefficient of the metal, and dp/dT is computed along the vapor-pressure saturation line. The subscript n denotes that derivatives are computed while the metal is kept electrically neutral, as it should be.

In the limit of 0°K, ϕ is equal to $-\mu_e$, and therefore α is given by the relation

$$\alpha = -\left(\frac{\partial \mu_e}{\partial T}\right)_n = -\left(\frac{\partial \mu_e}{\partial T}\right)_{p,n} - \left(\frac{\partial \mu_e}{\partial p}\right)_{T,n} \frac{dp}{dT}$$

$$= s_e^g - v_e^g \frac{dp}{dT} \quad \text{for } T \to 0 \tag{9.54}$$

where

$$s_e^g = -\left(\frac{\partial \mu_e}{\partial T}\right)_{p,n} \quad \text{and} \quad v_e^g = -\left(\frac{\partial \mu_e}{\partial p}\right)_{T,n}$$

The quantities s_e^g and v_e^g are the partial entropy and the partial volume, respectively, of the electrons of the electrically neutral vapor phase (Eqs. 7.147 and 7.148).

Since in the limit $T = 0$ the state of the metal and its vapor is described by the ground state only (i.e., a pure state), it follows that

$$s_e^g = 0 \quad \text{for } T = 0 \tag{9.55}$$

Moreover, it can be shown that

$$\frac{dp}{dT} = 0 \quad \text{for } T = 0 \tag{9.56}$$

Combining Eqs. 9.54 to 9.56 yields that

$$\alpha = 0 \quad \text{for } T = 0 \tag{9.57}$$

At finite temperatures the temperature coefficient α can be expressed in terms of a number of partial entropies and partial volumes of various components by substitution of the second part of Eq. 9.43 into the last part of Eq. 9.53. The result is not useful, however, because the values of several terms appearing in it cannot be computed individually.

At temperatures greater than the Debye temperature of the metal, an estimate of α can be obtained by means of the first part of Eq. 9.44:

$$\alpha = \frac{1}{2}k\left[\ln\frac{(kT)^{5/2}}{p}\frac{g_a(2\pi m_a)^{3/2}}{g_i^s h^3} - \frac{2\bar{\beta}\Delta\epsilon}{3k} - \frac{\pi^2 kT}{6\Delta\epsilon} + \frac{5}{2} - \frac{d\ln p}{d\ln T}\right]$$

$$+\frac{d}{dT}(V_D + \Delta x)$$

$$\approx \frac{1}{2}k\left[\ln\frac{(kT)^{5/2}}{p}\frac{g_a(2\pi m_a)^{3/2}}{g_i^s h^3} - \frac{2\bar{\beta}\Delta\epsilon}{3k} + \frac{5}{2} - \frac{d\ln p}{d\ln T}\right] \tag{9.58}$$

since the term $(\pi^2 kT/6\Delta\epsilon)$ is much smaller than unity, Δx is small, and dV_D/dT is found[1] to be a small fraction of the Boltzmann constant k. Approximate values of α computed by means of the last part of Eq. 9.58 and data discussed in Sec. 9.3.7 are listed in the last column of Table 9.12.

9.4 Electron Emission from a Pure Nonuniform Surface

9.4.1 General Remarks. A *pure nonuniform surface* of a crystalline substance is defined as the surface of the substance consisting of parts, each of which has a

Table 9.12 Approximate Temperature Coefficients α Computed by Means of the Second Part of Eq. 9.58

Metal	T, °K	p (Fig. 9.12), torr	$\bar{\beta}$ (Ref. 25), 10^{-6}/°K	$\frac{d\ln p}{d\ln T}$ (Fig. 9.12)	α/k (Eq. 9.58)
Potassium	300	2.2×10^{-8}	250	37.3	1.3
Cesium	285	3×10^{-7}	290	32.9	1.8
Tungsten	2100	3×10^{-11}	20	46.9	0.6
Rhenium	2100	1×10^{-9}	30	45.3	0.4

different structure but can be regarded as a pure uniform surface. It is also called a *polycrystalline surface.* The uniform parts can have a variety of linear dimensions ranging from macroscopic to atomic and can be distributed in a variety of patterns. The shape of a polycrystalline surface can either be flat or appear as a succession of hills and valleys resulting from the intersections of planes of different crystallographic orientations or from the juxtaposition of various patterns of surface atoms, or both.

Unless special precautions are taken during preparation, surfaces of experimental specimens even of pure metals are usually nonuniform. This fact is amply supported by many experimental results related to thermionic, field, and photoelectric emission from pure metal surfaces and to adsorption properties, contact potentials (interelectrode motives), and surface structures of such metal surfaces.[1]

Surface characteristics of pure nonuniform emitter surfaces depend on both the relative size and the distribution of uniform parts. For purposes of thermionic conversion, the characteristics depend also on the relative distance between emitter and collector in a setup such as that shown in Fig. 9.2.

In general, it is not possible to associate a single work function with a nonuniform surface to serve for determining the interelectrode motive, the saturation current density, and the energy flux density as is done for a pure uniform surface.

Three types of pure nonuniform surfaces can be distinguished. In the first type the linear dimensions of the uniform parts are much larger than both the distance of the points just outside each part from the surface and the distance between emitter and collector. For this type each part can be analyzed as an independent pure uniform surface by means of the procedures of Secs. 9.2 and 9.3. This implies that, in a stable equilibrium state, the value of the electrochemical potential μ_e is common to all parts but each part has its own interfacial motive and work function. Moreover, this implies that juxtaposition of parts of different work functions has a negligible effect on the value of the work function of each part; in other words, exchange, correlation, and double-layer effects of each part are much stronger than effects resulting from the juxtaposition of parts. Thus the nonuniform surface can be characterized by a number of work functions and a corresponding number of fractional areas associated with the various parts. These values, in turn, can be used to compute various emission characteristics. For example, the electron-saturation-current density J_S is given by the relation

$$J_S = \sum_j A f_j (1 - \bar{r}_j) \exp\left(-\frac{\phi_j}{kT}\right) \tag{9.59}$$

where f_j, $\bar{r}_j$, and ϕ_j are the area per unit area, the average reflection coefficient, and the work function, respectively, of the jth uniform part and $\Sigma_j f_j = 1$.

In the second type of nonuniform surface, the linear dimensions of the uniform parts are smaller than the distance of the points just outside each part from the surface. For this type, adjacent parts are strongly coupled and cannot be analyzed independently of each other. Accordingly, the nonuniform surface must be treated like a surface with a complicated structure and must be characterized by a single work function in a manner analogous to that used for pure uniform surfaces. Calculating the value of this work function is much more involved than calculating that of a pure uniform surface. Nevertheless, correlation of experimental and theoretical results yields the same qualitative features as those discussed in Secs. 9.2 and 9.3.

Finally, in the third type of pure nonuniform surface, the linear dimensions of each part are: (1) sufficiently large so that, to zero order, an interfacial motive and a work function can be associated with each part and (2) sufficiently small so that effects of differences in work functions of adjacent uniform parts on the interfacial motives cannot be neglected. Specifically, the linear dimensions are comparable with the distances from the surfaces of the points just outside each part, and the effects of work-function differences modify the interfacial motives in the vicinity of the surfaces just outside.

The third type of nonuniform surface is called *patchy*, and each uniform part is called a *patch*. Moreover, the electrostatic fields resulting from the inequality in work functions of adjacent patches are called *patch fields*. Patch fields are superimposed on the fields resulting from exchange and correlation effects.

Sections 9.4.2 to 9.4.4 present the effects of patch fields on the motive, electron current emission, and electron velocity distribution.

9.4.2 Effects of Patch Fields on the Motive. We will consider the system of the two electrodes in Fig. 9.2 to be in a stable equilibrium state at temperature T and to be such that: (1) the emitter electrode is enclosed by a patchy surface; (2) the linear dimensions of the patches are much smaller than the interelectrode spacing; (3) ψ_j is the motive at the points just outside the jth patch in the absence of other patches; and (4) f_j is the area per unit area for the jth patch.

Typical motives for two arbitrary emitter patches of different structures, graphed vs. x partly through the interelectrode space, are shown in Fig. 9.14a. They are also shown superimposed in Fig. 9.14b. Included in Fig. 9.14 are the common value μ_e of the electrochemical potential of all the patches and the minimum one-electron energy level ϵ_{e0} in the emitter.

We see from Fig. 9.14b that, in the interfacial region between the points just inside each patch and the *points just outside the patchy surface*, defined later, the motives are different. The reasons are as follows: Very near each patch the motive is determined primarily by exchange, correlation, and double-layer

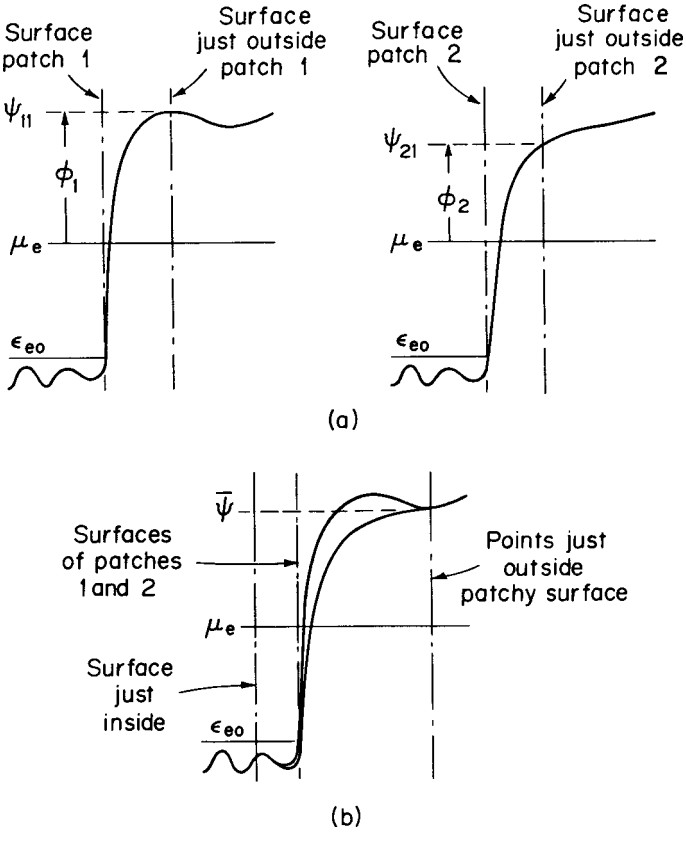


Fig. 9.14 Interfacial motives for two typical patches. (a) Separate. (b) Superimposed.

effects of the patch regarded as a pure uniform surface. These effects are predominant up to the points just outside the patch (Fig. 9.14a). At these points, the motive of the jth patch is ψ_{j1} and a work function ϕ_j can be defined by the relation

$$\phi_j = \psi_{j1} - \mu_e \tag{9.60}$$

Between the points just outside each patch and the points just outside the patchy surface, the patch fields are effective. In other words, in this region differences in motives ψ_{j1} or, equivalently, work functions ϕ_j of adjacent patches give rise to charge redistributions along the surface (dipole moments

along the surface) which create electrostatic fields that alter the motives. The patch fields vary exponentially with distance from the surface.[1] Their net effect is a change of the motives from their individual values ψ_j to the common value $\bar{\psi}$ of the points just outside the patchy surface (Fig. 9.14b). Simple electrostatic considerations yield that the value $\bar{\psi}$ is given by the relation[1]

$$\bar{\psi} = \sum_j f_j \psi_{j1} \tag{9.61}$$

i.e., by an area-weighted average of the motives at the points just outside each patch.

We also see from Fig. 9.14b that beyond the points just outside the patchy surface the values of the motives are identical. The reasons are that, beyond these points, the effects of the exponentially decreasing patch fields are negligible and the motive is described by the image-charge law. The motive $-e^2/16\pi\varepsilon_0 x$ for the image charge varies more slowly than exponentially with x and is independent of the detailed structure of the emitter surface. Of course, as the distance increases beyond the points just outside the patchy surface, space-charge effects become predominant.

It follows from the preceding discussion that the points just outside the patchy surface are defined as the points beyond which patch fields need not be considered and near which the motive is determined by the image-charge law.

The average motive $\bar{\psi}$ is used in a variety of ways. For example, it is the boundary value at the emitter for calculating the interelectrode motive by means of the methods described in Chap. 10. Again, in the limit of 0°K, its value $\bar{\psi}^*$ defines a work function $\bar{\phi}^*$ by the relation

$$\bar{\phi}^* = \bar{\psi}^* - \mu_e^* = \sum_j f_j \phi_j^* \tag{9.62}$$

where

$$\phi_j^* = \psi_j^* - \mu_e^* \tag{9.63}$$

The work function $\bar{\phi}^*$ can be measured by means of the Kelvin method for determining the contact potential between two surfaces.

On the other hand, we will see that $\bar{\psi}$ and μ_e do not uniquely characterize the electron-emission capabilities of the patchy surface.

9.4.3 Effects of Patch Fields on Electron Emission. As discussed in connection with Schottky plots for pure uniform surfaces, electron thermionic emission depends on the value of the applied electrostatic field. Patchy surfaces are no exception. They differ from uniform surfaces, however, in that the relations

between electron emission and surface characteristics, such as ψ_{j1}, ϕ_j, and f_j, are more involved by virtue of the effects of patch fields. Specifically, these relations are drastically affected by the relative magnitude of the applied field E with respect to the patch fields. They can be classified into three cases, depending on whether the value of the applied field E is smaller than, comparable with, or greater than the absolute magnitude $|\overline{E}_p|$ of the average patch field $\overline{E}_p$.

Qualitatively, the three cases can be represented by the Schottky plot of $\ln J$ vs. $E^{1/2}$, shown in Fig. 9.15, where J is the electron-emission-current density. The small graph above each case in the figure includes the motives of two representative patches of the emitter surface.

We see from Fig. 9.15 that, at both very small and very large values of E, the Schottky plot is linear, as would be expected for the same plot for a uniform surface. The two linear relations differ from each other, however, and define different intercepts with the ln J-axis. For intermediate values of E, the Schottky plot is nonlinear. This phenomenon is usually referred to as the *anomalous Schottky effect.*

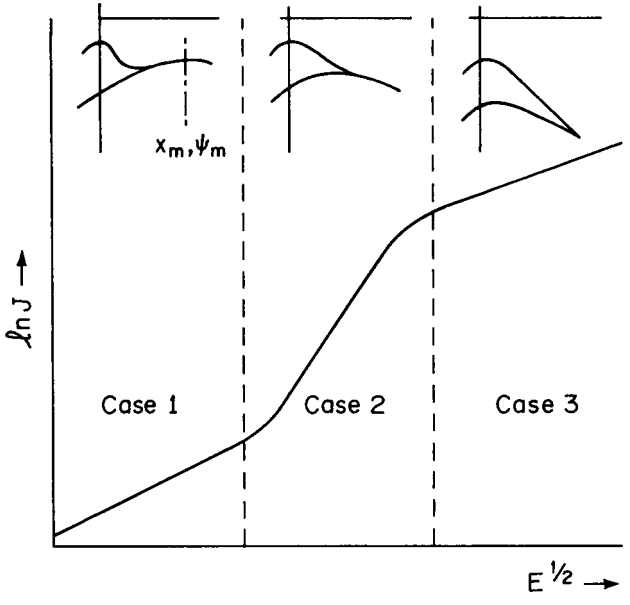


Fig. 9.15 Schottky plot for a patchy surface, including the shapes of two motives in the vicinity of the points just outside the patchy surface.

Quantitatively, each of the three cases can be described as follows:

Case 1. *Weak Applied Fields.* If the applied field E is such that a motive maximum occurs at the position x_m in front of yet near the points just outside the patchy surface (Fig. 9.15, case 1), then the motive maximum is beyond the range of effectiveness of the patch fields E_p but still in a region where space-charge effects are negligible.

A measure of the average patch field $\bar{E}_p$ is given by the relation

$$|\bar{E}_p| \approx \frac{\delta\phi}{\Delta y} \tag{9.64}$$

where $\delta\phi$ is the maximum difference in work functions of patches and Δy is a linear dimension of a typical patch. Thus case 1 corresponds to an applied field E such that

$$E \ll |\bar{E}_p| \tag{9.65}$$

Around the position x_m, the dependence of the motive on distance is described by the image-charge law, i.e., by $-e^2/16\pi\varepsilon_0 x$. Accordingly, the effect of sufficiently weak applied fields is the formation of a motive maximum ψ_m at x_m that differs from $\bar{\psi}$ by the amount $\Delta\psi$ given by Eq. 9.17.

Under these conditions all electrons emitted from patches whose work functions ϕ_j are smaller than $\psi_m - \mu_e$ must pass over a motive of height ψ_m. Thus, if only energies with positive velocity components normal to the surface are considered, these patches can be characterized by a work function $\bar{\phi} - \Delta\psi$, where $\bar{\phi}$ is given by the relation

$$\bar{\phi} = \bar{\psi} - \mu_e = \sum_j f_j \phi_j \tag{9.66}$$

and ϕ_j is defined by Eq. 9.60. On the other hand, all electrons emitted from patches whose work functions are larger than $\psi_m - \mu_e$ must pass over a motive of height larger than ψ_m and, therefore, are practically unaffected by the applied field. It follows that the Schottky plot is a straight line. The intercept of this line with the ln J-axis does not yield a zero-field current density characterized by $\bar{\phi}$.

An exception to the preceding conclusion occurs when the patchy surface consists of only two types of patches having approximately equal areas but substantially different work functions. For this surface most of the electron emission originates from the low-work-function patch, and the surface can be characterized by a work function equal to $\bar{\phi}$. The zero-field steady-state current density J_0 is given by the relation

$$J_0 = Af_{\min}(1 - \bar{r}) \exp\left(-\frac{\bar{\phi}}{kT}\right) \tag{9.67}$$

where $f_{\min}$ is the area per unit area of the low-work-function patch equal to about 0.5.

Case 2. *Applied Fields Comparable with Patch Fields.* When the applied field is comparable with the patch fields, the anomalous Schottky effect mentioned earlier is observed. Here the Schottky plot is nonlinear and the motives are as shown in Fig. 9.15, case 2. Specifically, maxima of motives of patches for which ψ_{j1} is smaller than $\bar{\psi}$ occur at positions between the points just outside each patch and the points just outside the patchy surface, namely, in the region where patch fields are appreciable. In this region, since the gradient of the motive is the sum of the image-charge field and the patch fields, and since the patch fields depend exponentially on distance from the surface, the maxima of the motives decrease faster with the value of the applied field than if patch fields were absent. Hence the average slope of the Schottky plot for case 2 is larger than that for case 1. Approximate expressions for the anomalous Schottky effect are given by Herring and Nichols.[1]

Case 3. *Applied Fields Stronger than Patch Fields.* When the applied field E is sufficiently strong ($E \gg |\bar{E}_p|$) so that the shapes of two representative motives near the emitter are as shown in Fig. 9.15, case 3, then the motives are independent of patch fields, and emission from each patch is independent of the presence of the other patches.

Here the reduction $\Delta\psi$ of each interfacial motive is given by Eq. 9.17, and the steady-state current density J is given by the relation

$$J = \sum_j Af_j(1 - \bar{r}_j) \exp\left(-\frac{\phi_j}{kT}\right) \exp\left[\frac{e\,(eE/16\pi\varepsilon_0)^{1/2}}{kT}\right] \tag{9.68}$$

We see from Eq. 9.68 that the Schottky plot for case 3 is a straight line. The intercept of this line with the ln J-axis defines the zero-field steady-state current density J_0 given by the relation

$$J_0 = \sum_j Af_j(1 - \bar{r}_j) \exp\left(-\frac{\phi_j}{kT}\right) \tag{9.69}$$

9.4.4 Effects of Patch Fields on the Velocity Distribution of Emitted Electrons. Figure 9.16 shows representative motives of the system of the two electrodes in Fig. 9.2 in a steady state, with the collector kept at a very low temperature, such that the interelectrode motive is retarding (decelerating) for electrons originating from the emitter. Figure 9.16a corresponds to a patchy

emitter surface and a uniform collector surface, and Fig. 9.16b corresponds to patchy emitter and patchy collector surfaces. In both motives the interelectrode spacing is much larger than the range of effectiveness of the patch fields.

If the effect of velocity components parallel to the emitter surface is neglected, electrons originating from patches with work functions larger than $\overline{\phi}$ (Fig. 9.16) will have a shifted half-Maxwellian distribution at the points just outside the patchy emitter, and electrons originating from emitter patches with work functions smaller than $\overline{\phi}$ will have a half-Maxwellian distribution; the overall distribution, therefore, is not half-Maxwellian.

If the effect of velocity components parallel to the emitter surface is included, these components can contribute to deviations from the half-Maxwellian through the process of *channeling*. Specifically, the equimotive surfaces in the interfacial region of a patchy emitter are not planar. The wavy character of the equimotive surfaces can result in changes of velocity directions from parallel to normal to the polycrystalline surface. These changes, in turn, contribute to deviations from the half-Maxwellian distribution and are said to be caused by the process of channeling.

A quantitative but approximate discussion of the effect of patch fields on electron distribution in terms of reflection coefficients is given by Herring and Nichols.[1]

Deviations from the half-Maxwellian affect the amount of current entering the collector both when the collector is uniform and when it is patchy. They also affect the expression for the energy flux density compared with that given for pure uniform surfaces.

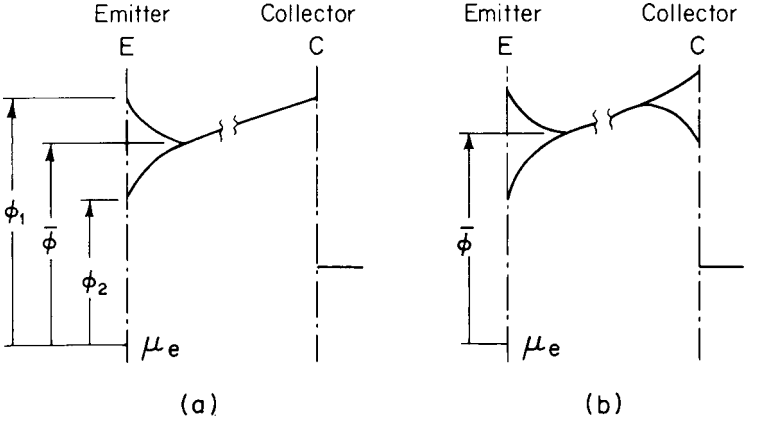


Fig. 9.16 Retarding interelectrode motives between (a) patchy emitter and uniform collector surfaces and (b) patchy emitter and patchy collector surfaces.

9.5 Composite Surfaces—Definitions

A pure uniform or pure nonuniform surface of a substance covered by a monolayer (layer of atomic thickness) of adsorbed particles of a different material is defined as a *composite surface.* For example, a tungsten surface covered by a monolayer of cesium particles is called a composite surface. The crystalline substance is called the *substrate,* and the adsorbed substance is called the *adsorbate.*

Three types of composite surfaces can be distinguished. In the first type the electronegativity of the free atoms of the adsorbate is smaller than the work function of the surface of the substrate, and the composite surface will therefore be called an *electropositive surface.* For example, a cesiated surface of tungsten is an electropositive surface because the electronegativity of cesium is about 2.2 ev whereas the work function of tungsten is about 5 ev.

In the second type the electronegativity of the free atoms of the adsorbate is larger than the work function of the surface of the substrate, and the composite surface will therefore be called an *electronegative surface.* For example, an oxygenated surface of tungsten is an electronegative surface because the electronegativity of oxygen is about 9.65 ev whereas the work function of tungsten is about 5 ev.

Finally, in the third type the adsorbate consists of a mixture of particles of various materials with atomic electronegativities different from each other and different from the work function of the surface of the substrate. In this case the composite surface will be called a *mixed surface.* For example, a cesiated and oxygenated surface of tungsten is a mixed surface.

For electropositive and electronegative surfaces, the expectation value of the number of adsorbate particles per unit surface area of the substrate will be denoted by σ, the maximum possible value of σ by σ_f, and the ratio σ/σ_f by θ.

$$\theta = \frac{\sigma}{\sigma_f} \tag{9.70}$$

The quantity σ_f is called the *adsorbate surface density of a full monolayer.* The ratio θ is called the *degree of coverage of the substrate surface* or simply the *coverage*. When the surface is partially covered by the adsorbate, θ is smaller than unity, whereas for a full monolayer θ is equal to unity.† Analogous definitions can be given for mixed surfaces.

For a given substrate–adsorbate combination, no universally acceptable definition of a full monolayer is available; in other words, different investigators define different values of σ_f of σ for which θ is equal to unity. For the purposes

†By definition, conditions under which θ is greater than unity do not correspond to a monolayer, i.e., a layer of atomic thickness.

of this book, values of σ_f will be determined primarily on the basis of geometric considerations, such as the atomic radii of the adsorbate and the substrate. For example, since the atomic radius of cesium is about twice as large as the atomic radius of tungsten, σ_f for a cesiated monocrystalline tungsten surface will be taken equal to $\sigma_m/4$, where σ_m is the number of tungsten atoms per unit area of the tungsten surface.

Composite surfaces can also be characterized as *uniform* or *nonuniform* depending on whether the substrate surface is uniform or nonuniform and the coverage θ is larger than a lower limit. The lower limit of θ, about 10^{-3}, corresponds to a stable equilibrium state in which the distance between absorbate particles is comparable to the distance of the points just outside from the surface. For example, a cesiated (1 1 0) surface of tungsten with a coverage θ greater than about 10^{-3} is a uniform electropositive surface. For degrees of coverage smaller than about 10^{-3}, the composite surface must be considered as patchy for the reasons discussed in Sec. 9.4.1.

Emission characteristics of electropositive surfaces are discussed in Sec. 9.6. A brief discussion of electronegative surfaces and of mixed surfaces is given in Sec. 9.7.

9.6 Emission Characteristics of Electropositive Surfaces

9.6.1 General Remarks. This section presents both theoretical and experimental results on emission characteristics of electropositive surfaces. The emphasis of the presentation is, in general, on uniform surfaces because they provide an essential starting point for the discussion of other electropositive surfaces and, in particular, on cesiated refractory metal surfaces because of their importance in thermionic conversion.

The experimental setup for the study of emissions from uniform electropositive surfaces can be represented by the two-electrode-plus-battery system shown in Fig. 9.17. In this figure the emitter is a refractory metal enclosed by a pure uniform surface and is immersed in the vapor of the adsorbate. The electronegativity of the adsorbate is smaller than the work function of the pure uniform surface of the substrate. The pressure of the adsorbate vapor in the interelectrode space, p^g, is maintained by the liquid adsorbate reservoir appended to the adsorbate vapor container. The temperature T_R of the reservoir is smaller than the temperature T of the emitter and the temperature of the collector and is controlled independently of these two temperatures. Finally, the pressure p^g may be different from the vapor pressure of the liquid adsorbate at temperature T_R (see Sec. 8.5.2, Eq. 8.39).

In general, the pressure p^g is small but is much greater than the vapor pressures of the pure emitter and pure collector materials. Under this condition any free

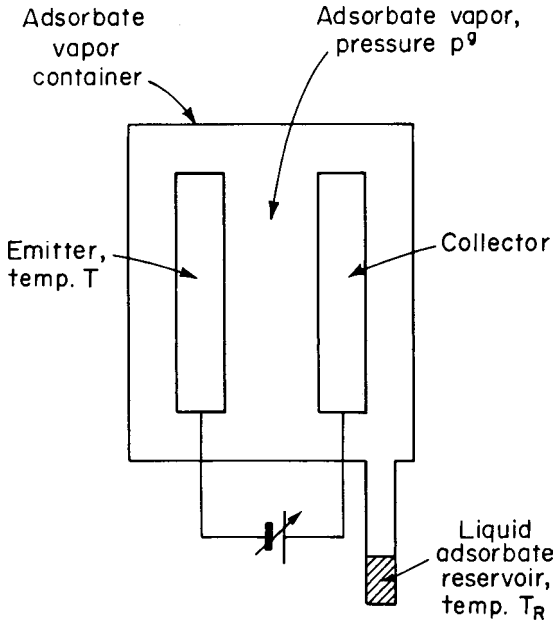


Fig. 9.17 Schematic of setup for the study of emissions from composite surfaces.

neutral atoms and free positive ions of the electrode materials in the interelectrode space can be neglected in the analysis of the phenomena in that space.

When the two electrodes and the vapor in the interelectrode space are either in a steady state or in a state of mutual stable equilibrium, each of the electrode surfaces is covered by a monolayer of particles of the adsorbate. These particles are held on each substrate surface by forces that are relatively strong, and therefore they are a part of an adsorbate–substrate combination. The formation of the adsorbate layer on the substrate is analogous to the formation of a diatomic molecule by the bonding of two dissimilar atoms.

The geometric arrangement of the adsorbate particles on the surface of the substrate of the emitter is shown schematically in Fig. 9.18. In this figure each cross (+) denotes the expectation value of the position of the center of the nucleus of a particle, and each particle is represented by a sphere of radius equal to the atomic radius of the particle. The geometric surface defined by the centers of the nuclei of the adsorbate is called the *adsorbate surface.*

As a result of the strong binding forces between adsorbate and substrate, the emission characteristics of an electropositive surface can differ substantially

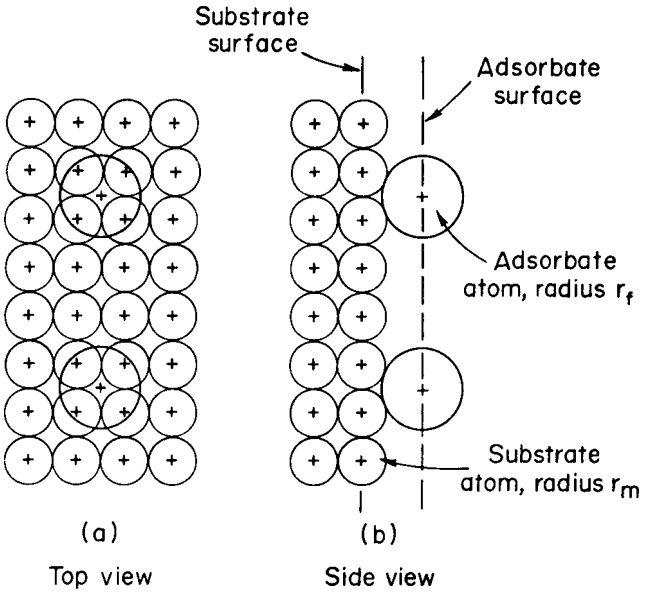


Fig. 9.18 Geometric arrangement of absorbate particles on a substrate.

from those of the substrate surface in the absence of the adsorbate. For example, the electropositive surface can be a copious emitter of free atoms and free positive ions of the adsorbate material, in addition to electrons. It follows that, in general, adsorbate atom and ion emissions cannot be neglected from the analysis as might be done in some analyses of pure uniform surfaces.

The emission characteristics of a uniform electropositive emitter can be discussed along the lines indicated in Sec. 9.1 and in terms of thermodynamic potentials identical to those used for the discussion of pure uniform surfaces in Sec. 9.3.7. For example, an electron work function can be defined to serve for evaluating the electron interelectrode motive, electron-saturation-current density, and electron-energy-flux density. Again, an ion work function can be defined to serve for evaluating the ion interelectrode motive, ion-saturation-current density, and ion-energy-flux density. Here, however, the values of the thermodynamic potentials depend on both the properties of the substrate and the properties of the adsorbate vapor in a state of mutual stable equilibrium. For example, the value of the electron work function ϕ_e of a cesiated tungsten surface (see also Vol. I, Chap. 4) can be smaller by several electron volts than the value of the electron work function ϕ_{em} of the surface of the pure tungsten, the so-called *bare work function.* Among other things, the decrease depends on the

crystallographic orientation of the surface of tungsten and the pressure of the cesium vapor in which the tungsten is immersed.

In view of these remarks, it follows that analyzing the emission characteristics is equivalent to calculating various thermodynamic potentials. Rigorous analysis of the thermodynamic potentials of an electropositive surface is even more difficult than that of pure uniform surfaces, however, because of the added complexity introduced by the dissimilarity between adsorbate and substrate particles. This dissimilarity makes calculating force fields and, therefore, energy eigenstates in the interfacial region of an electropositive surface an extremely complicated mathematical task.

The first detailed experimental and theoretical study of electropositive surfaces was performed by Langmuir and Taylor,[49] who investigated the emission properties of a cesiated-tungsten surface. Subsequently, much experimental and theoretical work has been presented by many investigators. For thermionic conversion, however, the most practical work is that of Rasor and Warner[50] and of Gyftopoulos and Levine.[51,52]

The theories of Langmuir and Taylor, Rasor and Warner, and Gyftopoulos and Levine have many common features, but they also have substantial differences. Some of the salient points of the three theories are summarized here.

All three theories assume that: (1) a two-dimensional system, a surface system, consisting of mobile adsorbate particles can be identified; and (2) electrons are transferred from the surface system to the surface of the substrate, thus causing the creation of electrostatic fields that, in turn, result in a reduction of the work function of the substrate. However, whereas Langmuir and Taylor assume that one electron per adsorbate particle is transferred to the substrate, Rasor and Warner and Gyftopoulos and Levine introduce in their respective theories the possibility that less than one electron per adsorbate particle is transferred on the average to the substrate surface. This innovation, introduced in different forms in the two theories, provides one possible explanation of the inconsistency observed by Langmuir and Taylor at high coverages between the change in work function of the substrate calculated from electron-emission data and that calculated from atom-emission data.

Langmuir and Taylor[49] used some thermodynamic relations supplemented by several correlations of experimental results and derived atom-, ion-, and electron-emission characteristics of the cesiated tungsten surface they investigated.

Rasor and Warner[50] also used some thermodynamic relations supplemented by Langmuir's ion-emission data and derived atom-, ion-, and electron-emission characteristics of cesiated refractory-metal surfaces in general. In their derivations they assumed that the dipole moment and the heat of desorption of the ionic species (cesium particles in an "ionic state") are independent of the nature

of the substrate. Moreover, they assumed that the change of the heat of desorption of the ionic species is proportional to the change of the electron work function and that the constant of proportionality is a function of the coverage θ only.

On the other hand, Gyftopoulos and Levine[51,52] adapted to surfaces empirical correlations that have been developed in the theory of molecular bonds and combined these correlations with thermodynamic relations to derive atom-, ion-, and electron-emission characteristics for any electropositive surface.

For purposes of thermionic conversion and for degrees of coverage greater than about 0.5, no satisfactory general theory exists at the present time. On the other hand, for degrees of coverage smaller than about 0.5, a complete and rigorous thermodynamic analysis of electropositive surfaces can be performed without use of either experimental emission data or the empirical correlations of Gyftopoulos and Levine. Such an analysis is possible because Langmuir and Taylor showed experimentally that, for θ smaller than 0.5, the interparticle forces between cesium atoms adsorbed on tungsten are of the dipole–dipole type and that the work-function change caused by adsorption can be fully accounted for by electrostatic effects only. Input information required in the analysis, which is presented in Secs. 9.6.2 to 9.6.4, is the ionic radius of the adsorbate and the bare work function. Analytical results are compared with experimental data in Sec. 9.6.5. Finally, nonuniform electropositive surfaces are discussed briefly in Sec. 9.6.6.

9.6.2 Model for Moderate Coverages ($\theta < 0.5$). If a finite piece of a refractory metal enclosed by a uniform electropositive surface of coverage θ is in mutual stable equilibrium with the vapor of the adsorbate at temperature T and at pressure p^g much greater than the vapor pressure of the substrate at T, then points just inside and points just outside the electropositive surface can be defined in a manner analogous to that used for pure uniform surfaces (Sec. 9.3.2). The force fields in the vapor beyond the points just outside depend on space-charge effects only; the force fields in the interfacial region between points just outside and points just inside depend on properties of and interactions between substrate and adsorbate particles; and the force fields in the bulk of the substrate up to the points just inside depend on properties of the substrate only.

Because p^g is much greater than the vapor pressure of the substrate, any free positive ions and free neutral atoms of the substrate can be neglected from the analysis of the vapor of the adsorbate. Accordingly, the vapor beyond the points just outside can be regarded as a heterogeneous mixture of electrons and free positive ions and free neutral atoms of the adsorbate. At temperatures of practical interest, each component and species of the mixture behaves like a perfect gas, and, to an excellent approximation, the densities of positive ions

other than singly charged ions are negligibly small. It follows that many properties of the phases of the mixture can be expressed in terms of p^g and T by means of the perfect-gas relations. For example, the total potential μ_a^g for atoms is given by the relation

$$\mu_a^g = kT \ln \frac{p^g}{(kT)^{5/2}} \frac{h^3}{g_a^g (2\pi m_f)^{3/2}} + \bar{\varepsilon}_{oa} \tag{9.71}$$

where m_f, g_a^g and $\bar{\varepsilon}_{oa}$ are the mass, degeneracy, and energy, respectively, of internal structure of the free atoms of the mixture. Because of the conditions for mutual stable equilibrium, μ_a^g equals the total potential of atoms in the interfacial region and in the bulk of the substrate.

Consistent with the preceding approximation for the mixture, the interfacial region and the bulk of the substrate can be regarded as an electrically neutral, two-component, open system consisting of electrons and positive ion cores. These two components can interact chemically with each other to form an electrically neutral species that has the same total potential as the free atoms of the mixture. The chemical reaction is analogous to that between electrons and free positive ions in the vapor beyond the points just outside.

This conclusion and the observation that the bulk of the substrate is not affected by the presence of the adsorbate suggest that the interfacial region and the bulk may be replaced approximately by an electrically neutral, two-dimensional, open system (i.e., a surface system) consisting of electrons and singly charged positive-ion cores of the adsorbate and existing in the force fields of a structureless mixture of electrons and ion cores of the substrate.

In the surface system the expectation value of the number of either electrons or singly charged adsorbate positive ions per unit surface area of the substrate is equal to σ. Moreover, one effect of the substrate force fields is that each electron is strongly correlated with a singly charged adsorbate positive ion so that the pair can be regarded as an electrically neutral species. Each pair, called an *ionic particle,* is characterized by an internal structure with a partition function q_a^s and is interacting with the other ionic particles. Other effects of the substrate force fields are discussed later.

By virtue of the condition for mutual stable equilibrium between the surface system and the vapor, the total potential μ_a^s of the ionic particles must satisfy the relation

$$\mu_a^s = \mu_a^g \tag{9.72}$$

We will see subsequently that, for given interaction laws between ionic particles, Eq. 9.72 determines the dependence of σ (or θ) on p^g and T.

Quantitatively, a rigorous definition of ionic particle and interparticle forces in the surface system is not available. To proceed with the analysis, we must

introduce an approximate and tractable model. The model presented here will be based on the assumptions that follow.

Assumption 1. The net effect of interactions between adsorbate and substrate is the formation of the ionic particles of the surface system.

Assumption 2. Each ionic particle has two translational degrees of freedom along the surface, i.e., it is *mobile,* and it has one vibrational degree of freedom normal to the surface.

Assumption 3. The exchange-plus-correlation effects for electrons near the surface system are identical to the exchange-plus-correlation effects for electrons near the surface of the substrate in the absence of the adsorbate.

Assumption 4. The electrostatic effects of the surface system are equal to the sum of the electrostatic effects of the ionic particles plus the electrostatic effects of the surface of the substrate in the absence of the adsorbate.

Assumption 5. For monovalent adsorbates the ionic particle is defined as a neutral adsorbate atom A_1 such that: (1) in the absence of interactions with other adsorbate particles, A_1 has a permanent dipole moment M_{01}, which is defined subsequently, and a ground-state energy ϵ_{01}; (2) A_1 is polarizable and has a polarizability α_1; (3) interactions of A_1 with other particles of its species are described by the dipole–dipole force law; and (4) the vibrations of A_1 are represented by a one-dimensional harmonic oscillator with g_1^s-fold degenerate energy levels and a fundamental frequency ν_1.

The values of the dipole moment M_{01}, energy ϵ_{01}, polarizability α_1, and frequency ν_1 are determined by the following arguments.

The expectation values of the positions of the centers of the nuclei of the adsorbate and the substrate can be represented by the schematic shown in Fig. 9.18. In this figure each adsorbate nucleus is nested symmetrically between substrate nuclei. The distances between the centers of substrate nuclei at the substrate surface are determined by the crystallographic orientation of this surface and correspond to a density of σ_m nuclei per unit area. The distances between the centers of adsorbate nuclei are determined by σ. Finally, the distance between nearest adsorbate and substrate centers is equal to the sum $r_f + r_m$ of the atomic radii r_f of the adsorbate and r_m of the substrate.

To define ionic particles of the surface system, we replace the substrate by a perfect conductor enclosed by a structureless surface at a distance r_1 from the adsorbate surface, where r_1 is taken equal to the radius of the free singly charged positive ion of the adsorbate, the so-called ionic radius of the adsorbate.

The ground state of the ionic particle is regarded as the state in which one valence electron of the adsorbate is distributed along the surface of the perfect conductor according to the image-force law and a singly charged positive ion of the adsorbate is within a sphere of radius r_1. This state of charge distribution is shown schematically in Fig. 9.19a. For electrostatic calculations valid only

outside the perfect conductor, this state can also be represented by the positive and negative spherical charges shown in Fig. 9.19b. For example, we see from Fig. 9.19b that the dipole moment of the ionic particle without self-depolarization is equal to er_1. We can readily verify that the dipole moment M_{01} with self-depolarization is given by the relation

$$M_{01} = \frac{er_1}{1 + [\alpha_i/4\pi\varepsilon_0(2r_1)^3]} = \frac{8}{9}er_1 \tag{9.73}$$

The second part of Eq. 9.73 results from the estimate of the polarizability α_i of the singly charged positive ion of the adsorbate given by the relation[51]

$$\alpha_i = 4\pi\varepsilon_0 r_1^3 \tag{9.74}$$

Again, we see from Fig. 9.19b that, with respect to a zero-energy reference level at the points just outside, the energy ϵ_{01} is given by the relation

$$\epsilon_{01} = -\frac{e^2}{16\pi\varepsilon_0 r_1} + \bar{\varepsilon}_{oi} \tag{9.75}$$

where $\bar{\varepsilon}_{oi}$ is the energy of the internal structure of a free singly charged adsorbate positive ion.

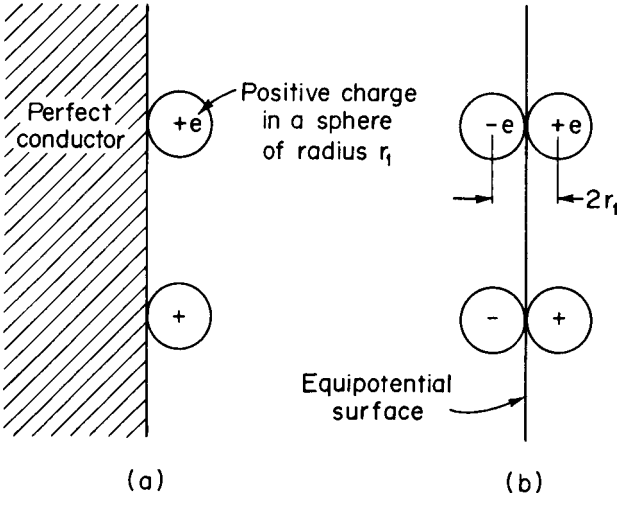


Fig. 9.19 Representation of ionic particle.

Finally, Luke[53] has shown that, consistent with the preceding definition of the ionic particle, the polarizability α_1 is given by the relation

$$\alpha_1 = \frac{4}{3}\alpha_i \tag{9.76}$$

We can obtain an estimate of the value of the fundamental frequency ν_1 by assuming that the energy and amplitude of oscillation of the harmonic oscillator are equal to $-e^2/16\pi\varepsilon_0 r_1$ and r_1, respectively (Ref. 52, I). Thus we find that ν_1 is given by the relation

$$\nu_1 = \frac{1}{2\pi r_1}\left(\frac{e^2/16\pi\varepsilon_0 r_1}{m_f}\right)^{1/2} \tag{9.77}$$

The partition function q_a^s can be regarded as the partition function q_i^s of a vibrating ion core when the electron is distributed along the perfect conductor surface as specified in the definition of the ground state of the ionic species. It follows from the discussion in Sec. 7.3.5 that q_i^s is given by the relation

$$q_a^s = q_i^s = g_1^s \frac{\exp(-\epsilon_{01}/kT)}{1 - \exp(-h\nu_1/kT)} \tag{9.78}$$

At temperatures of practical interest, kT is much greater than $h\nu_1$, and Eq. 9.78 can be approximated by the relation

$$q_a^s = q_i^s = g_1^s \frac{kT}{h\nu_1} \exp\left(-\frac{\epsilon_{01}}{kT}\right) \tag{9.79}$$

This completes the presentation of the model that will be used to analyze electropositive surfaces involving monovalent adsorbates at degrees of coverage $\theta < 0.5$. Thermodynamic relations that are consistent with the model are discussed in Sec. 9.6.3.

9.6.3 Equations for Stable Equilibrium States. This section presents expressions for thermodynamic potentials and emission characteristics consistent with the surface model introduced in Sec. 9.6.2.

The expressions for thermodynamic potentials depend on the forces between particles of the surface system. These forces have been assumed to be long range (dipole–dipole type), and therefore they are effective for all degrees of coverage except in the limit of θ arbitrarily close to zero.

For θ very close to zero, the surface can be regarded as a semiperfect two-dimensional gas, and the total potential μ_a^s can be computed by means of the relations given in Sec. 7.11.3. Specifically, μ_a^s can be written in the form

$$\mu_a^s = \mu_i^s + \mu_e^s \tag{9.80a}$$

where μ_i^s and μ_e^s are the electrochemical potentials of singly charged adsorbate positive ions and electrons in the surface, respectively. The values of μ_i^s and μ_e^s can be computed explicitly only when the interfacial motive is known. To avoid the need for the interfacial motive, we can write Eq. 9.80a in the alternative form

$$\mu_a^s = (\mu_{ci}^s)_1 + (\mu_{ce}^s)_1 \tag{9.80b}$$

where $(\mu_{ci}^s)_1$ and $(\mu_{ce}^s)_1$ are the chemical potentials of singly charged positive ions and of electrons, respectively, with respect to the points just outside. For θ very close to zero, $(\mu_{ci}^s)_1$ is given by the relations (Eq. 7.263)

$$\begin{aligned}(\mu_{ci}^s)_1 &= kT \ln \sigma_f \theta - kT \ln T - kT \ln q_i^s - kT \ln \frac{2\pi m_f k}{h^2} \\ &= kT \ln \frac{\nu_1 \sigma_f \theta}{g_1^s} \frac{h^3}{2\pi m_f (kT)^2} - \frac{e^2}{16\pi\varepsilon_0 r_1} + \bar{\varepsilon}_{oi} \qquad \text{for } \theta \approx 0\end{aligned} \tag{9.81}$$

where use has been made of Eqs. 9.75 and 9.79. On the other hand, $(\mu_{ce}^s)_1$ is given by the relation

$$(\mu_{ce}^s)_1 = -\phi_{em} \qquad \text{for } \theta \approx 0 \tag{9.82}$$

since in the limit of very small θ the work function of the surface is equal to the bare work function ϕ_{em} and $(\mu_{ce}^s)_1$ is always equal to the negative of the work function. Combining Eqs. 9.80b and 9.82 yields that

$$\mu_a^s = kT \ln \frac{\nu_1 \sigma_f \theta}{g_1^s} \frac{h^3}{2\pi m_f (kT)^2} - \frac{e^2}{16\pi\varepsilon_0 r_1} + \bar{\varepsilon}_{oi} - \phi_{em} \qquad \text{for } \theta \approx 0 \tag{9.83}$$

For θ greater than some very small value, the forces between the particles of the surface system become effective and, therefore, the surface must be regarded as nondilute (Sec. 7.11.4). For nondilute surfaces the expression for the total potential μ_a^s of the ionic particles depends on the effects of the interparticle forces on the spreading force τ.

The spreading force τ can be expressed by the relation

$$\tau = \tau_0 + \tau_1 \tag{9.84}$$

where τ_0 and τ_1 are the components of τ caused by short- and long-range forces, respectively, between the ionic particles. Using Eq. 9.84 in Eq. 7.267 yields that μ_a^s is given by the relation

$$\mu_a^s = \mu_a^s(\sigma_f\theta^*) + \int_{\theta^*}^{\theta} \frac{1}{\sigma_f\theta}\left(\frac{\partial\tau_1}{\partial\theta}\right)_T d\theta + \int_{\theta^*}^{\theta} \frac{1}{\sigma_f\theta}\left(\frac{\partial\tau_0}{\partial\theta}\right)_T d\theta \tag{9.85}$$

where $\mu_a^s(\sigma_f\theta^*)$ is the value of μ_a^s in the limit of a degree of coverage θ^* arbitrarily close to zero; in other words, $\mu_a^s(\sigma_f\theta^*)$ can be computed by means of Eq. 9.83.

The component τ_0 of τ can be expressed as a function of θ by use of the van der Waals excluded area assumption. Thus

$$\tau_0 = \frac{\sigma_f\theta kT}{1-\theta} = \frac{\sigma kT}{1-\theta} \tag{9.86}$$

On the other hand, owing to dipole–dipole interactions between polarizable ionic particles, two estimates of the component τ_1 of τ have been computed by Langmuir.[54] He obtained one estimate, τ_{1v}, from the second virial of the virial expansion for the nondilute surface. This estimate is given by the relation

$$\tau_{1v} = D_v T^{1/3} M_1^{4/3} \ (\sigma_f\theta)^2 = D_v T^{1/3} M_1^{4/3} \ \sigma^2 \tag{9.87}$$

where

$$D_v = 9.287 \times 10^{19} \qquad (\text{ev-cm}^{2/3})/(\text{coulomb}^{4/3}\text{-}^\circ\text{K}^{1/3}) \tag{9.88}$$

and M_1 is the depolarized dipole moment of the ionic particles. If the average depolarizing field is assumed equal to $\Delta\phi/er_1$, where $\Delta\phi$ is the change in electron work function given by Eq. 9.101b, M_1 is given by the relation

$$M_1 = \frac{M_{01}}{1+b\sigma_f\theta} = \frac{M_{01}}{1+b\sigma} = \frac{(8/9)er_1}{1+(16\pi r_1^2/3)\sigma} \tag{9.89}$$

where

$$b = \frac{\alpha_1}{\varepsilon_0 r_1} = \frac{16\pi}{3} r_1^2 \tag{9.90}$$

and use has been made of Eqs. 9.73, 9.74, and 9.76.

Langmuir[54] obtained a second estimate τ_{1a} of τ_1 based on the assumption that the ionic particles are arranged on a hexagonal array on the surface of the substrate so that the surface density is σ. The estimate τ_{1a} is given by the relation

$$\tau_{1a} = D_a M_1^2 (\sigma_f \theta)^{5/2} = D_a M_1^2 \sigma^{5/2} \tag{9.91}$$

where

$$D_a = 7.52 \times 10^{31} \text{ ev-cm/coulomb}^2 \tag{9.92}$$

For a given θ, both τ_{1v} and τ_{1a} are underestimates of τ_1. The second virial can be interpreted as representing effects of binary collisions only. It follows that τ_{1v} is smaller than τ_1 because effects of collisions other than binary are neglected. The difference between τ_1 and τ_{1v} is smaller, the smaller the value of θ. The hexagonal array has the smallest energy with respect to other configurations of the ionic particles. It follows that τ_{1a} is smaller than τ_1 because contributions of arrays with energies higher than the energy of the hexagonal are neglected. For a given temperature the difference between τ_1 and τ_{1a} is smaller, the larger the value of θ.

Langmuir[54] suggested a specific combination of τ_{1v} and τ_{1a} as an improved estimate of τ_1. His suggestion, however, results in some thermodynamic inconsistencies that will not be discussed here. For the purposes of this book, we will assume that τ_1 is given by the relation

$$\tau_1 = (1 - \theta)\tau_{1v} + \tau_{1a} \tag{9.93}$$

Substituting Eqs. 9.86 and 9.93 into Eq. 9.85 and using Eq. 9.83, we find that μ_a^s is given by the relation

$$\mu_a^s = kT \ln \frac{\nu_1 \sigma_f}{g_1^s} \frac{h^3}{2\pi m_f (kT)^2} - \frac{e^2}{16\pi\varepsilon_0 r_1} + \bar{\varepsilon}_{oi} - \phi_{em} + kT\left(\ln \frac{\theta}{1-\theta} + \frac{\theta}{1-\theta}\right) + \Delta\mu \tag{9.94}$$

where

$$\Delta\mu = \frac{D_v M_{01}^{4/3} T^{1/3}}{b} \left\{3 - \frac{3 + 2b\sigma}{(1 + b\sigma)^{1/2}} + \frac{1}{b\sigma_f}\left[4.5 - \frac{2 + b\sigma}{(1 + b\sigma)^{1/2}} - 2.5\,(1 + b\sigma)^{3/2}\right]\right\} + \frac{D_a M_{01}^2}{b^{3/2}}\left[\arctan (b\sigma)^{1/2} - \frac{(b\sigma)^{1/2}}{(1 + b\sigma)^2}\right] \tag{9.95}$$

Equation 9.94 is independent of θ^*, as it should be, and reduces to Eq. 9.93 when θ is very small. It depends explicitly on the bare work function ϕ_{em} and not on the electron work function ϕ_e of the surface system.

Equation 9.94 can be used in the condition for mutual stable equilibrium (Eq. 9.72) to obtain a relation between θ, T, and p^g or between θ, T, and the neutral-atom random-flux density Γ_a. Specifically, using Eqs. 9.71, 9.72, and 9.94, we obtain the relation

$$A_a - \frac{B_a}{kT} = \ln \frac{p^g}{(2\pi m_f kT)^{1/2}} = \ln \Gamma_a \tag{9.96}$$

where

$$A_a \equiv \ln \Gamma^s_{a0} + \ln \frac{2\theta}{1-\theta} + \frac{\theta}{1-\theta} \tag{9.97}$$

$$B_a \equiv \frac{e^2}{16\pi\varepsilon_0 r_1} + \phi^*_{em} - V_i - \Delta\mu \tag{9.98}$$

$$\Gamma_a \equiv \frac{p^g}{(2\pi m_f kT)^{1/2}} \tag{9.99}$$

$$\Gamma^s_{a0} \equiv \frac{\sigma_f \nu_1 g^g_a}{2 \exp(\alpha_m/k) g^s_1} \tag{9.100}$$

Here ϕ^*_{em} is the value of ϕ_{em} in the limit of 0°K and α_m is the temperature coefficient of ϕ_{em} such that

$$\phi_{em} = \phi^*_{em} + \alpha_m T$$

Equation 9.96 is expressed in terms of A_a and B_a for several reasons. The quantity A_a is independent of temperature, and the difference $A_a - \ln \Gamma^s_{a0}$ is almost equal to zero in the range of θ between 0.1 and 0.5. It will be seen later that this feature of A_a suggests an accurate and practical correlation between the work function ϕ_e of the surface system and the properties of the adsorbate vapor. With regard to B_a, it is shown in Appendix 9.B that, in the limit of 0°K, B_a is a function of θ or σ only and is equal to the *enthalpy of desorption of neutral adsorbate atoms* from the surface.

Expressions similar to Eq. 9.96 can be derived for electrons and singly charged adsorbate positive ions. Indeed, by virtue of assumptions 3 and 4 of Sec. 9.6.2, the work function ϕ_e of the surface system is equal to the bare work function ϕ_{em} plus the electrostatic effects of the ionic particles evaluated at the points just outside. The electrostatic effects are caused by the depolarized dipoles. With respect to the points just outside, the dipoles can be regarded as a continuous

double layer[55] with dipole density $\sigma_f\theta$ or σ. Thus ϕ_e can be expressed, in terms of M_{01} and θ, by means of the relation

$$\phi_e = \phi_{em} - \frac{e\sigma_f\theta}{\varepsilon_0}\frac{M_{01}}{1+b\sigma_f\theta} = \phi_{em} - \frac{eM_{01}}{\varepsilon_0}\frac{\sigma}{1+b\sigma} \tag{9.101a}$$

or

$$\Delta\phi \equiv \phi_{em} - \phi_e = \frac{e\sigma_f\theta}{\varepsilon_0}\frac{M_{01}}{1+b\sigma_f\theta} = \frac{(8e^2r_1/9\varepsilon_0)\sigma}{1+(16\pi r_1^2/3)\sigma} \tag{9.101b}$$

Using ϕ_e in the expression for the electron-random-current density $(e\Gamma_e)_1$ at the points just outside (Eq. 9.7), we obtain the relation

$$A_e - \frac{B_e}{kT} = \ln (e\Gamma_e)_1 \tag{9.102}$$

where

$$A_e \equiv \ln AT^2 \exp\left(-\frac{\alpha_m}{k}\right) \tag{9.103}$$

$$B_e \equiv \phi_{em}^* - \frac{eM_{01}}{\varepsilon_0}\frac{\sigma_f\theta}{1+b\sigma_f\theta} \tag{9.104}$$

and $A = 120\ \text{amp/cm}^2\text{-}^\circ\text{K}^2$.

Finally, for singly charged positive ions, the chemical potential $(\mu_{ci}^s)_1$ at the points just outside is given by a combination of Eqs. 9.80b, 9.94, and 9.101a, namely,

$$(\mu_{ci}^s)_1 = \mu_a^s - (\mu_{ce}^s)_1 = \mu_a^s + \phi_e$$

$$= kT \ln \frac{\nu_1\sigma_f}{g_1^s}\frac{h^3}{2\pi m_f (kT)^2} - \frac{e^2}{16\pi\varepsilon_0 r_1} + \bar{\varepsilon}_{oi} + kT\left(\ln\frac{\theta}{1-\theta} + \frac{\theta}{1-\theta}\right)$$

$$+ \Delta\mu - \Delta\phi \tag{9.105}$$

where use has been made of the equality between $(\mu_{ce}^s)_1$ and $-\phi_e$. On the other hand, $(\mu_{ci}^s)_1$ must also be equal to the chemical potential $(\mu_{ci}^g)_1$ for singly

charged positive ions of a phase of the mixture at the points just outside (Eq. 7.238), namely,

$$(\mu_{ci}^{g})_1 = kT \ln \frac{(p_i^g)_1}{(kT)^{5/2}} \frac{h^3}{g_i^g (2\pi m_f)^{3/2}} + \bar{\varepsilon}_{oi} \tag{9.106}$$

where $(p_i^g)_1$ is the partial pressure of the singly charged positive ions of the phase at the points just outside. Combining Eqs. 9.105 and 9.106 we find that the ion-random-flux density $(\Gamma_i)_1$ at the points just outside is given by the relation

$$A_i - \frac{B_i}{kT} = \ln (\Gamma_i)_1 \tag{9.107}$$

where

$$A_i \equiv \ln \Gamma_{io}^s + \ln \frac{2\theta}{1-\theta} + \frac{\theta}{1-\theta} \tag{9.108}$$

$$B_i \equiv \frac{e^2}{16\pi\varepsilon_0 r_1} + \Delta\phi - \Delta\mu \tag{9.109}$$

$$(\Gamma_i)_1 \equiv \frac{(p_i^g)_1}{(2\pi m_f kT)^{1/2}} \tag{9.110}$$

and

$$\Gamma_{i0}^s \equiv \frac{\sigma_f \nu_1 g_i^g}{2g_1^s} \tag{9.111}$$

Comparing Eqs. 9.108 and 9.109 with Eqs. 9.97 and 9.98, respectively, we obtain the relations

$$A_i = A_a - \ln \frac{g_a^g}{g_i^g} + \frac{\alpha_m}{k} \tag{9.112}$$

and

$$B_i = B_a - (\phi_{em}^* - \Delta\phi) + V_i = B_a - \phi_e + \alpha_m T + V_i \tag{9.113}$$

Equations 9.112 and 9.113 indicate that information on the neutral-atom and electron random fluxes, i.e., information on the quantities A_a, B_a, and ϕ_e, also

yields information on the positive-ion random flux. Moreover, we see from the first part of Eq. 9.113 that, in the limit of T equal to 0°K, when B_a is a function of θ only, B_i is also a function of θ only. For this condition B_i represents the *enthalpy of desorption of singly charged adsorbate positive ions* from the surface in the limit of 0°K. Indeed, in this limit B_a and V_i are the enthalpy of desorption and the first ionization energy, respectively, of neutral adsorbate atoms; $-(\phi_{em}^* - \Delta\phi)$ is the work required to return an electron to the surface; and therefore, B_i is the enthalpy of desorption of singly charged adsorbate positive ions.

Finally, combining Eqs. 9.96 and 9.107, we find that the ratio $(\Gamma_i)_1/\Gamma_a$ of the ion-random-flux density $(\Gamma_i)_1$ and the atom-random-flux density Γ_a at the points just outside is given by the relation

$$\frac{(\Gamma_i)_1}{\Gamma_a} = \frac{g_i^g}{g_a^g} \exp\left(\frac{\phi_e - V_i}{kT}\right) \tag{9.114}$$

This relation is called the *Saha–Langmuir equation.*

This completes the derivation of expressions for thermodynamic potentials and emission characteristics consistent with the model assumed in Sec. 9.6.2. For a given uniform electropositive surface in a stable equilibrium state specified by the values of the bare work function ϕ_{em} of the substrate, the ionic radius r_1 of the absorbate, the first ionization energy V_i of the adsorbate, the temperature T, and the vapor pressure p^g, these expressions can be used to calculate all quantities of interest as functions of T and p^g. For example, Eq. 9.96 can be used to determine θ or σ as a function of T and p^g. Thus the neutral-atom random-flux density Γ_a (Eq. 9.99), the electron work function ϕ_e (Eq. 9.101), the electron-random-current density $(e\Gamma_e)_1$ (Eq. 9.102), and the ion-random-flux density (Eq. 9.107) can be expressed as functions of T and p^g.

An approximate relation between the expectation value σ of the adsorbate density per unit substrate surface area and T and p^g can be derived from Eqs. 9.96 and 9.97. Combining the two equations, we obtain the relation

$$kT\left(\frac{B_a}{kT} - \ln\frac{2\theta}{1-\theta} - \frac{\theta}{1-\theta}\right) = kT\ln\frac{\Gamma_{a0}^s}{\Gamma_a}$$

$$= kT\ln\frac{\sigma_f \nu_1 g_a^g}{2\exp(\alpha_m/k)\, g_1^s}\,\frac{(2\pi m_f kT)^{1/2}}{p^g} \tag{9.115}$$

For electropositive surfaces of interest to thermionic conversion in the range of θ between 0.1 and 0.5 and with T a few hundred degrees centered about some practical temperature, the ratio B_a/kT is much greater than unity, whereas the

quantity $\ln [2\theta/(1-\theta)] + \theta/(1-\theta)$ is of the order of unity. In addition, the change of B_a with T is relatively small (Eqs. 9.95 and 9.98). It follows that the left-hand side of Eq. 9.115 is approximately equal to B_a and can be regarded as a function of θ or σ only and, therefore, that for the specified conditions θ or σ is approximately a function of $kT \ln (\Gamma_{a0}^s/\Gamma_a)$ only. Moreover, since $\Delta\phi$ is a function of θ only (Eq. 9.101), it follows that $\Delta\phi$ is approximately a function of $kT \ln (\Gamma_{a0}^s/\Gamma_a)$ only.

For cesium as an adsorbate, another approximate relation between σ and vapor properties can be derived from Eq. 9.115. When the Knudsen effect is negligible (Sec. 8.5.2), the pressure p_{Cs} of the cesium vapor can be expressed in terms of the temperature T_R of the liquid-cesium reservoir by means of the empirical relation (Vol. I, Eq. 4.1)

$$p_{Cs} = A_{Cs} T_R^{-1/2} \exp\left(-\frac{B_{Cs}}{kT_R}\right)$$

$$= 2.45 \times 10^8 \, T_R^{-1/2} \exp\left(-\frac{8910}{T_R}\right) \text{ torr} \quad \text{for } T_R \text{ in } ^\circ\text{K} \tag{9.116}$$

Using Eq. 9.116 for p^g in Eq. 9.115, we obtain the relation

$$kT\left[\frac{B_a}{kT} - \ln\frac{2\theta}{1-\theta} - \frac{\theta}{1-\theta} - \ln\frac{\sigma_f \nu_1 g_a^g}{2 \exp(\alpha_m/k)\, g_1^s} \frac{(2\pi m_f kTT_R)^{1/2}}{A_{Cs}}\right]$$

$$= B_{Cs} \frac{T}{T_R} \tag{9.117}$$

For typical cesiated refractory metal surfaces used as emitters in thermionic converters, namely, for θ between 0.1 and 0.5 and T between 1400 and 2000°K, the left-hand side of Eq. 9.117 can be regarded as a function of only θ or σ because its change with T is small. It follows that σ is approximately a function of T/T_R only. Similarly, $\Delta\phi$ is a function of T/T_R only. The approximate dependence of $\Delta\phi$ of cesiated surfaces on T/T_R has been derived by several investigators.[50,52] A graph of $\Delta\phi$ vs. T/T_R is called a *Rasor plot.* It is noteworthy that the dependence of σ on T/T_R is less accurate than that of σ on $kT \ln (\Gamma_{a0}^s/\Gamma_a)$. The reason is that the left-hand side of Eq. 9.117 is a stronger function of T than the left-hand side of Eq. 9.115.

Universal graphs of $\Delta\phi$ vs. σ (Eq. 9.101b) for cesium, potassium, sodium, and lithium on any uniform refractory metal surface are shown in Fig. 9.20. The only input information needed for each of these graphs is the ionic radius r_1 of the adsorbate. Ionic radii are tabulated.

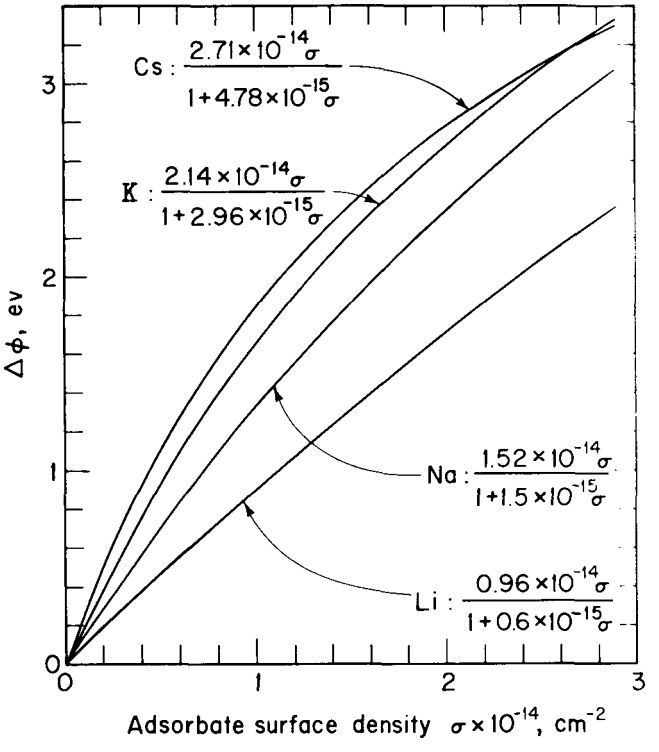


Fig. 9.20 Theoretical results for $\Delta\phi$ vs. σ.

For each electropositive surface, $\Delta\phi$ can also be presented in the form of a graph of $\Delta\phi$ vs. $\theta(\sigma/\sigma_f)$, provided that the surface density σ_f is known. The value of σ_f depends on the crystallographic orientation of the surface of the substrate. Graphs of $\Delta\phi$ vs. θ for three values of σ_f for a cesiated-tungsten surface are shown in Fig. 9.21.

Again, for each electropositive surface, A_a can be presented in the form of a graph of A_a vs. θ (Eq. 9.97), provided that Γ_{a0}^s (Eq. 9.100) is specified. The value of Γ_{a0}^s can be computed given the value of the temperature coefficient α_m of the substrate surface in addition to the value of σ_f. Knowledge of the degeneracies g_a^g and g_1^s in Eq. 9.100 is not necessary because usually it is assumed that $g_a^g = g_1^s$.

Again, for each electropositive surface and each surface temperature, B_a can be presented in the form of a graph of B_a vs. θ or σ (Eqs. 9.98 and 9.95), provided that the value of the ionization energy V_i of the adsorbate, the value of the bare work function ϕ_{em}^* in the limit of 0°K, and the values of r_1 and σ_f are specified.

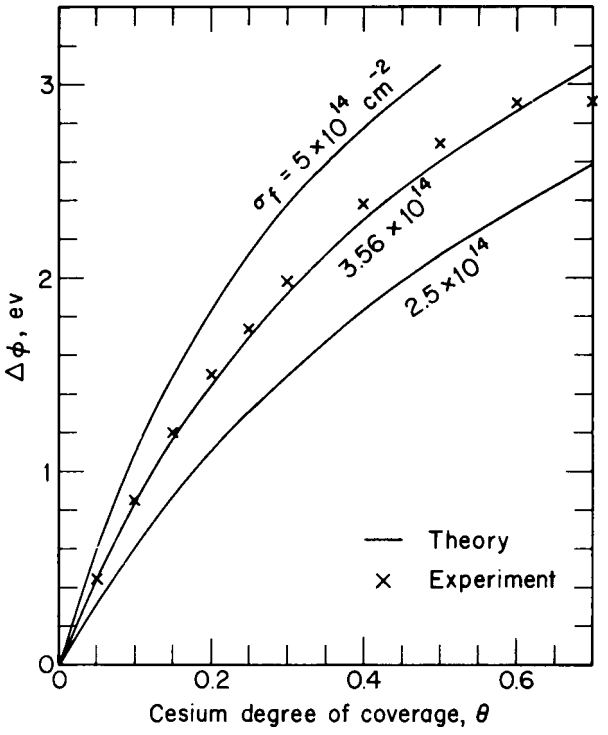


Fig. 9.21 Theoretical and experimental results of $\Delta\phi$ vs. θ for tungsten.

The value of ϕ^*_{em} depends on the crystallographic orientation of the surface of the substrate. Graphs of B_a vs. θ for two values of T for a cesiated-tungsten surface are shown in Fig. 9.22. We see from this figure that the dependence of B_a on T is indeed weak.

Finally, for each electropositive surface and each temperature, eliminating θ from Eqs. 9.101 and 9.115, we find a relation either between $\Delta\phi$ and $kT \ln (\Gamma^s_{a0}/\Gamma_a)$ or between ϕ_e and $kT \ln (\Gamma^s_{a0}/\Gamma_a)$. The two graphs of $\Delta\phi$ vs. $kT \ln (\Gamma^s_{a0}/\Gamma_a)$ (Fig. 9.23) and ϕ_e vs. $kT \ln (\Gamma^s_{a0}/\Gamma_a)$ (Fig. 9.24) are useful because they can be related to readily measurable quantities, as discussed in Sec. 9.6.4.

9.6.4 Equations for Steady States. This section relates the results of Sec. 9.6.3 to experimental results obtained from uniform electropositive surfaces in steady states.

If the emitter surface in Fig. 9.17 is in a steady state at temperature T and degree of coverage θ, we will assume that the electron work function ϕ_e, the

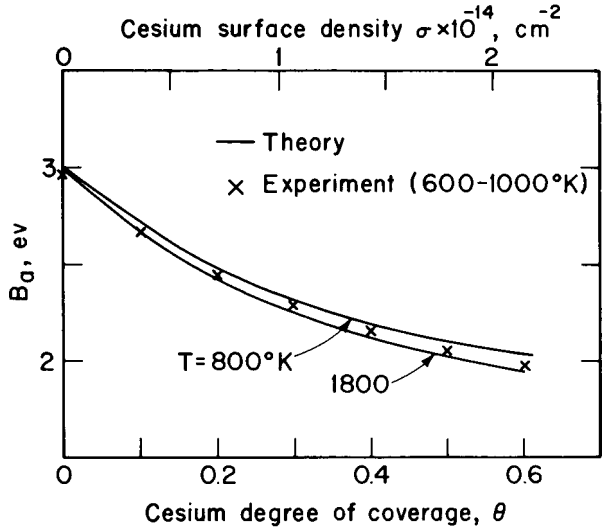


Fig. 9.22 Comparison of theoretical and experimental values of B_a vs. θ. Assumed surface density $\sigma_f = 3.56 \times 10^{14}$ cm^{-2}.

electron-saturation-current density J_S, the adsorbate positive-ion saturation-current density $\tilde{J}_S$, and the adsorbate neutral-atom emission-rate density Γ^s_{va} have the same values as the electron work function and corresponding random-flux densities, respectively, of the surface at the same T and θ but in a state of mutual stable equilibrium with the adsorbate vapor. Thus the evaluation of the various emission properties reduces to that of the degree of coverage θ.

In steady state the degree of coverage θ is determined not only by T and the adsorbate pressure p^g but also by the field conditions prevailing in the vicinity of the points just outside the emitter. It is such that the total emission rate of adsorbate particles (positive ions plus neutral atoms) is equal to the total arrival rate of positive ions plus neutral atoms. The problem has been investigated by Maugis.[5,6]

When the field conditions are such that all adsorbate positive ions are accelerated away from the emitter, the only particles arriving at the emitter surface are neutral atoms. If the neutral-atom arrival-rate density is denoted by $\Gamma^g_{\mu a}$, then θ is determined by the steady-state relation

$$\Gamma^s_{va} + \frac{\tilde{J}_S}{e} = \Gamma^g_{\mu a} \tag{9.118}$$

By virtue of the assumptions made about the emission rates, $\ln \Gamma^s_{va}$ and $\ln (\tilde{J}_S/e)$ are given by the left-hand sides of Eqs. 9.96 and 9.107, respectively. On the

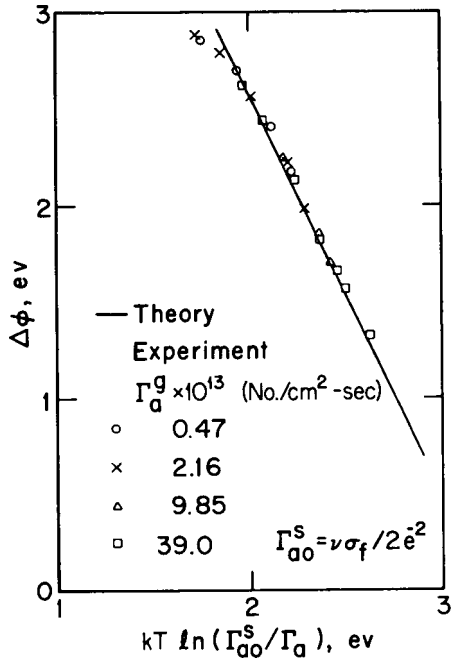


Fig. 9.23 Comparison of theoretical and experimental values of $\Delta\phi$ vs. $kT \ln (\Gamma^s_{ao}/\Gamma_a)$.

other hand, the arrival rate $\Gamma^g_{\mu a}$ will be assumed to be equal to the random-flux density in an adsorbate vapor at pressure p^g and temperature T^g, namely,

$$\Gamma^g_{\mu a} = \Gamma_a \equiv \frac{p^g}{(2\pi m_f k T^g)^{1/2}} \tag{9.119}$$

For given θ and T, Eq. 9.118 indicates that the value of $\Gamma^g_{\mu a}$ must be larger than the neutral-atom random-flux density that would prevail under conditions of mutual stable equilibrium between surface and vapor. Moreover, since $\tilde{J}_S/e\Gamma^s_{\nu a}$ has been assumed to be given by the Saha–Langmuir equation, combining Eqs. 9.114, 9.118, and 9.119, we find that $\tilde{J}_S$ satisfies the relation

$$\tilde{J}_S = \frac{ep^g}{(2\pi m_f k T^g)^{1/2} \{1 + (g^g_a/g^g_i) \exp [(V_i - \phi_e)/kT]\}} \tag{9.120}$$

namely, a relation analogous to that used in Vol. I, Chap. 2, to calculate ion emission from cesiated surfaces.

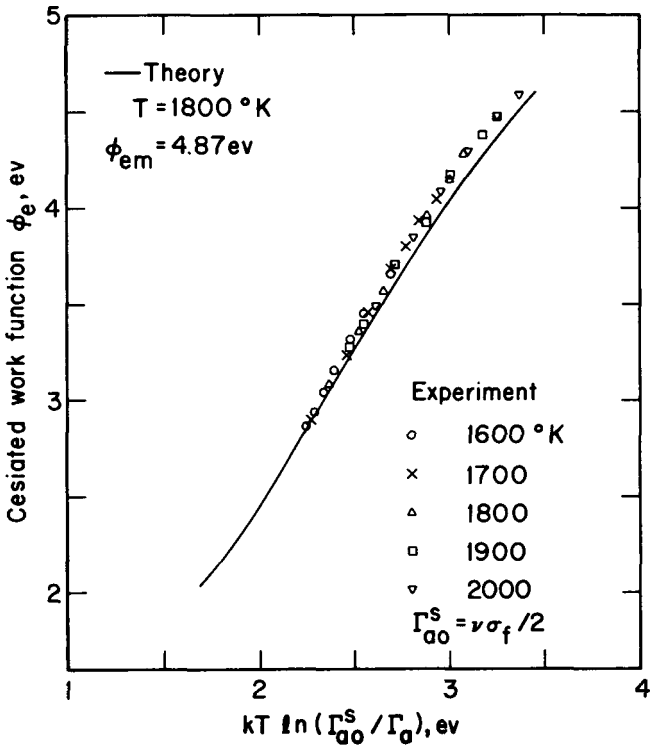


Fig. 9.24 Comparison of theoretical and experimental values of ϕ_e vs. $kT \ln (\Gamma^s_{ao}/\Gamma_a)$.

When the field conditions are such that all electrons are accelerated away from the emitter, the particles arriving at the emitter surface are neutral atoms plus positive ions. If the positive-ion arrival-rate density is denoted by $\Gamma^g_{\mu i}$, then θ is determined by the steady-state relation

$$\Gamma^s_{\nu a} + \frac{\tilde{J}_S}{e} = \Gamma^g_{\mu a} + \Gamma^g_{\mu i} \tag{9.121}$$

For practical adsorbates and adsorbate vapor temperatures, $\Gamma^g_{\mu i}$ can be assumed approximately equal to $\tilde{J}_S/e$, and therefore

$$\Gamma^s_{\nu a} \approx \Gamma^g_{\mu a} \tag{9.122}$$

Since ln $\Gamma^s_{\nu a}$ has been assumed to be given by Eq. 9.96, it follows that Eq. 9.122 can be written in the form

$$A_a - \frac{B_a}{kT} = \ln \Gamma^g_{\mu a} = \ln \frac{p^g}{(2\pi m_f kT^g)^{1/2}} \tag{9.123}$$

where A_a and B_a are given by Eqs. 9.97 and 9.98, respectively. Using Eq. 9.97 in Eq. 9.123, we obtain the relation

$$kT\left(\frac{B_a}{kT} - \ln\frac{2\theta}{1-\theta} - \frac{\theta}{1-\theta}\right) = kT \ln \frac{\Gamma^s_{a0}}{\Gamma^g_{\mu a}}$$
$$= kT \ln \frac{\sigma_f \nu_1}{2 \exp(\alpha_m/k)} \frac{(2\pi m_f kT^g)^{1/2}}{p^g} \tag{9.124}$$

for $g^g_a = g^s_1$. This relation and considerations similar to those discussed in connection with Eq. 9.115 suggest that in steady state θ or σ and therefore $\Delta\phi$ and ϕ_e are functions of $kT \ln (\Gamma^s_{a0}/\Gamma^g_{\mu a})$ only. Accordingly, theoretical results on $\Delta\phi$ and ϕ_e vs. $kT \ln (\Gamma^s_{a0}/\Gamma_a)$ can be directly compared with corresponding experimental data on $\Delta\phi$ and ϕ_e vs. $kT \ln (\Gamma^s_{a0}/\Gamma^g_{\mu a})$ taken over a limited range of T.

9.6.5 Comparing Experimental and Theoretical Results. Langmuir and Taylor[49] made extensive experimental studies of an almost uniform tungsten surface covered by cesium and in the temperature range between 600 and 1000°K. They report the following values for σ_f, ϕ^*_{em}, and ϕ_{em}: $\sigma_f = 3.56 \times 10^{14}$ cm^{-2}, $\phi^*_{em} = 4.76$ ev, and $\phi_{em} = 4.62$ ev for $T = 600$ to 1000°K. The bare work function values ϕ^*_{em} and ϕ_{em} imply an average temperature coefficient α_m equal to $-2k$.

The experimental results of Langmuir and Taylor on $\Delta\phi$ vs. θ, B_a vs. θ, and $\Delta\phi$ vs. $kT \ln (\Gamma^s_{a0}/\Gamma^g_{\mu a})$ are superimposed on Figs. 9.21, 9.22, and 9.23, respectively. We see from these figures that the results are in good agreement with the theoretical predictions corresponding to the input values reported by Langmuir and Taylor in the range of coverage between 0 and about 0.5.

Hatsopoulos and Rufeh[57] reported data on another almost uniform cesiated-tungsten surface for which $\phi_{em} = 4.87$ ev and $T = 1600$ to 2000°K. They did not report values for σ_f and α_m. Their results on ϕ_e vs. $kT \ln (\Gamma^s_{a0}/\Gamma^g_{\mu a})$ are superimposed on Fig. 9.24, in which the value of Γ^s_{a0} (Eq. 9.100) is computed for $\sigma_f = 3.56 \times 10^{14}$ cm^{-2} and $\alpha_m = 0$. The agreement between theory and experiment is satisfactory but is limited to values of coverage up to about 0.5.

The relatively small spread in $\Delta\phi$ or ϕ_e data graphed vs. $kT \ln (\Gamma^s_{a0}/\Gamma^g_{\mu a})$ is achieved only when the range of T is limited to a few hundred degrees. Otherwise the spread can be larger; i.e., $\Delta\phi$ or ϕ_e depends both on $kT \ln (\Gamma^s_{a0}/\Gamma^g_{\mu a})$ and T. The dependence of ϕ_e on T is illustrated by the theoretical

results shown in Fig. 9.25. These results are based on the same values of σ_f, ϕ^*_{em}, and ϕ_{em} as those of the cesiated-tungsten surface investigated by Langmuir and Taylor.[49]

9.6.6 Nonuniform Electropositive Surfaces. As discussed in Sec. 9.4.1, substrate surfaces are usually nonuniform, and therefore electropositive surfaces in practical devices are nonuniform.

In principle, the analysis of nonuniform electropositive surfaces can be approached in a manner analogous to that discussed in Sec. 9.4 depending on the relative sizes of the various uniform parts of the substrate. The results can be fairly complicated, however.

In practice, experimental data on electron emission are often correlated with a unique work function for the entire surface, and the work function is correlated with the ratio T/T_R of the emitter temperature T and the reservoir temperature T_R.

9.7 Electronegative Surfaces and Mixed Surfaces

Electronegative surfaces, i.e., surfaces for which the electronegativity of the free atoms of the adsorbate is larger than the work function of the surface of the

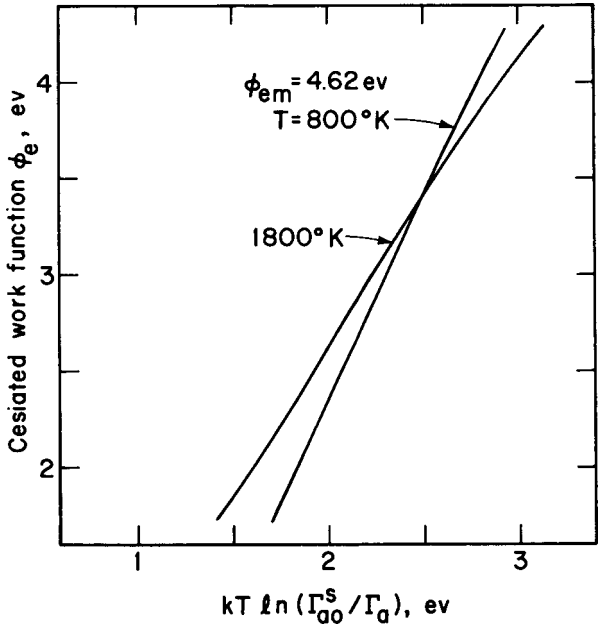


Fig. 9.25 Theoretical results on ϕ_e vs. $kT \ln (\Gamma_{ao}/\Gamma_a)$ for T = 800 and 1800°K.

substrate, have been studied by many researchers. For example, refractory-metal surfaces covered by minute quantities of oxygen have been studied by Engelmaier and Stickney,[58] Alleau and Bacal,[59] and Desplat.[60] An electronegative surface has a work function that is, in general, larger than that of the surface of the substrate.

Mixed surfaces, i.e., surfaces covered by particles of various materials with atomic electronegativities different from each other and different from the work function of the surface of the substrate (both larger and smaller) have also been studied,[61,62] because, when used in thermionic converters, they hold the promise of substantially improving the performance of converters.

Despite many excellent experimental results, the basic phenomena controlling the emission characteristics of electronegative and mixed surfaces are still poorly understood, and much more work needs to be done. For example, in oxygenated-tungsten surfaces, strong evidence exists that surface rearrangement occurs at temperatures typical of those of thermionic emitters. Again, oxygen can interchange position with surface tungsten atoms, and faceting of the original surface may occur. Finally, bulk oxidation is possible. In oxygenated and cesiated surfaces, the adsorbates may appear on the surface in the forms of various oxides.

Because the theories of these surfaces are in a state of evolution and the results obtained so far are not sufficiently accurate for the purposes of thermionic conversion, they will not be discussed.

Appendix 9.A Emission Properties

9.A.1 Calculating Work Functions as Functions of Temperature. Let us consider a finite piece of a metal enclosed by a pure uniform surface and in a state of mutual stable equilibrium with its own vapor. In general, the metal should be regarded as a two-component open system consisting of ν_0 valence electrons per atom and of positive ion cores each with ν_0 units of charge. The vapor can be regarded as a heterogeneous mixture of neutral free atoms of the metal, electrons, and free positive ions.

At temperatures of practical interest, each component of the mixture behaves like a perfect gas. Moreover, to an excellent approximation, the densities of positive ions in the mixture, other than singly charged, are negligibly small. For this reason, only singly charged free positive ions need be considered in the analysis.

In the context of the "one-particle approximation" for both electrons and ions, two motives are necessary to analyze the stable state described; one applies to electrons and the other to singly charged positive ions. For a surface that is a copious emitter of electrons rather than ions, representative graphs of the two motives vs. x are superimposed on Fig. 9.10.

For these graphs each atom in the metal is regarded as consisting of one valence electron and a singly charged positive ion core. This atom is defined so that it is energetically equivalent to an atom with ν_0 valence electrons and an ion core with ν_0 positive units of charge. The equivalence is determined by the requirement that the atom total potential μ_a has the same value for either type of atom in the metal. Details are given later.

When both electron and ion motives are considered, the points just outside are determined by the points at which one of the motives is maximum. To the right of the points just outside (Fig. 9.10), the forces on the charged particles are controlled by space-charge effects only. Moreover, in this region the two motives are equal and opposite.

As discussed in Sec. 9.3.7, for the conditions shown in Fig. 9.10, the electron work function ϕ is given by the relation

$$\phi = -\frac{1}{2}\mu_a - \frac{1}{2}(\zeta^s_{e0} - \zeta^s_{i0}) + V_D + \Delta x \tag{9.A.1}$$

In Eq. 9.A.1 the values of μ_a can be expressed in terms of properties of the mixture by means of the perfect-gas relations applied to an electrically neutral differential phase of the mixture. The values of ζ^s_{e0} can be expressed in terms of quantities obtained from a good one-electron calculation of the ground state of the metal, the volume expansion coefficient of the metal, and the temperature coefficient of the electrochemical potential of the Sommerfeld gas model. Finally, the values of ζ^s_{i0} can be computed by means of the Debye theory of lattice vibrations of the metal.

At the position of the zero-energy reference level (Fig. 9.10), a differential phase of the mixture is electrically neutral, and μ_e, μ_i, and μ_a are given by the perfect-gas relations

$$\mu_e = kT \ln \frac{(p_e)_0}{(kT)^{5/2}} \frac{h^3}{g_e (2\pi m_e)^{3/2}} \tag{9.A.2}$$

$$\mu_i = kT \ln \frac{(p_i)_0}{(kT)^{5/2}} \frac{h^3}{g_i (2\pi m_i)^{3/2}} + \bar{\varepsilon}_{oi} \tag{9.A.3}$$

and

$$\mu_a = kT \ln \frac{(p_a)_0}{(kT)^{5/2}} \frac{h^3}{g_a (2\pi m_a)^{3/2}} + \bar{\varepsilon}_{oa} \tag{9.A.4}$$

where $(p_s)_0$, m_s, g_s, and $\bar{\varepsilon}_{os}$ for $s = e$, i, and a are the partial pressure, mass,

degeneracy, and energy, respectively, of internal structure of electrons (e), singly charged ions (i), and neutral atoms (a);

$$\bar{\varepsilon}_{oi} - \bar{\varepsilon}_{oa} = V_i \tag{9.A.5}$$

and V_i is the first ionization energy of the free atom of the metal. The energy $\bar{\varepsilon}_{oe}$ for the internal structure of the electron is assumed to be equal to zero.

The partial pressure $(p_e)_0$ is equal to $(p_i)_0$. Moreover, for all practical purposes the partial pressure $(p_a)_0$ is equal to the vapor pressure p of the metal surface at temperature T. Accordingly, we find that μ_a is given by the relation

$$\mu_a = kT \ln \frac{p}{(kT)^{5/2}} \frac{h^3}{g_a (2\pi m_a)^{3/2}} - V_i + \bar{\varepsilon}_{oi} \tag{9.A.6}$$

As discussed in Sec. 9.3.7, the values of ζ^s_{e0} are given by the approximate relations

$$\zeta^s_{e0} = \Delta\epsilon - \frac{2}{3}\Delta\epsilon \int_0^T \beta \, dT - \frac{(\pi kT)^2}{12\,\Delta\epsilon} = \Delta\epsilon - \frac{2}{3}\epsilon\bar{\beta}\,\Delta T - \frac{(\pi kT)^2}{12\,\Delta\epsilon} \tag{9.A.7}$$

Finally, the values of ζ^s_{i0} can be computed by means of the Debye theory of lattice vibrations. For example, for temperatures T greater than the Debye temperature θ_D of the metal, ζ^s_{i0} is given by the relation (Sec. 7.12)

$$\zeta^s_{i0} = \bar{\varepsilon}_{oi} + 3kT\left[\ln\frac{\theta_D}{T} - \frac{1}{3} - \frac{3\theta_D}{8T} + \frac{1}{40}\left(\frac{\theta_D}{T}\right)^2\right] - kT \ln g_i^s + pv_i^s$$

$$\text{for } T \gg \theta_D \tag{9.A.8}$$

where g_i^s and v_i^s are the degeneracy and the partial volume of the singly charged ions in the metal. For problems of interest to thermionic conversion, the terms $(\theta_D/T)^2/40$ and pv_i^s can be neglected from Eq. 9.A.8. For temperatures T smaller than θ_D, the expression for ζ^s_{i0} is different from that given by Eq. 9.A.8.

Combining Eqs. 9.A.1, 9.A.6, 9.A.7, and 9.A.8, we obtain Eq. 9.44.

9.A.2 Equivalence of Multivalent and Monovalent Atoms in a Metal. For a multivalent metal with ν_0 valence electrons per atom and with respect to a zero-energy reference level very far from the surface of the metal, the total potential μ_a is given by the relation

$$\mu_a = -\epsilon_e(\nu_0) + \nu_0\, \zeta^s_{e0}(\nu_0) - \epsilon_{i0} + \bar{\varepsilon}_{oi}(\nu_0) + \zeta^s_{i0} \tag{9.A.9}$$

where $-\epsilon_e(\nu_0)$ is the minimum one-electron energy of the metal, $\zeta^s_{e0}(\nu_0)$ is the electron Gibbs free energy with respect to $-\epsilon_e(\nu_0)$, $-\epsilon_{i0}$ is the minimum ion energy, $\bar{\varepsilon}_{oi}(\nu_0)$ is the energy associated with the internal structure of the multiply charged ions in the metal, and ζ^s_{i0} is the ion Gibbs free energy with respect to $-\epsilon_{i0}$. It is noteworthy that the quantities $-\epsilon_e(\nu_0)$, $\zeta^s_{e0}(\nu_0)$, and $\bar{\varepsilon}_{oi}(\nu_0)$ are different from the quantities $-\epsilon_{e0}$, ζ^s_{e0}, and $\bar{\varepsilon}_{oi}$ defined in Sec. 9.3.7 because of the different number of valence electrons.

Let $\zeta^s_{e0}(\nu_0 - 1)$ be the one-electron energy for which the number of electrons in the metal in one-electron-energy levels between $\zeta^s_{e0}(\nu_0 - 1)$ and infinity is equal to unity. The energy $\zeta^s_{e0}(\nu_0 - 1)$ is a function of temperature.

Now, in terms of $\zeta^s_{e0}(\nu_0 - 1)$, Eq. 9.A.9 can be expressed in the forms given by the relations

$$\begin{aligned}\mu_a &= -\epsilon_e(\nu_0) + \zeta^s_{e0}(\nu_0 - 1) + \zeta^s_{e0} - \epsilon_{i0} + \bar{\varepsilon}_{oi}(\nu_0) + (\nu_0 - 1)\,\zeta^s_{e0}(\nu_0 - 1) + \zeta^s_{i0} \\ &= -\epsilon_{e0} + \zeta^s_{e0} - \epsilon_{i0} + \bar{\varepsilon}_{oi} + \zeta^s_{i0}\end{aligned} \tag{9.A.10}$$

where

$$-\epsilon_{e0} \equiv -\epsilon_e(\nu_0) + \zeta^s_{e0}(\nu_0 - 1) \tag{9.A.11}$$

$$\zeta^s_{e0} \equiv \nu_0\, [\zeta^s_{e0}(\nu_0) - \zeta^s_{e0}(\nu_0 - 1)] \tag{9.A.12}$$

and

$$\bar{\varepsilon}_{oi} \equiv \bar{\varepsilon}_{oi}(\nu_0) + (\nu_0 - 1)\,\zeta^s_{e0}(\nu_0 - 1) \tag{9.A.13}$$

From the preceding definitions, it follows that $-\epsilon_{e0}$ can be interpreted as the minimum one-electron energy of a monovalent metal, ζ^s_{e0} can be interpreted as the electron electrochemical potential of a monovalent metal, and $\bar{\varepsilon}_{oi}$ can be interpreted as the energy associated with the internal structure of a singly charged ion core in the metal. Accordingly, the second part of Eq. 9.A.10 can also be regarded as the expression for the total potential of a monovalent metal that is energetically equivalent to the metal with ν_0 valence electrons per atom.

9.A.3 Comparing Various Emission Rates from Refractory Metal Surfaces. Equation 9.44 can be rewritten in the form

$$\ln \frac{(\Gamma_e)_1}{\Gamma_a} = \frac{\phi - V_i + \zeta_{e0}^g - 2(V_D + \Delta x)}{kT} - \ln \frac{g_a m_a}{g_i^s g_e m_e} + 3\left[\ln \frac{T}{\theta_D} + \frac{1}{3} + \frac{3\theta_D}{8T}\right] \quad (9.A.14)$$

where $(\Gamma_e)_1$ is the *electron-random-flux density* just outside the surface. For refractory metals, substituting typical values in the right-hand side of Eq. 9.A.14, we find that the ratio $(\Gamma_e)_1/\Gamma_a$ is much greater than unity. For this reason neutral atoms can be neglected in some analyses of refractory metal surfaces. Similar remarks can be made about the ratio $(\Gamma_e)_1/(\Gamma_i)_1$, where $(\Gamma_i)_1$ is the ion-random-flux density just outside the surface.

The magnitude of the ratio $(\Gamma_e)_1/\Gamma_a$ can also be estimated by the relation

$$\frac{(\Gamma_e)_1}{\Gamma_a} = \frac{AT^2 \exp(-\phi/kT)}{ep/(2\pi m_a kT)^{1/2}} \quad (9.A.15)$$

and the vapor pressure data given in Fig. 9.12.

Appendix 9.B Calculating Enthalpy of Desorption

In the limit of 0°K, the enthalpy of desorption λ_a of neutral adsorbate atoms is defined by the relation

$$\lambda_a = (h_a^g - h_a^s) \qquad \text{for } T = 0 \quad (9.B.1)$$

where h_a^g and h_a^s denote the enthalpies of the atoms in the vapor and on the surface, respectively.

Since the vapor can be regarded as a perfect gas, from the discussion in Sec. 7.8.13, it follows that h_a^g is given by the relation

$$h_a^g = \frac{5}{2}kT + \bar{\varepsilon}_{oa} \quad (9.B.2)$$

The enthalpy h_a^s can be evaluated as follows: Adapting Eq. 7.150 to a surface yields that h_a^s is given by the relation

$$h_a^s = -T^2 \left(\frac{\partial(\mu_a^s/T)}{\partial T}\right)_\tau \tag{9.B.3}$$

Combining Eqs. 9.71, 9.72, and 9.96, we find that the ratio μ_a^s/T is given by the relation

$$\frac{\mu_a^s}{T} = kA_a - \frac{B_a}{T} + k \ln \frac{h^3}{2\pi m_f (kT)^2 g_a^g} + \frac{\bar{\varepsilon}_{oa}}{T} \tag{9.B.4}$$

In this relation μ_a^s can be regarded as a function of either T and τ or T and σ. Accordingly,

$$d\left(\frac{\mu_a^s}{T}\right) = \left[\frac{\partial(\mu_a^s/T)}{\partial T}\right]_\tau dT + \left[\frac{\partial(\mu_a^s/T)}{\partial \tau}\right]_T d\tau \tag{9.B.5a}$$

or

$$\left[\frac{\partial(\mu_a^s/T)}{\partial T}\right]_\sigma = \left[\frac{\partial(\mu_a^s/T)}{\partial T}\right]_\tau + \left[\frac{\partial(\mu_a^s/T)}{\partial \tau}\right]_T \left(\frac{\partial \tau}{\partial T}\right)_\sigma \tag{9.B.5b}$$

In addition, by virtue of

$$\left(\frac{\partial \mu_a^s}{\partial \tau}\right)_T = \frac{1}{\sigma} \tag{9.B.6}$$

and of Eqs. 9.84, 9.86, 9.87, 9.91, and 9.93,

$$\left(\frac{\partial \tau}{\partial T}\right)_\sigma = \frac{\sigma k}{1-\theta} + \frac{(1-\theta)\tau_{1v}}{3T} \tag{9.B.7}$$

Combining Eqs. 9.97, 9.98, and 9.B.3 with 9.B.7, we find the relation

$$h_a^s = -B_a + \bar{\varepsilon}_{oa} + kT\left(\frac{\partial B_a}{k\,\partial T} + 2 - \frac{1}{1-\theta}\right) + \frac{(1-\theta)\tau_{1v}}{3\sigma} \tag{9.B.8}$$

The enthalpy of desorption h_a^{sg} at any temperature is given by the relation

$$h_a^{sg} \equiv h_a^g - h_a^s = B_a + kT\left(-\frac{\partial B_a}{k\,\partial T} + \frac{1}{2} + \frac{1}{1-\theta}\right) - \frac{(1-\theta)\tau_{1v}}{3\sigma} \tag{9.B.9}$$

and, therefore, in the limit of 0°K,

$$\lambda_a = h_a^{sg}(T=0) = B_a \tag{9.B.10}$$

since both $T(\partial B_a/\partial T)$ and $\tau_{1\nu}$ become zero for T equal to zero.

The enthalpy of desorption is frequently called the heat of desorption because it is equal to the heat that must be added to the surface per atom desorbed to maintain the surface at constant temperature. It follows that λ_a is the heat of desorption for $T \to 0$.

The energy of desorption e_a^{sg} is related to the enthalpy of desorption by the expression

$$e_a^{sg} = h_a^{sg} + \frac{\tau}{\sigma} - p^g v^g = h_a^{sg} + \frac{\tau}{\sigma} - kT \tag{9.B.11}$$

which, in the limit of 0°K, becomes

$$e_a^{sg}(T=0) = \lambda_a + \frac{\tau_1(T=0)}{\sigma} \tag{9.B.12a}$$

or, by virtue of Eqs. 9.87, 9.91, and 9.93,

$$e_a^{sg}(T=0) = \lambda_a + D_a M_1^2 \sigma^{3/2} \tag{9.B.12b}$$

We see from Eq. 9.B.12b that, for very dilute surfaces ($\sigma \ll 1$ or $\theta \ll 1$), $e_a^{sg}(T=0)$ approaches λ_a but for nondilute surfaces $e_a^{sg}(T=0)$ and λ_a differ by a finite amount.

References

1. C. Herring and M. H. Nichols, Thermionic Emission, *Rev. Mod. Phys.*, 21: 185-270 (1949).
2. J. C. Slater, A Generalized Self-Consistent Field Method, *Phys. Rev.*, 91: 528-530 (1953).
3. E. P. Wigner and F. Seitz, On the Constitution of Metallic Sodium. II, *Phys. Rev.*, 46: 509-524 (1934); also E. P. Wigner, On the Interaction of Electrons in Metals, *Phys. Rev.*, 46: 1002-1011 (1934).
4. J. Bardeen, Theory of the Work Function. II. The Surface Double Layer, *Phys. Rev.*, 49: 653-663 (1936).
5. H. J. Juretschke, Exchange Potential in the Surface Region of a Free-Electron Metal, *Phys. Rev.*, 92: 1140-1144 (1953).

6. T. L. Loucks and P. H. Cutler, The Effect of Correlation on the Surface Potential of a Free Electron Metal, *J. Phys. Chem. Solids,* 25: 105-113 (1964).
7. J. Gadzuk, Many-Body Theory of a Rapidly Varying Inhomogeneous Gas, Report No. NASA-PM-67, National Aeronautics and Space Administration, 1967.
8. J. C. Slater, *Insulators, Semiconductors and Solids,* Vol. 3 of *Quantum Theory of Molecules and Solids,* p. 263, McGraw-Hill Book Company, New York, 1967.
9. J. R. Smith, Self-Consistent Many-Electron Theory of Electron Work Functions and Surface Potential Characteristics for Selected Metals, *Phys. Rev.,* 181: 522-529 (1969).
10. M. F. Koskinen, A Comparison of Two Theories of Cesium Adsorption, in *Proceedings of the First International Conference on Thermionic Electrical Power Generation, London, Sept. 20–25, 1965,* Paper 6-7, 1965.
11. C. Herring and M. H. Nichols, Thermionic Emission, *Rev. Mod. Phys.,* 21: Appendix III (1949).
12. P. W. Bridgman, *The Thermodynamics of Electrical Phenomena in Metals and a Condensed Collection of Thermodynamic Formulas,* Dover Publications, Inc., New York, 1960.
13. J. F. Kenney, Energy Bands of the Alkali Metals. I. Lithium and Sodium, in Quarterly Progress Report of the Solid State and Molecular Theory Group, Massachusetts Institute of Technology, p. 53, 1964.
14. J. F. Kenney, Energy Bands of the Alkali Metals. II. Rubidium and Cesium, Quarterly Progress Report of the Solid State and Molecular Theory Group, Massachusetts Institute of Technology, p. 66, 1967.
15. I. F. Mattheis, Electronic Structure of Nb and Ta, *Phys. Rev., A,* 1: 373-380 (1970).
16. I. F. Mattheis, Fermi Surface in Tungsten, *Phys. Rev.,* 139(6A): 1893-1904 (1965).
17. I. F. Mattheis, Bond Structure and Fermi Surface for Rhenium, *Phys. Rev.,* 151: 450-464 (1966).
18. E. Wigner and J. Bardeen, Theory of the Work Functions of Monovalent Metals, *Phys. Rev.,* 48: 84-87 (1935).
19. F. Seitz, *The Modern Theory of Solids,* McGraw-Hill Book Company, New York, 1940.
20. E. P. Wigner, On the Interaction of Electrons in Metals, *Phys. Rev.,* 46: 1002-1011 (1934); also Effects of the Electron Interaction on the Energy Levels of Electrons in Metals, *Trans. Faraday Soc.,* 34: 678-685 (1938).
21. J. C. Slater, *Insulators, Semiconductors and Solids,* Vol. 3 of *Quantum Theory of Molecules and Solids,* p. 247, McGraw-Hill Book Company, New York, 1967.
22. J. Bardeen, Electron Correlation and Screening Effects in Relation to Surface Physics, *Surface Sci.,* 2: 381-388 (1964).
23. G. N. Hatsopoulos and E. P. Gyftopoulos, On the Relation Between Work Function and Other Thermophysical Properties, in *Proceedings of Thermionic Conversion Specialists Conference, Miami Beach, Fla., Oct. 26–29, 1970,* pp. 203-209, 1970.
24. R. E. Honig, *Vapor-Pressure Curves of the Elements,* Radio Corporation of America, 1962.
25. D. Gray (Ed.), *American Institute of Physics Handbook,* 2nd ed., McGraw-Hill Book Company, New York, 1963.
26. V. S. Fomenko, *Handbook of Thermionic Properties,* G. V. Samsonov (Ed. and Tr.), Plenum Publishing Corporation, New York, 1966.
27. R. Smoluchowski, Anisotropy of the Electronic Work Function of Metals, *Phys. Rev.,* 60: 661-674 (1941).
28. D. Steiner and E. P. Gyftopoulos, An Equation for the Prediction of Bare Work Functions, in *Proceedings of the 27th Annual Conference on Physical Electronics, Cambridge, Mass., Mar. 20–22, 1967,* pp. 160-168, 1967; also E. P. Gyftopoulos,

Comments on Work Function Theories, in *Proceedings of the Second International Conference on Thermionic Electrical Power Generation, Stressa, Italy, May 27–31, 1968,* pp. 1225-1248, 1968.
29. G. V. Samsonov, Y. B. Paderno, and V. S. Fomenko, Thermionic Emission Characteristics of the Transition Metals and Their Compounds, *Sov. Phys.–Tech. Phys. (Engl. Transl.),* 11: 1070-1080 (1967).
30. N. S. Rasor and C. Warner, Correlation of Electron, Ion, and Atom Emission Energies, Report No. AI-6799, Atomics International, November 1961.
31. C. V. Bocciarelli, An Empirical Relationship Between Fermi Level and Electron Affinity in Metals, *Solid State Commun.,* 5: 821-823 (1967).
32. W. Gordy and W. Thomas, Electronegativities of the Elements, *J. Chem. Phys.,* 24: 439-444 (1956).
33. W. Sachtler, Halbempirische Methode zur Berechrung des Elektronen Austrittpotentials von Metallen, *Z. Elektrochem.,* 59: 119-122 (1955).
34. W. H. Adams, On the Solution of the Hartree–Fock Equation in Terms of Localized Orbitals, *J. Chem. Phys.,* 14: 89-102 (1961).
35. C. Edmiston and K. Ruedenberg, Localized Atomic and Molecular Orbitals, *Rev. Mod. Phys.,* 35: 457-465 (1963).
36. E. P. Gyftopoulos and G. N. Hatsopoulos, Quantum Thermodynamic Meaning of Electronegativity and Work Function, in *Proceedings of the Second International Conference on Thermionic Electrical Power Generation, Stressa, Italy, May 27–31, 1968,* pp. 1249-1260, 1968.
37. W. E. Gordy, A New Method of Determining Electronegativity from Other Atomic Properties, *Phys. Rev.,* 69: 604-607 (1946).
38. H. O. Pritchard and H. A. Skinner, The Concept of Electronegativity, *Chem. Rev.,* 55: 745-786 (1955).
39. J. C. Slater, *Symmetry and Energy Bands in Crystals,* Vol. 2 of *Quantum Theory of Molecules and Solids,* Chap. 4, McGraw-Hill Book Company, New York, 1965.
40. L. Pauling, *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry,* 3rd ed., pp. 393-448, Cornell University Press, Ithaca, N. Y., 1960.
41. G. F. Smith, Thermionic and Surface Properties of Tungsten Crystals, *Phys. Rev.,* 94: 295-308 (1954).
42. E. D. Protopopov et al., Emission Parameters of Tantalum and Molybdenum Single Crystals, *Sov. Phys.–Solid State (Engl. Transl.),* 8: 909-914 (1966).
43. U. V. Azizov and G. N. Shuppe, Emission and Adsorption Characteristics of Faces of a Tungsten Single Crystal, *Sov. Phys.–Solid State (Engl. Transl.),* 7: 1591-1594 (1966).
44. U. V. Azizov et al., Emission Properties of a Molybdenum Single Crystal, *Sov. Phys.–Solid State (Engl. Transl.),* 7: 2232-2234 (1966).
45. E. M. Savitskii et al., Anisotropic Work Function of Molybdenum Single Crystals, *Sov. Phys.–Tech. Phys. (Engl. Transl.),* 11: 974-976 (1967).
46. R. Wichner and T. H. Pigford, Work Function of Monocrystalline and Polycrystalline Rhenium, *Proceedings of Thermionic Conversion Specialists Conference, Houston, Tex., Nov. 2–4, 1966,* pp. 405-412, 1966.
47. T. A. Callcott and A. V. MacRae, Photoemission from Clean and Cesium-Covered Nickel Surfaces, *Phys. Rev.,* 178: 966-978 (1969).
48. B. J. Hopkins, T. J. Lee, and C. B. Williams, Temperature Coefficients of the Work Functions of the (1 1 0) and (1 0 0) Surfaces of Tungsten Single Crystals and of Polycrystalline Tungsten Foil, *J. Appl. Phys.,* 40: 1728-1732 (1969).
49. I. Langmuir and J. B. Taylor, The Evaporation of Atoms, Ions, and Electrons from

Cesium Films on Tungsten, *Phys. Rev.*, 44: 423-458 (1933).
50. N. S. Rasor and C. Warner, Correlation of Emission Processes for Adsorbed Alkali Films on Metal Surfaces, *J. Appl. Phys.*, 35: 2589-2600 (1964).
51. E. P. Gyftopoulos and J. D. Levine, Work Function Variation of Metals Coated by Metallic Films, *J. Appl. Phys.*, 33: 67-73 (1962).
52. J. D. Levine and E. P. Gyftopoulos, Adsorption Physics of Metals Partially Covered by Metallic Particles. I, *Surface Sci.*, 1: 171-193 (1964); II, *Surface Sci.*, 1: 225-241 (1964); and III, *Surface Sci.*, 1: 349-360 (1964).
53. K. P. Luke, Thermionic Work Function of a Single Crystal, Sc. D. Thesis, Massachusetts Institute of Technology, Cambridge, 1964.
54. I. Langmuir, Vapor Pressures, Evaporation, Condensation, and Adsorption, *J. Amer. Chem. Soc.*, 54: 2798-2832 (1932).
55. J. R. MacDonald and C. A. Barlow, Jr., Theory of Work Function Change on Adsorption of Polarizable Films, *J. Chem. Phys.*, 44: 202-206 (1966).
56. D. Maugis, Étude thermodynamique de la fonction metal-plasma thermique. Contribution a l'étude de l'ionization de surface, *Ann. Radioélect.*, 23(93): 211 (1968).
57. G. N. Hatsopoulos and F. Rufeh, Thermodynamic Correlation of Work Function, *Energy Convers.*, 10: 135 (1970).
58. W. Engelmaier and R. E. Stickney, Thermionic and Adsorption Characteristics of a Single Crystal Tungsten Filament Exposed to Oxygen, *Surface Sci.*, 11: 370 (1968).
59. T. Alleau, and M. Bacal, Influence of Oxygen Adsorption and Oxygen–Cesium Co-adsorption on (1 0 0) Tungsten Work Function, in *Proceedings of Thermionic Conversion Specialists Conference, Miami Beach, Fla., Oct. 26–29, 1970*, pp. 434-440.
60. J. L. Desplat, W (1 0 0) Work Function Change During Adsorption of Oxygen, Cesium, and Oxygen–Cesium Co-adsorption, in *Proceedings of the Third International Conference on Thermionic Electrical Power Generation, Juelich, Germany, June 5–9, 1972*, pp. 1397-1407.
61. L. K. Hansen and C. Warner, Basic Research in Thermionic Energy Conversion, Final Technical Summary Report, Dec. 1, 1971–Nov. 30, 1972, USAEC Report AD-755955, Atomics International, January 1973.
62. B. Gunther, D. Lieb, C. Richmond, and F. Rufeh, Summary Report on Oxygen Additives, Report TEE-3056-2, Thermo Electron Corp., May 1974.

10

Collisionless Transport Phenomena

10.1 General Remarks

This chapter presents various analyses of collisionless transport phenomena, i.e., analyses of transport of charged particles through the interelectrode space of converters under the influence of long-range electrostatic forces and in the absence of short-range collisional forces.

As stated in Vol. I, Chaps. 3 and 4, collisionless transport phenomena are predominant throughout the interelectrode space of vacuum and low-pressure converters and in the emitter and collector sheaths of high-pressure converters.

The analysis of these phenomena consists essentially in finding the electron motive or electrostatic potential energy as a self-consistent solution of Poisson's equation. Solving this equation requires knowledge of the charged-particle densities. The densities in turn depend on both the conditions imposed by the boundaries of the region in which the transport of particles is collisionless and the values of the motive. The dependence of the densities on the values of the motive imposes the requirement of self-consistency.

Different solutions of Poisson's equation are found depending on the assumptions made about the boundary conditions, namely, the sources and sinks of charged particles, and on the approximations introduced in the equations describing a particular problem. Each of these solutions is qualified as a *space-charge theory.*

The first space-charge theory was developed by Langmuir[1] in the early part of the century for electrons moving between two infinite, parallel electrodes in vacuum. In this theory the electrons are assumed to be emitted from one of the electrodes with negligible velocity and to be collected at the other. Child[2] investigated a similar problem in which, however, the charged particles were heavy positive ions.

Later Langmuir and others modified the analysis by assuming that the electrons are emitted with a half-Maxwellian distribution of velocities or with constant velocity different from zero and normal to the electrode surfaces. Finally, space-charge theories involving both electrons and positive ions have been developed by Langmuir and his co-workers as well as other investigators.

The various space-charge theories relevant to thermionic conversion are discussed in order of increasing numerical complexity. The discussions are restricted to collisionless transport phenomena occurring between parallel-plane

electrodes. For space-charge theories applicable to other electrode geometries, the reader is referred to the literature.[3]

An interesting feature of some of the self-consistent solutions of Poisson's equation is that they are unstable with respect to small perturbations. For example, a certain interelectrode motive in a low-pressure cesium converter can be altered substantially by a small localized perturbation in charge densities. Some solutions can stabilize into a time-oscillatory relation between the output current and the voltage across a resistive load connected to the electrodes. Oscillatory current–voltage relations have been observed experimentally but have not proved practical.

10.2 The Child–Langmuir Theory

In his first space-charge theory, Langmuir derived the motive or electrostatic potential energy distribution in the space between two infinite parallel-plane electrodes shown schematically in Fig. 10.1. Electrode E is a thermionic emitter of electrons only, and electrode C does not emit any particles. The motive difference between the points just outside E and the points just outside C is negative and is maintained by means of an externally applied voltage supply. The current density flowing from E to C through the interelectrode space is J. The interelectrode space is evacuated except for the electrons that are in transit from E to C.

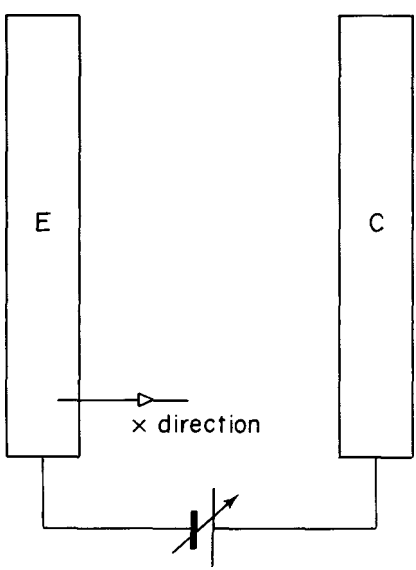


Fig. 10.1 Parallel-plane electrode arrangement.

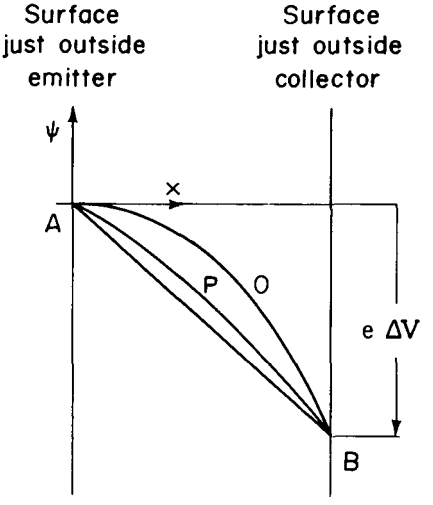


Fig. 10.2 Parabolic motive distributions (*APB* and *AOB*) due to constant-velocity electron space charge.

When the temperature of E is sufficiently small, very few electrons are emitted and J is very small. These electrons have practically no effect on the interelectrode motive, and therefore the motive is a linear function of position, as shown by line AB in Fig. 10.2.

As the temperature of E is raised, J increases and the number of electrons in transit in the interelectrode space is appreciable. These electrons affect both the motive distribution and the current flow in the interelectrode space.

An estimate of the effect of the electron space charge on the motive can be obtained as follows: The electron speed v_{ex} in the x direction (Fig. 10.1) depends on the motive. Nevertheless, as a zero-order approximation, the magnitude v_e of the speed can be assumed to be constant. Thus the electron density n_e throughout the interelectrode space becomes independent of position and is given by the relation

$$n_e = \frac{J}{ev_e} \tag{10.1}$$

Moreover, since the current flow is one dimensional, Poisson's equation for the motive ψ is given by the relation

$$\frac{d^2\psi}{dx^2} = -\frac{eJ}{\varepsilon_0 v_e} \tag{10.2}$$

The general solution of Eq. 10.2 is a parabolic function of position. When the origin of the energy and position coordinates is taken at the points just outside the emitter, ψ is given by the relation

$$\psi = -\frac{1}{2}\frac{eJ}{\varepsilon_0 v_e} x^2 + \left(\frac{d\psi}{dx}\right)_E x \tag{10.3}$$

where $(d\psi/dx)_E$ is the gradient of the motive just outside the emitter E. The value of $(d\psi/dx)_E$ depends on the values of J, of the interelectrode spacing d, and of the motive difference $-e\,\Delta V$ between the points just outside E and C. For given values of J, d, and ΔV, $(d\psi/dx)_E$ is larger than the slope of the line AB in Fig. 10.2 because of the presence of electrons in transit in the interelectrode space.

For a relatively small value of J, $(d\psi/dx)_E$ is negative, and Eq. 10.3 can be represented by the parabola APB shown in Fig. 10.2. Under this condition all the electron current emitted from E is accelerated toward C since the gradient of the motive is negative everywhere.

If the value of J is increased above the value corresponding to the interelectrode motive APB (i.e., if the temperature of E is increased), $(d\psi/dx)_E$ increases. It follows that a value J_0 of the current density exists for which

$$\left(\frac{d\psi}{dx}\right)_E = 0 \tag{10.4}$$

and therefore

$$\psi = -\frac{1}{2}\frac{eJ_0}{\varepsilon_0 v_e} x^2 \tag{10.5}$$

For the current density J_0, the interelectrode motive is represented by the parabola AOB in Fig. 10.2, and, for x equal to d, Eq. 10.5 yields the relation

$$J_0 = \frac{2\varepsilon_0 v_e}{d^2}(-\Delta V) \tag{10.6}$$

For fixed values of ΔV (<0) and d, Eq. 10.6 gives an estimate of the maximum current density that can flow between E and C. If the emitted current density were larger than J_0, $(d\psi/dx)_E$ would be positive, and therefore an electron decelerating field would exist near the surface just outside electrode E. Such a field would reduce the current density from E to C to a value lower than the emitted current density but equal to J_0.

This approximate analysis means that the current flow between E and C is ultimately limited by the electron space charge. The same conclusion is reached if the analysis is carried out without the simplifying approximation that v_e is constant.

For relatively small values of J, the speed v_e of the electrons at each position in the interelectrode space is given by the relation

$$\frac{1}{2} m_e v_e^2 \approx -\psi \tag{10.7}$$

provided that the kinetic energy of the electrons at the points just outside E is much smaller than $-\psi$. Thus, for a given current density J, the electron number density n_e and Poisson's equation for the motive ψ become

$$n_e = \frac{J}{ev_e} = \frac{J}{e(-2\psi/m_e)^{1/2}} \tag{10.8}$$

and

$$\frac{d^2\psi}{dx^2} = -\frac{eJ}{\varepsilon_0(-2\psi/m_e)^{1/2}} \tag{10.9}$$

Multiplying Eq. 10.9 by $2\, d\psi/dx$ and integrating the result from 0 to x yield

$$\left(\frac{d\psi}{dx}\right)^2 - \left(\frac{d\psi}{dx}\right)_E^2 = \frac{2eJ}{\varepsilon_0}(-2m_e\psi)^{1/2} \tag{10.10}$$

As already discussed, the value of $(d\psi/dx)_E$ depends on the values of J, d, and $-e\,\Delta V$. For given values of d and ΔV and for a relatively small value of J, $(d\psi/dx)_E$ is negative. Then, by virtue of Eq. 10.9, $d\psi/dx$ must be negative for all values of x. Under these conditions all the current emitted thermionically from E reaches C.

On the other hand, a value J_0 of the current density exists for which $(d\psi/dx)_E = 0$ although $d\psi/dx$ is negative throughout the interelectrode space. For this value Eq. 10.10 becomes

$$\frac{d\psi}{dx} = -\left(\frac{2eJ_0}{\varepsilon_0}\right)^{1/2}(-2m_e\psi)^{1/4} \tag{10.11}$$

Integrating Eq. 10.11 from x equal to zero to x equal to d yields the relation

$$J_0 = \frac{4\sqrt{2}}{9}\varepsilon_0\left(\frac{e}{m_e}\right)^{1/2}\frac{(-\Delta V)^{3/2}}{d^2}$$

$$= 2.330 \times 10^{-6}\,\frac{|\Delta V|^{3/2}}{d^2}\ \text{amp/cm}^2 \qquad \text{for } \Delta V \text{ in volts and } d \text{ in cm} \qquad (10.12)$$

For given values of d and ΔV (<0), Eq. 10.12 gives another estimate of the maximum current density that can flow between E and C. The reason for this conclusion is similar to that discussed in connection with Eq. 10.6.

Equation 10.12 is the approximate *Child–Langmuir equation* for electron-space-charge-limited current flow between two infinite parallel-plane electrodes in vacuum. It is called the Child–Langmuir equation because Langmuir derived it for electrons and Child derived it for ions. Of course, for ion currents the mass m_i of the ions must be used, and therefore the numerical coefficient in the second part of Eq. 10.12 will be $2.330 \times 10^{-6}\,(m_e/m_i)^{1/2}$.

Equation 10.12 is also called the three-halves power relation between current and voltage. When the electrode work functions are approximately equal, the internal voltage drop ΔV is approximately equal to the applied voltage V and the current between the two parallel-plane electrodes varies in proportion to the three-halves power of the voltage.* This proportionality holds when the current density is not temperature limited, i.e., when J_0 is smaller than the saturation-current density and when the initial velocity of either electrons or positive ions at the points just outside the emitter is negligible compared with that produced by the acceleration of the particles in the interelectrode space. On the other hand, when the initial velocity of the thermionically emitted particles is not negligible, Eq. 10.12 does not hold. A space-charge theory that includes initial electron velocities is discussed in Sec. 10.3.

10.3 The Langmuir Theory

10.3.1 Interelectrode Motive. The Child–Langmuir equation is a poor approximation for the current vs. internal-voltage-drop relation in a thermionic converter. For example, when ΔV equals zero, Eq. 10.12 yields that J_0 is equal to zero for all converters, a result that is not valid. For this reason the interelectrode motive between two infinite parallel-plane electrodes, such as those shown in Fig. 10.1, is derived here without neglecting the distribution of initial velocities of the thermionically emitted electrons. This distribution is a half-Maxwellian in the positive x direction. The results of the derivation are

*It can be easily shown by means of simple electrostatic arguments that the approximate three-halves power relation holds for any electrode geometry.

applicable to vacuum diodes operated in the space-charge-limited mode and for which the electron flow is approximately one dimensional and collisionless and for which back emission is negligible. The derivation was first obtained by Adams but was published by Langmuir.[4] Their method is followed here with some minor modifications.

The shape of the interelectrode motive of a vacuum diode which satisfies the conditions stated is shown in Fig. 10.3. This motive is obtained as a solution of the equation

$$\frac{d^2\psi}{dx^2} = -\frac{e^2 n_e}{\varepsilon_0} \tag{10.13}$$

The electron number density n_e at each position x (at a plane surface at a distance x from the surface just outside the emitter) is given by the integral of the electron distribution function $f_e(x,v_e)$ over all velocities. The distribution function, in turn, is determined by the values of the motive ψ. To solve Eq. 10.13, we must first express $f_e(x,v_e)$ in terms of ψ and then n_e in terms of $f_e(x,v_e)$.

We see from Fig. 10.3 that thermionically emitted electrons with a total energy ϵ_x associated with the velocity v_{ex} in the positive x direction will be returned to the emitter when ϵ_x is smaller than the maximum motive ψ_m. It follows that, in the space between the surface just outside the emitter and the

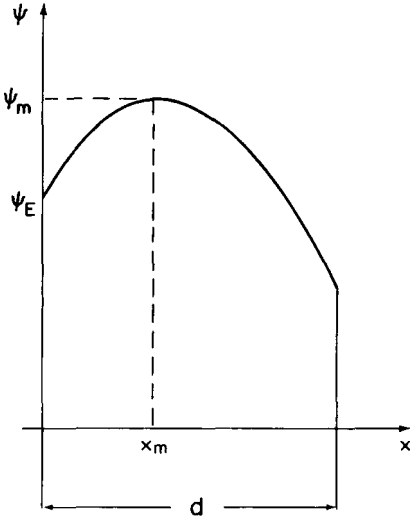


Fig. 10.3 Interelectrode motive in a vacuum diode.

surface characterized by motive ψ_m, electrons move both toward and away from the emitter. In this space the density n_e is determined by both types of electrons since the space-charge effect of a charged particle does not depend on the direction of its motion. Alternatively, the electron distribution function in this space is different from zero for electrons with velocity components in either x direction.

On the other hand, in the space between the surface of motive ψ_m and the surface just outside the collector, all electrons are accelerated toward the collector. Therefore in this space n_e is determined only by electrons emitted with initial energy such that ϵ_x is greater than ψ_m. Here the distribution function is zero for electrons with velocity components in the negative x direction.

If the origin of energy and space coordinates is at the points just outside the emitter (Fig. 10.4), then at

$$x = x_m \qquad \psi = \psi_m > 0 \qquad \frac{d\psi}{dx} = 0 \tag{10.14}$$

the emitter and collector coordinates are $(0, 0)$ and (x_C, ψ_C), respectively; the space in which electrons move in all directions corresponds to $x \leqslant x_m$; and the space in which electrons move only toward the collector corresponds to $x \geqslant x_m$.

The electron distribution function $f_e(x, \mathbf{v}_e)$ is a solution of the Boltzmann equation (Eq. 8.140) without the collision integral. This equation, also called the *Vlasov equation,* is given by the relation

$$v_{ex} \frac{\partial f_e}{\partial x} - \frac{1}{m_e} \frac{d\psi}{dx} \frac{\partial f_e}{\partial v_{ex}} = 0 \tag{10.15}$$

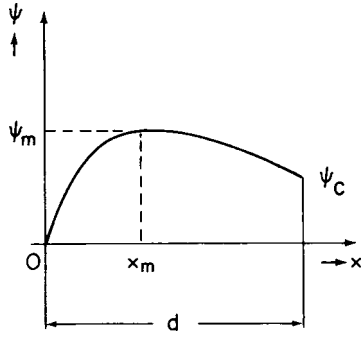


Fig. 10.4 Interelectrode motive in a vacuum diode. Coordinates originate at a point just outside the emitter.

since the electron flow is one-dimensional and the force on an electron is equal to the negative gradient of the motive.

The independent variables x and v_{ex} in Eq. 10.15 are separable. Because of this separability, we can readily verify that the general solution of the equation is given by the relation

$$f_e(x,\mathbf{v}_e) = \sum_{j=0}^{\infty} A_j \exp\left[-B_j\left(\psi + \frac{1}{2}m_e v_e^2\right)\right] \tag{10.16}$$

where A_j and B_j are arbitrary integration constants to be subsequently determined from the boundary conditions and the electron speed v_e satisfies the relation

$$v_e^2 = v_{ex}^2 + v_{ey}^2 + v_{ez}^2 \tag{10.17}$$

At the maximum motive the distribution function $f_e(x_m,\mathbf{v}_e)$ is half-Maxwellian at the emitter temperature T_E, because at this motive the minimum value of the velocity component along the x-axis is equal to zero, and all electrons originate from the emitter and reach ψ_m without experiencing any collisions. It follows that $f_e(x_m,\mathbf{v}_e)$ is given by the relation

$$f_e(x_m,\mathbf{v}_e) = \begin{cases} 2n_e(x_m)\left(\dfrac{m_e}{2\pi kT_E}\right)^{3/2} \exp\left[-\dfrac{m_e}{2kT_E}\,(v_{ex}^2 + v_{ey}^2 + v_{ez}^2)\right] & \text{for } v_{ex} \geqslant 0 \\ 0 & \text{for } v_{ex} < 0 \end{cases} \tag{10.18}$$

or, equivalently,

$$f_e(x_m,\mathbf{v}_e) = 2n_e(x_m)\left(\frac{m_e}{2\pi kT_E}\right)^{3/2} \exp\left(-\frac{m_e}{2kT_E}\,v_e^2\right) u(v_{ex}) \tag{10.19}$$

where $u(t)$ is the unit step function* and $n_e(x_m)$, the electron number density at x equal to x_m, is given by the relation

$$\begin{aligned} n_e(x_m) &\equiv \int \mathbf{v}_e\, f_e(x_m,\mathbf{v}_e)\, d\mathbf{v}_e \\ &= \int_0^{\infty} dv_{ex} \int_{-\infty}^{\infty} dv_{ey} \int_{-\infty}^{\infty} dv_{ez}\, f_e(x_m,\mathbf{v}_e) \end{aligned} \tag{10.20}$$

*The unit step function $u(t)$ is equal to zero and unity for negative and nonnegative values, respectively, of its argument t. In symbols

$u(t) \equiv 0$ for $t < 0$ and $u(t) \equiv 1$ for $t \geqslant 0$

It is shown later that $n_e(x_m)$ is uniquely determined by the output-current density J and the emitter temperature T_E.

Equation 10.19 can be regarded as a boundary condition for the general solution given by Eq. 10.16. At $x = x_m$, where $\psi = \psi_m$, the right-hand sides of Eqs. 10.16 and 10.19 must be identical, and therefore

$$A_0 = 2n_e(x_m)\left(\frac{m_e}{2\pi kT_E}\right)^{3/2} \exp\left(-\frac{\psi_m}{kT_E}\right)$$
$$B_0 = \frac{1}{kT_E} \quad \text{and} \quad A_j = 0 \quad \text{for } j \geqslant 1 \tag{10.21}$$

The values of the coefficients A_j and B_j given by relations 10.21 are applicable at all other positions because: (1) The right-hand side of Eq. 10.16 can be thought of as a series expansion for the distribution function in terms of the linearly independent exponential functions of v_e. (2) The linearly independent functions are different from zero for all values of v_e. (3) If a particular exponential function is not needed to represent the distribution function at one position, it is not needed at any other position. To complete the derivation of $f_e(x,\mathbf{v}_e)$, we need only to determine the range of values of velocity applicable at x.

For values of x greater than x_m, electrons are accelerated by the negative space charge. In other words, in this region (Fig. 10.4) the gradient of the motive is negative. The minimum value $(v_{ex})_{\min}$ of the velocity component v_{ex} is that of electrons for which, at x equal to x_m, v_{ex} equals zero. These electrons pass through a motive difference equal to $\psi_m - \psi$, and therefore $(v_{ex})_{\min}$ is given by the relation

$$(v_{ex})_{\min} = \left[\frac{2(\psi_m - \psi)}{m_e}\right]^{1/2} \tag{10.22}$$

It follows that the range of values of v_{ex} applicable at positions x greater than x_m is given by the relation

$$\left[\frac{2(\psi_m - \psi)}{m_e}\right]^{1/2} \leqslant v_{ex} < \infty \quad \text{for } x \geqslant x_m \tag{10.23}$$

For values of x smaller than x_m, the electrons are decelerated by the negative space charge. Electrons without appropriate total initial energy to reach ψ_m are returned to the emitter. Hence the value of v_{ex} can be positive or negative. The minimum value is negative and corresponds to the value of electrons reaching ψ_m with v_{ex} equal to 0 and then returning to a position of motive ψ.

Accordingly the range of values of v_{ex} applicable at positions smaller than x_m is given by the relation

$$-\left[\frac{2(\psi_m - \psi)}{m_e}\right]^{1/2} \leqslant v_{ex} < \infty \qquad \text{for } x < x_m \tag{10.24}$$

Combining Eqs. 10.16 and 10.21 with ranges 10.23 and 10.24 yields that the distribution function is given by the relations:

$$f_e(x,\mathrm{v}_e) = 2n_e(x_m)\left(\frac{m_e}{2\pi kT_E}\right)^{3/2} \exp\left(\frac{\psi_m - \psi}{kT_E} - \frac{m_e v_e^2}{2kT_E}\right) \times u[v_{ex} - (v_{ex})_{\min}] \qquad \text{for } x \geqslant x_m \tag{10.25}$$

$$f_e(x,\mathrm{v}_e) = 2n_e(x_m)\left(\frac{m_e}{2\pi kT_E}\right)^{3/2} \exp\left(\frac{\psi_m - \psi}{kT_E} - \frac{m_e v_e^2}{2kT_E}\right) \times u[v_{ex} + (v_{ex})_{\min}] \qquad \text{for } x \leqslant x_m \tag{10.26}$$

The next task is to compute n_e from the general identity

$$n_e \equiv \int_{-\infty}^{\infty} dv_{ex} \int_{-\infty}^{\infty} dv_{ey} \int_{-\infty}^{\infty} dv_{ex}\, f_e(x,\mathrm{v}_e) \tag{10.27}$$

By virtue of Eq. 10.25, Eq. 10.27 yields

$$n_e = \frac{2n_e(x_m)}{\pi^{1/2}} \exp(\gamma) \int_{\gamma^{1/2}}^{\infty} \exp(-t^2)\, dt$$

$$= n_e(x_m) \exp(\gamma)\,[1 - \operatorname{erf}(\gamma^{1/2})] \qquad \text{for } x \geqslant x_m \tag{10.28}$$

where the dimensionless quantities γ and t and the tabulated error integral $\operatorname{erf}(\gamma^{1/2})$ are defined by the relations

$$\gamma = \frac{\psi_m - \psi}{kT_E} \tag{10.29}$$

$$t^2 = \frac{m_e v_e^2}{2kT_E} \tag{10.30}$$

and

$$\operatorname{erf}(\gamma^{1/2}) \equiv \frac{2}{\pi^{1/2}} \int_0^{\gamma^{1/2}} \exp(-t^2)\, dt \tag{10.31}$$

and use has been made of the identities erf $(\infty) \equiv 1$ and erf $(0) \equiv 0$. Similarly, by virtue of Eq. 10.26, Eq. 10.27 yields

$$n_e = n_e(x_m) \exp(\gamma) \left[1 + \operatorname{erf}(\gamma^{1/2})\right] \qquad \text{for } x < x_m \tag{10.32}$$

Substituting Eqs. 10.28 and 10.32 into Eq. 10.13 results in a dimensionless differential equation for γ:

$$2\frac{d^2\gamma}{d\xi^2} = \exp(\gamma) \left[1 \pm \operatorname{erf}(\gamma^{1/2})\right] \qquad \text{for } \xi \gtrless 0 \tag{10.33}$$

where the plus sign of $\pm$ applies for ξ negative ($x < x_m$), and the minus sign applies for ξ positive ($x > x_m$) and where the dimensionless distance

$$\xi = \frac{x - x_m}{x_0} \tag{10.34}$$

and the *normalization length* x_0 is given by the relation

$$x_0^2 \equiv \frac{\varepsilon_0 k T_E}{2e^2 n_e(x_m)} \tag{10.35}$$

The initial conditions for Eq. 10.33 are given by the relations

$$\gamma = 0 \quad \text{and} \quad \frac{d\gamma}{d\xi} = 0 \qquad \text{for } \xi = 0 \tag{10.36}$$

(see Eqs. 10.14 and 10.29). Therefore the first integral of Eq. 10.33 is of the form

$$\left(\frac{d\gamma}{d\xi}\right)^2 = \left[\exp(\gamma) - 1 \pm \exp(\gamma) \operatorname{erf}(\gamma^{1/2}) \mp 2\left(\frac{\gamma}{\pi}\right)^{1/2}\right] \qquad \text{for } \xi \lesseqgtr 0 \tag{10.37}$$

where the upper signs in $\pm$ and $\mp$ apply for ξ negative and the lower signs for ξ positive. Subsequent integration of Eq. 10.37 results in the implicit relation

$$\xi = \mp \int_0^{\gamma} \frac{dt}{\left[\exp(t) - 1 \pm \exp(t) \operatorname{erf}(t^{1/2}) \mp 2(t/\pi)^{1/2}\right]^{1/2}} \qquad \text{for } \xi \lesseqgtr 0 \tag{10.38}$$

with the same convention regarding upper and lower signs in $\pm$ and $\mp$.

The integral on the right-hand side of Eq. 10.38 can be evaluated numerically. The results are shown in Fig. 10.5, which is essentially a combination of

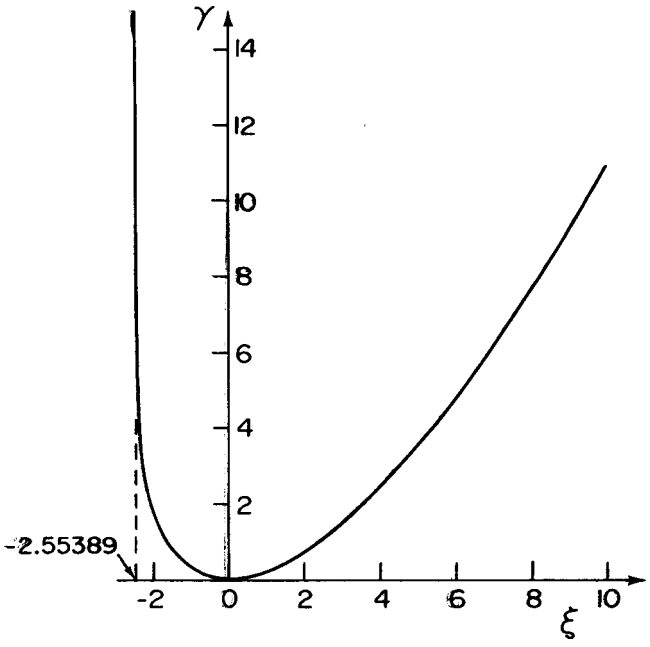


Fig. 10.5 Dimensionless motive γ $[(\psi_m - \psi)/kT_E]$ vs. dimensionless distance ξ $[(x - x_m)/x_0]$.

Figs. 3.7a and 3.7b, Chap. 3, Vol. I, except for a minor change in notation. Figure 10.5 shows that γ varies from infinity to zero to infinity as ξ varies from -2.55389 to zero to infinity. For extremely large values of γ, the results are not physically valid, because for such values the motion of the electrons is relativistic and relativistic effects are not included in this analysis. This restriction is of no practical consequence, however, since very large values of γ never occur in vacuum converters.

Equation 10.38 and the numerical results in Fig. 10.5 represent Langmuir's space-charge theory, which gives the interelectrode motive in a vacuum diode operated in the space-charge-limited mode. Adapting the theory to the calculation of output characteristics, which is partially discussed in Vol. I, Chap. 3, is discussed in detail in Sec. 10.3.2.

10.3.2 Output-Current Characteristics of Vacuum Diodes. The output-current density J of a vacuum diode operating in the space-charge-limited mode with

negligible back emission is due solely to electrons going past the maximum motive ψ_m (Fig. 10.4). This current is given by the relation

$$J = e \int_0^\infty dv_{ex} \int_{-\infty}^{\infty} dv_{ey} \int_{-\infty}^{\infty} dv_{ez}\, v_{ex} f_e(x_m, \mathbf{v}_e) \tag{10.39}$$

Substituting Eq. 10.19 into Eq. 10.39 yields, after some algebra,

$$J = en_e(x_m)\left(\frac{2kT_E}{\pi m_e}\right)^{1/2} \tag{10.40}$$

We see from Eq. 10.40 that J and $n_e(x_m)$ are interrelated.

The electron-saturation-current density J_{ES} is given by Richardson's equation and by the relation

$$J_{ES} = e \int_0^\infty dv_{ex} \int_{-\infty}^{\infty} dv_{ey} \int_{-\infty}^{\infty} dv_{ez}\, v_{ex} f_e(0, \mathbf{v}_e) \tag{10.41}$$

In words, at the surface just outside the emitter ($x = 0$), only electrons with positive velocity components along the x-axis ($v_{ex} > 0$) contribute to J_{ES}. Substituting Eq. 10.26 into Eq. 10.41 yields

$$J_{ES} = en_e(x_m)\left(\frac{2kT_E}{\pi m_e}\right)^{1/2} \exp(\gamma_E) = J \exp(\gamma_E) \tag{10.42}$$

where

$$\gamma_E = \frac{\psi_m}{kT_E} \quad \text{(Fig. 10.4)}$$

$$= \frac{\psi_m - \psi_E}{kT_E} \quad \text{(Fig. 10.3)}$$

By virtue of Eq. 10.40, the normalization length x_0 defined by Eq. 10.35, can also be expressed in terms of J. Specifically

$$x_0 = \left(\frac{\varepsilon_0^2 k^3}{2\pi m_e e^2}\right)^{1/4} \frac{T_E^{3/4}}{J^{1/2}}$$

$$= 1.09 \times 10^{-6} \frac{T_E^{3/4}}{J^{1/2}} \quad \text{cm} \qquad \text{for } T_E \text{ in } ^\circ\text{K and } J \text{ in amp/cm}^2 \tag{10.43}$$

For given electrode properties and output-current density, γ_E and x_0 are uniquely determined by Eqs. 10.42 and 10.43, respectively. Thus, if Langmuir's

space-charge theory is used to analyze the interelectrode motive, steps 1 to 7, given in Vol. I, Chap. 3, Sec. 3.3.3, yield the output-current characteristics of a vacuum converter in the space-charge-limited range of electrode output voltage. This range extends from the critical point to the saturation point. At the critical point the maximum motive occurs just outside the collector, namely, at a distance d from the emitter. By virtue of Eqs. 10.42 and 10.43, it follows that the critical output-current density J_R must satisfy the relations

$$\gamma_{ER} = \ln \frac{J_{ES}}{J_R} \tag{10.44}$$

$$\xi_{ER} = 9.186 \times 10^5 \frac{J_R}{T_E^{3/4}} \tag{10.45}$$

$$\xi_{CR} = 0 \tag{10.46}$$

where γ_{ER} is the normalized interelectrode motive difference between electrodes and ξ_{ER} and ξ_{CR} are the dimensionless distances of the maximum motive from the emitter and the collector, respectively, at the critical point of operation. It follows that J_R can be found from Fig. 10.5 by means of a trial-and-error procedure. For practical emitter temperatures T_E and interelectrode spacings d, the distance ξ_{ER} is close to its limiting value (about 2.554). To a good approximation, substituting this value in Eq. 10.45 yields that J_R is given by the relation

$$J_R = 7.728 \times 10^{-12} \frac{T_E^{3/2}}{d^2} \quad \text{amp/cm}^2 \qquad \text{for } T_E \text{ in } ^\circ\text{K and } d \text{ in cm} \tag{10.47}$$

and the critical voltage V_R is given by the expression

$$eV_R = \phi_C + kT_E \ln \frac{AT_E^2}{J_R} \tag{10.48}$$

where A = 120 amp/cm^2-°K^2.

At the saturation point the maximum motive occurs just outside the emitter. Therefore the dimensionless distances of this maximum from the electrodes are

$$\xi_{ES} = 0 \qquad \xi_{CS} = 9.186 \times 10^5 \frac{J_{ES}^{1/2} d}{T_E^{3/4}} \tag{10.49}$$

and the motive difference γ_{CS} between electrodes can be obtained from the γ vs. ξ relation for ξ equal to ξ_{CS}. Thus the electrode output voltage V_S is given by the relation

$$eV_S = \phi_E - \phi_C - \gamma_{CS} kT_E, \tag{10.50}$$

For small values of $\phi_E - \phi_C$ compared with $\gamma_{CS} kT_E$, Eqs. 10.49 and 10.50 are equivalent to the Child–Langmuir equation.

10.4 Electron-Space-Charge Theory in the Presence of Back Emission

Langmuir's space-charge theory can be readily extended to the analysis of the interelectrode motive of vacuum diodes in which back emission is appreciable.

For the electrode system shown in Fig. 10.1, in the presence of back emission and when the current flow is space-charge limited, the shape of the interelectrode motive is the same as that shown in Fig. 10.4. The distribution function $f_e(x,\mathbf{v}_e)$ satisfies the Vlasov equation (Eq. 10.15), and the general form of $f_e(x,\mathbf{v}_e)$ is given by Eq. 10.16. Here, however, at the maximum motive the distribution function $f_e(x_m,\mathbf{v}_e)$ is the sum of two half-Maxwellian distribution functions because of back emission. One of these two half-Maxwellian distributions corresponds to electrons that originate from the emitter at temperature T_E and reach the maximum motive with a minimum value of v_{ex} equal to zero. All these electrons go past the maximum motive and are accelerated by the negative space charge toward the collector. The other half-Maxwellian corresponds to electrons that originate from the collector at temperature T_C and reach the maximum motive with a maximum value of v_{ex} along the negative x direction (minimum absolute value of the velocity component along the x direction) equal to zero. All these electrons originating from the collector go past the maximum motive and are accelerated by the negative space charge toward the emitter. It follows that the distribution function $f_e(x_m,\mathbf{v}_e)$ is given by the relation

$$f_e(x_m,\mathbf{v}_e) = \frac{2n_e(x_m)}{1+\alpha}\left(\frac{m_e}{2\pi kT_E}\right)^{3/2} \exp\left(-\frac{m_e v_e^2}{2kT_E}\right) u(v_{ex}) + \frac{2\alpha\, n_e(x_m)}{1+\alpha}\left(\frac{m_e}{2\pi kT_C}\right)^{3/2} \exp\left(-\frac{m_e v_e^2}{2kT_C}\right) u(-v_{ex}) \tag{10.51}$$

where α is the ratio of the electron number density at x_m caused by electrons originating from the collector to that caused by electrons originating from the emitter, i.e.,

$$\alpha = \frac{\int_{-\infty}^{0} dv_{ex} \int_{-\infty}^{\infty} dv_{ey} \int_{-\infty}^{\infty} dv_{ez}\, f_e(x_m,\mathbf{v}_e)}{\int_{0}^{\infty} dv_{ex} \int_{-\infty}^{\infty} dv_{ey} \int_{-\infty}^{\infty} dv_{ez}\, f_e(x_m,\mathbf{v}_e)} \tag{10.52}$$

and $n_e(x_m)$ is the electron number density at the maximum motive. The density $n_e(x_m)$ is given by the integral of $f_e(x_m,\mathbf{v}_e)$ over all $\mathbf{v}_e$ (Eq. 10.27). We see from

Eq. 10.51 that, for both α equal to zero and α equal to infinity, the equation reduces to Eq. 10.19, as it should. Of course, for α equal to infinity the roles of the emitter and the collector are interchanged, and consequently T_E and v_{ex} in Eq. 10.19 should be replaced by T_C and $-v_{ex}$, respectively.

Comparing the general form for $f_e(x,\mathrm{v}_e)$ (Eq. 10.16), evaluated at $x = x_m$ and $\psi = \psi_m$, with Eq. 10.51 yields that the coefficients A_j and B_j are given by the relations

$$A_0 = \frac{2n_e(x_m)}{1+\alpha}\left(\frac{m_e}{2\pi kT_E}\right)^{3/2}\exp\left(\frac{\psi_m}{kT_E}\right) \qquad B_0 = \frac{1}{kT_E}$$

$$A_1 = \frac{2\alpha\, n_e(x_m)}{1+\alpha}\left(\frac{m_e}{2\pi kT_C}\right)^{3/2}\exp\left(\frac{\psi_m}{kT_C}\right) \qquad B_1 = \frac{1}{kT_C} \qquad A_j = 0 \quad \text{for } j \geqslant 2 \tag{10.53}$$

These coefficients are applicable at all positions x.

For electrons originating from the emitter, the range of values of velocity at each position x is the same as that presented in Sec. 10.3.1. The range is given either by relation 10.23 for $x > x_m$ or by relation 10.24 for $x < x_m$. On the other hand, for electrons originating from the collector, the range of values of velocity at each position x can be readily established by using the arguments presented in Sec. 10.3.1. Thus we find that

$$-\infty < v_{ex} < \left[\frac{2(\psi_m - \psi)}{m_e}\right]^{1/2} \qquad \text{for } x \geqslant x_m \tag{10.54}$$

and

$$-\infty < v_{ex} < -\left[\frac{2(\psi_m - \psi)}{m_e}\right]^{1/2} \qquad \text{for } x \leqslant x_m \tag{10.55}$$

Combining Eqs. 10.16 and 10.53 and ranges 10.23, 10.24, 10.54, and 10.55 yields that the distribution function $f_e(x,\mathrm{v}_e)$ is given by the relations:

$$f_e(x,\mathrm{v}_e) = \frac{2n_e(x_m)}{1+\alpha}\left(\frac{m_e}{2\pi kT_E}\right)^{3/2}\exp\left(\frac{\psi_m - \psi}{kT_E} - \frac{m_e v_e^2}{2kT_E}\right)u[v_{ex} - (v_{ex})_{\min}]$$

$$+ \frac{2\alpha\, n_e(x_m)}{1+\alpha}\left(\frac{m_e}{2\pi kT_C}\right)^{3/2}\exp\left(\frac{\psi_m - \psi}{kT_C} - \frac{m_e v_e^2}{2kT_C}\right)$$

$$\times\, u[-v_{ex} + (v_{ex})_{\min}] \qquad \text{for } x \geqslant x_m \tag{10.56}$$

$$f_e(x,\mathbf{v}_e) = \frac{2n_e(x_m)}{1+\alpha}\left(\frac{m_e}{2\pi kT_E}\right)^{3/2} \exp\left(\frac{\psi_m - \psi}{kT_E} - \frac{m_e v_e^2}{2kT_E}\right) u[v_{ex} + (v_{ex})_{\min}]$$

$$+ \frac{2\alpha\, n_e(x_m)}{1+\alpha}\left(\frac{m_e}{2\pi kT_C}\right)^{3/2} \exp\left(\frac{\psi_m - \psi}{kT_C} - \frac{m_e v_e^2}{2kT_C}\right)$$

$$\times\, u[-v_{ex} - (v_{ex})_{\min}] \qquad \text{for } x \leqslant x_m \tag{10.57}$$

where $(v_{ex})_{\min}$ is defined by Eq. 10.22.

The electron number density, n_e, can be obtained by means of Eq. 10.27. By virtue of Eqs. 10.56 and 10.57, we find that n_e is given by the relation

$$n_e = \frac{n_e(x_m)}{1+\alpha}\{\exp(\gamma)\,[1 \pm \operatorname{erf}(\gamma^{1/2})] + \alpha \exp(\delta\gamma)\,[1 \mp \operatorname{erf}(\delta^{1/2}\gamma^{1/2})]\}$$

$$\text{for } x \lesseqgtr x_m \tag{10.58}$$

where

$$\gamma = \frac{\psi_m - \psi}{kT_E} \qquad \delta = \frac{T_E}{T_C}$$

and where the upper signs in $\pm$ and $\mp$ are used for $x < x_m$ and the lower signs are used for $x > x_m$. For α equal to zero, Eq. 10.58 reduces to the corresponding equation derived for n_e due to emitter electrons only (Eqs. 10.28 and 10.32).

Substituting Eq. 10.58 into Eq. 10.13 yields the differential equation

$$2\frac{d^2\gamma}{d\xi^2} = \exp(\gamma)\,[1 \pm \operatorname{erf}(\gamma^{1/2})] + \alpha \exp(\delta\gamma)\,[1 \mp \operatorname{erf}(\delta^{1/2}\gamma^{1/2})]$$

$$\text{for } \xi \lesseqgtr 0 \tag{10.59}$$

where

$$\xi = \frac{x - x_m}{x_0'} \tag{10.60}$$

and*

$$x_0'^2 = \frac{\varepsilon_0 kT_E(1+\alpha)}{2e^2\, n_e(x_m)} \tag{10.61}$$

*The definitions of x_0' and of x_0 (Eq. 10.35) are physically the same. Both $[n_e(x_m)]/(1+\alpha)$ in Eq. 10.61 and $n_e(x_m)$ in Eq. 10.35 represent the electron number density at the maximum motive contributed by electrons originating from the emitter.

The initial conditions for Eq. 10.59 are given by relations 10.36, and therefore the first integral of Eq. 10.59 is of the form

$$\left(\frac{d\gamma}{d\xi}\right)^2 = \exp(\gamma) - 1 \pm \exp(\gamma)\,\mathrm{erf}(\gamma^{1/2}) \mp 2\left(\frac{\gamma}{\pi}\right)^{1/2} + \frac{\alpha}{\delta}\left[\exp(\delta\gamma) - 1 \mp \exp(\delta\gamma)\,\mathrm{erf}(\delta^{1/2}\gamma^{1/2}) \pm 2\left(\delta\frac{\gamma}{\pi}\right)^{1/2}\right] \quad \text{for } \xi \lesseqgtr 0 \qquad (10.62)$$

Subsequent integration of Eq. 10.62 results in an implicit relation between γ and ξ similar to that derived in Langmuir's theory. In contrast to that theory, however, here the values of ξ are bounded both from below and from above as γ increases indefinitely. The exact values of the lower (negative) and upper (positive) bounds depend on the parameters α and δ.

The integration of Eq. 10.62 has been performed numerically with the aid of a digital computer.[5] Typical results for some values of the parameters δ and α/δ are shown in Fig. 10.6. For more-detailed listings of the computations, the reader is referred to Ref. 5. These results can be used to calculate output-current characteristics of vacuum diodes with appreciable back emission.

When both electrodes in Fig. 10.1 have the same work function ($\alpha = 1$) and are at the same temperature ($\delta = 1$), we can integrate Eq. 10.62 and find the relation

$$\xi = \sqrt{2}\,\arccos\exp\left(-\frac{\gamma}{2}\right) \qquad (10.63)$$

10.5 Space-Charge Theories in the Presence of Both Electrons and Positive Ions

The presence of positive ions in the interelectrode space of a thermionic converter affects the performance of the converter. The effects of positive ions on the interelectrode motive of vacuum diodes, the interelectrode motive of low-vapor-pressure diodes, or the sheaths of high-vapor-pressure diodes depend on both the nature of the ion sources and the location of the sources with respect to the electrodes. Accordingly a number of space-charge theories applicable to different combinations of electron and ion sources are discussed in the subsequent sections.

10.5.1 Planar Electron and Ion Sources. Langmuir[6] investigated the interelectrode motive between two infinite parallel-plane electrodes shown schematically in Fig. 10.7. In this figure electrode E is a thermionic emitter of electrons only. At the points just outside E, the velocity of the electrons is negligible. Electrode C does not emit any particles and is positively charged with respect to E by means of an externally applied voltage supply. Singly charged positive ions with

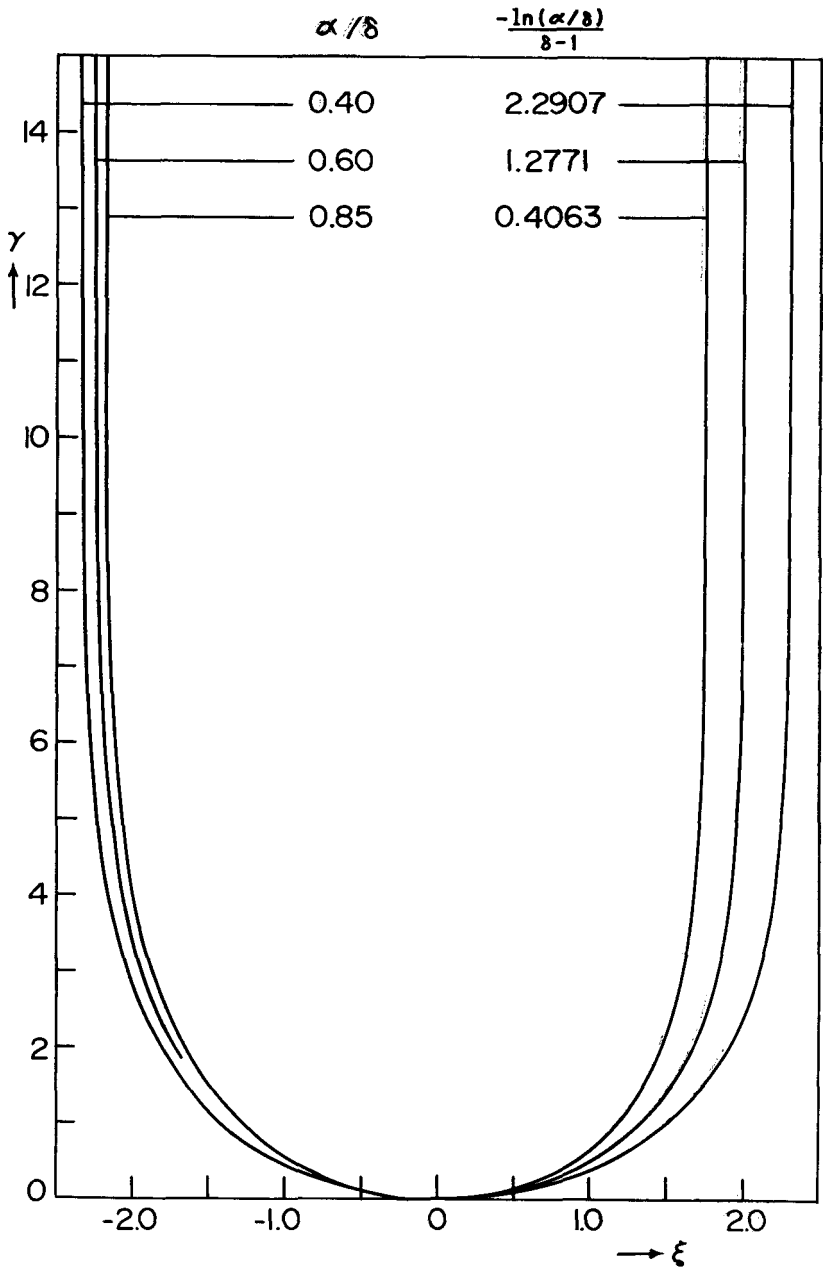


Fig. 10.6 Graph of dimensionless motive γ vs. dimensionless distance ξ for vacuum diode with back emission[5] and $\delta = 1.4$.

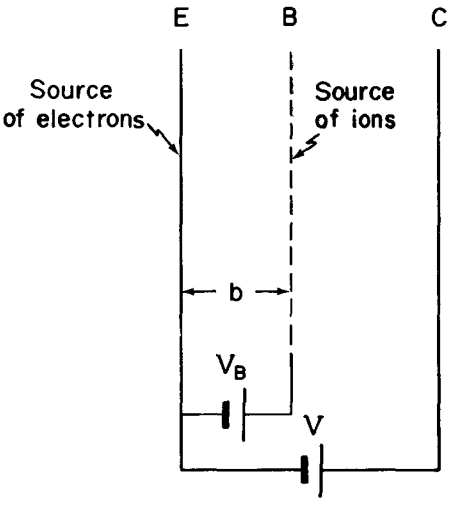


Fig. 10.7 Parallel-plane electrode arrangement with positive-ion source at *B*.

negligible initial velocity are produced at the source *B* located between the two electrodes. The interelectrode space is evacuated except for the electrons and positive ions in transit.

In the absence of a positive-ion source and for a given internal voltage drop $\Delta V < 0$, the current flow from *E* to *C* is limited by the electron space charge, and the current density J_0 is given by the Child–Langmuir equation (Eq. 10.12), provided that J_0 is smaller than the electron-saturation-current density from *E*.

When the ion source is introduced and the interelectrode motive has the shape shown in Fig. 10.8, an ion current flows from *B* to *E*.* Because of this current the electron space charge is partially neutralized, and the electron current density *J* from *E* to *C* is larger than J_0. For the given interelectrode motive shape, no ion current flows between *B* and *C* since the initial velocity of the ions is assumed to be negligible.

The value of *J* depends on the distance *b* of the ion source from *E*, and on the ion current density $\tilde{J}_B$ flowing from *B* to *E*. These dependencies can be found as follows: The electron and ion velocities v_e and v_i at a position *x* (Fig. 10.8) satisfy the relations

$$\frac{1}{2} m_e v_e^2 = -\psi \qquad \text{for all } \psi \tag{10.64}$$

*The ion source must be kept at the motive ψ_B with respect to *E* by means of an external power supply.

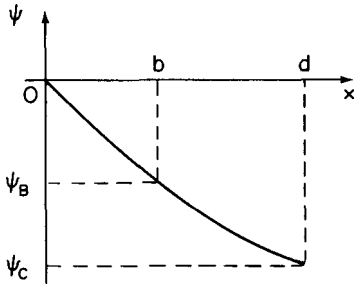


Fig. 10.8 Interelectrode motive for electrode system in Fig. 10.7.

and

$$\frac{1}{2} m_i v_i^2 = \psi - \psi_B \qquad \text{for } \psi - \psi_B \geqslant 0 \tag{10.65}$$

By virtue of these relations, the electron and ion number densities n_e and n_i, respectively, and Poisson's equation for the interelectrode motive are given by the relations

$$n_e = \frac{J}{e\,(-2\psi/m_e)^{1/2}} \qquad \text{for all } \psi \tag{10.66}$$

$$n_i = \begin{cases} \dfrac{\tilde{J}_B}{e\,[2(\psi - \psi_B)/m_i]^{1/2}} & \text{for } \psi - \psi_B \geqslant 0 \\ 0 & \text{for } \psi - \psi_B < 0 \end{cases} \tag{10.67}$$

and

$$\frac{d^2\psi}{dx^2} = \begin{cases} -\dfrac{eJ}{\varepsilon_0\,(-2\psi/m_e)^{1/2}} + \dfrac{e\tilde{J}_B}{\varepsilon_0\,[(2\psi - \psi_B)/m_i]^{1/2}} & \text{for } \psi - \psi_B \geqslant 0 \\ -\dfrac{eJ}{\varepsilon_0\,(-2\psi/m_e)^{1/2}} & \text{for } \psi - \psi_B < 0 \end{cases} \tag{10.68}$$

Introducing the dimensionless quantities

$$\beta = \left(\frac{m_i}{m_e}\right)^{1/2} \frac{\tilde{J}_B}{J} \qquad \gamma = \frac{\psi}{e\,|\Delta V|} \qquad \xi = \frac{x}{d} \tag{10.69}$$

into Eq. 10.68 yields the relation

$$\frac{d^2\gamma}{d\xi^2} = \begin{cases} \frac{4}{9}\frac{J}{J_0}\left[\gamma^{-1/2} - \beta\,(\gamma_B - \gamma)^{-1/2}\right] & \text{for } \gamma_B - \gamma \geqslant 0 \quad (10.70a) \\ \frac{4}{9}\frac{J}{J_0}\,\gamma^{-1/2} & \text{for } \gamma_B - \gamma < 0 \quad (10.70b) \end{cases}$$

where J_0 is given by Eq. 10.12.

If electron emission from E is limited only because of space charge, the initial conditions for Eq. 10.70, parts a and b, are given by the relations

$$\gamma = 0 \quad \text{and} \quad \frac{d\gamma}{d\xi} = 0 \quad \text{for } \gamma = 0 \tag{10.71}$$

Therefore the first integral of Eq. 10.70a is of the form

$$\frac{d\gamma}{d\xi} = \frac{4}{3}\left(\frac{J}{J_0}\right)^{1/2}\{\gamma^{1/2} + \beta[(\gamma_B - \gamma)^{1/2} - \gamma_B^{1/2}]\}^{1/2} \quad \text{for } \gamma_B - \gamma \geqslant 0 \tag{10.72}$$

Further integration of Eq. 10.72 is facilitated if the location of the ion source is explicitly stated. Two cases can be distinguished:

1. *Ion source at the points just outside electrode C* ($\gamma_B = 1$). When the ion source is located at the points just outside C, its normalized motive γ_B is equal to unity. It follows from Eq. 10.72 that the gradient of the motive at the points just outside C is proportional to $(1 - \beta)^{1/2}$, and therefore

$$\beta \leqslant 1 \tag{10.73}$$

since the gradient must be real.

When β is equal to unity, the gradient of the motive at the position of the ion source is equal to zero. For this condition both electron and ion current flows are limited by the space charge in the interelectrode space and not by the emission capabilities of electrode E and ion source B. In other words, the ion current flowing from B to E cannot be more than the fraction $(m_e/m_i)^{1/2}$ of the electron current from E to C. Moreover, for β equal to 1 and γ_B equal to 1, Eqs. 10.70a and 10.72 remain invariant when γ is replaced by $1 - \gamma$. It follows that under these conditions the interelectrode motive is symmetrical about the midplane defined as ξ equal to 0.5 and γ equal to 0.5. In the half space between E and the midplane, the net space charge is negative; in the remaining half space, the net space charge is positive. Regions in which the net space charge has both signs have been called *double sheaths* by Langmuir.

position of the positive-ion source, maximum electron current flows from E to C when β is equal to unity, i.e., when both the electron and ion currents are space charge limited.

For β equal to unity, graphs of $(J/J_0)^{1/2}\,\xi_B$, $(J/J_0)^{1/2}\,(1-\xi_B)$, and $(J/J_0)^{1/2}$ vs. γ_B are shown in Fig. 10.9. We see from this figure that the largest maximum ratio $(J/J_0)_{max}$ is obtained when

$$\gamma_B = (\gamma_B)_m = 0.61$$

The value $(\gamma_B)_m$ corresponds to

$$(\xi_B)_m = 0.44 \qquad \left(\frac{J}{J_0}\right)_{max} = 4.56$$

In words, the maximum effectiveness of the positive ions is achieved when the ion source is located at a distance b_m from E equal to $0.44d$ because then the electron current is 4.56 times larger than the current that would flow in the absence of ions.

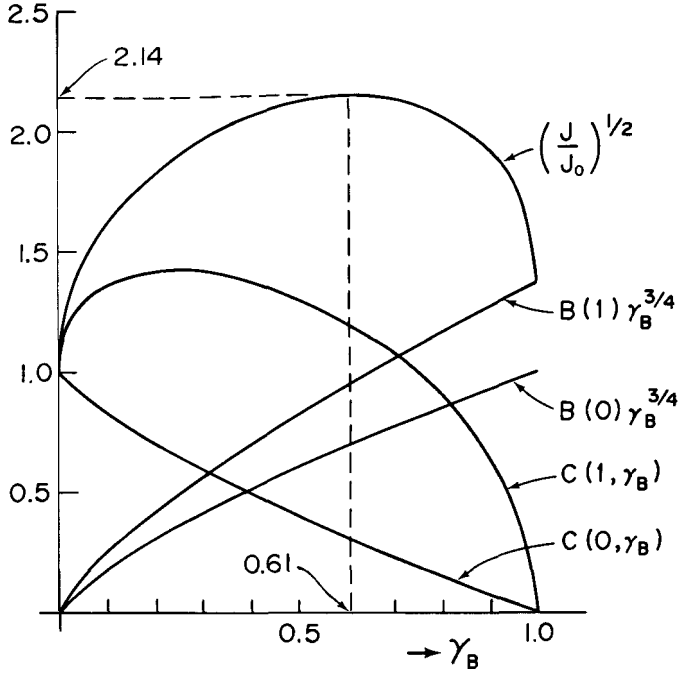


Fig. 10.9 Graphs for determining the position of maximum effectiveness of positive-ion source in Fig. 10.7.

In conclusion, the preceding simplified analysis illustrates the effects of negative and positive space charges on the current flow between two electrodes. Because of its simplicity this analysis provides a good insight into the physical phenomena involved. Nevertheless, neglecting initial velocities and considering an idealized positive-ion source render the quantitative results inapplicable to thermionic converters. A more realistic theory is discussed in Sec. 10.5.2.

10.5.2 Planar Electron and Ion Sources at Identical Positions—Half-Maxwellian Distributions of Initial Velocities. In low-vapor-pressure diodes, electrons and ions are emitted thermionically from the hot electrode with half-Maxwellian velocity-distribution functions. Accordingly the interelectrode motive and the output-current characteristics of these diodes must be analyzed by means of a space-charge theory that accounts for the initial velocities. Similar remarks can be made about the motive in the vicinity of a cold electrode in a high-vapor-pressure diode, namely, a collector sheath (see, however, Chap. 11). Analyses of the type described have been performed by Auer[7] and McIntyre,[8] whose results, with some minor modifications, are summarized here.

For the system of the two infinite parallel-plane electrodes, shown in Fig. 10.1, we will assume that (1) electrode E is a thermionic emitter of both electrons and singly charged positive ions; (2) at the points just outside E and in the positive x direction, both types of charged particles have half-Maxwellian velocity distributions; (3) electrode C does not emit any charged particles; and (4) the vapor pressure in the interelectrode space is sufficiently low so that effects of particle collisions of any type can be neglected. This system is a good representation for a low-vapor-pressure diode for which the linear dimensions of the electrodes are much larger than the interelectrode spacing d and for which back emission is negligible.

The calculation of the electron interelectrode motive ψ can proceed in a manner analogous to that presented in Sec. 10.3.1. First, the distribution functions $f_e(x,\mathrm{v}_e)$ of the electrons and $f_i(x,\mathrm{v}_i)$ of the positive ions are expressed in terms of the motive ψ. Second, the charged-particle number densities n_e and n_i are derived from the distribution functions. Third, these densities are replaced in Poisson's equation to obtain a differential relation between motive and position x in the interelectrode space. Finally, Poisson's equation is integrated to obtain the ψ vs. x relation.

Each of the distribution functions $f_e(x,\mathrm{v}_e)$ and $f_i(x,\mathrm{v}_i)$ is a solution of a Vlasov equation. For the one-dimensional geometry under consideration, the two Vlasov equations are given by the relations

$$v_{ex}\frac{\partial f_e}{\partial x} - \frac{1}{m_e}\frac{d\psi}{dx}\frac{\partial f_e}{\partial v_{ex}} = 0 \tag{10.80}$$

and

$$v_{ix}\frac{\partial f_i}{\partial x}+\frac{1}{m_i}\frac{\partial \psi}{\partial x}\frac{\partial f_i}{\partial v_{ix}}=0 \tag{10.81}$$

since the force $(-d\psi/dx)$ on an electron is equal and opposite to the force $d\psi/dx$ on a singly charged positive ion.

The general solutions of Eqs. 10.80 and 10.81 are given by the relations

$$f_e(x,\mathbf{v}_e)=\sum_{j=0}^{\infty} A_{ej}\exp\left[-B_{ej}\left(\psi+\frac{1}{2}m_e v_e^2\right)\right] \tag{10.82}$$

and

$$f_i(x,\mathbf{v}_i)=\sum_{j=0}^{\infty} A_{ij}\exp\left[B_{ij}\left(\psi-\frac{1}{2}m_i v_i^2\right)\right] \tag{10.83}$$

where A_{ej}, B_{ej}, A_{ij}, and B_{ij} are integration constants to be subsequently determined (see also Sec. 10.3.1).

At the points just outside the emitter, both electrons and positive ions are emitted from E with half-Maxwellian velocity distribution functions at the temperature T_E of the emitter. If the origin of energy and space coordinates is taken at the points just outside the emitter, it follows that $f_e(0,\mathbf{v}_e)$ and $f_i(0,\mathbf{v}_i)$ for $v_{ex}>0$ are given by the relations

$$f_e(0,\mathbf{v}_e)=2n_e^+(0)\left(\frac{m_e}{2\pi kT_E}\right)^{3/2}\exp\left(-\frac{m_e v_e^2}{2kT_E}\right)u(v_{ex}) \tag{10.84}$$

and

$$f_i(0,\mathbf{v}_i)=2n_i^+(0)\left(\frac{m_i}{2\pi kT_E}\right)^{3/2}\exp\left(\frac{m_i v_i^2}{2kT_E}\right)u(v_{ix}) \tag{10.85}$$

where $n_\alpha^+(0)$ for $\alpha=e$ and i is the number density corresponding to those particles of species α which move in the positive x direction at x equal to zero. In general, $n_\alpha^+(0)$ is different from the actual number density $n_\alpha(0)$ for $\alpha=e$ and i because $n_\alpha(0)$ is determined by particles of species α moving in both the positive and the negative x directions.

Equations 10.84 and 10.85 are boundary conditions for Eqs. 10.82 and 10.83, respectively. Accordingly the values of the integration constants A_{ij} and B_{ij} are given by the relations

$$A_{e0} = 2n_e^+(0)\left(\frac{m_e}{2\pi kT_E}\right)^{1/2} \qquad B_{e0} = \frac{1}{kT_E} \qquad A_{ej} = 0 \qquad \text{for } j \geq 1 \tag{10.86}$$

and

$$A_{i0} = 2n_i^+(0)\left(\frac{m_i}{2\pi kT_E}\right)^{1/2} \qquad B_{i0} = \frac{1}{kT_E} \qquad A_{ij} = 0 \qquad \text{for } j \geq 1 \tag{10.87}$$

The values of the constants define the forms of the distribution functions. Next we must determine the appropriate ranges of velocities v_{ex} and v_{ix} at each position x. These ranges must be found in a way consistent with the expected shape of the motive.

For example, if the electron motive is as shown in Fig. 10.10, the positive ions reduce but do not completely eliminate the electron space charge (the net space charge is negative throughout the interelectrode space), and the motive at the points just outside the collector is positive; then, from the discussion in Sec. 10.3.1, it follows that the range of electron velocities is given by the relation

$$\mp (v_{ex})_{\min} = \mp\left[\frac{2(\psi_m - \psi)}{m_e}\right]^{1/2} \leq v_{ex} < \infty \qquad \text{for } x \lesseqgtr x_m \tag{10.88}$$

The upper signs in $\mp$ apply in the space between E and the maximum motive, and the lower signs apply between the maximum motive and C. Combining Eqs. 10.82, 10.86, and 10.88 yields that $f_e(x,\mathrm{v}_e)$ is given by the relation

$$f_e(x,\mathrm{v}_e) = 2n_e^+(0)\left(\frac{m_e}{2\pi kT_E}\right)^{1/2} \exp\left[-\left(\frac{\psi}{kT_E} + \frac{m_e v_e^2}{2kT_E}\right)\right] u[v_{ex} \pm (v_{ex})_{\min}]$$

$$\text{for } x \lesseqgtr x_m \tag{10.89}$$

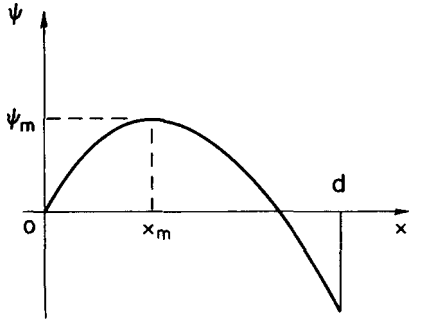


Fig. 10.10 Interelectrode motive shape for negative net space charge and a positive motive ψ_C at the points just outside the collector.

The distribution function is formally identical with that given by Eqs. 10.25 and 10.26 because

$$n_e^+(0) = n_e(x_m) \exp\left(\frac{\psi_m}{kT_E}\right)$$

The range of velocities for the positive ions is derived by means of arguments similar to those used for the electrons. Wherever electrons are accelerated by the space charge, ions are decelerated and vice versa. Consequently, for values of x smaller than x_m (Fig. 10.10), ions are accelerated and move only along the positive x direction. Since the minimum ion velocity at the points just outside the emitter and in the positive x direction is equal to zero, it follows that the minimum ion velocity $(v_{ix})_{\min}$ and the range of ion velocities are given by the relations

$$(v_{ix})_{\min} = \left(\frac{2\psi}{m_i}\right)^{1/2} \quad \text{for } x \leqslant x_m \tag{10.90}$$

and

$$\left(\frac{2\psi}{m_i}\right)^{1/2} \leqslant v_{ix} < \infty \quad \text{for all } x \tag{10.91}$$

The velocity range (Eq. 10.91) is applicable also for $x > x_m$ because in this region ions are decelerated, but the net motive difference between the points just outside the emitter and any position x is positive and is equal to ψ.

By virtue of Eq. 10.91, combining Eqs. 10.83 and 10.87 yields that $f_i(x,\mathbf{v}_i)$ is given by the relation

$$f_i(x,\mathbf{v}_i) = 2n_i^+(0)\left(\frac{m_i}{2\pi kT_E}\right)^{3/2} \exp\left(\frac{\psi}{kT_E} - \frac{m_i v_i^2}{2kT_E}\right) u[v_{ix} - (v_{ix})_{\min}] \tag{10.92}$$

The number densities n_e of electrons and n_i of ions can be found as integrals of the corresponding distribution functions over all velocities. By virtue of Eqs. 10.89 and 10.92, these densities are given by the relations

$$n_e = \int_{\mathbf{v}_e} d\mathbf{v}_e \, f_e(x,\mathbf{v}_e) = n_e^+(0)\, N_e \tag{10.93}$$

and

$$n_i = \int_{\mathbf{v}_i} d\mathbf{v}_i \, f_i(x,\mathbf{v}_i) = n_i^+(0)\, N_i \tag{10.94}$$

where

$$N_e = \exp(-\gamma)\ \{1 \pm \text{erf}\ [(\gamma_m - \gamma)^{1/2}]\} \qquad \text{for } x \lesssim x_m \tag{10.95}$$

$$N_i = \exp(\gamma)\ [1 - \text{erf}\ (\gamma^{1/2})] \tag{10.96}$$

and

$$\gamma = \frac{\psi}{kT_E} \tag{10.97}$$

Substituting Eqs. 10.93 and 10.94 into Poisson's equation for the electron motive yields

$$2\frac{d^2\gamma}{d\xi^2} = -(N_e - \beta N_i) \tag{10.98}$$

where

$$\beta = \frac{n_i^+(0)}{n_e^+(0)} \tag{10.99}$$

$$\xi = \frac{x}{x_1} \tag{10.100}$$

$$x_1^2 \equiv \frac{\varepsilon_0 kT_E}{2e^2\, n_e^+(0)} \tag{10.101}$$

The definitions of both the normalization length x_1 (Eq. 10.101) and the dimensionless motive γ (Eq. 10.97) are different from the definitions of x_0 (Eq. 10.35) and γ (Eq. 10.29) given in Sec. 10.3.1.

The term β (Eq. 10.99) is identical to the ion-richness ratio discussed in Vol. I, Chap. 4. Indeed, the electron- and the ion-saturation-current densities J_{ES} and $\tilde{J}_{ES}$ are given by the relations

$$J_{ES} = e\int_0^\infty dv_{ex}\int_{-\infty}^\infty dv_{ey}\int_{-\infty}^\infty dv_{ez}\, v_{ex}\, f_e(0,\mathbf{v}_e) = e\, n_e^+(0)\left(\frac{2kT_E}{\pi m_e}\right)^{1/2} \tag{10.102}$$

and

$$\tilde{J}_{ES} = e\int_0^\infty dv_{ix}\int_{-\infty}^\infty dv_{iy}\int_{-\infty}^\infty dv_{iz}\, v_{ix}\, f_i(0,\mathbf{v}_i) = e\, n_i^+(0)\left(\frac{2kT_E}{\pi m_i}\right)^{1/2} \tag{10.103}$$

where use has been made of Eqs. 10.89 and 10.92. Consequently β is given by the relation

$$\beta = \frac{n_i^+(0)}{n_e^+(0)} \equiv \frac{\tilde{J}_{ES}}{J_{ES}} \left(\frac{m_i}{m_e}\right)^{1/2} = \text{ion-richness ratio}$$

The condition β smaller than unity is sufficient for the space charge to be negative in the vicinity of the emitter. Indeed, at the points just outside the emitter, the number densities $n_e(0)$ and $n_i(0)$ are given by the relations

$$n_e(0) = n_e^+(0)\,[1 + \text{erf}\,(\gamma_m^{1/2})] \qquad n_i(0) = n_i^+(0)$$

Therefore

$$\frac{n_e(0)}{n_i(0)} = \frac{1 + \text{erf}\,(\gamma_m^{1/2})}{\beta}$$

It follows that, when β is smaller than unity, $n_e(0)$ is larger than $n_i(0)$; namely, near the emitter the negative charges predominate over the positive charges.

At the maximum motive (Fig. 10.10)

$$\xi = \xi_m \qquad \gamma = \gamma_m \qquad \frac{d\gamma}{d\xi} = 0 \tag{10.104}$$

In view of these conditions, a first integral of Eq. 10.98 is given by the relation

$$\left(\frac{d\gamma}{d\xi}\right)^2 = F(\gamma) \tag{10.105}$$

where

$$F(\gamma) = \exp(-\gamma_m)\left\{\exp(\gamma_m - \gamma) \pm \exp(\gamma_m - \gamma)\,\text{erf}\,[(\gamma_m - \gamma)^{1/2}]\right.$$
$$\left. \mp 2\left(\frac{\gamma_m - \gamma}{\pi}\right)^{1/2} - 1\right\} + \beta\left[\exp(\gamma) - \exp(\gamma_m) - \exp(\gamma)\,\text{erf}\,(\gamma^{1/2})\right.$$
$$\left. + \exp(\gamma_m)\,\text{erf}\,(\gamma_m^{1/2}) + 2\left(\frac{\gamma}{\pi}\right)^{1/2} - 2\left(\frac{\gamma_m}{\pi}\right)^{1/2}\right]$$
$$\text{for } \xi \lesseqgtr \xi_m \text{ and } \beta < 1 \tag{10.106}$$

Equation 10.105 can be integrated numerically.

Relations similar to that given by Eq. 10.105 can be derived for shapes of the interelectrode motive other than that shown in Fig. 10.10. Each motive shape results in a specific form for the function $F(\gamma)$, and the motive can be found by means of numerical integration of Eq. 10.105. The various forms of $F(\gamma)$ will not be discussed. For these the reader is referred to the literature.[7]

Numerically computed families of γ vs. ξ for various interelectrode motives are shown in Figs. 10.11 and 10.12. These families of curves constitute the space-charge theory for low-vapor-pressure diodes in which electrons and positive ions are emitted thermionically from the hot electrode, the current flows are one-dimensional, collisions of any type in the interelectrode space are negligible, and back emission is practically zero. As already stated, this theory yields the interelectrode motive and can be combined with a systematic procedure somewhat similar to that described for vacuum diodes to obtain output-current and other characteristics.

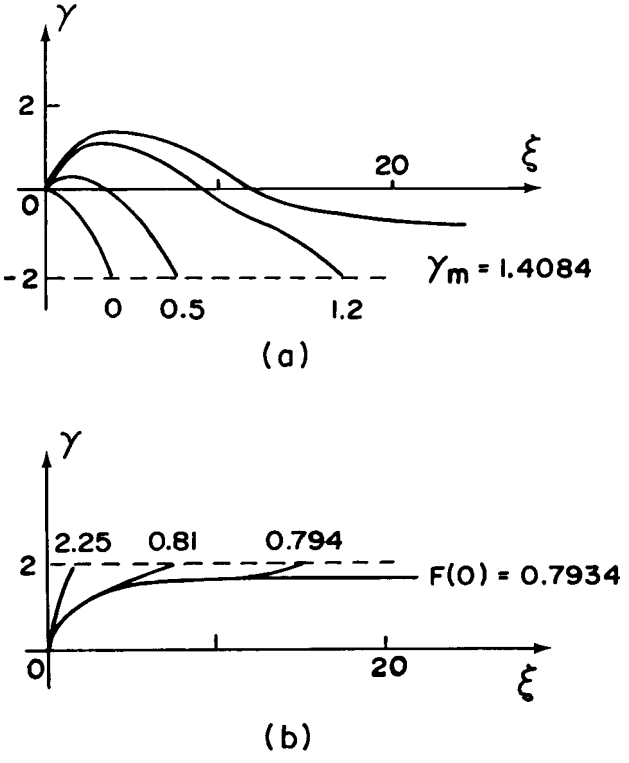


Fig. 10.11 Interelectrode motives under electron-rich conditions at the emitter.[8] (a) $\beta = 0.1$ and $\gamma_C = -2$. (b) $\beta = 0.9$ and $\gamma_C = 2$.

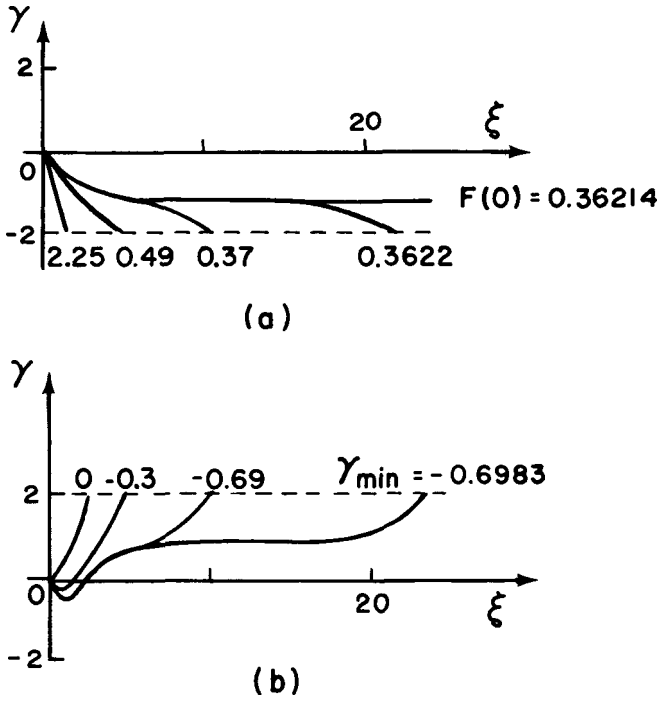


Fig. 10.12 Interelectrode motives under ion-rich conditions at the emitter.[8] (a) $\beta = 0.75$ and $\gamma_C = -2$. (b) $\beta = 5$ and $\gamma_C = 2$.

10.5.3 On Limiting Values of β for Negative-Space-Charge Neutralization. As discussed in Vol. I, Chap. 4, the limiting value of the ion-richness ratio β (Eq. 10.99), which separates the electron-rich from the ion-rich regime of operation, is not unique and is not the same for all low-vapor-pressure diodes. In general, a range of values of β around unity exists which has the following features: (1) the extent β_1 to β_2 of the range depends on the interelectrode spacing and on the internal voltage drop; (2) for values of β smaller than the lower limit β_1 of the range, the diode operates in the electron-rich regime; (3) for values of β larger than the upper limit β_2 of the range, the diode operates in the ion-rich regime; and (4) for values of β within the range, the diode characteristics are oscillatory. The problem of limiting values of β has not been analyzed completely.

For the shape of the interelectrode motive shown in Fig. 10.10 and for a zero value of the internal voltage drop, an approximate analysis of the limiting values of β has been performed[7] and the lower limiting value β_1 is found to be

$$\beta_1 = 0.809764 \tag{10.107}$$

For an interelectrode motive shape with a minimum and a zero value of the internal voltage drop, an approximate analysis of the limiting values of β has also been performed and the upper limiting value β_2 is found to be

$$\beta_2 = \frac{1}{\beta_1} = 1.2349 \tag{10.108}$$

We can verify that for $\beta > \beta_2$ a low-vapor-pressure diode with zero internal voltage drop has a stable motive and is operating in the ion-rich regime, while for β between unity and β_2 the diode has an unstable motive and its characteristics are oscillatory. Oscillatory diode characteristics are not discussed in this book. For these the reader is referred to the literature.[9]

10.5.4 Planar Electron and Ion Sources at Two Positions with Half-Maxwellian Distributions of Velocities. In high-vapor-pressure diodes, the sheath formed in the vicinity of the emitter can be regarded as a collisionless space-charge region between two planar sources E and P as shown schematically in Fig. 10.13. In this figure, source E is located at the points just outside the emitter and is supplying both electrons and singly charged positive ions with half-Maxwellian distributions of velocities at the emitter temperature T_E. Source P is located at the interface between either the sheath region and the transition region or the sheath region and the plasma region of the diode, depending on how the interelectrode system is analyzed (Vol. I, Chap. 4). Source P is supplying both electrons and singly charged positive ions to the sheath and represents the effects of collisions

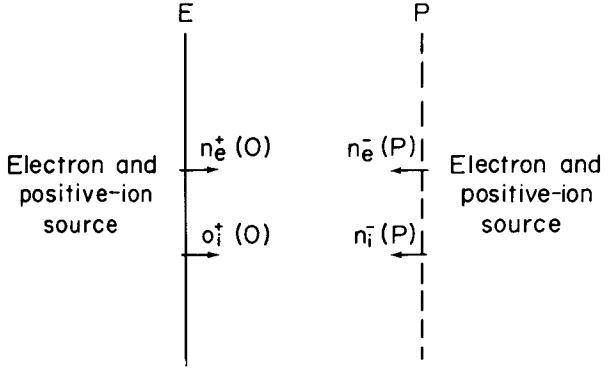


Fig. 10.13 Schematic of planar electron and ion sources at positions E and P.

in the transition and plasma regions. The distributions of velocities of the electrons and ions supplied by P are not known precisely. In many discussions they are assumed to be half-Maxwellian at the temperatures of the electrons (T_e) and of the ions (T_i) of the plasma region.

Motive distributions between the two sources shown in Fig. 10.13 can be obtained by means of the procedure described in Sec. 10.5.2. The distribution functions $f_e(x,\mathrm{v}_e)$ and $f_i(x,\mathrm{v}_i)$ are solutions of Eqs. 10.80 and 10.81 and are given by Eqs. 10.82 and 10.83, respectively. The number of terms in Eqs. 10.82 and 10.83 depends on the distributions of velocities of the particles supplied by the two sources. The ranges of velocities, and, therefore, the densities n_e and n_i, and the form of Poisson's equation at each position, including the positions of the sources, depend on the shape of the motive.

If the distributions of velocities of the particles supplied by the two sources are given by the relations:

$$f_e^+(0,\mathrm{v}_e) = 2n_e^+(0)\left(\frac{m_e}{2\pi kT_E}\right)^{3/2} \exp\left(-\frac{m_e v_e^2}{2kT_E}\right) u(v_{ex}) \qquad \text{for source at } E \qquad (10.109)$$

$$f_i^+(0,\mathrm{v}_i) = 2n_i^+(0)\left(\frac{m_i}{2\pi kT_E}\right)^{3/2} \exp\left(-\frac{m_i v_i^2}{2kT_E}\right) u(v_{ix}) \qquad \text{for source at } E \qquad (10.110)$$

$$f_e^-(P,\mathrm{v}_e) = 2n_e^-(P)\left(\frac{m_e}{2\pi kT_e}\right)^{3/2} \exp\left(-\frac{m_e v_e^2}{2kT_e}\right) u(-v_{ex}) \qquad \text{for source at } P \qquad (10.111)$$

$$f_i^-(P,\mathrm{v}_i) = 2n_i^-(P)\left(\frac{m_i}{2\pi kT_i}\right)^{3/2} \exp\left(-\frac{m_i v_i^2}{2kT_i}\right) u(-v_{ix}) \qquad \text{for source at } P \qquad (10.112)$$

where $n_e^-(P)$ and $n_i^-(P)$ are defined at the position of the source at P and for the negative x direction in a manner analogous to that used for $n_e^+(0)$ and $n_i^+(0)$, then Poisson's equation for the dimensionless electron motive γ (ψ/kT_E) is given by the relation

$$2\frac{d^2\gamma}{d\xi^2} = -N_e^+(\gamma) + \beta N_i^+(\gamma) - \frac{n_e^-(P)}{n_e^+(0)} N_e^-(\gamma) + \frac{n_i^-(P)}{n_e^+(0)} N_i^-(\gamma) \qquad (10.113)$$

where

$$\xi = \frac{x}{x_1} \qquad x_1^2 = \frac{\varepsilon_0 kT_E}{2e^2\, n_e^+(0)}$$

β is defined by Eq. 10.99 and the expressions for N depend on the shape of the motive. For example, for the motive shape shown in Fig. 10.14, N is given by the relations

$$\begin{aligned} N_e^+(\gamma) &= \exp(-\gamma)\,[1 - \operatorname{erf}(-\gamma)^{1/2}] \\ N_i^+(\gamma) &= \exp(\gamma)\,\{1 + \operatorname{erf}[(\gamma - \gamma_P)^{1/2}]\} \\ N_e^-(\gamma) &= \exp[\delta_1(\gamma_P - \gamma)]\,\{1 + \operatorname{erf}[(-\delta_1\gamma)^{1/2}]\} \end{aligned} \tag{10.114}$$

and

$$N_i^-(\gamma) = \exp[\delta_2(\gamma - \gamma_P)]\,\{1 - \operatorname{erf}[\delta_2^{1/2}(\gamma - \gamma_P)^{1/2}]\}$$

where

$$\delta_1 \equiv \frac{T_E}{T_e} \qquad \delta_2 = \frac{T_E}{T_i} \tag{10.115}$$

Analogous expressions for N can be found for other motive shapes.

For a specified motive shape, integrating Eq. 10.113 requires two initial conditions only. For example, the values of γ and $d\gamma/d\xi$ at E might be specified. These values, together with input-data values of T_E, T_e, T_i, $n_e^+(0)$, $n_i^-(0)$, $n_e^-(P)$, $n_i^-(P)$, and the distance between E and P would result, in general, in a solution γ vs. ξ, and, in particular, in values of γ, $d\gamma/d\xi$, and $d^2\gamma/d\xi^2$ at P. Of course, many other combinations of initial conditions and input-data values are possible.

Because of the large number of input parameters and the complexity of the problem, it is not easy to judge a priori whether a given combination of specifications will result in a physically meaningful solution, e.g., number densities n_e and n_i that are positive at all positions, or whether the combination is compatible with results obtained for the plasma region that defines the source at P. These observations have important implications about the performance

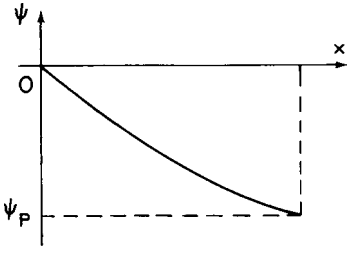


Fig. 10.14 Electron motive shape between sources E and P of Fig. 10.13.

characteristics of high-vapor-pressure converters and are discussed again in Chaps. 11 and 12. Here it suffices to say that solutions of Eq. 10.113 have not been studied in detail. Only some preliminary results have been obtained by Wang, Lieb, and Pigford.[10] Nevertheless, these results illustrate the difficulties and complications cited.

References

1. I. Langmuir, The Effect of Space Charge and Residual Gases on Thermionic Currents in High Vacuum, *Phys. Rev.*, **2**: 450-486 (1913).
2. C. D. Child, Discharge from Hot CaO, *Phys. Rev.*, **32**: 492-511 (1911).
3. H. F. Ivey, Space Charge Limited Currents, in *Advances in Electronics and Electron Physics*, Vol. VI, pp. 137-256, Academic Press, Inc., New York, 1954.
4. I. Langmuir, The Effect of Space Charge and Initial Velocities on the Potential Distribution and Thermionic Current Between Parallel Plane Electrodes, *Phys. Rev.*, **21**: 419-435 (1923).
5. A. Kniazzeh, Potential Distribution Between Two Parallel Thermionic Emitters with Application to the Thermo Electron Engine, B.S. Thesis, Massachusetts Institute of Technology, 1959.
6. I. Langmuir, The Interaction of Electron and Positive Ion Space Charges in Cathode Sheaths, *Phys. Rev.*, **33**: 954-989 (1929).
7. P. L. Auer, Potential Distributions in a Low-Pressure Converter, *J. Appl. Phys.*, **31**: 2096-2103 (1960).
8. R. G. McIntyre, Extended Space-Charge Theory in Low-Pressure Thermionic Converters, *J. Appl. Phys.*, **33**: 2485-2489 (1962).
9. W. T. Norris, Nature of Spontaneous Oscillations in a Cesium Diode Energy Converter, *J. Appl. Phys.*, **35**: 3260-3268 (1964); also P. Burger, Zero-Field Solutions and Their Stability in the One-Dimensional, Low-Pressure Cesium Diode, *J. Appl. Phys.*, **38**: 3360-3368 (1967).
10. C. Wang, D. Lieb, and T. H. Pigford, Extended Space-Charge Theory in the Emitter Sheath, in *IEEE Thermionic Conversion Specialists Conference, Carmel, Calif., Oct. 21–23, 1969*, Report IEEE-69-C-51-ED, pp. 228-236, Institute of Electrical and Electronics Engineers, 1969.

11

Collisional Transport Phenomena

11.1 General Remarks

This chapter presents various analyses of collisional transport phenomena in cesium converters, namely, analyses of transport of charged particles in the interelectrode space under the influence of both long-range electrostatic forces and short-range forces, such as those encountered in elastic and ionization interparticle collisions.

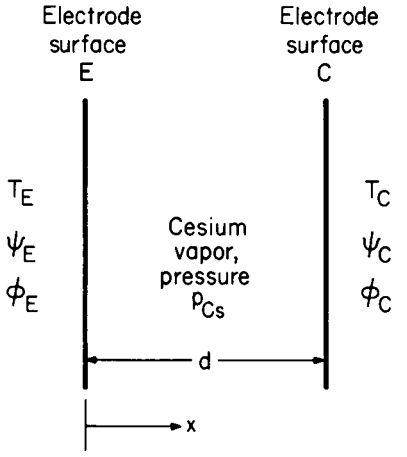


Fig. 11.1 Schematic of cesium vapor system.

Analyzing these phenomena involves considering an open system A consisting of cesium vapor confined between the surfaces just outside two parallel metal electrodes E and C, which are a distance d apart (Fig. 11.1). Since the distance d is much smaller than the linear dimensions of the electrode surfaces, all spatial dependences can be expressed in terms of the distance x from electrode E. The surfaces E and C are characterized by temperatures T_E and T_C, electron motives ψ_E and ψ_C, electron work functions ϕ_E and ϕ_C, and emission currents of electrons and singly charged positive cesium ions corresponding to the values T_E, ϕ_E, T_C, and ϕ_C. The cesium pressure p_{Cs} is greater than the vapor pressure of

either electrode material and is such that the product $p_{Cs}d$ is greater than 1 mil-torr.

The object of the analysis of collisional transport phenomena is to find such quantities as particle densities and electrostatic field as functions of position x. In turn, these functions can be used to obtain relations between charged-particle fluxes and the electron motive difference between the points just outside E and C. These relations can then be combined with results on surfaces, presented in Chap. 9, to obtain performance characteristics of high-pressure diodes. This combination is discussed in Chap. 12.

In steady state and under conditions of interest to thermionic conversion, the different species (electrons, ions, and neutrals) of the vapor system behave like perfect gases. As such, the properties of each species α at each position x can be obtained from a distribution function $f_\alpha(x_\alpha,\mathbf{v}_\alpha)$ in velocity space $\mathbf{v}_\alpha$. Each distribution function $f_\alpha(x_\alpha,\mathbf{v}_\alpha)$ satisfies a steady-state Boltzmann equation that includes, in separate terms, the effects of elastic and inelastic collisions and the effects of the electrostatic field. Each collisional term depends on the distribution functions of all the colliding species. The electrostatic field depends on the distribution functions for electrons and positive ions via Poisson's equation.

The steady-state Boltzmann equations and Poisson's equation for the distribution functions and the electrostatic field are nonlinear and are coupled because of the form of the collisional terms and of the dependence of Poisson's equation on the distribution functions. In addition, the collisional terms are quite involved because of the complexity of the dependence of cross sections on particle velocities. For these reasons, the steady-state Boltzmann equations are not amenable to exact numerical analysis. By making a number of simplifying assumptions, however, we can obtain approximate results that are valid under certain practical conditions of operation of cesium converters. In the discussion that follows, various approximate analyses are considered in order of increasing numerical complexity.

An exception to the numerical difficulties mentioned are the conditions for which the vapor system is in a state of stable equilibrium. For this state, all properties, such as particle densities and electrostatic field, can be computed relatively easily and are completely independent of all the details of the characteristics for elastic and inelastic collisions.

From the standpoint of performance, analysis of stable equilibrium states is impractical since it corresponds to zero output voltage and zero output current of cesium converters. Nevertheless, the analysis brings forth many salient features of the vapor system, such as the formation of sheath and plasma regions, and provides valuable insight and guidance for analyzing practical conditions of operation.

11.2 Stable Equilibrium States of the Cesium Vapor System

We will consider the vapor system A, shown schematically in Fig. 11.1, in a state of stable equilibrium characterized by temperature T and cesium pressure p_{Cs} such that $p_{Cs}d > 1$ mill-torr. In this condition A can be regarded as a heterogeneous mixture of only neutral cesium atoms, cesium positive ions, and electrons. At temperatures and pressures of practical interest, each component of the mixture behaves like a perfect gas, and, to an excellent approximation, cesium positive ions other than singly charged ones can be neglected in the analysis. Thus the cesium particles and the electrons satisfy the chemical equilibrium equation for the reaction

$$Cs \rightleftharpoons Cs^+ + e \tag{11.1}$$

and the mixture can be analyzed by adapting the method and results discussed in Secs. 7.8 to 7.10.

Specifically, if subscripts e, i, and a denote electrons, singly charged positive cesium ions, and neutral cesium atoms, respectively, and if the electron motive is measured with respect to some arbitrary zero reference level, the following results are applicable to the stable equilibrium state under consideration.

The distribution functions f_e, f_i, and f_a at each position x are given by the relations

$$f_e = n_e(0)\left(\frac{m_e}{2\pi kT}\right)^{3/2} \exp\left[\frac{\mu_e - (m_e v_e^2/2) - \psi + \psi_E}{kT}\right]$$

$$= n_e(x^*)\left(\frac{m_e}{2\pi kT}\right)^{3/2} \exp\left[\frac{\mu_e - (m_e v_e^2/2) - \psi + \psi^*}{kT}\right] \tag{11.2}$$

$$f_i = n_i(0)\left(\frac{m_i}{2\pi kT}\right)^{3/2} \exp\left[\frac{\mu_i - (m_i v_i^2/2) + \psi - \psi_E}{kT}\right]$$

$$= n_i(x^*)\left(\frac{m_i}{2\pi kT}\right)^{3/2} \exp\left[\frac{\mu_i - (m_i v_i^2/2) + \psi - \psi^*}{kT}\right] \tag{11.3}$$

$$f_a = n_a\left(\frac{m_a}{2\pi kT}\right)^{3/2} \exp\left[\frac{\mu_a - (m_a v_a^2/2) - V_i}{kT}\right] \tag{11.4}$$

where n_α, m_α, and μ_α are the number density, mass, and total potential, respectively, of species α for $\alpha = e$, i, and a; ψ_E and ψ^* are the motives at the points just outside electrode E and position x^*, respectively; and V_i is the first

ionization energy of a. In these relations the energies $\bar{\varepsilon}_{oe}$ and $\bar{\varepsilon}_{oi}$ for the internal structures of electrons and ions are taken equal to zero. The density n_a of the neutrals and, therefore, the distribution function f_a are uniform throughout the vapor system. The densities $n_e(0)$ and $n_i(0)$ are evaluated at $x = 0$, with ψ equal to ψ_E, and $n_e(x^*)$ and $n_i(x^*)$ are evaluated at $x = x^*$, with ψ equal to ψ^*. The dependence of f_e and f_i on position x is implicit through ψ.

The number densities n_e and n_i at each position x, Poisson's equation for the electron motive ψ, and the electrostatic field E are given by the relations

$$n_e = n_e(0) \exp\left(-\frac{\psi - \psi_E}{kT}\right) = n_e(x^*) \exp\left(-\frac{\psi - \psi^*}{kT}\right) \tag{11.5}$$

$$n_i = n_i(0) \exp\left(\frac{\psi - \psi_E}{kT}\right) = n_i(x^*) \exp\left(\frac{\psi - \psi^*}{kT}\right) \tag{11.6}$$

$$\frac{d^2\psi}{dx^2} = \frac{e^2}{\varepsilon_0}\left[n_i(0) \exp\left(\frac{\psi - \psi_E}{kT}\right) - n_e(0) \exp\left(-\frac{\psi - \psi_E}{kT}\right)\right] \tag{11.7}$$

and

$$E = \frac{1}{e}\frac{d\psi}{dx} \tag{11.8}$$

The total potentials μ_e, μ_i, and μ_a are uniform throughout the vapor and the electrodes and satisfy the chemical equilibrium equation

$$\mu_e + \mu_i = \mu_a$$

Their values are fixed by the values of the corresponding potentials of the electrodes; i.e., $\mu_e = \mu_e^E = \mu_e^C$, $\mu_i = \mu_i^E = \mu_i^C$, and $\mu_a = \mu_a^E = \mu_a^C$, where μ_α^E and μ_α^C are the total potentials of species α of electrodes E and C, respectively. The chemical potentials μ_{ce}, μ_{ci}, and μ_{ca} are such that $\mu_{ce} = \mu_e - \psi$, $\mu_{ci} = \mu_i + \psi$, $\mu_{ca} = \mu_a$, and $\mu_{ce} + \mu_{ci} = \mu_{ca}$.

The motives ψ_E and ψ_C at the points just outside E and C are fixed by the work functions ϕ_E and ϕ_C since $\psi_E = \mu_e + \phi_E$ and $\psi_C = \mu_e + \phi_C$. The densities $n_e(0)$ and $n_i(0)$ are fixed by the motive ψ_E. Equation 11.7 can be solved numerically subject to the boundary conditions $\psi(0) = \psi_E$ and $\psi(d) = \psi_C$.

For values of $\psi_E - \psi_C$ of the order of 1 ev and values of a typical electron number density n_e such that the length $[\varepsilon_0 kT/e^2 n_e]^{1/2}$ is much smaller than the interelectrode spacing d, it turns out that the solutions of Eq. 11.7 assume one of the shapes shown in Fig. 11.2, depending on the values of the ion-richness ratios β_E and β_C at E and C, respectively. For example, when both β_E and β_C are

greater than unity, the shape is represented by curve 1 in Fig. 11.2. In general, when both β_E and β_C are different from unity, the solutions have the following common features. In the central region of the vapor system, the motive is almost constant over a broad range of values of x. This feature implies that the electrostatic field is almost zero and that the electron and ion densities n_e and n_i are almost equal to each other and, therefore, that n_i/n_e at these positions is almost equal to unity. The region is called a plasma. On the other hand, in a region in the vicinity of each electrode surface the motive varies appreciably over a short distance of the order of $[\varepsilon_0 kT/e^2 n_e]^{1/2}$, indicating that the electrostatic field is strong and that the electron and ion densities differ appreciably from each other. Each such region is called a sheath. In a stable equilibrium state the sheaths are always monotonic functions of position, i.e., single valued.

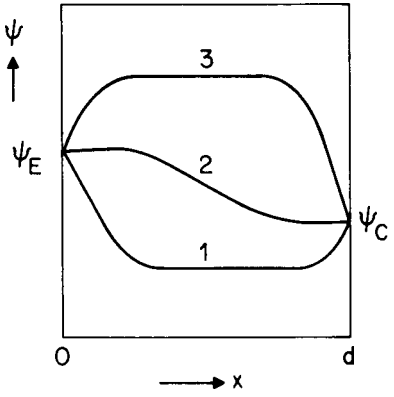


Fig. 11.2 Typical shapes of motives vs. x for cesium vapor system in stable equilibrium states.

As an example, an excellent approximate solution of Eq. 11.7 can be found explicitly under the following circumstances. We will assume that the ion-richness ratios $n_i(0)/n_e(0)$ and $n_i(d)/n_e(d)$ at E and C, respectively, are greater than unity. Then it can be verified by means of Poisson's equation and simple qualitative arguments that the shape of the graph ψ vs. x is that shown schematically in Fig. 11.3. This shape implies that n_i is greater than n_e throughout the vapor system. Starting from E, n_i decreases and n_e increases with distance from $x = 0$ up to a point x^* where the ratio $n_i(x^*)/n_e(x^*)$ is minimum but is greater than unity. At that point $(d\psi/dx)_{x=x^*}$ is equal to zero. Beyond x^*, n_i increases and n_e decreases with distance up to electrode C at $x = d$, where $n_i = n_i(d)$ and $n_e = n_e(d)$.

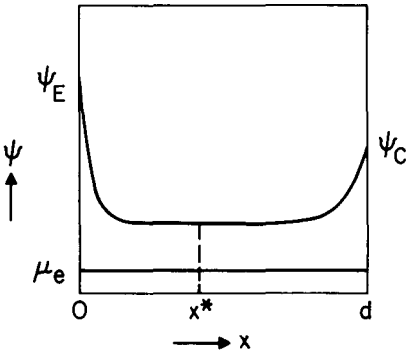


Fig. 11.3 Schematic of motive vs. x for cesium vapor system in a stable equilibrium state and for $\beta_E > 1$ and $\beta_C > 1$.

Introducing the dimensionless quantities

$$\gamma = \frac{\psi - \psi^*}{kT} \tag{11.9}$$

$$\xi = \frac{x}{\lambda_D} \tag{11.10}$$

$$\lambda_D = \left[\frac{\varepsilon_0 kT}{e^2 n_e(x^*)}\right]^{1/2} = 7.45 \times 10^2 \left[\frac{T}{n_e(x^*)}\right]^{1/2} \text{ cm} \quad \text{for } T \text{ in ev and } n_e \text{ in cm}^{-3} \tag{11.11}$$

and

$$\beta^* = \frac{n_i(x^*)}{n_e(x^*)} \tag{11.12}$$

into Poisson's equation (Eq. 11.7), we find

$$\frac{d^2\gamma}{d\xi^2} = \beta^* \exp(\gamma) - \exp(-\gamma) \tag{11.13}$$

Integrating Eq. 11.13, we obtain

$$\left(\frac{d\gamma}{d\xi}\right)^2 = 2[\beta^* \exp(\gamma) + \exp(-\gamma)] + C_1 \tag{11.14}$$

and

$$\int \frac{d\gamma}{\{2[\beta^* \exp(\gamma) + \exp(-\gamma)] + C_1\}^{1/2}} = \xi + C_2 \tag{11.15}$$

where the values of the constants of integration C_1 and C_2 are determined by the boundary conditions.

The quantity λ_D (Eq. 11.11) is called the *Debye length.* If the relation between n_e and p_{Cs} is taken as that given by the Saha equation for an electrically neutral cesium plasma in a stable equilibrium state (Sec. 7.10.1), we see from Eq. 11.11 that, for practical values of T and p_{Cs} of interest to thermionic conversion, λ_D is much smaller than the interelectrode spacing d of a cesium converter, which is of the order of 5 to 20 mils.

In view of the preceding discussion, approximating β^* by unity and applying condition $(d\gamma/d\xi)_{\xi=\xi^*} = 0$ to Eq. 11.14, we find that after some rearrangement Eq. 11.15 yields the approximate relation

$$\frac{\psi}{kT} = 4 \text{ arc tan } h\left[\exp(-\sqrt{2}\,\xi) \tan h\left(\frac{\psi_E}{4\,kT}\right)\right]$$

$$+ 4 \text{ arc tan } h\left\{\exp[\sqrt{2}\,(\xi - \xi_C)] \tan h\left(\frac{\psi_C}{4\,kT}\right)\right\} \tag{11.16}$$

where $\xi^* = x^*/\lambda_D$ and $\xi_C = d/\lambda_D \gg 1$. We see from Eq. 11.16 that for $\xi = 0$ the motive ψ is approximately equal to ψ_E and for $\xi = \xi_C$ the motive ψ is approximately equal to ψ_C. Superimposed on Fig. 11.3 is the value for μ_e, which is uniform throughout the electrodes and the vapor system. For given values of ϕ_E, ϕ_C, T, and p_{Cs}, the relative magnitude of ψ^* with respect to ψ_E and ψ_C can be found by computing the chemical potential μ^*_{ce} for electrons at position x^* and comparing $-\mu^*_{ce}$ with ϕ_E and ϕ_C. When $-\mu^*_{ce}$ is smaller than ϕ_E, then ψ^* is smaller than ψ_E, and, when $-\mu^*_{ce}$ is larger than ϕ_E, then ψ^* is larger than ψ_E. For $\lambda_D \ll d$, the cesium plasma is electrically almost neutral at $x = x^*$. Then μ^*_{ce} is given by the relation

$$\mu^*_{ce} = -\frac{1}{2}\left[V_i + kT \ln \frac{4\,(2\pi m_e kT)^{3/2}\,kT}{h^3 p_{Cs}}\right]$$

$$= -\frac{1}{2}\left(3.89 + T \ln \frac{1.4 \times 10^7\, T^{5/2}}{p_{Cs}}\right) \text{ ev} \qquad \text{for } T \text{ in ev and } p_{Cs} \text{ in torr}$$

For $0.05 < T < 0.2$ ev ($500 < T < 2300°K$) and $0.1 < p_{Cs} < 10$ torr, we see from this relation that μ_{ce}^* is in the range between -3 and -2 ev.

The primary reason for the development of sheath regions is the disparity between the electron and ion densities imposed by the emission properties of the electrodes. This disparity exists also under conditions of steady state, and therefore sheaths develop also under these conditions. The thickness of each sheath is small because, whenever the charged-particle densities differ appreciably, the electrostatic field is strong and tends to equalize the densities by attracting charges of the proper sign from the environs. The equalization occurs over a distance of the order of a Debye length.

The model for the analysis of steady-state conditions is discussed in Sec. 11.4. The analysis of steady states is much more involved than that of stable equilibrium states and is strongly dependent on the values of the cross sections for the various collisional processes. Results can be obtained only by means of approximations. In this respect, the stable-equilibrium-state results provide helpful guidance for the development of approximate theories, presented in Sec. 11.5 and thereafter. Collision cross sections and related quantities needed in the analysis of steady states are discussed in Sec. 11.3.

11.3 Collisional Phenomena

11.3.1 General Remarks. This section presents theoretical and experimental results on cross sections and related quantities of importance to cesium-vapor systems.

When two or more particles come within the range of their mutual force fields, they interact with each other. Hence particles flowing through the cesium-vapor system interact with the particles of the system. The interactions can be regarded as collisions during which the natures and the states of the colliding particles can be altered.

Collisions can be divided into two major categories, elastic and inelastic. An elastic collision between two particles does not affect the natures and states of internal structures of the colliding particles. Thus the total translational kinetic energy and momentum of the two particles is the same before and after the collision. An inelastic collision entails a change in internal structure of one or more of the colliding particles. Thus either the total translational kinetic energy (and momentum) is not the same before and after the collision or the natures of the colliding particles are not the same before and after the collision, or both. Inelastic collisions must be specifically identified in terms of the particles involved before and after the collisions. For example, an ionization collision is defined as the event in which an electron and an atom interact to produce two electrons and a positive ion.

Analyzing the effects of different types of collisions involves calculating reaction rates. Reaction rates can be expressed in terms of physical constants, such as cross sections, each of which is characteristic of a particular collision.

The evaluation of cross sections is a difficult quantum-mechanical problem and is not discussed in this book. Instead, we present the various characteristic aspects of collisional processes of interest by using a mixture of classical and quantum mechanical pictures.

11.3.2 Dynamics of Elastic Binary Collisions. Here we consider elastic collisions between two spherically symmetric particles of masses m_1 and m_2. We assume that the force which either particle exerts on the other is directed along the line joining the centers 1 and 2 of the particles and that it is a function of the distance between 1 and 2.

The terms "before the collision" and "after the collision" denote times during which the particles do not essentially influence each other. The velocities of the particles before and after the collision have definite values denoted by $\mathbf{v}_1$ and $\mathbf{v}_2$ and $\mathbf{v}_1'$ and $\mathbf{v}_2'$, respectively.

The initial and final velocities are interrelated by virtue of conservations of momentum and energy. They satisfy the relations

$$m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = m_1\mathbf{v}_1' + m_2\mathbf{v}_2' \tag{11.17}$$

and

$$\frac{1}{2}(m_1 v_1^2 + m_2 v_2^2) = \frac{1}{2}(m_1 v_1'^2 + m_2 v_2'^2) \tag{11.18}$$

If g and g' denote the magnitudes of the relative velocities $\mathbf{v}_1 - \mathbf{v}_2$ and $\mathbf{v}_1' - \mathbf{v}_2'$, respectively, then elementary algebraic manipulations of Eqs. 11.17 and 11.18 yield that

$$g = g' \tag{11.19}$$

i.e., that the relative velocity $\mathbf{g}_{21} = \mathbf{v}_2 - \mathbf{v}_1$ is changed in direction but not in magnitude as a result of the elastic collision. It follows that the dynamic effect of a collision is known when the change in direction of $\mathbf{g}_{21}$ is determined.

Conservation of momentum and conservation of energy alone are not sufficient to determine the final direction of $\mathbf{g}_{21}' = \mathbf{v}_2' - \mathbf{v}_1'$ with respect to that of $\mathbf{g}_{21}$. It will be seen that this direction depends not only on the initial values $\mathbf{v}_1$ and $\mathbf{v}_2$ but also on two geometric factors.

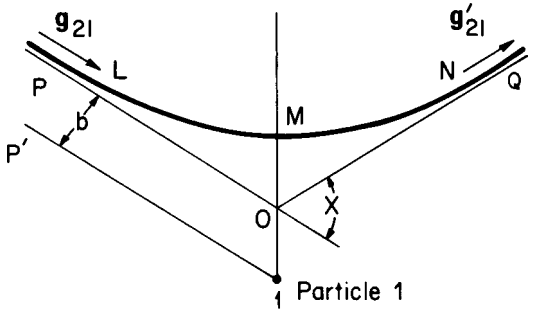


Fig. 11.4 Schematic trajectory (*LMN*) of motion of particle 2 relative to particle 1.

The motion of center 2 of the second particle relative to center 1 of the first particle is shown schematically in Fig. 11.4. Because of the assumed central force field, it is confined to a plane through 1 that contains the vector $\mathbf{g}'_{2\,1}$. The trajectory *LMN* described by 2 relative to 1 approaches asymptotically the lines *PO* and *OQ*, which are directed along the directions $\mathbf{g}_{2\,1}$ and $\mathbf{g}'_{2\,1}$ of the initial and final relative velocities.

The plane of the trajectory *LMN* can be defined by the direction of the line $P'1$, which is drawn parallel to $\mathbf{g}_{2\,1}$, and the polar orientation of the plane about $P'1$. The direction of $P'1$ is fixed by the initial velocities $\mathbf{v}_1$ and $\mathbf{v}_2$. The polar orientation of the plane about $P'1$ is independent of initial velocities, however. It is one of the two additional geometric variables necessary to specify the collision. It will be denoted by the *polar angle* ζ between the plane of *LMN* and a plane containing $P'1$ and a direction fixed in space.

Finally, the angle X through which $\mathbf{g}_{2\,1}$ is deflected depends, in general, on the magnitude g of $\mathbf{g}_{2\,1}$ and on the distance b between $P'1$ and PO (Fig. 11.4). The distance b, called the *impact parameter,* is the second geometric variable of the collision.

The functional relation between X, b, and g depends on the force law of interaction between 1 and 2. Two results corresponding to two special types of interactions are discussed in Secs. 11.3.4 and 11.3.5. We will see that, if the states of the two colliding particles are as prescribed by the specified kinematic and geometric variables, one elastic collision will occur. Quantum-mechanical considerations, however, prohibit the prescription of the state of a particle with such specificity. Instead, the state must be prescribed in terms of probabilities, and therefore the collisional event cited will have a probability associated with its occurrence. This probability can be evaluated as follows:

We will assume that the steady state of particle 1 is described by a distribution function $f_1(\mathbf{r}, \mathbf{v}_1)$ such that $f_1 d\mathbf{v}_1\, d^3\mathbf{r}_1$ is the probability that 1 will be in a small volume $d^3\mathbf{r}_1$ around some position $\mathbf{r}_1$ and with a velocity between $\mathbf{v}_1$ and

$\mathbf{v}_1 + d\mathbf{v}_1$. Similarly, we will assume that the steady state of particle 2 is described by a distribution function $f_2(\mathbf{r},\mathbf{v}_2)$. For a collision of the type discussed previously to occur in a small interval of time dt, particle 2 must occupy a volume equal to $(g\,dt)\,b\,db\,d\zeta$, as shown schematically in Fig. 11.5. It follows that the probability for particle 2 to be in such a volume with velocity between $\mathbf{v}_2$ and $\mathbf{v}_2 + d\mathbf{v}_2$ is given by the expression $f_2\,d\mathbf{v}_2\,gb\,db\,d\zeta\,dt$.

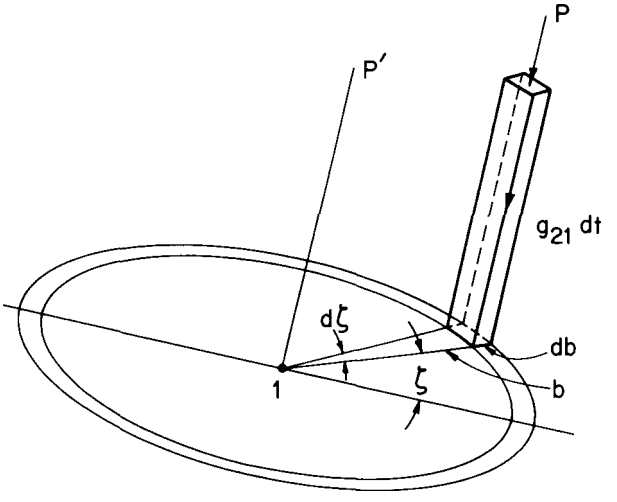


Fig. 11.5 Schematic of volume occupied by particle 2 colliding with particle 1.

If the states of 1 and 2 are statistically independent, the probability of the elastic collision under consideration is the product of the probabilities for the two states. Accordingly, per unit volume and per unit velocity range of 1 and per unit time, the occurrence of the elastic collision must be weighted by the factor dC given by the relation

$$dC = f_1 f_2\,d\mathbf{v}_2\,gb\,db\,d\zeta \tag{11.20}$$

This factor can be integrated with respect to $\mathbf{v}_2$, b, and ζ to yield the expected reaction rate per unit volume and per unit velocity range of 1, as was done in Chap. 8 in connection with the Boltzmann equation.

Expressions similar to Eq. 11.20 can be written for other collisions, elastic or inelastic, except that the geometric factor $b\,db\,d\zeta$, which has the dimensions of area, will have a different form and a different operational interpretation. Each of these geometric factors defines a differential cross section.

11.3.3 Cross Section, Free Path, and Collision Frequency. The geometric factor $b\,db\,d\zeta$ that appears in the expression dC for an elastic collision (Eq. 11.20) defines a *differential scattering cross section* $\sigma'(g,X)$ over a differential solid angle $d^2\Omega$ by means of the relations

$$\sigma'(g,X)\,d^2\Omega = b\,db\,d\zeta \tag{11.21}$$

and

$$d^2\Omega = \sin X\,dX\,d\zeta \tag{11.22}$$

or equivalently

$$\sigma'(g,X) = \frac{b}{\sin X}\frac{db}{dX} \tag{11.23}$$

The dimension of $\sigma'(g,X)$ is that of area.

The integral $\sigma(g)$ of $\sigma'(g,X)$ over all solid angles $d^2\Omega$ is called the *total microscopic elastic scattering cross section* and is given by the relation

$$\sigma(g) = \int d^2\Omega\,\sigma'(g,X) = 2\pi \int \sigma'(g,X)\sin X\,dX \tag{11.24}$$

The quantity $\sigma(g)$ has the dimension of area.

Another cross section of practical interest is that for momentum transfer during an elastic collision. Relative to 1, the momentum change of 2 in the direction $\mathbf{g}_{21}$ is given by the expression $[m_1m_2/(m_1 + m_2)]\,\mathbf{g}_{12}\,(1 - \cos X)$. Accordingly, a cross section $\sigma_m(g)$ for momentum transfer can be defined by the relation

$$\sigma_m(g) = \int d^2\Omega\,\sigma'(g,X)\,(1 - \cos X) \tag{11.25}$$

Total microscopic cross sections $\sigma_j(g)$ can be defined for various collisional processes, one for each type j of elastic or inelastic collision between two particles. The sum of all these determines the overall total microscopic cross section $\sigma_t(g)$, namely,

$$\sigma_t(g) = \sum_j \sigma_j(g) \tag{11.26}$$

Microscopic cross sections can be given a number of simple interpretations. If all particles 1 and 2 have speeds 0 and v and densities N and n, respectively, and if their interactions are characterized by a microscopic cross section $\sigma(v)$, then f_1

and f_2 in Eq. 11.20 are delta functions in velocity space. Hence, integrating Eq. 11.20, we find that the reaction rate dn/dt per unit volume is given by the relations

$$\frac{dn}{dt} = \sigma(v)\,(Nnv) \tag{11.27a}$$

$$= \frac{nv}{1/[N\,\sigma(v)]} \tag{11.27b}$$

$$= n[vN\,\sigma(v)] \tag{11.27c}$$

Equation 11.27a indicates that the number of reactions per unit volume per unit length $v\,dt$ is proportional to the densities N and n. The constant of proportionality is $\sigma(v)$, and therefore $\sigma(v)$ can be interpreted as the number of interactions per unit length $v\,dt$, per particle 1, and per particle 2.

Equation 11.27b indicates that the reaction rate per unit volume can be found by dividing the total length nv traveled by n particles 2 per unit volume and per unit time by the length $\lambda(v)$ defined by the relation

$$\lambda(v) = \frac{1}{N\,\sigma(v)} \tag{11.28}$$

assuming that particles 2 experience one collision every time they travel a distance λ. The quantity λ is called a *free path.* Various free paths can be defined, one for each collisional process j corresponding to a cross section $\sigma_j(v)$. By virtue of Eq. 11.26, free paths add in parallel, i.e.,

$$\frac{1}{\lambda_t} = \sum_j \frac{1}{\lambda_j} \tag{11.29}$$

Finally, Eq. 11.27c suggests that the reaction rate per unit volume can be found by multiplying the density n by the frequency $\nu(v)$ defined by the relation

$$\nu(v) = vN\,\sigma(v) = \frac{v}{\lambda(v)} \tag{11.30}$$

The quantity $\nu(v)$ is called the *collision frequency.* Collision frequencies of different processes are additive.

11.3.4 Hard-Sphere Elastic Collisions. Here we consider elastic collisions in which the force acts only upon contact of the colliding particles and in which the particles are represented by spheres of diameters d_1 and d_2. These are called *hard-sphere elastic collisions.*

When the colliding particles contact each other, the relative velocity $\mathbf{g}_{12}$ changes into the direction $\mathbf{g}'_{12}$. Here the distance d_{12} between the centers of the spheres at collision, the diameters d_1 and d_2, and the impact parameter b are related by the expressions

$$d_{12} = \frac{1}{2}(d_1 + d_2) \tag{11.31}$$

and

$$b = d_{12} \cos \frac{X}{2} \tag{11.32}$$

This collision model is unique in that X depends only on b and not on g.

Substituting Eq. 11.32 into Eqs. 11.23 to 11.25, we find that

$$\sigma' = \frac{d_{12}^2}{4} \tag{11.33}$$

$$\sigma = \pi d_{12}^2 = \frac{\pi}{4}(d_1 + d_2)^2 \tag{11.34}$$

and

$$\sigma_m = \pi d_{12}^2 \tag{11.35}$$

In words, in the hard-sphere elastic collision model the total microscopic cross sections σ and σ_m are equal to the area corresponding to a circle with a radius equal to the sum of the radii of the colliding particles. For identical spherical particles, σ will be the cross sectional area of the particle.

11.3.5 Coulomb Scattering. Here we consider two particles interacting via a central force, i.e., the forces $\mathbf{F}_1$ and $\mathbf{F}_2$ on the two particles of masses m_1 and m_2 at positions $\mathbf{r}_1$ and $\mathbf{r}_2$, given by

$$\mathbf{F}_1 = m_1 \ddot{\mathbf{r}}_1 \quad \text{and} \quad \mathbf{F}_2 = m_2 \ddot{\mathbf{r}}_2 = -\mathbf{F}_1 \tag{11.36}$$

where the double dot indicates the second derivative with respect to time t.

Multiplying Eqs. 11.36 by m_2 and m_1, respectively, and subtracting the results, we obtain

$$\mathbf{F}_1 = m_r(\ddot{\mathbf{r}}_1 - \ddot{\mathbf{r}}_2) \tag{11.37}$$

where m_r is the *reduced mass* given by the relation

$$m_r = \frac{m_1 m_2}{m_1 + m_2} \tag{11.38}$$

We see from Eq. 11.37 that the scattering process can be described by the motion of a particle with a reduced mass m_r about a fixed point at a distance $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ and acted upon by the central force.

The cross vector product $\mathbf{F}_1 \times \mathbf{r}$ is equal to 0 because the force is along the direction $\mathbf{r}$. This product can be rewritten in the form

$$\frac{d}{dt}(\mathbf{r} \times \dot{\mathbf{r}}) = 0 \tag{11.39}$$

and therefore

$$\mathbf{r} \times \dot{\mathbf{r}} = \text{constant} \tag{11.40}$$

where $\dot{\mathbf{r}}$ is the relative velocity, i.e., $\dot{\mathbf{r}} = \dot{\mathbf{r}}_1 - \dot{\mathbf{r}}_2$. Equation 11.40 provides a formal proof of the statement made earlier that throughout the collision the two particles move in the same plane; here the plane is perpendicular to the constant vector $\mathbf{r} \times \dot{\mathbf{r}}$. Hence only a two-dimensional problem need be solved.

The geometry of the scattering process is shown schematically in Fig. 11.6.

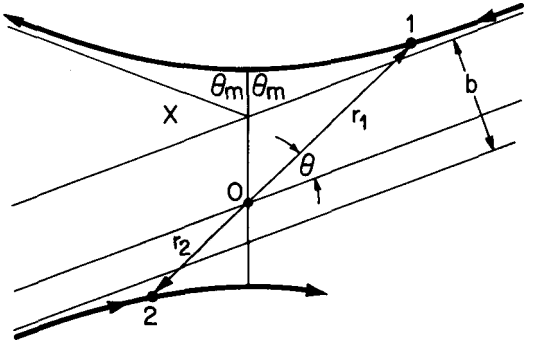


Fig. 11.6 Schematic of trajectories of two particles in a central-field scattering collision.

Point O is the center of gravity of the two particles and moves with velocity

$$\mathbf{v} = \frac{m_1}{m_1 + m_2}\,\dot{\mathbf{r}}_1 + \frac{m_2}{m_1 + m_2}\,\dot{\mathbf{r}}_2$$

Particles 1 and 2 are separated by a total distance $\mathbf{r}$, and the ratio of the lengths of the segments $O1$ and $O2$ is equal to m_2/m_1. The angle between the line joining the two particles when closest together and the line between the two particles initially is denoted by θ_m, and the deflection X of particle 1 with respect to particle 2 is given by the relation

$$X = \pi - 2\theta_m \tag{11.41}$$

The total energy and total angular momentum about point O are constant at all times and are equal to their respective values before the initiation of the collision. In polar coordinates r and θ, these two invariants are given by the relations

$$\frac{1}{2} m_r(\dot{r}^2 + r^2\dot{\theta}^2) + \phi(r) = \frac{1}{2} m_r g^2 \tag{11.42}$$

and

$$m_r r^2 \dot{\theta} = m_r b g \tag{11.43}$$

where $\phi(r)$ is the potential energy of the central force field and g the initial relative speed. Eliminating the time variable between Eqs. 11.42 and 11.43, we find the equation of the trajectory in polar coordinates, i.e.,

$$\frac{dr}{d\theta} = \pm \frac{r^2}{b}\left[1 - \frac{b^2}{r^2} - \frac{2\,\phi(r)}{m_r g^2}\right]^{1/2} \tag{11.44}$$

The trajectory is symmetric about the line of closest approach. The minus sign in $\pm$ applies when the particles are approaching each other, and the plus sign applies when they are receding from each other.

At the position of closest approach the distance $r_{\min}$ is the minimum positive root of the relation

$$1 - \frac{b^2}{r^2} - \frac{2\,\phi(r)}{m_r g^2} = 0 \tag{11.45}$$

Thus the angle θ_m and the deflection X are given by the relations

$$\theta_m = \int_{r_{\min}}^{\infty} dr \frac{d\theta}{dr} \tag{11.46}$$

and

$$X(g,b) = \pi - 2b \int_{r_{\min}}^{\infty} \frac{dr/r^2}{\{1 - (b^2/r^2) - [2\,\phi(r)/m_r g^2]\}^{1/2}} \tag{11.47}$$

For elastic collisions between an electron of mass m_e and a singly charged, heavy positive ion of mass m_i, the reduced mass m_r is approximately equal to m_e, the potential energy $\phi(r)$ is equal to $-e^2/4\pi\varepsilon_0 r$, the relative velocity g is approximately equal to v_e, and Eq. 11.47 yields

$$\cot \frac{X}{2} = \frac{4\pi\varepsilon_0 m_e v_e^2 b}{e^2} \tag{11.48}$$

Substituting b and db/dX from Eq. 11.48 into Eq. 11.23, we find that

$$\sigma'(v_e,X) = \frac{e^4}{[8\pi\varepsilon_0 m_e v_e^2 \sin^2 (X/2)]^2} \tag{11.49}$$

This is the so-called Rutherford scattering relation for a Coulomb elastic collision between an electron and a singly charged positive ion.

Similarly, substituting Eq. 11.49 into Eq. 11.25, we find that the total cross section σ_m for momentum transfer is given by the relation

$$\sigma_m = \frac{1}{4\pi} \int_{X=\delta}^{\pi} \frac{e^4}{(\varepsilon_0 m_e v_e^2)^2} \frac{d \sin (X/2)}{\sin (X/2)}$$

$$= \frac{e^4}{4\pi (\varepsilon_0^2 m_e v_e^2)^2} \ln \frac{1}{\sin (\delta/2)} \tag{11.50}$$

We see from Eq. 11.50 that, if the smallest scattering angle δ is equal to 0 (i.e., if the maximum value of b is infinity), then the integral for σ_m diverges logarithmically. The divergence is due to the assumed infinite range of the Coulomb field. For the problems of interest here, this assumption cannot be valid for all interparticle distances.

When electrons and ions are in states near stable equilibrium, the force on each charged particle can be obtained from the motive that obeys Poisson's equation. This force is not Coulombic and does not contribute to scattering events. Moreover, it is valid when correlation and exchange effects are negligible, i.e., when the effects of a test particle (electron or ion) on the charge distributions surrounding it can be neglected. Correlation and exchange effects can be neglected when the mutual interaction between far-distant charged particles is evaluated. On the other hand, when two charged particles are near each other, their charge distributions are highly correlated and correlation effects cannot be neglected. Following Bohm and Pines,[1] we will assume that correlations between electrons and ions extend over a spherical region of radius equal to a Debye length determined by the density n_e and temperature T_e of the corresponding stable equilibrium state of the electrons. Within this sphere, the interaction is described approximately by Coulomb's law.

By virtue of these remarks, the Debye length can be used in Eq. 11.48 as the maximum value of the impact parameter that yields the minimum scattering angle δ. Thus δ and σ_m are given by the relations

$$\delta = \left(\sin\frac{X}{2}\right)_{\min} = \frac{e^2}{4\pi\varepsilon_0 m_e v_e^2}\left(\frac{e^2 n_e}{\varepsilon_0 kT_e}\right)^{1/2} = \frac{e^2}{12\pi\varepsilon_0 kT_e}\left(\frac{e^2 n_e}{\varepsilon_0 kT_e}\right)^{1/2} \tag{11.51}$$

and

$$\sigma_m = \frac{e^4 \ln \Lambda_e}{4\pi\varepsilon_0^2 (m_e v_e^2)^2} \tag{11.52}$$

where

$$\Lambda_e = \frac{12\pi (\varepsilon_0 kT_e/e^2)^{3/2}}{n_e^{1/2}} \tag{11.53}$$

and $\ln \Lambda_e$ is called the *Coulomb logarithm for electrons.* In writing the second part of Eq. 11.51, we have replaced $m_e v_e^2$ with $3kT_e$. The reasons for this are that σ_m depends logarithmically on δ and that the preceding semiclassical derivation of σ_m is not valid for small values of v_e where quantum-mechanical effects predominate.

A similar procedure can be used for ion–ion Coulomb scattering. Then we find that σ_m is given by the relation

$$\sigma_m = \frac{e^4 \ln \Lambda_i}{\pi \varepsilon_0^2 (m_i g^2)^2} \tag{11.54}$$

where

$$\ln \Lambda_i = \frac{12\pi (\varepsilon_0 k T_i / e^2)^{3/2}}{n_i^{1/2}} \tag{11.55}$$

11.3.6 Experimental and Computed Cesium Cross Sections. This section presents experimental and computed values of cesium cross sections of interest to thermionic conversion. The discussion follows the review prepared by Houston.[2]

(a) Neutral–Neutral Elastic Scattering. Experimental values of neutral–neutral cross section for cesium–cesium collisions have been reviewed by Houston,[2] who concluded that a cross section of about 110 Å^2 is appropriate in analysis of cesium-vapor systems.

(b) Ion–Neutral Elastic Scattering. Most experimental values of ion–neutral cross sections are derived from measurements of the ion mobility M_i (see Secs. 11.5.4 and 11.5.5). The relation between mobility M_i and the average ion–neutral momentum transfer cross section $\bar{\sigma}_{ia}$ is given by the expression

$$M_i = \frac{3\pi^{1/2} e}{8 n_a \bar{\sigma}_{ia} (m_i k T_i)^{1/2}}$$

This expression is practically identical to that given in Table 11.5, Sec. 11.5.5, for hard-sphere ion–neutral collisions with a cross section $\sigma_{ia} = \bar{\sigma}_{ia}$.

Most published values for the mobility of cesium ions[2] are near 0.07 cm^2/volt-sec. This agreement is somewhat deceptive, however, since M_i is a function of the ion temperature T_i. The results of Sheldon's study[3] of the temperature dependence of M_i, expressed as ratios with respect to M_i at 300°K, are given in the second column of Table 11.1; the corresponding values of M_i and $\bar{\sigma}_{ia}$ are given in columns 3 and 4 of the table.[2]

(c) Electron–Neutral Elastic Scattering. Many experimental values and theoretical calculations of the electron–neutral cesium-atom elastic scattering cross section have been reported in the literature. From his review of the data, Houston[2] concluded that the value of 1400 collisions per centimeter, corresponding to a cross section of about 400 Å^2, is a reasonable average of the theoretical and experimental results in the electron temperature range from

Table 11.1 Temperature Dependence of Cs^+ Mobility M_i

T, °K	$\frac{M_i(T)}{M_i(300°K)}$†	$M_i(T)$,‡ cm²/volt-sec	$\bar{\sigma}_{ia}$,‡ Å²
300	1.000	0.089	1470
600	0.782	0.070	1320
800	0.702	0.063	1270
1000	0.646	0.058	1240
1200	0.605	0.054	1210
1400	0.572	0.051	1190
1600	0.544	0.049	1160
1800	0.520	0.047	1140
2000	0.501	0.045	1130

†Data from Ref. 3.
‡Data from Ref. 2.

2000 to 4000°K. The cross section appears to fall as the electron temperature is decreased below 2000°K, and a value of about 300 Å² might be more appropriate at 1000°K. There is still a large spread in the available data, however, and further high-quality experimental work is needed.

(d) Electron–Neutral Ionization Collisions. In cesium-vapor systems of interest to thermionic conversion, very few electrons have energies above the ionization threshold of about 3.9 ev. For this reason the most important portion of the ionization-cross-section vs. electron-energy curve is that corresponding to the energy range between 4 and 6 ev. This portion can be approximated by a straight line. After reviewing both experimental and theoretical data, Houston[2] concluded that about 4 Å²/volt is a reasonable value for the slope of this line.

11.3.7 Average Values, Relaxation Times, and Mean Free Paths. In general, cross sections and, therefore, free paths and collision frequencies depend on the relative speed g of the colliding particles of masses m_1 and m_2. For particles in states described by distribution functions in velocity space, many expressions characteristic of collisional processes (e.g., collision rates) involve integrals over velocities of some function $G(g)$ of the relative speed g times the distribution functions of the colliding particles. The evaluation of such integrals is facilitated by the following considerations.

An integral C_{12} for a collisional process between particles 1 and 2 can be expressed in the form

$$C_{12} = n_1 n_2 \overline{g\, G(g)} \tag{11.56}$$

where

$$\overline{g\,G(g)} = \begin{cases} \frac{1}{2}\int_{\mathbf{v}_1} d^3\mathbf{v}_1 \int_{\mathbf{v}_2} d^3\mathbf{v}_2\, f_1(\mathbf{v}_1) f_2(\mathbf{v}_2)\, g\, G(g) & \text{for } m_1 = m_2 \quad (11.57) \\ \int_{\mathbf{v}_1} d^3\mathbf{v}_1 \int_{\mathbf{v}_2} d^3\mathbf{v}_2\, f_1(\mathbf{v}_1) f_2(\mathbf{v}_2)\, g\, G(g) & \text{for } m_1 \neq m_2 \quad (11.58) \end{cases}$$

and where n_1 and n_2 are the number densities and f_1 and f_2 the distribution functions per particle of particles 1 and 2, respectively. For example, if $G(g)$ is the total cross section, C_{12} is the total collision rate. This rate can be expressed as the product of the densities times the so-called *average value* $\overline{g\,G(g)}$. The factor $\frac{1}{2}$ appears in the expression for $m_1 = m_2$ because the integral counts each collision twice.

When both f_1 and f_2 are Maxwellian distributions at temperature T, such that

$$f_\alpha = \left(\frac{m_\alpha}{2\pi kT}\right)^{3/2} \exp\left(-\frac{m_\alpha v_\alpha^2}{2kT}\right) \quad \text{for } \alpha = 1 \text{ and } 2$$

then, after some simple algebraic manipulations, we find that

$$\overline{g\,G(g)} = \begin{cases} \dfrac{1}{2\,(\pi m_1)^{1/2}\,(kT)^{3/2}} \displaystyle\int_0^\infty dU\, U\, G(U) \exp\left(-\frac{U}{2kT}\right) & \text{for } m_1 = m_2 \quad (11.59) \\ \dfrac{4}{(2\pi m_1)^{1/2}\,(kT)^{3/2}} \displaystyle\int_0^\infty dU\, U\, G(U) \exp\left(-\frac{U}{kT}\right) & \text{for } m_1 \ll m_2 \quad (11.60) \end{cases}$$

where $U \equiv m_1\, g^2/2$. For $m_1 \ll m_2$, Eq. 11.60 is entirely equivalent to finding the average value by integrating over the distribution f_1 only, because then g is approximately equal to v_1.

A useful application of Eqs. 11.59 and 11.60 is in the calculation of relaxation times. For a given collisional process, a *relaxation time* can be defined as the average value of the time between collisions, i.e., the average value of the inverse collision frequency.† It is a measure of the time required for a nonequilibrium distribution to approach a stable-equilibrium-state distribution as a result of the collisional process in question.

†For some problems it may be more appropriate to define relaxation time as the inverse of the average value of the collision frequency.

For momentum-transfer Coulomb collisions, the inverse collision frequency $1/\nu_{12}$ and the relaxation time t_{m12} are given by the relations

$$\frac{1}{\nu_{12}} = \frac{1}{n_2 \sigma_m g} = \frac{4\pi\varepsilon_0^2 m_r^2 g^3}{e^4 n_2 \ln \Lambda_1} \tag{11.61}$$

and

$$t_{m12} = \int_{v_1} d^3 v_1 \int_{v_2} d^3 v_2 \frac{1}{\nu_{12}} f_1(v_1) f(v_2) \tag{11.62}$$

Substituting $G(g) \propto g^2$ in Eqs. 11.59 and 11.60, we find that the relaxation times t_{mee} for electron–electron, t_{mii} for ion–ion, and t_{mei} for electron–ion momentum-transfer Coulomb collisions are given by the relations

$$t_{mee} = \frac{16(\pi)^{1/2} \varepsilon_0^2 m_e^{1/2} (kT)^{3/2}}{e^4 n_e \ln \Lambda_e} \tag{11.63}$$

$$t_{mii} = \frac{16(\pi)^{1/2} \varepsilon_0^2 m_i^{1/2} (kT)^{3/2}}{e^4 n_i \ln \Lambda_i} \tag{11.64}$$

and

$$t_{mei} = \frac{32(2\pi)^{1/2} \varepsilon_0^2 m_e^{1/2} (kT)^{3/2}}{e^4 n_i \ln \Lambda_e} \tag{11.65}$$

Comparing t_{mee} with t_{mii}, we conclude that electrons relax much faster by means of electron–electron collisions than ions do by means of ion–ion collisions. For example, electrons relax about 500 times as fast as cesium ions when both are at the same temperature and have the same densities. Similarly, comparing t_{mee} with t_{mei}, we conclude that t_{mee} is $\frac{1}{2}\sqrt{2}$ as large as t_{mei} for the same electron and ion densities. Finally, we can show that t_{mie} is approximately equal to $(m_i/m_e)t_{mei}$ and, therefore, momentum transfer from ions to electrons takes much longer than momentum transfer from electrons to ions.

Relaxation times for energy transfer by means of Coulomb collisions can also be defined, but the calculations are much more involved and are not discussed here. Nevertheless, it is noteworthy that electron–ion energy exchange occurs over a much longer time than electron–electron momentum transfer.

For hard-sphere elastic collisions between electrons and heavy particles characterized by a cross section σ_{ea}, the relaxation time t_{ea} is given by the relation

$$t_{ea} = \frac{(2m_e)^{1/2}}{n_a \sigma_{ea} (\pi kT)^{1/2}} \tag{11.66}$$

For practical conditions this time is much larger than t_{mee}.

Each relaxation time t_{12} can be used to define a corresponding average collision frequency $\bar{\nu}_{12}$ and an average or mean free path $\bar{\lambda}_{12}$ by means of the relations

$$\bar{\nu}_{12} = \frac{1}{t_{12}} \tag{11.67}$$

and

$$\bar{\lambda}_{12} = \bar{g} t_{12} \tag{11.68}$$

where $\bar{g}$ is the average value of g. For $m_1 = m_2$, $\bar{g}$ is equal to $4(kT/\pi m_1)^{1/2}$, and for $m_1 \ll m_2$, $\bar{g}$ is equal to $(8kT/\pi m_1)^{1/2}$.

These definitions of average values of collision frequencies and free paths are not unique. For example, an average collision frequency $\bar{\nu}_{12}$ and an average free path $\bar{\lambda}_{12}$ can be defined by means of the relations

$$\bar{\nu}_{12} = n_2 \bar{\sigma}_{12} g \quad \text{and} \quad \bar{\lambda}_{12} = \frac{1}{n_2 \bar{\sigma}_{12}} \tag{11.69}$$

where $\bar{\sigma}_{12}$ is the average cross section. Which average is appropriate depends on the problem and the form in which the solution of the problem is sought.

11.4 General Formulation of Steady-State Transport Phenomena

We will consider the cesium-vapor system A, shown in Fig. 11.1, in a steady state defined by practical values of T_E, ϕ_E, T_C, ϕ_C, and p_{Cs}. For such values, A can be regarded as a heterogeneous mixture of perfect gases of electrons (e), singly charged positive cesium ions (i), and neutral cesium atoms (a). The properties of each species α can be obtained from the distribution function $f_\alpha(x_\alpha, v_\alpha)$. Each distribution function satisfies a steady-state Boltzmann equation.

The Boltzmann equations for f_e, f_i, and f_a are given by the relations

$$v_{ex} \frac{\partial f_e}{\partial x_e} - \frac{eE}{m_e} \frac{\partial f_e}{\partial v_{ex}} = C_{ee} + C_{ei} + C_{ea} + C^{\text{in}} \tag{11.70}$$

$$v_{ix} \frac{\partial f_i}{\partial x_i} + \frac{eE}{m_i} \frac{\partial f_i}{\partial v_{ix}} = C_{ie} + C_{ii} + C_{ia} + C^{\text{in}} \tag{11.71}$$

and

$$v_{ax} \frac{\partial f_a}{\partial x_a} = C_{ae} + C_{ai} + C_{aa} - C^{\text{in}} \tag{11.72}$$

where $v_{\alpha x}$ is the component of velocity $\mathbf{v}_\alpha$ in the x direction; $C_{\alpha\beta}$ is the collision integral for elastic collisions between species α and β for $\alpha,\beta = e$, i, and a; and C^{in} is the inelastic collision integral for the formation of electrons and singly charged positive ions from neutral cesium atoms. The elastic collision integrals and the inelastic collision integral are functionals of the cross sections and the distribution functions.

The number densities n_e, n_i, and n_a at each position x, Poisson's equation for the electron motive ψ, and the electrostatic field E are given by the relations

$$n_\alpha = \int_{\mathbf{v}_\alpha} f_\alpha \, d^3\mathbf{v}_\alpha \qquad \text{for } \alpha = e, i, \text{ and } a \tag{11.73}$$

$$\frac{d^2\psi}{dx^2} = \frac{e^2}{\varepsilon_0} (n_e - n_i) \tag{11.74}$$

and

$$E = \frac{1}{e} \frac{d\psi}{dx} \tag{11.75}$$

Solving the coupled, nonlinear set of integrodifferential equations (11.70 to 11.75) would yield the distribution functions f_e, f_i, and f_a and the motive ψ. These functions would in turn specify the relation between a fixed net current density J and the internal voltage drop ΔV given by the expressions

$$J = J_e - J_i \tag{11.76}$$

and

$$e \, \Delta V = \psi_E - \psi_C \tag{11.77}$$

where

$$J_\alpha = e \int_{\mathbf{v}_\alpha} v_{\alpha x} f_\alpha \, d^3\mathbf{v}_\alpha \qquad \text{for } \alpha = e \text{ and } i \tag{11.78}$$

In principle, Eqs. 11.70 to 11.75 can be solved once the boundary conditions are specified. For electrons and ions the boundary conditions at the electrode surfaces E and C (Fig. 11.1) are given by the relations

$$f_\alpha(0,\mathbf{v}_\alpha) = n_\alpha(0)\left(\frac{m_\alpha}{2\pi kT_E}\right)^{3/2} \exp\left(-\frac{m_\alpha v_\alpha^2}{2kT_E}\right) u(v_{\alpha x}) \qquad \text{for } \alpha = e \text{ and } i \tag{11.79}$$

and

$$f_\alpha(d,\mathbf{v}_\alpha) = n_\alpha(d)\left(\frac{m_\alpha}{2\pi kT_C}\right)^{3/2} \exp\left(-\frac{m_\alpha v_\alpha^2}{2kT_C}\right) u(-v_{\alpha x}) \qquad \text{for } \alpha = e \text{ and } i \tag{11.80}$$

where $u(x)$ is the unit step function defined by the relation

$$u(x) = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x < 0 \end{cases} \tag{11.81}$$

and the densities $n_\alpha(0)$ and $n_\alpha(d)$ are determined by the emission properties of E and C. Specifically, if the electron- and ion-emission-flux densities at E are denoted by Γ_{ve}^E and Γ_{vi}^E, respectively, and at C by Γ_{ve}^C and Γ_{vi}^C, respectively, then $n_\alpha(0)$ and $n_\alpha(d)$ for $\alpha = e$ and i are given by the relations

$$n_\alpha(0) = \left(\frac{2\pi m_\alpha}{kT_E}\right)^{1/2} \Gamma_{v\alpha}^E \qquad \text{for } \alpha = e \text{ and } i \tag{11.82}$$

and

$$n_\alpha(d) = \left(\frac{2\pi m_\alpha}{kT_C}\right)^{1/2} \Gamma_{v\alpha}^C \qquad \text{for } \alpha = e \text{ and } i \tag{11.83}$$

The electron-emission-flux densities are determined from the Richardson equation:

$$e\Gamma_{ve}^E = AT_E^2 \exp\left(-\frac{\phi_E}{kT_E}\right) \quad \text{and} \quad e\Gamma_{ve}^C = AT_C^2 \exp\left(-\frac{\phi_C}{kT_C}\right)$$

where $A = 120$ amp/cm^2-°K^2. The ion-emission-flux densities are determined from the Saha–Langmuir equation:

$$\Gamma_{vi}^{E} = \frac{p_{Cs}}{(2\pi m_a kT_E)^{1/2}\{1 + 2\exp[(V_i - \phi_E)/kT_E]\}}$$

and

$$\Gamma_{vi}^{C} = \frac{p_{Cs}}{(2\pi m_a kT_C)^{1/2}\{1 + 2\exp[(V_i - \phi_C)/kT_C]\}}$$

The following considerations greatly simplify the analysis of neutral cesium atoms. For conditions of interest to thermionic conversion, only a small fraction of cesium atoms is ionized and the net neutral-atom flux is equal to zero. It follows that the density n_a is much greater than both n_e and n_i throughout the vapor system. In addition, the exchange of energy between electrons and neutrals is negligible because m_e is much smaller than m_a. For these reasons the values of the collision integrals C_{ae}, C_{ai}, and C^{in} can be shown to be much smaller than the value of C_{aa}, and therefore C_{ae}, C_{ai}, and C^{in} can be neglected from the right-hand side of Eq. 11.72. When these approximations are introduced, the Boltzmann equation for the neutrals is decoupled from the others and can be solved independently. A zero-order approximation f_{a0} for the distribution function of the neutrals is given by the relation

$$f_{a0} = n_a\left(\frac{m_a}{2\pi kT_a}\right)^{3/2}\exp\left(-\frac{m_a v_a^2}{2kT_a}\right) \tag{11.84}$$

where the density n_a and temperature T_a are functions of position x only. The dependence of n_a and T_a on x can be determined from the hydrodynamic equations (Sec. 8.8.6). When C_{ae}, C_{ai}, and C^{in} are neglected, the equations for the particle flux density Γ_a, pressure p_a, and energy-flux density w_a become

$$\frac{d\Gamma_a}{dx} = 0 \qquad \frac{dp_a}{dx} = 0 \qquad \frac{dw_a}{dx} = 0 \tag{11.85}$$

with

$$p_a = n_a kT_a \quad \text{and} \quad w_a = -k_a\frac{dT_a}{dx} \tag{11.86}$$

where k_a is the thermal conductivity of the neutral cesium atoms. Integrating Eqs. 11.85 subject to the conditions $\Gamma_a = 0$, $p_a = p_{Cs}$, $T_a(0) = T_E$, and $T_a(d) = T_C$, we find

$$\Gamma_a = 0 \tag{11.87}$$

$$p_{Cs} = n_a k T_a \quad \text{and} \quad w_a = \frac{1}{d}\int_{T_C}^{T_E} k_a \, dT_a = -k_a \frac{dT_a}{dx} \tag{11.88}$$

Equations 11.88 determine the dependence of n_a and T_a on x and, therefore, the state of the neutrals via Eq. 11.84.

In summary, Eqs. 11.70 to 11.75 can be solved for f_e and f_i subject to boundary conditions 11.79 and 11.80 and for f_{a0} given by Eq. 11.84. Unfortunately it has not been possible to carry out this program, and no general solution has ever been found. To obtain solutions we must introduce a number of additional approximations. These approximations are discussed in the remaining sections of this chapter.

11.5 Solutions of Linearized Boltzmann Equations

In this section we obtain solutions of the Boltzmann equations (11.70 and 11.71) under conditions for which all or some of the phases of the cesium-vapor system A (Fig. 11.1) are in steady states near the corresponding stable equilibrium states. We will show in Sec. 11.6 that under practical conditions of operation such solutions are valid for the plasma but not for the sheath regions of A.

11.5.1 Linearized Boltzmann Equations. As discussed in Sec. 8.8.7, Eqs. 11.70 and 11.71 can be linearized and solved subject to four restrictions. The first restriction is that the distributions f_e, f_i, and f_a can be approximated by the expressions

$$f_e = f_{e0}[1 + \tilde{\phi}_e(x_e, \mathbf{v}_e)] \qquad f_i = f_{i0}[1 + \tilde{\phi}_i(x_i, \mathbf{v}_i)] \qquad f_a = f_{a0} \tag{11.89}$$

where

$$f_{\alpha 0} = n_\alpha \left(\frac{m_\alpha}{2\pi k T_\alpha}\right)^{3/2} \exp\left(-\frac{m_\alpha v_\alpha^2}{2kT_\alpha}\right) \quad \text{for } \alpha = e, i, \text{ and } a \tag{11.90}$$

$$n_\alpha = \int f_\alpha \, d^3\mathbf{v}_\alpha \qquad 3n_\alpha k T_\alpha = \int m_\alpha v_\alpha^2 f_\alpha \, d^3\mathbf{v}_\alpha \qquad p_\alpha = n_\alpha k T_\alpha$$
$$\text{for } \alpha = e, i, \text{ and } a \tag{11.91}$$

and $\tilde{\phi}_\alpha(x_\alpha, \mathbf{v}_\alpha)$ is a slowly varying scalar function of x_α and $\mathbf{v}_\alpha$ such that

$$|\tilde{\phi}_\alpha(x_\alpha, \mathbf{v}_\alpha)| \ll 1 \quad \text{for } \alpha = e \text{ and } i \tag{11.92}$$

In general, the particle density n_α, temperature T_α, and pressure p_α are functions of position x and are characteristic of the corresponding stable equilibrium state at each position x.

The other three restrictions in Sec. 8.8.7 refer to approximate expressions for the collision integrals and the equality of T_i and T_a.

Substituting distributions 11.89 into Eqs. 11.70 and 11.71, using the approximations resulting from the restrictions, and neglecting second-order terms in $\tilde{\phi}_\alpha$ and terms involving gradients of $\tilde{\phi}_\alpha$, we find

$$\frac{v_{ex} f_{e0}}{p_e}\left[\frac{dp_e}{dx} + en_eE + \left(U_e^2 - \frac{5}{2}\right)n_e k\frac{dT_e}{dx}\right] = -\left[\nu_{ei}(v_e) + \nu_{ea}(v_e)\right] f_{e0}\tilde{\phi}_e - n_e I_e(\tilde{\phi}_e) \quad (11.93)$$

$$\frac{v_{ix} f_{i0}}{p_i}\left[\frac{dp_i}{dx} - en_iE + \left(U_i^2 - \frac{5}{2}\right)n_i k\frac{dT_i}{dx}\right] = -n_i^2\, I_i(\tilde{\phi}_i) - n_i n_a\, I_{ia}(\tilde{\phi}_i) \quad (11.94)$$

where

$$\mathbf{U}_a = \left(\frac{m_\alpha}{2kT_\alpha}\right)^{1/2} \mathbf{v}_\alpha \qquad \text{for } \alpha = e \text{ and } i \quad (11.95)$$

The linear integral operators I_e, I_i, and I_{ia} are defined by the relations

$$I_\alpha[G(\mathbf{v}_\alpha)] \equiv \frac{1}{n_\alpha^2}\int_\Omega d^2\Omega \int_{\mathbf{v}} d^3\mathbf{v}\, f_{\alpha 0}(x,\mathbf{v}_\alpha) f_{\alpha 0}(x,\mathbf{v}) \times \left[G(\mathbf{v}_\alpha) + G(\mathbf{v}) - G(\mathbf{v}'_\alpha) - G(\mathbf{v}')\right] \times |\mathbf{v}_\alpha - \mathbf{v}|\, \sigma_{\alpha\alpha}(|\mathbf{v}_\alpha - \mathbf{v}|, X) \qquad \text{for } \alpha = e \text{ and } i \quad (11.96)$$

and

$$I_{ia}[G(\mathbf{v}_i)] \equiv \frac{1}{n_i n_a}\int_\Omega d^2\Omega \int_{\mathbf{v}_a} d^3\mathbf{v}_a\, f_{i0}(x,\mathbf{v}_i) f_{a0}(x,\mathbf{v}_a) \times \left[G(\mathbf{v}_i) - G(\mathbf{v}'_i)\right]\, |\mathbf{v}_i - \mathbf{v}_a|\, \sigma_{ia}(|\mathbf{v}_i - \mathbf{v}_a|, X) \quad (11.97)$$

$G(\mathbf{v}_\alpha)$ can be any scalar or vector function of $\mathbf{v}_\alpha$, $\sigma_{\alpha\alpha}$ and σ_{ia} are elastic collision cross sections, and $\nu_{e\beta}$ is the collision frequency defined by Eq. 8.171. It is noteworthy that the operators I_α and $I_{\alpha\beta}$ satisfy the symmetry relation

$$\int_{\mathbf{v}_\alpha} H(\mathbf{v}_\alpha)\, I\,[G(\mathbf{v}_\alpha)]\, d^3\mathbf{v}_\alpha = \int_{\mathbf{v}_\alpha} G(\mathbf{v}_\alpha)\, I\,[H(\mathbf{v}_\alpha)]\, d^3\mathbf{v}_\alpha \qquad \text{for } I = I_\alpha \text{ and } I_{\alpha\beta} \quad (11.98)$$

where $H(\mathbf{v}_\alpha)$ is any scalar or vector function of $\mathbf{v}_\alpha$.

The expressions in the left-hand sides of Eqs. 11.93 and 11.94 suggest that the solutions $\tilde{\phi}_e$ and $\tilde{\phi}_i$ can be written in the forms

$$\tilde{\phi}_e(x,\mathbf{v}_\alpha) = -\frac{1}{n_e p_e}\left(\frac{2kT_e}{m_e}\right)^{1/2}\left[A_e(U_e)\left(\frac{dp_e}{dx} + en_eE\right)\right.$$

$$\left. + B_e(U_e)\, n_e k \frac{dT_e}{dx}\right]\frac{U_{ex}}{U_e}$$

$$= -\frac{1}{n_e p_e}\left(\frac{2kT_e}{m_e}\right)^{1/2}\left[\mathbf{A}_e \cdot \mathbf{F}_e + \mathbf{B}_e \cdot n_e k\, \nabla T_e\right] \tag{11.99}$$

and

$$\tilde{\phi}_i(x,\mathbf{v}_i) = -\frac{1}{n_i p_i}\left(\frac{2kT_i}{m_i}\right)^{1/2}\left[A_i(U_i)\left(\frac{dp_i}{dx} - en_iE\right) + B_i(U_i)\, n_i k \frac{dT_i}{dx}\right]\frac{U_{ix}}{U_i}$$

$$= -\frac{1}{n_i p_i}\left(\frac{2kT_i}{m_i}\right)^{1/2}\left[\mathbf{A}_i \cdot \mathbf{F}_i + \mathbf{B}_i \cdot n_i k\, \nabla T_i\right] \tag{11.100}$$

where $A_\alpha(U_\alpha)$ and $B_\alpha(U_\alpha)$ for $\alpha = e$ and i are scalar functions of the magnitude of $\mathbf{U}_\alpha$, $\mathbf{A}_\alpha \equiv A_\alpha(U_\alpha)\, \mathbf{U}_\alpha/U_\alpha$, $\mathbf{B}_\alpha \equiv B_\alpha(U_\alpha)\, \mathbf{U}_\alpha/U_\alpha$, and $\mathbf{F}_e$ and $\mathbf{F}_i$ are vectors having components only along the x direction such that

$$F_{ex} \equiv \frac{dp_e}{dx} + en_eE \quad \text{and} \quad F_{ix} \equiv \frac{dp_i}{dx} - en_iE \tag{11.101}$$

and the gradients ∇T_e and ∇T_i are also along the x direction only. Substituting Eqs. 11.99 and 11.100 into Eqs. 11.93 and 11.94, we find that

$$\frac{\mathbf{U}_e f_{e0}}{n_e} = I_e(\mathbf{A}_e) + \frac{f_{e0}}{n_e^2}\left[\nu_{ei}(v_e) + \nu_{ea}(v_e)\right]\mathbf{A}_e \tag{11.102}$$

$$\frac{\mathbf{U}_e f_{e0}}{n_e}\left(U_e^2 - \frac{5}{2}\right) = I_e(\mathbf{B}_e) + \frac{f_{e0}}{n_e^2}\left[\nu_{ei}(v_e) + \nu_{ea}(v_e)\right]\mathbf{B}_e \tag{11.103}$$

$$\frac{\mathbf{U}_i f_{i0}}{n_i} = I_i(\mathbf{A}_i) + \frac{n_a}{n_i} I_{ia}(\mathbf{A}_i) \tag{11.104}$$

and

$$\frac{\mathbf{U}_i f_{i0}}{n_i}\left(U_i^2 - \frac{5}{2}\right) = I_i(\mathbf{B}_i) + \frac{n_a}{n_i} I_{ia}(\mathbf{B}_i) \tag{11.105}$$

It follows that the problem of determining the distribution functions f_e and f_i is reduced to that of finding four scalar functions, $A_e(U_e)$, $B_e(U_e)$, $A_i(U_i)$, and $B_i(U_i)$, which are solutions of Eqs. 11.102 to 11.105. When electron–electron collisions are negligible, i.e., when the integral I_e is negligible, Eqs. 11.102 and 11.103 can be readily solved for $A_e(U_e)$ and $B_e(U_e)$, and f_e can therefore be readily computed. These solutions, which correspond to the so-called Lorentz plasmas, are discussed in Sec. 11.5.4.

In general, the solutions of Eqs. 11.102 to 11.105 cannot be obtained explicitly, and therefore it is necessary to resort to techniques for generating approximate solutions. One such technique, which involves the expansion of $A_\alpha(U_\alpha)$ and $B_\alpha(U_\alpha)$ into series of Sonine polynomials, is discussed in Sec. 11.5.2.

11.5.2 Sonine Polynomial Expansion for Distribution Functions.[4] To find approximate solutions to Eqs. 11.102 to 11.105, which, in turn, yield the distribution functions f_e and f_i, we will expand the scalar functions $A_\alpha(U_\alpha)/U_\alpha$ and $B_\alpha(U_\alpha)/U_\alpha$ into series of the complete orthonormal Sonine polynomials of order 3/2:

$$\frac{A_\alpha(U_\alpha)}{U_\alpha} = \sum_{n=0}^{\infty} a_{\alpha n} S_n^{3/2}(U_\alpha^2) \tag{11.106}$$

and

$$\frac{B_\alpha(U_\alpha)}{U_\alpha} = \sum_{n=0}^{\infty} b_{\alpha n} S_n^{3/2}(U_\alpha^2) \tag{11.107}$$

where $a_{\alpha n}$ and $b_{\alpha n}$ are expansion coefficients and $S_n^{3/2}(z)$ is the nth Sonine polynomial of order 3/2 and argument z.

Substituting series 11.106 and 11.107 into Eqs. 11.102 to 11.105, vector-dot-multiplying the results by $S_m^{3/2}(U_\alpha^2)\,\mathbf{U}_\alpha$ for $m = 0, 1, 2, \ldots$, and then integrating over all velocity space $\mathbf{v}_\alpha$, we find four infinite sets of linear algebraic equations of the forms

$$\sum_{n=0}^{\infty} \delta_e(m,n)\, a_{en} = \beta_{em} \tag{11.108}$$

$$\sum_{n=0}^{\infty} \delta_e(m,n)\, b_{en} = \gamma_{em} \tag{11.109}$$

$$\sum_{n=0}^{\infty} \delta_i(m,n)\, a_{in} = \beta_{im} \tag{11.110}$$

$$\sum_{n=0}^{\infty} \delta_i(m,n)\, b_{in} = \gamma_{im} \tag{11.111}$$

where $\delta_\alpha(m,n)$, $\beta_{\alpha m}$, and $\gamma_{\alpha m}$ for $\alpha = e$ and i are known quantities given by the relations

$$\delta_e(m,n) \equiv \int_{\mathbf{v}_\alpha} S_m^{(1/2)}(U_e^2)\, \mathbf{U}_e \cdot I_e\, [S_n^{(1/2)}(U_e^2)\, \mathbf{U}_e]\, d^3\mathbf{v}_e$$

$$+ \frac{1}{n_e^2} \int_{\mathbf{v}_\alpha} U_e^2\, f_{e0}\, [\nu_{ei}(v_e) + \nu_{ea}(v_e)]\, S_m^{(1/2)}(U_e^2)\, S_n^{(1/2)}(U_e^2)\, d^3\mathbf{v}_e \tag{11.112}$$

$$\delta_i(m,n) \equiv \int_{\mathbf{v}_\alpha} S_m^{(1/2)}(U_i^2)\, \mathbf{U}_i \cdot I_i\, [S_n^{(1/2)}(U_i^2)\, \mathbf{U}_i]\, d^3\mathbf{v}_i$$

$$+ \frac{n_a}{n_i} \int_{\mathbf{v}_\alpha} S_m^{(1/2)}(U_i^2)\, \mathbf{U}_i \cdot I_{ia}\, [S_n^{(1/2)}(U_i^2)\, \mathbf{U}_i]\, d^3\mathbf{v}_i \tag{11.113}$$

$$\beta_{\alpha m} \equiv \frac{1}{n_\alpha} \int_{\mathbf{v}_\alpha} S_m^{(1/2)}(U_\alpha^2)\, U_\alpha^2 f_{\alpha 0}\, d^3\mathbf{v}_\alpha$$

$$= \begin{cases} 3/2 & \text{for } m = 0 \\ 0 & \text{for } m \neq 0 \end{cases} \quad \text{for } \alpha = e \text{ and } i \tag{11.114}$$

and

$$\gamma_{\alpha m} \equiv \frac{1}{n_\alpha} \int_{\mathbf{v}_\alpha} S_m^{(1/2)}(U_\alpha^2)\, U_\alpha^2 \left(U_\alpha^2 - \frac{5}{2}\right) f_{\alpha 0}\, d^3\mathbf{v}_\alpha$$

$$= \begin{cases} -15/4 & \text{for } m = 1 \\ 0 & \text{for } m \neq 1 \end{cases} \quad \text{for } \alpha = e \text{ and } i \tag{11.115}$$

By virtue of Eqs. 11.108 to 11.111, the problem of solving the integral equations (11.102 to 11.105) for $A_\alpha(U_\alpha)$ and $B_\alpha(U_\alpha)$ is reduced to that of solving four infinite sets of algebraic equations for the values of the expansion coefficients a_{en}, b_{en}, a_{in}, and b_{in}, for $n = 0, 1, 2, \ldots$. Approximate solutions to any desired degree of accuracy can be obtained by truncating the Sonine polynomial series expansions after N terms and finding the $4N$ coefficients involved.

In principle, the matrix elements $\delta_\alpha(m,n)$ can be evaluated once the collision laws are specified. Specifically, the first term in the expression for $\delta_e(m,n)$ and

both terms in the expression for $\delta_i(m,n)$ are special cases of a general set of collision integrals evaluated and tabulated by Chapman and Cowling[5] and by Hirshfelder, Curtiss, and Bird.[6] The remaining terms in the expression for $\delta_e(m,n)$ can be evaluated relatively easily.

11.5.3 Transport Differential Equations. As discussed in Sec. 8.8.7, the only hydrodynamic equations that are consistent with the restrictions leading to linearization of the Boltzmann equations for f_e and f_i are the equation for the electron kinetic energy flux w_e and those for the electron and ion current densities J_e and J_i. They are given by Eqs. 8.180 and 8.181, which are repeated here for convenience:

$$\frac{dw_e}{dx} = -J_e E + \int \frac{m_e v_e^2}{2} C^{\text{in}} \, d^3\mathbf{v}_e \tag{11.116}$$

and

$$\frac{dJ_e}{dx} = \frac{dJ_i}{dx} = e \int C^{\text{in}} \, d^3\mathbf{v}_e \tag{11.117}$$

The densities w_e, J_e, and J_i can be written in terms of particle densities, temperatures, and electrostatic field and a set of transport coefficients. When the distribution functions are given by Eqs. 11.89 and 11.99 to 11.101, then

$$w_e = \int \frac{m_e v_e^2}{2} v_{ex} f_e \, d^3\mathbf{v}_e = \left(\frac{5}{2} kT_e + k_e^T kT_e\right) \frac{J_e}{e} - \kappa_e \frac{dT_e}{dx} \tag{11.118}$$

$$J_e \equiv e \int v_{ex} f_e \, d^3\mathbf{v}_e = -M_e \left[\frac{dp_e}{dx} + en_e E + k_e^T n_e k \frac{dT_e}{dx}\right] \tag{11.119}$$

and

$$J_i \equiv e \int v_{ix} f_i \, d^3\mathbf{v}_i = -M_i \left[\frac{dp_i}{dx} - en_i E + k_i^T n_i k \frac{dT_i}{dx}\right] \tag{11.120}$$

where M_α and k_α^T are the mobility and thermal diffusion ratio for species α, respectively, and κ_e is the electron thermal conductivity. The transport coefficients M_α, k_α^T, and κ_e are defined by the relations

$$M_\alpha \equiv \frac{2e}{3n_\alpha^2 m_\alpha} \int_{\mathbf{v}_\alpha} U_\alpha A_\alpha(U_\alpha) f_{\alpha 0} \, d^3\mathbf{v}_\alpha \tag{11.121}$$

$$k_\alpha^T \equiv \frac{2e}{3n_\alpha^2 m_\alpha M_\alpha} \int_{\mathbf{v}_\alpha} U_\alpha B_\alpha(U_\alpha) f_{\alpha 0} \, d^3\mathbf{v}_\alpha \tag{11.122}$$

and

$$\kappa_e \equiv \frac{M_e n_e k^2 T_e}{e} \left[\frac{2e}{3n_e^2 m_e M_e} \int_{\mathbf{v}_\alpha} U_e^3 B_e(U_e) f_{\alpha 0} \, d^3\mathbf{v}_\alpha - k_e^T \left(\frac{5}{2} + k_e^T \right) \right] \tag{11.123}$$

In the derivation of Eqs. 11.118 to 11.120, use has been made of the relation

$$\int_{\mathbf{v}_\alpha} \left(U_\alpha^2 - \frac{5}{2} \right) f_{\alpha 0} \mathbf{A}_\alpha \cdot \mathbf{U}_\alpha \, d^3\mathbf{v}_\alpha = \int_{\mathbf{v}_\alpha} f_{\alpha 0} \mathbf{B}_\alpha \cdot \mathbf{U}_\alpha \, d^3\mathbf{v}_\alpha \tag{11.124}$$

which follows from Eqs. 11.102 to 11.105 and the symmetry relation 11.98.

When the distribution functions are expressed in terms of Sonine polynomial series, then the transport coefficients M_α, k_α^T, and κ_e are given by the relations

$$M_\alpha = \frac{e a_{\alpha 0}}{n_\alpha m_\alpha} \qquad k_\alpha^T = \frac{b_{\alpha 0}}{a_{\alpha 0}} \qquad \text{and} \qquad \kappa_e = \frac{5k^2 T_e}{2m_e} \left(b_{e1} + \frac{2}{5} k_e^T b_{e0} \right) \tag{11.125}$$

It is noteworthy that the transport coefficients depend explicitly on only a small number of Sonine expansion coefficients regardless of how many terms are retained in the Sonine polynomial expansions, but the values of these coefficients depend on the number of terms retained in the expansions. The simplicity of Eqs. 11.125 is a consequence of the use of Sonine polynomials of order 3/2; if other orthogonal polynomials had been used, Eqs. 11.125 would contain terms corresponding to an infinite number of expansion coefficients.

The values of the expansion coefficients and, therefore, of the transport coefficients depend on the collision cross sections. Some representative results are discussed in Sec. 11.5.4.

Finally, before we establish the range of validity of the solutions of the linearized Boltzmann equations, it is instructive to express the particle and energy fluxes in terms of thermodynamic potentials or forces, i.e., to write the phenomenological equations.

The total potentials μ_e and μ_i are given by the relations (Eq. 7.218):

$$\mu_e = -kT_e \ln \frac{2(2\pi m_e)^{3/2}(kT_e)^{5/2}}{p_e h^3} + \psi = -kT_e \ln \frac{2(2\pi m_e kT_e)^{3/2}}{n_e h^3} + \psi$$

and

$$\mu_i = -kT_i \ln \frac{(2\pi m_i)^{3/2}(kT_i)^{5/2}}{p_i h^3} - \psi = -kT_i \ln \frac{(2\pi m_i kT_i)^{3/2}}{n_i h^3} - \psi$$

Using these expressions in the equations for the energy-flux density u_e and particle-flux densities Γ_e and Γ_i, we find, after some elementary algebra, that

$$u_e = w_e + \psi\Gamma_e = \frac{M_e n_e T_e}{e}\left\{\left[\left(\frac{5}{2}kT_e + k_e^T kT_e + \psi\right)^2 + k_e T_e^2\right]\tau_e' - \left(\frac{5}{2}kT_e + k_e^T kT_e + \psi\right)(\mu_e\tau_e)'\right\} \quad (11.126)$$

$$\Gamma_e = \frac{J_e}{e} = \frac{M_e n_e T_e}{e}\left[\left(\frac{5}{2}kT_e + k_e^T kT_e + \psi\right)\tau_e' - (\mu_e\tau_e)'\right] \quad (11.127)$$

and

$$\Gamma_i = \frac{J_i}{e} = \frac{M_i n_i T_i}{e}\left[\left(\frac{5}{2}kT_i + k_i^T kT_i - \psi\right)\tau_i' - (\mu_i\tau_i)'\right] \quad (11.128)$$

where $\tau_\alpha = 1/T_\alpha$ and a prime denotes a derivative with respect to x. These are the phenomenological equations excluding the ion-energy-flux density. The coefficients of the forces τ_α' and $(-\mu_\alpha\tau_\alpha)'$ are the conductivities L_{ij}. It is seen that the coefficient of $(-\mu_e\tau_e)'$ in the electron-energy-flux equation (Eq. 11.126) and the coefficient of τ_e' in the electron-flux equation (Eq. 11.127) are equal, i.e., they satisfy the necessary Onsager reciprocal relation.

11.5.4 Evaluation of Transport Coefficients. This section presents values of transport coefficients evaluated for special types of collisional processes that are representative of certain conditions of operation of the cesium-vapor system.

When electron–electron elastic collisions are negligible compared with other types of collisions, then we can neglect $I_e(\mathbf{A}_e)$ and $I_e(\mathbf{B}_e)$ from Eqs. 11.102 and 11.103, and we find that the electron perturbation function $\tilde{\phi}_e(x,\mathbf{v}_e)$ and the transport coefficients M_e, k_e^T, and κ_e are given by the relations

$$\begin{aligned}\tilde{\phi}_e(x,\mathbf{v}_e) &= -\frac{v_{ex}}{\nu_e(v_e)}\left[\frac{1}{p_e}\frac{dp_e}{dx} + \frac{eE}{kT_e} + \left(U_e^2 - \frac{5}{2}\right)\frac{1}{T_e}\frac{dT_e}{dx}\right] \\ &= \frac{v_{ex}}{k\,\nu_e(v_e)}\left[\left(\psi + \frac{m_e v_e^2}{2}\right)\tau_e' - (\mu_e\tau_e)'\right] \\ &= \frac{v_{ex}}{k\,\nu_e(v_e)}\left[\left(\frac{m_e v_e^2}{2} - \mu_{ce}\right)\tau_e' - \tau_e\mu_e'\right] \end{aligned} \quad (11.129)$$

$$M_e = \frac{e}{p_e} \int_{\mathbf{v}_e} \frac{v_e^2}{3\nu_e(v_e)} f_{e0}\, d^3\mathbf{v}_e \tag{11.130}$$

$$k_e^T = \frac{e}{M_e p_e} \int_{\mathbf{v}_e} \frac{v_e^2}{3\nu_e(v_e)} \left(\frac{m_e v_e^2}{2kT_e} - \frac{5}{2}\right) f_{e0}\, d^3\mathbf{v}_e \tag{11.131}$$

and

$$\kappa_e = \frac{M_e k p_e}{e} \left[\frac{e}{M_e p_e} \int_{\mathbf{v}_e} \frac{v_e^2}{3\nu_e(v_e)} \left(\frac{m_e v_e^2}{2kT_e}\right)^2 f_{e0}\, d^3\mathbf{v}_e - \left(\frac{5}{2} + k_e^T\right)^2\right] \tag{11.132}$$

where the collision frequency $\nu_e(v_e)$ is defined by the relation

$$\nu_e(v_e) \equiv \nu_{ei}(v_e) + \nu_{ea}(v_e) \tag{11.133}$$

These results could also be expressed in terms of the free path

$$\lambda_e(v_e) = \frac{v_e}{\nu_e(v_e)}$$

The last part of Eq. 11.129 indicates that $|\tilde{\phi}_e(x,\mathbf{v}_e)|$ is much smaller than unity when each of the two gradient-dependent terms is much smaller than unity or, equivalently,

$$\left|\frac{v_{ex}}{k\,\nu_e(v_e)} \left(\frac{m_e v_e^2}{2} - \mu_{ce}\right) \tau_e'\right| = \left|\frac{v_{ex}\,\lambda_e(v_e)}{kv_e} \left(\frac{m_e v_e^2}{2} - \mu_{ce}\right) \tau_e'\right| \ll 1 \tag{11.134}$$

and

$$\left|\frac{v_{ex}}{k\,\nu_e(v_e)}\, \tau_e \mu_e'\right| = \left|\frac{v_{ex}\,\lambda_e(v_e)}{kv_e}\, \tau_e \mu_e'\right| \ll 1 \tag{11.135}$$

For a given collision frequency, Eqs. 11.134 and 11.135 indicate that the limitations on the gradients τ_e' and μ_e' are less restrictive for small values of velocity than they are for large values of velocity. On the other hand, for the steady states of interest, the fraction of electrons that have large velocities is small, and therefore large velocities are unimportant.

Table 11.2 gives values of the transport coefficients M_e, k_e^T, and κ_e computed by means of Eqs. 11.121 to 11.123 for three so-called Lorentz plasmas for which collisions occur between electrons and one species of heavy particles. The heavy particles for the three plasmas are: (1) neutrals with a collision frequency $\nu_e(v_e)$ = constant = ν_{ea};(2) neutrals with a free path $\lambda_e(v_e)$ = constant = λ_{ea}, i.e.,

Table 11.2 Transport Coefficients for Lorentz Plasmas

Transport coefficient	Heavy-particle species and collision law		
	Neutrals, constant ν_{ea}	Neutrals, constant λ_{ea}	Ions, Coulomb force law
M_e	$\frac{e}{m_e\nu_{ea}}$	$\frac{e}{m_e\hat{\nu}_{ea}}$	$\frac{e}{m_e\hat{\nu}_{ei}}$
k_e^T	0	$-1/2$	$-3/2$
κ_e	$\frac{5n_ek^2T_e}{2m_e\nu_{ea}}$	$\frac{2n_ek^2T_e}{m_e\hat{\nu}_{ea}}$	$\frac{4n_ek^2T_e}{m_e\hat{\nu}_{ei}}$

the collision frequency $\nu_e(v_e)$ is proportional to v_e; and (3) ions with a collision force described by Coulomb's force law. The results for cases 2 and 3 are presented in terms of effective electron–neutral and electron–ion collision frequencies $\hat{\nu}_{ea}$ and $\hat{\nu}_{ei}$, respectively, defined by the relations

$$\hat{\nu}_{ea} \equiv \frac{3\pi^{1/2}n_a\sigma_{ea}(kT_e)^{1/2}}{2^{3/2}m_e^{1/2}} \qquad \hat{\nu}_{ei} \equiv \frac{2^{1/2}n_ie^4 \ln \Lambda_e}{128\pi^{1/2}\varepsilon_0^2 m_e^{1/2}(kT_e)^{3/2}} \tag{11.136}$$

where σ_{ea} is the electron–neutral hard-sphere collision cross section and $\ln \Lambda_e$ is the Coulomb logarithm for electrons.

Truncation errors associated with terminating the Sonine polynomial expansions after N terms can be estimated by using the Sonine expansion technique to evaluate approximate values for the transport coefficients of a Lorentz plasma and comparing the results with the exact values given in Table 11.2. The comparison for a Lorentz plasma with constant λ_{ea} reveals (Table 11.3) that retaining three terms in each of the Sonine expansions yields values for the three transport coefficients, M_e, k_e^T, and κ_e, which are within 10% of the exact values given in Table 11.2.

Table 11.3 Transport Coefficients for a Lorentz Plasma with Constant λ_{ea}

Number of terms in Sonine expansion	$\frac{M_e}{(M_e)_{\text{exact}}}$	$\frac{k_e^T}{(k_e^T)_{\text{exact}}}$	$\frac{\kappa_e}{(\kappa_e)_{\text{exact}}}$
$N = 1$	0.88	0	0
$N = 2$	0.95	0.77	0.85
$N = 3$	0.98	0.90	0.93

When electron–electron elastic collisions are not negligible, values of the transport coefficients have been computed by means of the Sonine polynomial expansion technique for a mixture of electrons, ions, and neutrals such that: (1) the elastic collisions between charged particles can be treated as Coulomb interactions; (2) the charged-particle–neutral-particle elastic collisions can be treated as hard-sphere interactions; and (3) $n_e \approx n_i$. Under these conditions the matrices $[\delta_e(m,n)]$ and $[\delta_i(m,n)]$ of Eqs. 11.108 to 11.111 can be readily evaluated. If three terms ($N = 3$) are retained in each of the Sonine polynomial expansions, $[\delta_e(m,n)]$ and $[\delta_i(m,n)]$ are given by the relations

$$[\delta_e(m,n)] = \frac{\hat{\nu}_{ea}}{n_e}\begin{bmatrix} 1.70 + 5.09\, r_e & -0.85 + 7.64\, r_e & -0.21 + 9.55\, r_e \\ -1.85 + 7.64\, r_e & 5.52 + 23.8\, r_e & -2.44 + 27.4\, r_e \\ -0.21 + 9.55\, r_e & -2.44 + 27.4\, r_e & 11.5 + 54.7\, r_e \end{bmatrix} \tag{11.137}$$

$$[\delta_i(m,n)] = \frac{\hat{\nu}_{ia}}{n_i}\begin{bmatrix} 1.70 & -0.42 & -0.05 \\ -0.42 & 6.25 + 10.2\, r_i & -1.85 + 7.64\, r_i \\ -0.05 & -1.87 + 7.64\, r_i & 13.8 + 28.6\, r_i \end{bmatrix} \tag{11.138}$$

where the effective ion–neutral and ion–ion collision frequencies $\hat{\nu}_{ia}$ and $\hat{\nu}_{ii}$ are defined by the relations

$$\hat{\nu}_{ia} \equiv \frac{3\pi^{1/4} n_a \sigma_{ia} (kT_i)^{1/2}}{4m_i^{1/2}} \qquad \hat{\nu}_{ii} \equiv \frac{n_i e^4 \ln \Lambda_i}{128\pi^{1/2}\varepsilon_0^2 m_i^{1/2} (kT_i)^{3/2}} \tag{11.139}$$

and σ_{ia} = ion–neutral hard-sphere cross section
$\ln \Lambda_i$ = Coulomb logarithm for ions
$r_e \equiv \hat{\nu}_{ei}/\hat{\nu}_{ea}$
$r_i \equiv \hat{\nu}_{ii}/\hat{\nu}_{ia}$

Using these matrices in Eqs. 11.108 to 11.111, we find the values of a_{e0}, b_{e0}, b_{e1}, a_{i0}, and b_{i0}, which, in turn, can be used in Eqs. 11.125 to obtain the values of the transport coefficients. These values are listed in Tables 11.4 and 11.5.

(a) Electron Transport Coefficients. Column 2 of Table 11.4 lists the values of the electron transport coefficients in the limit of $r_e = 0$, i.e., when no ions are present. They are denoted by the subscript 0. Column 3 in the table lists the values of the electron transport coefficients in the limit $r_e = \infty$, i.e., for a fully ionized plasma. They are denoted by the subscript ∞ and are in excellent agreement with results reported in the literature. Comparing these transport coefficients with those given in Table 11.2 for an electron–ion Lorentz plasma reveals that the effect of electron–electron collisions in a fully ionized plasma is

Table 11.4 Electron Transport Coefficients

Transport coefficient	Limit $r_e \ll 1$	Limit $r_e \gg 1$	General expression
M_e	$M_{e0} = \dfrac{0.98e}{m_e \hat{\nu}_{ea}}$	$M_{e\infty} = \dfrac{0.58e}{m_e \hat{\nu}_{ei}}$	$M_e = \dfrac{M_{e0} M_{e\infty}}{M_{e0} + M_{e\infty}} M_e(r_e)$
k_e^T	$k_{e0}^T = -0.45$	$k_{e\infty}^T = 0.71$	$k_e^T(r_e)$
κ_e	$\kappa_{e0} = \dfrac{1.88 n_e k^2 T_e}{m_e \hat{\nu}_{ea}}$	$\kappa_{e\infty} = \dfrac{0.93 n_e k^2 T_e}{m_e \hat{\nu}_{ei}}$	$\kappa_e = \dfrac{\kappa_{e0} \kappa_{e\infty}}{\kappa_{e0} + \kappa_{e\infty}} \kappa_e(r_e)$

to reduce M_e and κ_e to about 0.5 of the value obtained in the absence of electron–electron collisions. Column 4 of Table 11.4 lists expressions for the electron transport coefficients for $0 \leqslant r_e \leqslant \infty$. They are stated in terms of the functions $M_e(r_e)$, $k_e^T(r_e)$, and $\kappa_e(r_e)$, which are given by the relations

$$M_e(r_e) = \frac{1.00 + 14.1 r_e + 30.6 r_e^2 + 16.3 r_e^3}{1.00 + 21.1 r_e + 37.4 r_e^2 + 16.3 r_e^3} \tag{11.140}$$

$$k_e^T(r_e) = \frac{-0.45 + 3.07 r_e + 6.83 r_e^2}{1.00 + 12.4 r_e + 9.83 r_e^2} \tag{11.141}$$

$$\kappa_e(r_e) = \frac{1.00 + 25.5 r_e + 176 r_e^2 + 454 r_e^3 + 471 r_e^4 + 158 r_e^5}{1.00 + 31.5 r_e + 284 r_e^2 + 662 r_e^3 + 562 r_e^4 + 158 r_e^5} \tag{11.142}$$

It is noteworthy that $M_e(r_e)$ is approximately unity for all values of r_e. This implies that to a good approximation the contributions to the electron mobility from electron–neutral collisions and from electron–charged-particle collisions can be added in parallel. Similar conclusions apply to the contributions to the thermal conductivity of the electrons. These conclusions are, of course, restricted to the collision laws under consideration and do not necessarily apply to other collision laws.

(b) Ion Transport Coefficients. The ion transport coefficients M_i and k_i^T are given in Table 11.5 in a form analogous to that for the electrons (Table 11.4). The functions $M_i(r_i)$ and $k_i^T(r_i)$ are defined by the relations

$$M_i(r_i) = \frac{1.00 + 4.20 r_i + 2.86 r_i^2}{1.00 + 4.24 r_i + 2.91 r_i^2} \tag{11.143}$$

$$k_i^T(r_i) = \frac{0.18 + 0.35 r_i}{1.00 + 4.20 r_i + 2.86 r_i^2} \tag{11.144}$$

The ion mobility is practically independent of r_i since the function $M_i(r_i)$ is within 2% of unity in the entire range $0 \leqslant r_i \leqslant \infty$.

Table 11.5 Ion Transport Coefficients

Transport coefficient	Limit $r_i \ll 1$	Limit $r_i \gg 1$	General expression
M_i	$M_{i0} = \dfrac{0.90e}{m_i \hat{\nu}_{ia}}$	$M_{i\infty} = \dfrac{0.89e}{m_i \hat{\nu}_{ia}}$	$M_i = M_{i0}\, M_i(r_i)$
k_i^T	$k_{i0}^T = -0.18$	$k_{i\infty}^T = 0$	$k_i^T(r_i)$

If applicable, the solutions of the linearized Boltzmann equations, together with Poisson's equation, would provide a tractable model for analyses of cesium-vapor systems. The three hydrodynamic equations (Eqs. 11.116 and 11.117) and Poisson's equation (Eq. 11.74), together with the phenomenological equations (Eqs. 11.118 to 11.120), constitute a set of four second-order differential equations for the unknown quantities n_e, T_e, n_i, and ψ as functions of x. Seven of the eight boundary conditions necessary for solving the equations are provided by the values $n_e(0)$, $T_e(0) = T_E$, $n_i(0)$, ψ_E, $n_e(d)$, $T_e(d) = T_C$, and $n_i(d)$. The missing eighth boundary condition is provided by specifying a value for either the net current density J or the motive ψ_C at $x = d$.

11.6 Range of Validity of Linear Approximations for Cesium Vapor Systems

11.6.1 Conditions on T_e' and μ_e'. In general, the conditions for the validity of the linearized Boltzmann equations (Eqs. 11.134 and 11.135) are velocity dependent. For the steady states of interest, however, very few electrons have high velocities, and the distribution functions peak around the average kinetic energy, which is about kT_e. Accordingly, the term $m_e v_e^2/2$ in Eq. 11.134 can be replaced by kT_e and, therefore, can be neglected in comparison with the negative of the chemical potential $(-\mu_{ce})$, which is usually at least one order of magnitude larger than kT_e. Moreover, the ratio $\nu_{ex}/\nu_e(v_e)$ can be approximated by $\hat{\lambda}_e = \hat{v}_e/\hat{\nu}_e$, i.e., the ratio of an effective speed $\hat{v}_e$ and an effective collision frequency $\hat{\nu}_e$. Thus conditions 11.134 and 11.135 become

$$\left|\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\right| \ll \frac{kT_e}{|\mu_{ce}|} \tag{11.145}$$

and

$$\left|\frac{\hat{\lambda}_e}{\mu_{ce}}\frac{d\mu_e}{dx}\right| \ll \frac{kT_e}{|\mu_{ce}|} \tag{11.146}$$

For conditions of operation of interest to thermionic conversion, kT_e is approximately equal to 0.2 ev, and $|\mu_{ce}|$ is approximately equal to 2 to 3 ev. For example, at electrodes E and C the ratio $kT_e/|\mu_{ce}|$ is equal to kT_E/ϕ_E and kT_C/ϕ_C, respectively, and usually

$$\frac{kT_E}{\phi_E} \approx \frac{kT_C}{\phi_C} \approx \frac{1}{20}$$

It follows from conditions 11.145 and 11.146 that the linear approximations are valid when the fractional changes of T_e and μ_e per effective free path λ_e are about 1/100. For a cesium-vapor system about 10 effective free paths thick ($d/\hat{\lambda}_e \approx 10$) and with $T_e \approx 2000°K$ and $\mu_{ce} \approx 2$ ev, the linear approximations are valid if the temperature and total potential changes across the vapor system are of the order of a few hundred degrees Kelvin and a few tenths of an electron volt, respectively.

Since, in practice, $T_E - T_C$ is approximately 1000°K, the preceding discussion indicates that the linear approximations cannot be valid across a cesium-vapor system under conditions of operation of thermionic converters. Section 11.6.2 shows that the linear approximations fail primarily in the sheath regions.

11.6.2 Failure of Linear Approximations in Sheath Regions. We will consider the cesium-vapor system in a steady state. Because of the disparity between the electron and ion emission rates from each of the electrodes E and C, sheath and plasma regions similar to those established for stable equilibrium states (Sec. 11.2) form also under conditions of steady state, and motive distributions are similar to those given in Fig. 11.2. This conclusion will become evident from the numerical results presented in Chap. 12.

We will assume that the shape of the motive is approximately as shown in Fig. 11.3 and that the cesium-vapor system can be analyzed by means of the relations resulting from the linearized Boltzmann equations. We will prove that the assumption is invalid in the sheath regions when the electron-flux density Γ_e is an appreciable fraction of the random flux density Γ_{re}.

For present purposes, elastic collisional processes will be approximated by an effective electron–neutral hard-sphere free path $\hat{\lambda}_e$, and to first order the effects of ionization collisions will be neglected. Both approximations have a negligible effect on the results.

Under these conditions and approximations, the energy-flux density u_e is given by the relation

$$u_e = \left(\frac{5}{2}kT_e + k_e^T kT + \psi\right)\Gamma_e - \kappa_e \frac{dT_e}{dx}$$

$$= (2kT_e + \psi)\,\Gamma_e - \frac{8}{3}\,kT_e\Gamma_{re}\left(\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\right) \tag{11.147}$$

where, in writing the second part of Eq. 11.147, we have made use of the expressions for k_e^T and κ_e given in Table 11.2 and of Eq. 11.136 for $\hat{\nu}_{ea}$.

Consistent with the assumption about ionization collisions, to first order, du_e/dx and $d\Gamma_e/dx$ are approximately equal to 0. Hence, differentiating Eq. 11.147 with respect to x, replacing derivatives of ψ and $(\hat{\lambda}_e/T_e)(dT_e/dx)$ by $\Delta\psi_D/\lambda_D$ and $\Delta[(\hat{\lambda}_e/T_e)(dT_e/dx)]_D/\lambda_D$, respectively (where $\Delta[\ldots]_D$ denotes change in value over a distance of a Debye length λ_D), and keeping first-order terms only, we find

$$\Delta\left(\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\right)_D = \frac{3}{8}\frac{\Gamma_e}{\Gamma_{re}}\frac{\Delta\psi_D}{kT_e} \tag{11.148}$$

In a sheath the value of the ratio $\Delta\psi_D/kT_e$ is about 3, and therefore the difference beween the values of $(\hat{\lambda}_e/T_e)(dT_e/dx)$ at two points a distance λ_D apart is about Γ_e/Γ_{re}. This difference would satisfy the condition of validity of the linear approximations (Eq. 11.145) if and only if Γ_e/Γ_{re} is smaller than about 1/100. Under practical conditions of operation, however, Γ_e/Γ_{re} in the sheath regions is larger than 1/100, and therefore the linear approximations are not valid in these regions. The failure of the linear approximations in the sheath regions introduces serious computational difficulties in the analysis of the cesium-vapor system. Some of these difficulties have not yet been resolved.

In a plasma region of the type under consideration (Fig. 11.3), both Γ_e/Γ_{re} and $\Delta\psi_D/kT_e$ are much smaller than unity, and therefore the difference $\Delta[(\hat{\lambda}_e/T_e)(dT_e/dx)]_D$ is compatible with the values of $(\hat{\lambda}_e/T_e)(dT_e/dx_e)$ acceptable for the validity of linear approximations. It will be shown in Sec. 11.6.4 that, indeed, the plasma can be analyzed by means of linear approximations, i.e., by models in which both T' and μ' satisfy conditions 11.145 and 11.146, respectively.

The preceding discussion indicates that, for purposes of analysis, the cesium-vapor system should be subdivided into at least theee regions—two sheath regions, one near each electrode, and a plasma region in the center. Each of the sheath regions is about one Debye length thick and is analyzed by

methods other than those involved in the linearization of the Boltzmann equations. In general, the plasma is almost as thick as the interelectrode spacing and is analyzed by means of linear approximations. The results of the two analyses are matched at the interfaces between sheath and electrode and sheath and plasma to satisfy as many continuity requirements as compatible with the models used in the analysis.

An elaborate method of sheath analysis is presented in Sec. 11.7. An approximate but often inconsistent technique for sheath analysis is given in Sec. 11.6.3.

11.6.3 Method for Preliminary Sheath Analysis—Plasma Boundary Conditions. A simple method for treating the cesium vapor system consists in characterizing each sheath region by a monotonic change in electron motive over a negligibly small distance λ_s ($\lambda_s \ll d$) of the order of a Debye length λ_D. Between the two sheaths, the motive of the plasma varies gradually practically throughout the interelectrode spacing d.

A schematic of a motive vs. distance encountered in the method cited is shown in Fig. 11.7, which also illustrates the nomenclature that will be employed.

The plasma transport differential equations resulting from the linearized Boltzmann equations can be integrated to yield the relations between the plasma properties at the two sheath–plasma interfaces. The integration introduces constants that must be determined by the *plasma boundary conditions.* The plasma boundary conditions are obtained from the sheath analysis as follows.

Because the sheath regions are assumed to be thin, the effects of collisions in these regions can be neglected. Moreover, if the distribution functions of electrons and ions supplied by the plasma to the sheaths at planes 1 and 2

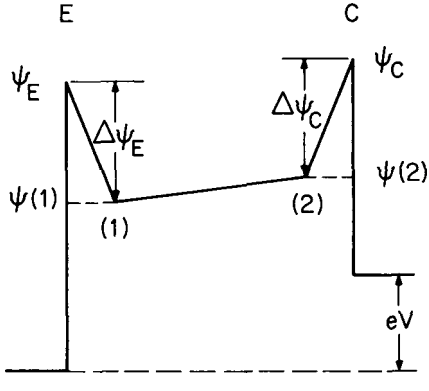


Fig. 11.7 Schematic of motive vs. distance.

(Fig. 11.7) are taken as half-Maxwellian at the conditions prevailing at the corresponding sheath–plasma interfaces, then the monotonic variations of the motive through the sheaths permit the derivation of the plasma boundary conditions without recourse to the tedious problem of finding the self-consistent solution of Poisson's equation. The plasma boundary conditions relate the emission fluxes from the electrodes to the fluxes through the plasma, and the electrode motives to those at the edges of the plasma. Their exact forms depend on whether electrons are accelerated or decelerated as they move through the sheaths in the direction from emitter to collector. If electrons are accelerated, the sheath is called accelerating, and, if they are decelerated, the sheath is called retarding.

In terms of the nomenclature indicated in Fig. 11.7, the following boundary conditions can be obtained by requiring conservaton of particles and energy.

1. *Accelerating emitter sheath* ($\Delta\psi_E < 0$).

$$\Gamma_e(1) = \Gamma_{ve}^E - \Gamma_{re}(1) \exp\left[-\frac{\psi(1) - \psi_E}{kT_e(1)}\right]$$

$$= \Gamma_{ve}^E - \Gamma_{re}(1) \exp\left[\frac{\Delta\psi_E}{kT_e(1)}\right] \tag{11.149}$$

$$\Gamma_i(1) = \Gamma_{vi}^E \exp\left(\frac{\Delta\psi_E}{kT_E}\right) - \Gamma_{ri}(1) \tag{11.150}$$

$$u_e(1) = \Gamma_{ve}^E (2kT_E + \psi_E) - [\Gamma_{ve}^E - \Gamma_e(1)]\,[2kT_e(1) + \psi_E] \tag{11.151}$$

2. *Retarding emitter sheath* ($\Delta\psi_E \geqslant 0$).

$$\Gamma_e(1) = \Gamma_{ve}^E \exp\left(-\frac{\Delta\psi_E}{kT_E}\right) - \Gamma_{re}(1) \tag{11.152}$$

$$\Gamma_i(1) = \Gamma_{vi}^E - \Gamma_{ri}(1) \exp\left(-\frac{\Delta\psi_E}{kT_E}\right) \tag{11.153}$$

$$u_e(1) = 2\Gamma_{re}(1)\,[kT_E - kT_e(1)] + \Gamma_e(1)\,[2kT_E + \psi(1)] \tag{11.154}$$

3. *Accelerating collector sheath* ($\Delta\psi_C \leqslant 0$).

$$\Gamma_e(2) = \Gamma_{re}(2) - \Gamma_{ve}^C \exp\left(\frac{\Delta\psi_C}{kT_C}\right) \tag{11.155}$$

$$\Gamma_i(2) = \Gamma_{ri}(2) \exp\left(\frac{\Delta\psi_C}{kT_C}\right) - \Gamma_{vi}^C \tag{11.156}$$

$$u_e(2) = 2\Gamma_{re}(2)\,[kT_e(2) - kT_C] + \Gamma_e(2)\,[2kT_C + \psi(2)] \tag{11.157}$$

4. *Retarding collector sheath* ($\Delta\psi_C \geqslant 0$).

$$\Gamma_e(2) = \Gamma_{re}(2) \exp\left[-\frac{\Delta\psi_C}{kT_e(2)}\right] - \Gamma_{ve}^C \tag{11.158}$$

$$\Gamma_i(2) = \Gamma_{ri}(2) - \Gamma_{vi}^C \exp\left(-\frac{\Delta\psi_C}{kT_C}\right) \tag{11.159}$$

$$u_e(2) = \Gamma_e(2)\,[2kT_e(2) + \psi_C] + \Gamma_{ve}^C\,[2kT_e(2) - 2kT_C] \tag{11.160}$$

In addition, continuity of motive through the system yields

$$\psi_E - \psi_C = \Delta\psi_E - \psi(2) + \psi(1) - \Delta\psi_C \tag{11.161}$$

Thus seven boundary conditions, six from two sheaths of specified polarity plus Eq. 11.161, are established. From them we get sufficient information to complete the analysis, i.e., to derive the relation J vs. $\psi_E - \psi_C$.

This method of analysis involves several inconsistencies and shortcomings that are discussed in the following paragraphs.

1. The distribution functions of particles flowing from the plasma to the sheaths are assumed to be half-Maxwellian, but the linear approximations used for analyzing the plasma involve perturbed half-Maxwellians. This inconsistency has serious implications in connection with comment 3.

2. Since the sheaths are assumed to be collisionless, particles from, say, electrode E reaching the plasma have a temperature T_E, but the temperature of the plasma at the emitter sheath–plasma interface is taken equal to $T_e(1)$. This inconsistency implies that the sheath cannot be analyzed without consideration of collisional processes.

3. Sheath motives are assumed to be monotonic and to extend over negligibly thin regions without consideration of Poisson's equation. Although this assumption may be valid under certain conditions of operation, experimental evidence exists which indicates that sheaths may be double valued or may extend over an appreciable fraction of the interelectrode spacing, or both. The question of whether the sheath is monotonic or double valued can be answered only by finding the self-consistent solution of Poisson's equation. As discussed in Chap. 10, when Poisson's equation is used, the motive function in the sheath cannot be matched to the motive in the plasma if the plasma is characterized as

being electrically neutral and as having Maxwellian distributions of charged particles at the edge of the sheath–plasma interfaces.

Because of these inconsistencies and shortcomings, a more general method of sheath analysis and more accurate plasma boundary conditions are presented in Sec. 11.8. It will be seen there that the preceding estimates of particle and energy fluxes are also grossly inaccurate. Nevertheless, in Sec. 11.6.4 we will use the preliminary results obtained here to prove that the linear approximations are indeed valid in the plasma region.

11.6.4 Validity of Linear Approximations in the Plasma Region. We will consider the cesium-vapor system in a steady state with an accelerating emitter sheath ($\Delta\psi_E < 0$ in Fig. 11.7). This state is representative of most practical conditions of operation. We will show that the linear approximations are valid in the plasma region by proving that the electron temperature gradient at the emitter sheath–plasma interface satisfies condition 11.145. Then, by virtue of Eq. 11.148, condition 11.145 is satisfied everywhere in the plasma.

We will approximate elastic collisional processes by a constant electron–neutral hard-sphere effective free path $\hat{\lambda}_e$. Therefore, at the emitter sheath–plasma interface, the electron-flux density $\Gamma_e(1)$ and the electron-energy-flux density $u_e(1)$ are given by the relations

$$\Gamma_e(1) = \left\{ \frac{M_e n_e T_e}{e} \left[(2kT_e - \mu_{ce})\, \tau_e' - \tau_e \mu_e' \right] \right\}_1$$

$$= \frac{4}{3} \left\{ \Gamma_{re} \frac{\mu_{ce}}{kT_e} \left[\left(1 - \frac{2kT_e}{\mu_{ce}} \right) \left(\frac{\hat{\lambda}_e}{T_e} \frac{dT_e}{dx} \right) - \frac{\hat{\lambda}_e}{\mu_{ce}} \frac{d\mu_e}{dx} \right] \right\}_1 \tag{11.162}$$

and

$$u_e(1) = \left[(2kT_e + \psi)\, \Gamma_e - \kappa_e \frac{dT_e}{dx} \right]_1$$

$$= \left\{ 2kT_e \left[\left(1 + \frac{\psi}{2kT_e} \right) \Gamma_e - \frac{4}{3} \left(\frac{\hat{\lambda}_e}{T_e} \frac{dT_e}{dx} \right) \Gamma_{re} \right] \right\}_1 \tag{11.163}$$

where, in writing Eqs. 11.162 and 11.163, we have made use of the values of M_e, k_e^T, and κ_e given in Table 11.2 and of Eq. 11.136 for $\hat{\nu}_{ea}$ and where subscript 1 denotes that all position-dependent quantities inside the bracket must be evaluated at the emitter sheath–plasma interface. On the other hand, $u_e(1)$ is also given by approximate boundary condition (Eq. 11.151):

$$u_e(1) = \left\{2kT_e\left[\left(\frac{T_E}{T_e} - 1\right)\Gamma_{ve}^E + \left(1 + \frac{\psi_E}{2kT_e}\right)\Gamma_e\right]\right\}_1 \tag{11.164}$$

Eliminating $u_e(1)$ between Eqs. 11.163 and 11.164, we find that

$$\left[\frac{\Delta\psi_E}{2kT_e}\Gamma_e - \left(1 - \frac{T_E}{T_e}\right)\Gamma_{ve}^E + \frac{4}{3}\left(\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\right)\Gamma_{re}\right]_1 = 0 \tag{11.165}$$

Since, for an accelerating sheath, $\Delta\psi_E$ is less than 0 and T_E is less than $T_e(1)$, Eq. 11.165 indicates that $(dT_e/dx)_1$ is greater than 0 and that $[(\hat{\lambda}_e/T_e)(dT_e/dx)]_1$ is of the order of $[\Gamma_{ve}^E/\Gamma_{re}]_1$ because $|\Delta\psi_E/2kT_e|_1$ is about equal to 2. When $(dT_e/dx)_1$ is greater than 0, Eq. 11.162 indicates that $(d\mu_e/dx)_1$ is less than 0.

Next, eliminating $\Gamma_e(1)$ between Eqs. 11.162 and 11.165, we find that

$$\begin{aligned}\left(\frac{\hat{\lambda}_e}{\mu_{ce}}\frac{d\mu_e}{dx}\right)_1 &= \left[1 - \frac{2kT_e}{\mu_{ce}}\left(1 - \frac{kT_e}{\Delta\psi_E}\right)\right]_1 \left(\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\right)_1 \\ &\quad + \frac{3}{4}\left[\frac{kT_e}{\mu_{ce}}\frac{kT_e}{\Delta\psi_E}\left(1 - \frac{T_E}{T_e}\right)\right]_1 \left(\frac{\Gamma_{ve}^E}{\Gamma_{re}}\right)_1 \\ &\approx \left[1 - \frac{2kT_e}{\mu_{ce}}\left(1 - \frac{kT_e}{\Delta\psi_E}\right)\right]_1 \left(\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\right)_1\end{aligned} \tag{11.166}$$

Equation 11.166 indicates that $[(\hat{\lambda}_e/\mu_{ce})(d\mu_e/dx)]_1$ is of the order of $[(\hat{\lambda}_e/T_e)(dT_e/dx)]_1$, which, in turn, is of the order of $[\Gamma_{ve}^E/\Gamma_{re}]_1$. Numerical results given in Chap. 12 indicate that, for $kT_e(1)$ about 0.2 ev, $[\Gamma_{ve}^E/\Gamma_{re}]_1$ is approximately equal to 0.01. It follows that both $[(\hat{\lambda}_e/T_e)(dT_e/dx)]_1$ and $[(\hat{\lambda}_e/\mu_{ce})(d\mu_e/dx)]_1$ are much smaller than $|kT_e/\mu_{ce}|_1$ and, therefore, that they satisfy conditions 11.145 and 11.146, respectively. By virtue of Eq. 11.148, we conclude that the linear approximations are valid in the plasma region.

11.7 Plasma Boundary Conditions—Monotonic Sheaths

This section presents two sets of plasma boundary conditions which are free of some of the inconsistencies discussed in Sec. 11.6.3. Each set is applicable to monotonic sheaths and is developed without reference to Poisson's equation. The questions of consistency with Poisson's equation and of double valuedness are examined in Sec. 11.8.1.

11.7.1 Boundary Conditions for Small Sheath Heights. At a sheath–plasma interface, the plasma distribution functions are perturbed Maxwellians

(Eqs. 11.89). Accordingly, each particle flux originating from the plasma and flowing into the sheath is given by the random flux plus a perturbation rather than the random flux only. Specifically, the net particle flux Γ, random flux Γ_r, and the directional fluxes Γ^+ and Γ^- in the plus and minus x directions are related by the expressions

$$\Gamma = \Gamma^+ - \Gamma^- \qquad 2\Gamma_r = \Gamma^+ + \Gamma^- \qquad \Gamma^+ = \Gamma_r + \frac{\Gamma}{2} \qquad \Gamma^- = \Gamma_r - \frac{\Gamma}{2}$$

These expressions have led several investigators to recommend that Γ^+ or Γ^-, whichever is appropriate, be used in the derivation of plasma boundary conditions for monotonic sheaths (Fig. 11.7) by means of the particle and energy balance equations across the sheaths. Thus the following set of boundary conditions are obtained:

1. *Accelerating emitter sheath* ($\Delta\psi_E < 0$).

$$\Gamma_e(1) = \Gamma_{ve}^E - \left[\Gamma_{re}(1) - \frac{\Gamma_e(1)}{2}\right] \exp\left[\frac{\Delta\psi_E}{kT_e(1)}\right] \tag{11.167}$$

$$\Gamma_i(1) = \Gamma_{vi}^E \exp\left(\frac{\Delta\psi_E}{kT_E}\right) - \left[\Gamma_{ri}(1) - \frac{\Gamma_i(1)}{2}\right] \tag{11.168}$$

$$u_e(1) = \Gamma_{ve}^E\,(2kT_E + \psi_E) - [\Gamma_{ve}^E - \Gamma_e(1)]\;[2kT_e(1) + \psi_E] \tag{11.169}$$

2. *Retarding emitter sheath* ($\Delta\psi_E > 0$).

$$\Gamma_e(1) = \Gamma_{ve}^E \exp\left(-\frac{\Delta\psi_E}{kT_E}\right) - \Gamma_{re}(1) + \frac{\Gamma_e(1)}{2} \tag{11.170}$$

$$\Gamma_i(1) = \Gamma_{vi}^E - \left[\Gamma_{ri}(1) - \frac{\Gamma_i(1)}{2}\right] \exp\left[-\frac{\Delta\psi_E}{kT_E}\right] \tag{11.171}$$

$$u_e(1) = \Gamma_{re}(1)\;[2kT_E - 2kT_e(1)] + \Gamma_e(1)\;[kT_E + kT_e(1) + \psi(1)] \tag{11.172}$$

3. *Accelerating collector sheath* ($\Delta\psi_C < 0$).

$$\Gamma_e(2) = \Gamma_{re}(2) + \frac{\Gamma_e(2)}{2} - \Gamma_{ve}^C \exp\left(\frac{\Delta\psi_C}{kT_C}\right) \tag{11.173}$$

$$\Gamma_i(2) = \left[\Gamma_{ri}(2) + \frac{\Gamma_i(2)}{2}\right] \exp\left(\frac{\Delta\psi_C}{kT_C}\right) - \Gamma_{vi}^C \tag{11.174}$$

$$u_e(2) = \Gamma_{re}(2)\,[2kT_e(2) - 2kT_C] + \Gamma_e(2)\,[kT_e(2) + kT_C + \psi(2)] \tag{11.175}$$

4. *Retarding collector sheath* ($\Delta\psi_C > 0$).

$$\Gamma_e(2) = \left[\Gamma_{re}(2) + \frac{\Gamma_e(2)}{2}\right] \exp\left[-\frac{\Delta\psi_C}{kT_e(2)}\right] - \Gamma_{ve}^C \tag{11.176}$$

$$\Gamma_i(2) = \Gamma_{ri}(2) + \frac{\Gamma_i(2)}{2} - \Gamma_{vi}^C \exp\left(-\frac{\Delta\psi_C}{kT_C}\right) \tag{11.177}$$

$$u_e(2) = \Gamma_e(2)\,[2\mathrm{kT}_e(2) + \psi_C] + \Gamma_{ve}^C\,[2kT_e(2) - 2kT_C] \tag{11.178}$$

These boundary conditions are better than those represented by Eqs. 11.149 to 11.160 because they are based on more accurate estimates for directional fluxes. Nevertheless, they are applicable only when the sheath heights are relatively small compared with kT and when the electrostatic field of the sheath–plasma interface is close to zero. These two remarks are clarified in Sec. 11.7.2.

11.7.2 Distribution Functions at Sheath–Plasma Interfaces. In the development of linear approximations (Sec. 11.5), the perturbation functions, $\Delta f_\alpha^{(1)} \equiv f_{\alpha 0}\tilde{\varphi}_\alpha$ for $\alpha = e$ and i, are obtained by neglecting terms involving gradients of $\Delta f_\alpha^{(1)}$. This neglect is justified in the plasma region where μ'_α, τ'_α, and the electrostatic field E are relatively small. It results in the phenomenological equations.

On the other hand, the phenomenological equations must be valid even in regions where the electrostatic field is strong provided that μ'_α and τ'_α are small. The reasons are that μ'_α and τ'_α enter in the criteria for the validity of the linear approximations whereas E does not and that μ'_α and τ'_α are not directly proportional to the value of E. For example, in a stable equilibrium state both μ'_α and τ'_α are equal to zero although E varies substantially throughout the vapor system.

If a region near a sheath–plasma interface is analyzed by means of linear approximations, neglecting gradients of the perturbation distribution functions is not justified because: (1) the magnitudes of the gradients may be comparable to those of other terms that are retained; (2) the gradients may alter the plasma boundary conditions appreciably; and (3) the gradients may have an important effect on the condition for the formation of a double-valued sheath.

An estimate of the contributions of the gradients can be obtained by means of an iterative procedure instead of the Sonine polynomial expansions employed in Sec. 11.5.

For present purposes, collisional processes near a sheath–plasma interface for both electrons and ions will be represented by effective collision frequencies $\hat{\nu}_e$ and $\hat{\nu}_i$, respectively. The values of $\hat{\nu}_e$ and $\hat{\nu}_i$ are determined by the relations

$$\hat{\nu}_e = \frac{e}{M_e m_e} \quad \text{and} \quad \hat{\nu}_i = \frac{e}{M_i m_i} \tag{11.179}$$

where M_e and M_i are the electron and ion mobilities employed in the plasma analysis. Under this condition, first estimates $\Delta f_e^{(1)}$ and $\Delta f_i^{(1)}$ of the perturbation distribution functions are given by the relations

$$\Delta f_e^{(1)} = \frac{v_{ex}}{k\hat{\nu}_e}\left[\left(\frac{m_e v_e^2}{2} - \mu_{ce}\right)\tau_e' - \tau_e \mu_e'\right] f_{e0} \tag{11.180}$$

and

$$\Delta f_i^{(1)} = \frac{v_{ix}}{k\hat{\nu}_i}\left[\left(\frac{m_i v_i^2}{2} - \mu_{ci}\right)\tau_i' - \tau_i \mu_i'\right] f_{i0} \tag{11.181}$$

where the superscript (1) is used to indicate that the estimates are obtained by neglecting the gradients in $\Delta f_\alpha^{(1)}$. (See also Eq. 11.129.)

Second estimates $\Delta f_\alpha^{(2)}$ of the perturbations can be obtained as follows: Writing each distribution function f_α in the form

$$f_\alpha = f_{\alpha 0} + \Delta f_\alpha^{(2)} \tag{11.182}$$

then substituting this form into the Boltzmann equation for species α (subject to the restrictions discussed in Sec. 11.5.1) and for an effective collision frequency $\hat{\nu}_\alpha$ only, using the perturbation distribution function $\Delta f_\alpha^{(1)}$ (Eqs. 11.180 and 11.181) to evaluate the gradients $\partial f_\alpha^{(2)}/\partial x$ and $\partial f_\alpha^{(2)}/\partial v_{\alpha x}$, neglecting nonlinear terms in μ_α' and τ_α' and second derivatives of μ_α and τ_α with respect to x, we find that $\Delta f_e^{(2)}$ and $\Delta f_i^{(2)}$ are given by the relations

$$\Delta f_e^{(2)} = \Delta f_e^{(1)}\left(1 + \frac{eE}{m_e \hat{\nu}_e v_{ex}}\right) \tag{11.183}$$

and

$$\Delta f_i^{(2)} = \Delta f_i^{(1)}\left(1 - \frac{eE}{m_i \hat{\nu}_i v_{ix}}\right) \tag{11.184}$$

Using Eqs. 11.183 and 11.184 in the expressions for the fluxes, we find that Γ_e, Γ_i, and u_e are given by the relations

$$\Gamma_e = \frac{M_e n_e T_e}{e}\left[\left(\frac{5}{2}kT_e - \mu_{ce}\right)\tau_e' - \tau_e\mu_e'\right] \tag{11.185}$$

$$\Gamma_i = \frac{M_i n_i T_i}{e}\left[\left(\frac{5}{2}kT_i - \mu_{ci}\right)\tau_i' - \tau_i\mu_i'\right] \tag{11.186}$$

and

$$u_e = \left(\frac{5}{2}kT_e + \psi\right)\Gamma_e + \frac{5n_e k^2 T_e^3}{2m_e\hat{\nu}_e}\tau_e' \tag{11.187}$$

Comparing Eqs. 11.185 to 11.187 with Eqs. 11.126 and 11.128 we see that the assumed effective collision frequencies yield transport differential equations identical to those obtained earlier except for the values of k_e^T and κ_e. Here $k_e^T = 0$ and $\kappa_e = 5n_e k^2 T_e/2m_e\hat{\nu}_e$. These values differ from those given in Tables 11.4 and 11.5. The differences prove to be unimportant for the subsequent development, however.

We see from Eqs. 11.183 and 11.184 that, depending on the value of the electrostatic field E, the terms $eE/m_e\hat{\nu}_e v_{ex}$ and $eE/m_i\hat{\nu}_i v_{ix}$ can be nonnegligible compared with unity and, therefore, that $\Delta f_\alpha^{(2)}$ can differ substantially from $\Delta f_\alpha^{(1)}$ although the phenomenological equations remain unaffected. It will be seen in Sec. 11.7.4 that these terms may also have an appreciable effect on the plasma boundary conditions.

The values of E for which Eqs. 11.183 and 11.184 are valid have an upper limit because the perturbations $\Delta f_e^{(2)}$ and $\Delta f_i^{(2)}$ violate the requirement of conservation of particles. Strictly speaking the integral of $\Delta f_\alpha^{(2)}$ over all velocities $\mathbf{v}_\alpha$ should be zero since $-\hat{\nu}_\alpha\,\Delta f_\alpha^{(2)}$ is an approximation for the elastic collision integral in the Boltzmann equation for species α. Here, however,

$$\Delta n_e = \int_{\mathbf{v}_e} \Delta f_e^{(2)}\, d\mathbf{v}_e = \frac{eEn_e}{m_e\hat{\nu}_e^2 k}\left[\left(\frac{3}{2}kT_e - \mu_{ce}\right)\tau_e' - \tau_e\mu_e'\right] \tag{11.188}$$

and

$$\Delta n_i = \int_{\mathbf{v}_i} \Delta f_i^{(2)}\, d\mathbf{v}_i = -\frac{eEn_i}{m_i\hat{\nu}_i^2 k}\left[\left(\frac{3}{2}kT_i - \mu_{ci}\right)\tau_i' - \tau_i\mu_i'\right] \tag{11.189}$$

The errors Δn_e and Δn_i will be consistent with the other approximations introduced in the problem if $\Delta n_\alpha \ll n_\alpha$ for $\alpha = e$ and i. Accordingly, the perturbations $\Delta f_e^{(2)}$ and $\Delta f_i^{(2)}$ are valid for values of E that satisfy the relations

$$\left|\frac{eE}{m_e \hat{\nu}_e^2 k}\left[\left(\frac{3}{2}kT_e - \mu_{ce}\right)\tau_e' - \tau_e \mu_e'\right]\right| \ll 1 \tag{11.190}$$

and

$$\left|\frac{eE}{m_i \hat{\nu}_i^2 k}\left[\left(\frac{3}{2}kT_i - \mu_{ci}\right)\tau_i' - \tau_i \mu_i'\right]\right| \ll 1 \tag{11.191}$$

11.7.3 Effects of Barriers on Non-Maxwellian Distributions. The perturbation distribution functions are non-Maxwellian. The effect of a barrier on the flux resulting from such a distribution is not, in general, represented by an exponential factor such as that applicable to a half-Maxwellian. If the perturbation distribution function is Δf, the charged particle flux $\Delta\Gamma$ that overcomes a barrier of height $\Delta\psi$ must be computed by means of the relation

$$\Delta\Gamma = \int_{v_{x\min}}^{\infty} dv_x \int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} dv_z \, v_x \, \Delta f \tag{11.192}$$

where

$$v_{x\min} = (2|\Delta\psi|/m_\alpha)^{1/2} \tag{11.193}$$

i.e., the minimum speed of the particles along the x direction which are capable of going over the barrier. In general, $\Delta\Gamma/\Gamma$ does not equal (1/2) exp $(-|\Delta\psi|/kT)$ which is used in Sec. 11.7.1. Analogous remarks apply to the energy flux Δu at the top of a motive barrier. This flux must be computed by means of the relation

$$\Delta u = \int_{v_{x\min}}^{\infty} dv_x \int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} dv_z \, v_x \left(\frac{mv^2}{2} + \psi_s\right) \Delta f_\alpha \tag{11.194}$$

where ψ_s is the motive at the bottom of the barrier.

Using the perturbation $\Delta f_e^{(2)}$ (Eq. 11.183) in Eqs. 11.192 and 11.194 for an emitter accelerating sheath, we find that

$$\left\{\Delta\Gamma_e = \left[\frac{\Gamma(3/2, \Delta\gamma_{Ee})}{\pi^{1/2}} + \frac{a_e}{\pi^{1/2}} \exp(-\Delta\gamma_{Ee})\right]\Gamma_e - \left[\Delta\gamma_{Ee}^{3/2} + a_e\left(\Delta\gamma_{Ee} - \frac{1}{2}\right)\right]\left(\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\right)\Gamma_{re} \exp(-\Delta\gamma_{Ee})\right\}_1 \tag{11.195}$$

and

$$\Delta u_e = \left[\frac{\Gamma(3/2,\Delta\gamma_{Ee})}{\pi^{1/2}} + \frac{a_e}{\pi^{1/2}}\exp(-\Delta\gamma_{Ee})\right]u_e + \left[\Delta\gamma_{Ee}^{1/2} + a_e\left(\Delta\gamma_{Ee} - \frac{1}{2}\right)\right]\frac{\exp(-\Delta\gamma_{Ee})}{\pi^{1/2}}kT_e\Gamma_e - \left\{(2kT_e + \psi_E)\left[\Delta\gamma_{Ee}^{1/2} + a_e\left(\Delta\gamma_{Ee} - \frac{1}{2}\right)\right] - \frac{1}{2}a_e kT_e\right\} \times \frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\Gamma_{re}\exp(-\Delta\gamma_{Ee})\Big|_1 \tag{11.196}$$

where the incomplete gamma function $\Gamma(\alpha,x)$ and the quantities a_α, $\Delta\gamma_{\alpha\beta}$, and $\hat{\lambda}_\alpha$ are defined by the relations

$$\Gamma(\alpha,x) = \int_x^\infty y^{\alpha-1}\exp(-y)\,dy$$

$$a_\alpha = \frac{e_\alpha E\hat{\lambda}_\alpha}{2kT_\alpha} \qquad e_i = e \qquad e_e = -e$$

$$\Delta\gamma_{\alpha\beta} = \frac{|\Delta\psi_\alpha|}{kT_\beta} \qquad \hat{\lambda}_\alpha = \left(\frac{2kT_\alpha}{m_\alpha}\right)^{1/2}\frac{1}{\hat{\nu}_\alpha}$$

and subscript 1 denotes that all position-dependent quantities of each equation must be evaluated at the sheath–plasma interface. It is noteworthy that, only for $\Delta\gamma_{Ee} = a_e = 0$, Eqs. 11.195 and 11.196 yield

$$\Delta\Gamma_e = \frac{\Gamma_e}{2} \quad \text{and} \quad \Delta u_e = \frac{u_e}{2}$$

i.e., the values used in Sec. 11.7.3. Expressions similar to Eqs. 11.195 and 11.196 can be found for other sheath configurations. Using these expressions in the particle- and energy-balance equations across monotonic sheaths, we find the plasma boundary conditions discussed in Sec. 11.7.4.

11.7.4 Boundary Conditions for Large Sheath Heights and Strong Interfacial Electrostatic Fields. For certain sheath configurations, the sheath height may be relatively large and the electrostatic field at the sheath–plasma interface may be nonnegligible although sufficiently small to satisfy Eqs. 11.190 and 11.191. Accordingly, the plasma boundary conditions must be derived by using the improved estimates of $\Delta\Gamma$ and Δu obtained in the preceding section. Thus we find the following results:

1. *Accelerating emitter sheath* ($\Delta\psi_E < 0$).

$$\{\Gamma_e = \Gamma_{ve}^E - \Gamma_{re}\exp(-\Delta\gamma_{Ee}) + \Delta\Gamma_e\}_1$$

or, by virtue of Eq. 11.195,

$$\left[\left[1 - \frac{\Gamma(3/2,\Delta\gamma_{Ee})}{\pi^{1/2}} - \frac{a_e}{\pi^{1/2}}\exp(-\Delta\gamma_{Ee})\right]\Gamma_e\right.$$
$$\left. = \Gamma_{ve}^E - \left\{1 + \left[\Delta\gamma_{Ee}^{3/2} + a_e\left(\Delta\gamma_{Ee} - \frac{1}{2}\right)\right]\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\right\}\Gamma_{re}\exp(-\Delta\gamma_{Ee})\right|_1 \quad (11.197)$$

$$\left[\left(1 - \frac{2a_i}{\pi^{1/2}}\right)\Gamma_i = 2\Gamma_{vi}^E\exp(-\Delta\gamma_{EE}) - \left(2 - a_i\frac{\hat{\lambda}_i}{T_i}\frac{dT_i}{dx}\right)\Gamma_{ri}\right]_1 \quad (11.198)$$

$$[u_e = (2kT_E + \psi_E)\Gamma_{ve}^E - (2kT_e + \psi_E)\Gamma_{re}\exp(-\Delta\gamma_{Ee}) + \Delta u_e]_1$$

or, by virtue of Eq. 11.196,

$$\left[\left[1 - \frac{\Gamma(3/2,\Delta\gamma_{Ee})}{\pi^{1/2}} - \frac{a_e}{\pi^{1/2}}\exp(-\Delta\gamma_{Ee})\right]u_e\right.$$
$$= (2kT_E - 2kT_e)\Gamma_{ve}^E + \frac{1}{2}a_e kT_e\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\Gamma_{re}\exp(-\Delta\gamma_{Ee})$$
$$+\left\{(2kT_e + \psi_E)\left[1 - \frac{\Gamma(3/2,\Delta\gamma_{Ee})}{\pi^{1/2}} - \frac{a_e}{\pi^{1/2}}\exp(-\Delta\gamma_{Ee})\right]\right.$$
$$\left.\left. + kT_e\left[\Delta\gamma_{Ee}^{3/2} + a_e\left(\Delta\gamma_{Ee} - \frac{1}{2}\right)\right]\frac{\exp(-\Delta\gamma_{Ee})}{\pi^{1/2}}\right\}\Gamma_e\right|_1 \quad (11.199)$$

2. *Retarding emitter sheath* ($\Delta\psi_E > 0$).

$$\left[\left(1 - \frac{2a_e}{\pi^{1/2}}\right)\Gamma_e = 2\Gamma_{ve}^E\exp(-\Delta\gamma_{EE}) - \left(2 - a_e\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\right)\Gamma_{re}\right]_1 \quad (11.200)$$

$$\left[\left[1 - \frac{\Gamma(3/2,\Delta\gamma_{EE})}{\pi^{1/2}} - \frac{a_i}{\pi^{1/2}}\exp(-\Delta\gamma_{EE})\right]\Gamma_i\right.$$
$$\left. = \Gamma_{vi}^E - \left\{1 + \left[\Delta\gamma_{EE}^{3/2} + a_i\left(\Delta\gamma_{EE} - \frac{1}{2}\right)\right]\frac{\hat{\lambda}_i}{T_i}\frac{dT_i}{dx}\right\}\Gamma_{ri}\exp(-\Delta\gamma_{EE})\right|_1 \quad (11.201)$$

$$\left\{\left(1 - \frac{2a_e}{\pi^{1/2}}\right)u_e = (4kT_E - 4kT_e)\Gamma_{ve}^E\exp(-\Delta\gamma_{EE})\right.$$

$$+\left[(2kT_e+\psi_E)\left(1-\frac{2a_e}{\pi^{1/2}}\right)-\frac{a_e}{\pi^{1/2}}kT_e\right]\Gamma_e+\frac{1}{2}a_ekT_e\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\Gamma_{re}\Bigg\}_1 \tag{11.202}$$

3. *Accelerating collector sheath* ($\Delta\psi_C < 0$).

$$\left[\left(1+\frac{2a_e}{\pi^{1/2}}\right)\Gamma_e=\left(2+a_e\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\right)\Gamma_{re}-2\Gamma_{ve}^C\exp(-\Delta\gamma_{CC})\right]_2 \tag{11.203}$$

$$\left[\left[1-\frac{\Gamma(3/2,\Delta\gamma_{Ce})}{\pi^{1/2}}+\frac{a_i}{\pi^{1/2}}\exp(-\Delta\gamma_{Ce})\right]\Gamma_i\right.$$

$$=\left\{1-\left[\Delta\gamma_{Ce}^{3/2}-a_i\left(\Delta\gamma_{Ce}-\frac{1}{2}\right)\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\right]\right\}\Gamma_{ri}\exp(-\Delta\gamma_{Ce})-\left.\Gamma_{vi}^C\right|_2 \tag{11.204}$$

$$\left\{\left(1+\frac{2a_e}{\pi^{1/2}}\right)u_e=(4kT_e-4kT_C)\Gamma_{ve}^C\exp(-\Delta\gamma_{CC})\right.$$

$$\left.+\left[(2kT_e+\psi_C)\left(1+\frac{2a_e}{\pi^{1/2}}\right)+\frac{a_e}{\pi^{1/2}}kT_e\right]\Gamma_e-\frac{1}{2}a_ekT_e\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\Gamma_{re}\right\}_2 \tag{11.205}$$

4. *Retarding collector sheath* ($\Delta\psi_C > 0$).

$$\left[\left[1-\frac{\Gamma(3/2,\Delta\gamma_{Ce})}{\pi^{1/2}}+\frac{a_e}{\pi^{1/2}}\exp(-\Delta\gamma_{Ce})\right]\Gamma_e\right.$$

$$=\left\{1-\left[\Delta\gamma_{Ce}^{3/2}-a_e\left(\Delta\gamma_{Ce}-\frac{1}{2}\right)\right]\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\right\}\Gamma_{re}\exp(-\Delta\gamma_{Ce})-\left.\Gamma_{ve}^C\right|_2 \tag{11.206}$$

$$\left[\left(1+\frac{2a_i}{\pi^{1/2}}\right)\Gamma_i=\left(2+a_i\frac{\hat{\lambda}_i}{T_i}\frac{dT_i}{dx}\right)\Gamma_{ri}-2\Gamma_{ve}^C\exp(-\Delta\gamma_{CC})\right]_2 \tag{11.207}$$

$$\left[\left[1-\frac{\Gamma(3/2,\Delta\gamma_{Ce})}{\pi^{1/2}}+\frac{a_e}{\pi^{1/2}}\exp(-\Delta\gamma_{Ce})\right]u_e\right.$$

$$=(2kT_e-2kT_C)\Gamma_{ve}^C-\frac{1}{2}a_ekT_e\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\Gamma_{re}\exp(-\Delta\gamma_{Ce})$$

$$+\left\{(2kT_e+\psi_C)\left[1-\frac{\Gamma(3/2,\Delta\gamma_{Ce})}{\pi^{1/2}}+\frac{a_e}{\pi^{1/2}}\exp(-\Delta\gamma_{Ce})\right]\right.$$

$$\left.\left.+kT_e\left[\Delta\gamma_{Ce}^{3/2}-a_e\left(\Delta\gamma_{Ce}-\frac{1}{2}\right)\right]\frac{\exp(-\Delta\gamma_{Ce})}{\pi^{1/2}}\right\}\Gamma_e\right|_2 \tag{11.208}$$

Boundary conditions 11.197 to 11.208 are more accurate than either of the two sets presented earlier. They include terms which, for a given problem, may be entirely negligible. Whether or not these terms are negligible, however, must be evaluated for each set of conditions and in a manner consistent with other approximations.

11.8 Double-Valued Sheaths

11.8.1 Criterion for the Formation of Double-Valued Sheaths. Up to this point in the development of plasma boundary conditions, we have assumed that sheaths are monotonic and collisionless. These assumptions result, among other things, in the evaluation of the particle distribution functions at the sheath–plasma interfaces. In each sheath region, these distribution functions, as well as those of particles emitted from the electrodes, are affected by the electrostatic field only, and therefore they can be used to calculate particle number densities n_e and n_i. In turn, particle number densities n_e and n_i can be used in Poisson's equation to evaluate whether or not the motive distribution in the sheath is monotonic.

Poisson's equation for the electron motive is given by Eq. 11.74. Introducing the dimensionless quantities

$$\gamma = \frac{\psi - \psi_s}{kT}$$

$$\xi = \frac{x_s - x}{\lambda_D}$$

$$\lambda_D = \left(\frac{\varepsilon_0 kT}{e^2 n_e(x_s)}\right)^{1/2}$$

and

$$\eta_\alpha = \frac{n_\alpha}{n_e(x_s)} \qquad \text{for } \alpha = e \text{ and } i$$

into Eq. 11.74, we find

$$\frac{d^2\gamma}{d\xi^2} = \eta_i - \eta_e \tag{11.209}$$

The normalization constants x_s, T, $n_e(x_s)$, and ψ_s at x_s are chosen depending on the sheath in question. Equation 11.209 can be formally integrated to yield

$$\frac{d\gamma}{d\xi} = \pm (\mathscr{E}_s^2 + \mathscr{F}_i - \mathscr{F}_e)^{1/2} \tag{11.210}$$

where $\mathscr{E}_s$ = the dimensionless electrostatic field at x_s $[(d\gamma/d\xi)_s]$

$$\mathscr{F}_e = 2\int_{\gamma_s}^{\gamma} \eta_e \, d\gamma$$

$$\mathscr{F}_i = 2\int_{\gamma_s}^{\gamma} \eta_i \, d\gamma$$

We see from Eqs. 11.209 and 11.210 that the electron-motive distribution will have an inflection point when $\eta_e = \eta_i$ and an extremum at $\gamma = \gamma_m$, where

$$\mathscr{E}_s^2 + \mathscr{F}_i(\gamma_m) - \mathscr{F}_e(\gamma_m) = 0 \tag{11.211}$$

If x_s is chosen at the sheath–plasma interface and no value γ_m exists (between zero and $\Delta\psi_E/kT$ or $\Delta\psi_C/kT$, depending on the sheath in question) which satisfies Eq. 11.211, then the sheath is indeed monotonic as assumed. On the other hand, if such a value of γ_m exists, then the sheath is double-valued and must be treated in a way that differs from that used for monotonic sheaths.

Equation 11.211 is a criterion for the formation of double-valued sheaths. It can also be expressed in the form of Gauss's law, i.e.,

$$\mathscr{E}_s + \int_0^{\xi_m} (\eta_i - \eta_e)\, d\xi = 0 \tag{11.212}$$

where ξ_m is the dimensionless distance of the extremum from the sheath–plasma interface. In the form of Eq. 11.212, the criterion indicates that the existence of an extremum requires a net positive charge between the sheath–plasma interface and the extremum which is equal and opposite to the surface charge at the sheath–plasma interface.

Unfortunately, attempts by many investigators to verify the monotonicity of certain sheaths have not been successful, and the question remains still unresolved. The reason for the lack of success of these attempts is basic and is related to the sudden total neglect of collisions in the sheath. This neglect decouples the distribution functions of particles originating from the electrode and of particles originating from the plasma and thus introduces an unmanageable discontinuity into the analysis. This point can be illustrated by reviewing again the work of Wang, Lieb, and Pigford,[7] discussed in Sec. 10.5.4.

11.8.2 Fundamental Difficulty with Analyses of Collisionless Sheaths. For a monotonic sheath of the form shown in Fig. 10.14 and half-Maxwellian

distribution functions† for both particles originating from the plasma and particles originating from the electrode, the dimensionless net charge density η $(\eta_i - \eta_e)$ is given by the right-hand side of Eq. 10.128, together with Eqs. 10.129. It can be readily verified that the derivative $d\eta/d\gamma$ at the sheath–plasma interface is infinite except when the vapor system (sheath and plasma) is in a stable equilibrium state. This infinite value arises from the derivative $d\eta_i/d\gamma$ at the sheath–plasma interface.

An infinite value for $d\eta/d\gamma$ at the sheath–plasma interface is found for other shapes of sheath motives, including those which are double valued. The infinite value may be due to either $d\eta_i/d\gamma$ or $d\eta_e/d\gamma$, depending on whether the sheath is accelerating or retarding for ions or electrons, respectively, originating from the plasma.

This nonphysical result arises from the abrupt neglect of collisions in the sheath analysis in comparison with that of the plasma. The neglect of collisions implies that charged particles of the proper sign are accelerated by the electrostatic field near the sheath–plasma interface without hindrance, so much so that $d\eta/d\gamma$ assumes an infinite value.

By virtue of this infinity, no satisfactory solution of Poisson's equation can be found, as illustrated by the work of Wang, Lieb, and Pigford.[7]

11.9 Program for Collisional Analyses of Sheaths

11.9.1 General Remarks. A program for the resolution of the fundamental difficulty discussed in Sec. 11.8.2 has been suggested by Keck. It consists in evaluating the distribution function of each charged species in the presence of both electrostatic fields and collisions characterized by an effective collision frequency. The evaluation is approached by means of an iterative numerical technique consistent with Poisson's equation. An interesting result of the numerical approximations is that in the vicinity of each electrode a collisionless region can be defined. This region, however, is separated from the collisional or plasma region by a curved rather than a planar boundary in phase space extending all the way from a point on the electrode surface to a point about an effective free path away from the surface. Each point on this surface signifies the interruption of collisions for particles originating from the plasma and having a specific energy.

A brief description of Keck's method is given in Secs. 11.9.2 and 11.9.3.

†The inclusion in the analysis of perturbation distribution functions for the particles originating from the plasma is not relevant to the present argument and does not alter the generality of the conclusion.

11.9.2 General Expression for Distribution Functions. Each side of the steady-state Boltzmann equation for the distribution function f_α of charged species α (Eqs. 11.70 to 11.72) is identically equal to $(df_\alpha/dt)_\epsilon$, namely, the total time derivation of f_α evaluated along a trajectory of constant energy ϵ defined by the relation

$$\epsilon = \frac{m_\alpha v_\alpha^2}{2} + \psi_\alpha(x) \qquad \text{for } \alpha = e \text{ and } i \tag{11.213}$$

Accordingly, the steady-state Boltzmann equation for f_α can be written in the forms

$$\left(\frac{df_\alpha}{dt}\right)_\epsilon \equiv v_{\alpha x}\frac{\partial f_\alpha}{\partial x} + \frac{e_\alpha E}{m_\alpha}\frac{\partial f_\alpha}{\partial v_{\alpha x}} \equiv \text{collision integrals} \tag{11.214}$$

For the purposes of this discussion and following Keck's suggestion, we will assume that all collisional processes for species α can be approximated by an effective collision frequency $\hat{\nu}_\alpha$ so that

$$\text{collision integrals} = \hat{\nu}_\alpha(f_{\alpha 0} - f_\alpha) \tag{11.215}$$

and therefore

$$\left(\frac{df_\alpha}{dt}\right)_\epsilon = \hat{\nu}_\alpha(f_{\alpha 0} - f_\alpha) \tag{11.216}$$

where $f_{\alpha 0}$ is a Maxwellian distribution function. Although it was used in earlier discussions of linear approximations (see Eq. 11.93), Eq. 11.215 is not based on linearization of the Boltzmann equation for species α. It can be given the interpretation that particles α enter a collision with distribution function f_α and emerge from a collision with a Maxwellian distribution function $f_{\alpha 0}$ and that the effective collision time $1/\hat{\nu}_\alpha$ is the same for both restituting and direct collisions (Sec. 8.8.3) of particles in the velocity range between $\mathbf{v}_\alpha$ and $\mathbf{v}_\alpha + d\mathbf{v}_\alpha$. The validity of the approximations inherent in this interpretation depends on the relative magnitudes of the free paths for momentum and energy transfers but is not discussed in this book.

For the cesium-vapor system shown in Fig. 11.1, integrating Eq. 11.216 along a trajectory of constant ϵ (Eq. 11.213). we find that

$$f_\alpha(\mathbf{v}_\alpha, x) = f_\alpha^{(b)}(\mathbf{v}_\alpha)\exp(-\zeta) + \int_0^\zeta f_{\alpha 0}(\epsilon, x')\exp(-\zeta')\,d\zeta' \tag{11.217}$$

where $f_\alpha^{(b)}(\mathbf{v}_\alpha)$ = distribution function of particles α entering from a boundary at $x = x_b$ ($b = E$ or C and $x_b = 0$ or d)

$f_{\alpha 0}(\epsilon,x')$ = Maxwellian distribution function at position x' that contributes particles with velocity v_α at x

$$\zeta = \hat{\nu}_\alpha \int_{x_b}^{x} \frac{dx''}{v_{\alpha x''}} \tag{11.218}$$

i.e., the product of the effective collision frequency and the time required by a particle with velocity v_α at x to move from x_b to x along a trajectory of constant ϵ; and

$$\zeta' = \hat{\nu}_\alpha \int_{x'}^{x} \frac{dx''}{v_{\alpha x''}} \tag{11.219}$$

i.e., the product of the effective collision frequency and the time required by a particle with velocity v_α at x to move from x' to x along a trajectory of constant ϵ. The boundary that must be considered depends on the values v_α of the velocity and $\psi_\alpha(x)$ of the motive.

The quantity ζ can be interpreted as the effective number of collisions experienced by a particle moving from the boundary at $x = x_b$ to position x along the trajectory of constant ϵ. Its value depends on the velocity v_α at x which, in turn, together with $\psi_\alpha(x)$, determines whether the trajectory originates from boundary E ($x_b = 0$) or C ($x_b = d$).

Similarly, the quantity ζ' can be interpreted as the effective number of collisions experienced by a particle moving between positions x and x'. Its values range from zero ($x' = x$) to ζ ($x' = x_b$).

For a given effective collision frequency $\hat{\nu}_\alpha$ and a given distance between points x and x', the number of collisions is larger the smaller the velocities $v_{\alpha x''}$ along the constant-ϵ trajectory that passes through x and x'. Moreover, it is seen from Eq. 11.217 that the larger the value of ζ', the less important the contribution of particles originating from x' to the distribution function $f_\alpha(v_\alpha,x)$ at x because of the exponential dependence of f_α on ζ'.

These observations result in considerable simplification of the solutions of integral Eq. 11.217. For example, for the plasma region of a cesium-vapor system of thickness d much greater than an effective free path $\hat{\lambda}_\alpha$ and for positions $\hat{\lambda}_\alpha < x < d - \hat{\lambda}_\alpha$, Eq. 11.217 can be reduced to the form obtained by means of linear approximations. Specifically, the distribution function $f_{\alpha 0}(\epsilon,x')$ can be approximated by the relation

$$f_{\alpha 0}(\epsilon,x') = f_{\alpha 0}(v_\alpha,x) + \left(\frac{\partial f_{\alpha 0}}{\partial x}\right)_\epsilon (x' - x) \tag{11.220}$$

namely, the first two terms of the Taylor series expansion of $f_{\alpha 0}(\epsilon,x')$ around x. The approximation is valid when the gradient of the Maxwellian distribution $f_{\alpha 0}$ with respect to x along the trajectory of constant ϵ is small. It will be seen that this requirement is satisfied when the gradients τ'_α and μ'_α obey the conditions obtained for the linear approximations.

In addition, the quantity $x' - x$ can be approximated by the relation

$$x' - x = -v_{\alpha x} \frac{\zeta'}{\hat{\nu}_\alpha} + \frac{1}{2} \frac{e_\alpha E}{m_\alpha} \frac{\zeta'^2}{\hat{\nu}_\alpha^2} \tag{11.221}$$

In this approximation we assume that along the constant-ϵ trajectory the acceleration $(e_\alpha E/m_\alpha)$ and the velocity $(v_{\alpha x})$ have values corresponding to position x. The assumption is valid because only relatively small values of ζ' and, therefore, $x' - x$ contribute appreciably to $f_\alpha(\mathbf{v}_\alpha,x)$.

Finally, for the positions of x under consideration and for all velocities up to some maximum value, ζ is much greater than unity for trajectories originating from both E and C, and thus for numerical results it can be assumed equal to infinity.

Substituting Eqs. 11.220 and 11.221 into Eqs. 11.217 and using the fact that ζ is much larger than unity, we find that $f_\alpha(\mathbf{v}_\alpha,x)$ is given by the approximate relation

$$\begin{aligned} f_\alpha(\mathbf{v}_\alpha,x) &= f_{\alpha 0}(\mathbf{v}_\alpha,x) - \left(\frac{\partial f_{\alpha 0}}{\partial x}\right)_\epsilon \left(\frac{v_{\alpha x}}{\hat{\nu}_\alpha} - \frac{e_\alpha E}{m_\alpha \hat{\nu}_\alpha^2}\right) \\ &= f_{\alpha 0}(\mathbf{v}_\alpha,x) + \frac{1}{k}\left[\left(\frac{m_a v_a^2}{2} - \mu_{c\alpha}\right)\tau'_\alpha - \tau_\alpha \mu'_\alpha\right] f_{\alpha 0}\left(\frac{v_{\alpha x}}{\hat{\nu}_\alpha} - \frac{e_\alpha E}{m_\alpha \hat{\nu}_\alpha^2}\right) \\ &= f_{\alpha 0} + \Delta f_\alpha^{(2)} \end{aligned} \tag{11.222}$$

where $\Delta f_\alpha^{(2)}$ is given by Eq. 11.183 for $\alpha = e$ and Eq. 11.184 for $\alpha = i$; therefore $f_\alpha(v_\alpha,x)$ is identical to the result obtained by means of linear approximations (Sec. 11.5), including relatively large electrostatic fields (Sec. 11.7.2) for the plasma and the sheath–plasma interface regions.

Simplification of Eq. 11.217 is also possible in a sheath region, and, more importantly, the results are free of the fundamental difficulty discussed in Sec. 11.8.2. This topic is discussed briefly in Sec. 11.9.3. Here it is noteworthy that, by virtue of Eq. 11.222, analyses of sheaths by means of Eq. 11.217 can be readily matched with results for the plasma region derived by means of linear approximations.

11.9.3 Collisional Sheath Analysis. Without loss of generality, we will consider an emitter sheath and will assume that the distribution function is given by Eq. 11.217.

For a cesium-vapor system such that $d \gg \hat{\lambda}_\alpha$, contributions to the distribution function originating directly from the collector (boundary C) are negligible. In other words, the first term in the right-hand side of Eq. 11.217, whenever present, will be associated with particles originating from the emitter. Whether or not the term is present depends, of course, on the values of v_α and x under consideration and on the motive distribution $\psi_\alpha(x)$. On the other hand, the integral terms in the right-hand side of Eq. 11.217 must be evaluated for contributions from positions x' throughout the system. Depending on the values of v_α and x and the motive $\psi_\alpha(x)$, the upper limit ζ of the integral will be either ζ_E or ζ_C, where

$$\zeta_E = \hat{\nu}_\alpha \int_0^x \frac{dx'}{v_{\alpha x'}} \quad \text{and} \quad \zeta_C = \hat{\nu}_\alpha \int_d^x \frac{dx'}{v_{\alpha x'}} \tag{11.223}$$

In other words, the constant-ϵ trajectory under consideration will originate either from the emitter ($\zeta = \zeta_E$) or from the collector ($\zeta = \zeta_C$). For $d \gg \hat{\lambda}_\alpha$, ζ_C is much greater than unity and for all numerical results can be assumed equal to infinity.

Integrating Eq. 11.217 over all velocities, we find that the particle density n_α is given by the relation

$$n_\alpha = \int_{v_\alpha} f_\alpha^E(v_\alpha) \exp(-\zeta_E)\, dv_\alpha + \int_{v_\alpha} dv_\alpha \int_0^\zeta f_{\alpha 0}(\epsilon, x') \exp(-\zeta')\, d\zeta' \tag{11.224}$$

In this relation the range of velocities applicable to each integral is determined by the motive.

The densities n_α, for $\alpha = e$ and i, can be used in Poisson's equation to obtain consistent solutions for the distribution functions, fluxes, and motive. Because of the unsurmountable numerical difficulties, Keck proposes to seek solutions by means of an iterative procedure. The procedure involves the following steps.

For zero-order solutions we assume that:

1. The plasma, including the region in which the electrostatic field is relatively strong, has been analyzed by means of linear approximations. At the emitter side this analysis is valid up to a position denoted by x_1, which is about a Debye length away from the emitter surface and whose value will be determined later. At x_1 the Maxwellian part of the distribution function and its gradient will be denoted by $(f_{\alpha 0})_1$ and $(\partial f_{\alpha 0}/\partial x)_1$, respectively, and the motive and its gradient by $\psi_{\alpha 1}$ and $(d\psi_\alpha/dx)_1$, respectively.

2. For positions different from x_1, the distribution function $f_{\alpha 0}(\epsilon, x')$ will be approximated initially by $f_{\alpha 0}^{(0)}(\epsilon, x')$ given by the relation

$$f_{\alpha 0}^{(0)}(\epsilon,x') = (f_{\alpha 0})_1 + \left(\frac{\partial f_{\alpha 0}}{\partial x}\right)_{1,\epsilon} (x' - x_1) \tag{11.225}$$

As discussed in Sec. 11.9.2, this approximation is valid for $x' > x_1$.

3. The motive ψ_α will be approximated by $\psi_\alpha^{(0)}$ given by the relations

$$\psi_\alpha^{(0)} - \psi_m = D\left[1 - \exp\left(\frac{x_m - x}{\lambda_D}\right)\right]^2 \qquad \text{for } x \leq x_1 \tag{11.226a}$$

and

$$\psi_\alpha^{(0)} - \psi_{\alpha 1} = \left(\frac{d\psi_\alpha}{dx}\right)_1 (x - x_1) \qquad \text{for } x \geq x_1 \tag{11.226b}$$

where x_m, ψ_m, and D are constants to be determined. It is noteworthy that Eq. 11.226a has an extremum (maximum or minimum) at $x = x_m$ and the exponential character typical of a motive distribution in a sheath of a stable equilibrium state for $x - x_m > \lambda_D$.

The values of x_m, x_1, ψ_m, and D can be determined from the continuity conditions at $x = 0$ and $x = x_1$. Specifically

$$\psi_{\alpha E} - \psi_m = D\left[1 - \exp\left(\frac{x_m}{\lambda_D}\right)\right]^2 \tag{11.227a}$$

$$\left(\frac{d\psi_\alpha}{dx}\right)_E = 2\frac{D}{\lambda_D}\left[1 - \exp\left(\frac{x_m}{\lambda_D}\right)\right]\exp\left(\frac{x_m}{\lambda_D}\right) \tag{11.227b}$$

$$\psi_{\alpha 1} - \psi_m = D\left[1 - \exp\left(\frac{x_m - x_1}{\lambda_D}\right)\right]^2 \tag{11.227c}$$

and

$$\left(\frac{d\psi_\alpha}{dx}\right)_1 = 2\frac{D}{\lambda_D}\left[1 - \exp\left(\frac{x_m - x_1}{\lambda_D}\right)\right]\exp\left(\frac{x_m - x_1}{\lambda_D}\right) \tag{11.227d}$$

It is clear that, for given values of $\psi_{\alpha E}$, $(d\psi_\alpha/dx)_E$, $\psi_{\alpha 1}$, and $(d\psi_\alpha/dx)_1$, Eqs. 11.227 can be solved for the four quantities x_m, x_1, ψ_m, and D. For example, for the important special case for which the extremum is at the emitter, $x_m = 0$, $(d\psi_\alpha/dx)_E = 0$, $\psi_m = \psi_E$,

$$D = \Delta\psi_{\alpha E}\left[1 + \frac{\lambda_D}{2\Delta\psi_{\alpha E}}\left(\frac{d\psi_\alpha}{dx}\right)_1\right]^2 \tag{11.228a}$$

and

$$x_1 = \lambda_D \ln\left[1 + \frac{2\Delta\psi_{\alpha E}}{\lambda_D}\left(\frac{d\psi_\alpha}{dx}\right)_1\right] \tag{11.228b}$$

where

$$\Delta\psi_{\alpha E} \equiv \psi_{\alpha 1} - \psi_E$$

Using the relations resulting from assumptions 1 to 3 in Eq. 11.224, we obtain the densities $n_\alpha^{(0)}(\psi)$ for the sheath. Then

$$n_e^{(0)} = \frac{2\,(2\pi m_e kT_e)^{3/2}}{h^2}\exp\left(\frac{\mu_e - \psi_e}{kT_e}\right)$$

$$n_i^{(0)} = \frac{(2\pi m_i kT_i)^{3/2}}{h^3}\exp\left(\frac{\mu_i + \psi_e}{kT_i}\right) \quad \text{for } T_i = T_a$$

and

$$\frac{3}{2}n_e^{(0)}\,kT_e = \int \frac{m_e v_e^2}{2} f_e\, d\mathbf{v}_e = \int \frac{m_e v_e^2}{2} f_{e0}\, d\mathbf{v}_e$$

can be used to find the first iterate $x^{(1)}(\psi)$, $\mu_e^{(1)}(\psi)$, $\mu_i^{(1)}(\psi)$, and $T_e^{(1)}(\psi)$, which, in turn, yield $f_{e0}^{(1)}(\epsilon,x)$, $f_{i0}^{(1)}(\epsilon,x)$, and $\psi_\alpha^{(1)}(x)$. This procedure may be repeated as many times as required to achieve the desired accuracy. It is anticipated, however, that the first iterate $\psi_\alpha^{(1)}$ will be sufficient for most practical purposes.

To carry out the preceding calculations in detail, we must consider specific possible shapes of the sheath motive. The numerical calculations are lengthy and will not be given in this book. It suffices to indicate only that the method does not suffer from the fundamental difficulty of Sec. 11.8.2 and that it is capable of disclosing the existence of double-valued sheaths.

References

1. D. Bohm and D. Pines, Collective vs. Individual Particle Aspect of the Interactions, *Phys. Rev.*, 85: 338-353 (1952).
2. J. M. Houston, Cross Section Values to Use in Analyzing the Cesium Thermionic Converter, Report on the Thermionic Conversion Specialists Conference, Cleveland, Ohio, Oct. 26–28, 1964, Institute of Electrical and Electronics Engineers, New York.
3. J. W. Sheldon, *J. Appl. Phys.*, 34: 444 (1963).
4. D. R. Wilkins and E. P. Gyftopoulos, Transport Phenomena in Low-Energy Plasmas, *J. Appl. Phys.*, 37: 3533-3540 (1966).

5. S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases,* Cambridge University Press, New York, 1961.
6. J. O. Hirshfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids,* John Wiley & Sons, Inc., New York, 1954.
7. C. Wang, D. Lieb, and T. H. Pigford, Extended Space-Charge Theory in the Emitter Sheath, *Proceedings of the Thermionic Conversion Specialists Conference, Carmel, Calif., Oct. 21–23, 1969,* Institute of Electrical and Electronics Engineers, New York.

12

Analysis of Performance Characteristics of High-Pressure Diodes

12.1 General Remarks

This chapter presents analyses of steady-state performance characteristics of high cesium pressure diodes, e.g., output current and efficiency of energy-conversion characteristics.

For a given diode (i.e., for given electrode surfaces a distance d apart), for given electrode temperatures, and for given cesium pressure, an analysis of performance characteristics involves the simultaneous and consistent consideration of both emission phenomena (Chap. 9) and transport phenomena (Chaps. 10 and 11) in the diode. The two types of phenomena are coupled because the emission properties of each electrode are affected by the electrostatic field (Schottky effect) created by the electric charges in transit in the interelectrode space.

In general, the analysis of performance characteristics is a very tedious numerical task that has not yet been completely achieved. There are several reasons for the difficulties: (1) the theoretical complexities of emission phenomena; (2) the lengthy calculations required for consistent coupling of emission and transport phenomena; and (3) the highly nonlinear nature of local transport phenomena such as those involved in ionization–recombination processes.

Performance characteristics have been analyzed, with varying degrees of sophistication, by several investigators. Invariably the procedure adopted by each investigator can be summarized as follows: (1) Either an analytical model or a tabulation of experimental data is employed to describe the zero-field emission properties of the cesiated electrode surfaces. (2) The interelectrode space of the diode, the cesium-vapor system, is subdivided into three regions, the emitter sheath, the plasma, and the collector sheath. (3) Each sheath region is analyzed as collisionless, and therefore the emission properties of the electrodes are related to the plasma properties at the sheath–plasma interface by means of boundary conditions and without consideration of Poisson's equation. For monotonic sheaths the boundary conditions employed are those given by either Eqs. 11.149 to 11.161 or Eqs. 11.167 to 11.178. (4) Nonzero-field electron and ion emission properties entering the boundary conditions are evaluated by use of the Schottky effect, which the investigators compute by using an approximate

value for the electrostatic field at the points just outside each electrode. (5) The plasma region is analyzed by means of linear approximations that result in a set of transport differential equations; these equations closely resemble those presented in Sec. 11.5.3. (6) The number of independent variables in the transport differential equations is reduced by one by assuming that the densities n_e of electrons and n_i of ions are approximately equal to each other at every position throughout the plasma. (7) Differences in approaches arise because of differences in additional assumptions made by various investigators regarding the integration of the transport differential equations.

In the following sections of this chapter, we summarize the analytical model used in plasma studies, present some numerical results related to performance characteristics, and, finally, discuss the validity of numerical results in the light of some general criteria of consistency and accuracy.

12.2 Analytical Model

The general analytical model, with minor modifications, has been used in many plasma analyses of thermionic converters.

We will consider a high-pressure diode in a steady state defined by the values T_E and T_C of the electrode temperatures, ϕ_{E0} and ϕ_{C0} of the cesiated electron work functions under zero-field conditions at the points just outside the electrodes, d of the interelectrode spacing, p of the cesium pressure, J of the output current density, and V of the electrode output voltage. The interelectrode spacing d is sufficiently small, compared with the linear dimensions of the electrode surfaces, so that transport phenomena can be regarded as one dimensional, and pd is sufficiently larger than 1 mil-torr so that collisional processes must be considered in the analysis.

By virtue of the discussions in Chap. 11, the vapor system in the interelectrode space will be subdivided into three regions, an emitter sheath, a plasma, and a collector sheath.

We will assume that the sheath regions can be analyzed as either collisionless or collisional by means of the techniques discussed in Chap. 11. The analysis of the sheath regions relates the conditions prevailing at the points just outside the electrodes to the conditions prevailing at the sheath–plasma interfaces. It provides boundary conditions for the solution of the transport differential equations that describe the plasma.

Next we will assume that the plasma region can be analyzed by means of linear approximations, i.e., by means of the transport differential equations developed in Sec. 11.5.3. These equations can be further simplified by taking the number densities of electrons and ions equal to each other ($n_e \approx n_i = n$) because of the following observation. One of the conditions that must be satisfied for the linear approximations to be valid is given by Eq. 11.148. For small temperature

gradients this equation implies that the motive change $\Delta\psi_D$ over a Debye length must be very small. By virtue of Poisson's equation, this implication can be satisfied if and only if the number densities n_e and n_i are almost equal to each other.

When $n_e \approx n_i = n$, Eqs. 11.116 to 11.120 and 11.75 can be rewritten as

$$\frac{d\Gamma_i}{dx} = \int C^{\text{in}}\, d^3\mathbf{v}_e \tag{12.1}$$

$$\frac{d\Gamma_e}{dx} = \int C^{\text{in}}\, d^3\mathbf{v}_e \tag{12.2}$$

$$\frac{dw_e}{dx} = -e\Gamma_e E + \int \frac{m_e v_e^2}{2}\, C^{\text{in}}\, d^3\mathbf{v}_e \tag{12.3}$$

$$\kappa_e \frac{dT_e}{dx} = \left(\frac{5}{2}kT_e + k_e^T\, kT_e\right)\Gamma_e - w_e \tag{12.4}$$

$$kT_e \frac{dn}{dx} = -\frac{e\Gamma_e}{M_e} - enE - n(1 + k_e^T)k\frac{dT_e}{dx} \tag{12.5}$$

$$\frac{d\psi}{dx} = eE \tag{12.6}$$

$$E = -\frac{1}{en(T_e + T_i)}\left\{T_i\left[\frac{e\Gamma_e}{M_e} + n(1 + k_e^T)k\,\frac{dT_e}{dx}\right]\right.$$
$$\left. -T_e\left[\frac{e\Gamma_i}{M_i} + n(1 + k_i^T)k\frac{dT_i}{dx}\right]\right\} \tag{12.7}$$

where all terms are as defined in Secs. 8.8.5 and 11.5. Equation 12.7 is obtained by combining Eqs. 11.119 and 11.120 for $n_e \approx n_i = n$.

By virtue of the assumption that $n_e \approx n_i = n$, Poisson's equation is not necessary and ψ can be found by integrating Eq. 12.6 for E given by Eq. 12.7. In addition, Eqs. 12.1 and 12.2 imply that

$$\Gamma = \Gamma_e - \Gamma_i = \text{constant} \quad \text{or} \quad J = J_e - J_i = \text{constant} \tag{12.8}$$

Since the value of the output current density J is one of the quantities that specifies the steady state of the diode, it follows from Eq. 12.8 that Eq. 12.1 need not be considered in the analysis of the plasma.

The inelastic-collision-rate integral $\int C^{in}\, d^3\mathbf{v}_e$ can be approximated by the relation[1-3]

$$\int C^{in}\, d^3\mathbf{v}_e = \nu_i n - \beta_r n^3 = \beta_r (n_{eq}^2 n - n^3) \tag{12.9}$$

where

$$\beta_r = k_r \left(\frac{e}{kT_e}\right)^{9/2} = \frac{5.6 \times 10^{-27}}{(T_e)^{9/2}} \quad \text{cm}^6/\text{sec} \qquad \text{for } T_e \text{ in ev} \tag{12.10}$$

$$n_{eq}^2 = n_a \left(\frac{2\pi m_e kT_e}{h^2}\right)^{3/2} \exp\left(-\frac{V_i}{kT_e}\right) \tag{12.11}$$

Because of the large differences in masses between electrons and cesium particles, the ionization energy of cesium can be regarded as only the energy carried off or supplied by an electron after a recombination or during an ionization collision; in other words, the energies of the heavy particles involved in the collision can be neglected. Accordingly, the integral $\int (m_e v_e^2/2)\, C^{in}\, d^3\mathbf{v}_e$ can be approximated by the relation

$$\int \frac{m_e v_e^2}{2} C^{in}\, d^3\mathbf{v}_e = -V_i \int C^{in}\, d^3\mathbf{v}_e = -V_i \beta_r (n_{eq}^2 n - n^3) \tag{12.12}$$

The electrode emission current densities required in the plasma boundary conditions are computed by means of the Richardson and Saha–Langmuir equations including Schottky effects. For example, for an accelerating emitter sheath, the emitter work function ϕ_E is given by the relations

$$\phi_E = \phi_{E0} - \Delta\psi \qquad \text{and} \qquad \Delta\psi = 3.4 \times 10^{-4} E_E^{1/2} \text{ ev} \quad \text{for } E_E \text{ in volts/cm} \tag{12.13}$$

where ϕ_{E0} is the zero-field cesiated emitter work function corresponding to the given T_E and p and E_E is the field at the emitter. The work function ϕ_E is used in the Richardson and Saha–Langmuir equations to evaluate J_{ES} and $\tilde{J}_{ES}$ (Γ_{ve}^E and $\Gamma_{\nu i}^E$), respectively. Specifically

$$J_{ES} = e\Gamma_{ve}^E = AT_E^2 \exp\left(-\frac{\phi_E}{kT_E}\right) \tag{12.14}$$

and

$$\tilde{J}_{ES} = e\Gamma_{\nu i}^E = \frac{ep}{(2\pi m_a kT_E)^{1/2}\,\{1 + 2\exp\,[(V_i - \phi_E)/kT_E]\}} \tag{12.15}$$

The zero-field cesiated work functions are evaluated by means of the procedures discussed in Chap. 9 or are taken from tabulated experimental data.

When we analyze the sheaths in a manner consistent with Poisson's equation, we obtain the electrostatic field at the points just outside each electrode surface. On the other hand, when this systematic procedure is not followed because of practical considerations, accurate evaluations of these fields are not possible. An estimate of the field just outside electrode β is given by the relation

$$E_\beta = \frac{|\Delta\psi_\beta|}{\lambda_D(\alpha)} \quad \text{for } \beta = E \text{ and } C \text{ and } \alpha = 1 \text{ and } 2 \tag{12.16}$$

where $|\Delta\psi_\beta|$ is the sheath height and $\lambda_D(\alpha)$ is the Debye length at the sheath–plasma interface α in question.

The temperature T_i of the ions is taken equal to the temperature T_a of the neutral atoms. The temperature T_a can be readily computed by integrating Eq. 11.88 from T_E to T_C for a constant neutral–neutral collision cross section σ_{aa}. Thus it is found that the thermal conductivity k_a, the temperature T_a, and the kinetic energy flux rate w_a (heat-conduction flux rate) are given by the relations

$$T_i = T_a = T_E \left\{1 - \frac{x}{d}\left[1 - \left(\frac{T_C}{T_E}\right)^{3/2}\right]\right\}^{2/3} \tag{12.17}$$

$$k_a = \frac{k}{2\sigma_{aa}}\left(\frac{\pi k T_a}{m_a}\right)^{1/2} \tag{12.18}$$

and

$$w_a = \frac{1}{3}\frac{kT_E}{\sigma_{aa}d}\left(\frac{\pi k T_E}{m_a}\right)^{1/2}\left[1 - \left(\frac{T_C}{T_E}\right)^{3/2}\right] \tag{12.19}$$

This model involves five first-order differential equations, Eqs. 12.2 to 12.6, for the five unknown functions of position (n, T_e, Γ_e, w_e, and ψ) and two unknown quantities for the sheath heights ($\Delta\psi_E$ and $\Delta\psi_C$). The differential equations can be integrated numerically subject to the six algebraic relations provided by the boundary conditions at the emitter and collector sheath–plasma interfaces. The result of this integration is the relation between the output-current density J and either the internal voltage drop ΔV or the electrode output voltage V, namely, an output current characteristic. In addition, this relation, together with energy balances, can be used to obtain other performance characteristics.

Numerical results obtained by means of the model are discussed later. Here it suffices to indicate that the computer-aided solutions of the differential equations present difficult and tedious numerical problems. The major difficulty arises from the nonlinear form of the ionization–recombination relation (Eq. 12.9) and its sensitivity to the value of the electron temperature. To avoid such difficulties, several investigators have suggested additional approximations. Some of these approximations can be best discussed by rewriting the model in dimensionless form.

12.3 Dimensionless Model

We will derive a set of dimensionless differential equations for the collisional transport phenomena by assuming that collisions can be described by constant effective collision frequencies $\hat{\nu}_e$ and $\hat{\nu}_i$. This assumption is not necessary but is made to simplify the form of the coefficients of the equations since, for constant $\hat{\nu}_e$ and $\hat{\nu}_i$ and $n_e \approx n_i = n$,

$$k_e^T = k_i^T = 0 \quad \text{and} \quad \kappa_e = \frac{5nk^2 T_e}{2m_e \hat{\nu}_e}$$

Thus after some elementary algebra we obtain the following set of dimensionless transport differential equations:

$$\frac{d\Gamma_e^*}{dx^*} = \frac{d\Gamma_i^*}{dx^*} = \frac{1}{\lambda_r^*} n^*(n_{eq}^{*2} - n^{*2}) \tag{12.20}$$

$$\frac{\lambda_e^*}{T_e^*}\frac{dT_e^*}{dx^*} = \frac{2}{5\sqrt{\pi}}\frac{1}{\Gamma_{re}^*}\left[\left(\frac{5}{2} + \frac{\psi^*}{T_e^*}\right)\Gamma_e^* - \frac{u_e^*}{T_e^*}\right] \tag{12.21}$$

$$\lambda_e^* \frac{d}{dx^*}\left(\frac{\mu_e^*}{T_e^*}\right) = -\frac{1}{\sqrt{\pi}\Gamma_{re}^*}\left\{\Gamma_e^* + \left(1 + \frac{2}{5}\frac{\psi^*}{T_e^*}\right)\left[\left(\frac{5}{2} + \frac{\psi^*}{T_e^*}\right)\Gamma_e^* - \frac{u_e^*}{T_e^*}\right]\right\} \tag{12.22}$$

$$\lambda_e^* \frac{d}{dx^*}[n^*(T_e^* + T_i^*)] = -\frac{T_e^{*1/2}}{\sqrt{\pi}}\left[1 + \left(1 + \frac{m_i\hat{\nu}_i}{m_e\hat{\nu}_e}\right)\Gamma_i^*\right]$$

$$\approx -\frac{T_e^{*1/2}}{\sqrt{\pi}}\left(1 + \frac{m_i\hat{\nu}_i}{m_e\hat{\nu}_e}\Gamma_i^*\right) \tag{12.23}$$

$$\lambda_e^* n^*(T_e^* + T_i^*)\frac{d\psi^*}{dx^*} = -\frac{T_e^{*1/2}}{\sqrt{\pi}}\left(T_i^*\Gamma_e^* - \frac{m_i\hat{\nu}_i}{m_e\hat{\nu}_e}T_e^*\Gamma_i^*\right)$$

$$-\lambda_e^* n^*\left(T_i^*\frac{dT_e^*}{dx^*} - T_e^*\frac{dT_i^*}{dx^*}\right) \quad (12.24)$$

where the dimensionless variables and parameters are defined by the relations

$$\Gamma_\alpha^* \equiv \frac{\Gamma_\alpha}{\Gamma} \quad \lambda_e^* \equiv \left(\frac{2kT_e}{m_e}\right)^{1/2}\frac{1}{\hat{\nu}_e d} \quad \lambda_r^* \equiv \left(\frac{kT_E}{2\pi m_e}\right)^{3/2}\frac{1}{\beta_r\Gamma^2 d}$$

$$\mu_\alpha^* \equiv \frac{\mu_\alpha}{kT_E} \quad n_\alpha^* \equiv \left(\frac{kT_E}{2\pi m_e}\right)^{1/2}\frac{n_\alpha}{\Gamma} \quad T_\alpha^* \equiv \frac{T_\alpha}{T_E} \quad (12.25)$$

$$u_\alpha^* \equiv \frac{u_\alpha}{kT_E\Gamma} \quad x^* \equiv \frac{x}{d} \quad \psi^* \equiv \frac{\psi}{kT_E}$$

and in the second part of Eq. 12.23 unity is omitted because $m_i\hat{\nu}_i \gg m_e\hat{\nu}_e$ for plasmas of interest for which $\lambda_e^* > \lambda_i^*$. It can be easily verified that the parts of Eq. 12.20 are equivalent to Eqs. 12.1 and 12.2, Eq. 12.21 is equivalent to Eq. 12.4, Eq. 12.22 is equivalent to an algebraic combination of Eqs. 11.126 and 11.127 in place of Eq. 12.3, Eq. 12.23 is equivalent to an algebraic combination of Eq. 12.5 and the corresponding equation for ions, and Eq. 12.24 is equivalent to Eq. 12.6. Other relations, such as boundary conditions, can be normalized in an analogous manner.

In the forms of Eqs. 12.20 to 12.24, the transport differential equations are amenable to certain simplifications that are discussed later.

12.4 Review of Plasma and Performance Analyses

Plasma analyses have been performed with various degrees of sophistication by several investigators. All analyses are based on descriptions of the plasma which are similar to the description provided by the transport differential equations given in Sec. 12.2 or 12.3.

Wilkins[4] developed a digital computer program called SIMCON for analyses of thermionic converters. SIMCON yields current–voltage characteristics that are in good qualitative and quantitative agreement with experimental data over a wide range of converter operating conditions. The program contains a number of mathematical and numerical deficiencies, however, which compromise its utility. For example, a constant electron temperature approximation is used in the integration of the transport differential equations. This approximation can lead

to errors in evaluating ionization and recombination rates and boundary conditions.

Subsequently, Wilkins, McCandless, and Derby[5] developed a more flexible computer program called THRIVE, which is free of most of the deficiencies of SIMCON. The plasma model in THRIVE is identical to that described in Sec. 12.2; the boundary conditions are those given by Eqs. 11.167 to 11.178; and the emission properties of the electrodes are described either by the empirical theory developed by Gyftopoulos, Levine, and Steiner (see Chap. 9) or by tabulated experimental data. The major faults of THRIVE are the collisionless treatment of the sheaths and the use of inaccurate boundary conditions.

Warner and Hansen[6] have made detailed studies of sheaths and plasma-property profiles (including electron temperatures) by means of both a digital computer program and a so-called algebraic technique. Their results are not so extensive as those obtained by Wilkins, McCandless, and Derby[5] by means of THRIVE and, in general, suffer the same deficiencies as those of THRIVE.

Baksht et al.[7] developed a program for analyzing current–voltage characteristics and property profiles for a number of converter operating points. In their mathematical model they assume a constant electron current equal to the output current. The plasma boundary conditions are not formulated to admit both accelerating and retarding sheaths adjacent to the electrode surfaces, and Schottky effects are not included in the numerical computations.

Several investigators have studied specialized areas important to the analysis of thermionic converter plasmas. Lieb and Bornhorst[8] investigated the effect of electron thermal conductivity and ionization on plasma property profiles. Curry et al.[9] made a detailed analysis and computation of recombination processes in a thermionic converter plasma. Dugan,[1] Harris,[2] and Hinnov and Hirschberg[3] also made significant contributions to the analysis of ionization–recombination processes. Bullis and Nighan,[10] Sheldon,[11] Houston,[12] and Stone and Reitz[13] have studied collisional phenomena in great detail. The models and information presented by each of these investigators provide valuable guidance in the development of analytical descriptions for thermionic converter plasmas.

12.5 Numerical Results

The most extensive analyses of performance characteristics of converters have been done by Wilkins and his coworkers using the THRIVE computer program. In each calculation the operating point of the converter is specified by the quantities ϕ_{E0}, T_E, ϕ_{C0}, T_C, d, p, and Γ or J. The object of each calculation is the evaluation of the electrode output voltage V given by the relation

$$eV = \phi_E + \Delta\psi_E + \Delta\psi_p + \Delta\psi_C - \phi_C \tag{12.26}$$

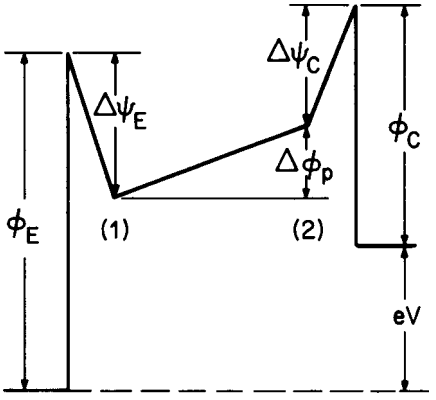


Fig. 12.1 Schematic of motive.

where all symbols are as shown in Fig. 12.1. All the quantities in the right-hand side of Eq. 12.26 can be evaluated when the operating point is as specified previously and the transport differential equations are solved subject to the plasma boundary conditions. The cesiated work functions ϕ_E and ϕ_C, including Schottky effects, and the sheath heights $\Delta\psi_E$ and $\Delta\psi_C$ are computed from the boundary conditions. The plasma motive difference, $\Delta\psi_p$, and the plasma property distributions vs. x are computed by integrating the transport differential equations.

In mathematical language the problem in question is a two-point boundary-value problem with nonlinear boundary conditions for a set of ordinary first-order nonlinear differential equations. Many methods of solution exist for such a problem. The method employed by Wilkins and coworkers is the *shooting technique.*

Application of the shooting technique to the solution of Eqs. 12.2 to 12.6, subject to boundary conditions 11.167 to 11.178, is illustrated in Fig. 12.2. Initiation of integration of the five transport differential equations requires knowledge of five initial conditions, i.e., $T_e(1)$, $n(1)$, $\Gamma_e(1)$, $u_e(1)$ [or $w_e(1)$], and $\psi(1)$. The three boundary conditions at the emitter sheath–plasma interface can yield only three of the initial conditions in terms of the remaining two. Therefore it is necessary to guess at the values of two initial conditions, say $T_e(1)$ and $n(1)$, and solve the boundary conditions for the remaining three, $\Gamma_e(1)$, $u_e(1)$, and $\psi(1)$. Thus the transport differential equations can be integrated from x_1 to x_2 as shown schematically in Fig. 12.2. At x_2 the results of the integration for $T_e(2)$ and $n(2)$ are used in the collector sheath–plasma interface boundary conditions to compute the values $\Gamma_e(2)$ and $u_e(2)$. These values will not coincide with the values $\Gamma_e(2)$ and $u_e(2)$ resulting from the

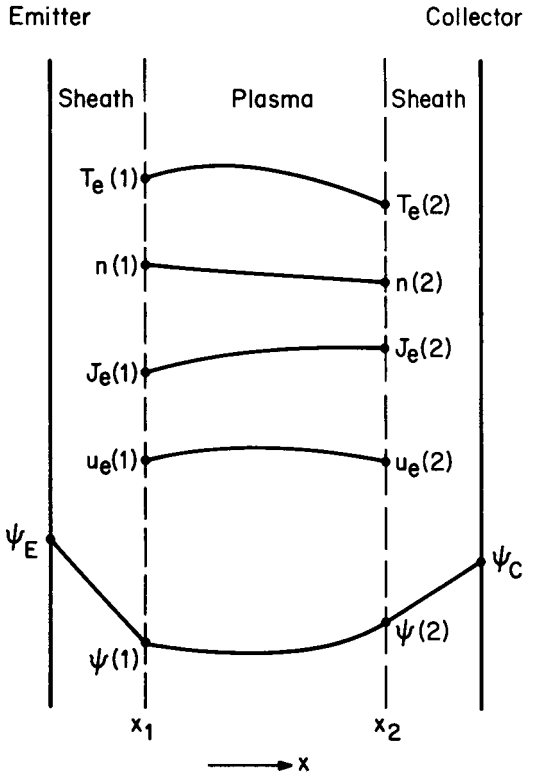


Fig. 12.2 Schematic representation of the solution of the transport equations.

$$T_e(x) = T_e(1) + \int_{x_1}^{x} \frac{dT_e}{dx}\,dx \qquad u_e(x) = u_e(1) + \int_{x_1}^{x} \frac{du_e}{dx}\,dx$$

$$n(x) = n(1) + \int_{x_1}^{x} \frac{dn}{dx}\,dx \qquad \psi(x) = \psi(1) + \int_{x_1}^{x} \frac{d\psi}{dx}\,dx$$

$$J_e(x) = J_e(1) + \int_{x_1}^{x} \frac{dJ_e}{dx}\,dx$$

integration of the transport differential equations since, in general, the assumed values $T_e(1)$ and $n(1)$ are not correct. Accordingly, new estimates of $T_e(1)$ and $n(1)$ are made, and the process is repeated until some satisfactory convergence is achieved between integration and boundary condition results at x_2.

For more details about the techniques used in THRIVE, the reader is referred to Ref. 5.

The THRIVE computer program has been used to obtain results in the following ranges of operating conditions:

$$1600 \leqslant T_E \leqslant 2100^\circ\text{K} \qquad 800 \leqslant T_C \leqslant 900^\circ\text{K}$$

$$571 \leqslant T_R \leqslant 635^\circ\text{K} \qquad 1 \leqslant d \leqslant 10 \text{ mils} \qquad 0 \leqslant J \leqslant 15 \text{ amp/cm}^2$$

For values of T_R, d, and J greater than the upper limits of these ranges, the program suffers from difficulties of numerical convergence.

The necessary coefficients of the transport differential equations are evaluated for hard-sphere collision cross sections such that $\sigma_{ea} = 400$ Å^2 and $\sigma_{ia} = 570$ Å^2. The ionization energy of cesium is taken as $V_i = 3.89$ ev, and the model for the emission properties of the cesiated electrodes is as described in Ref. 5.

Typical results obtained under various combinations of electrode temperatures, cesium pressures, and interelectrode spacings are shown in Figs. 12.3 to 12.12. The input conditions listed on each figure include the emitter ion-richness ratio β_E and the pd product. The figures are self-explanatory.

Some general conclusions reached by Wilkins and coworkers as a consequence of these studies are:

1. The plasma motive difference $\Delta\psi_p$ can be either positive or negative.

2. The charged-particle number density n vs. x has the following features: (a) in the ignited mode it peaks near the center of the interelectrode space; (b) its curvature decreases with current; (c) in the unignited mode it is monotonically decreasing from emitter to collector; (d) in the ignited mode its amplitude increases with current; (e) in the unignited mode its amplitude may or may not increase with current; (f) its peak tends to move toward the emitter as the current is raised; and (g) its gradient (dn/dx) is usually largest near the collector.

3. The electron temperature T_e vs. x has the following features: (a) its gradients are appreciable, particularly in the ignited mode of operation; (b) its gradient can be either positive or negative at the emitter sheath–plasma interface; (c) at open circuit it is monotonic, almost constant at the temperature of the emitter; (d) in the ignited mode it is significantly greater than the temperature of the cesium vapor; and (e) in the ignited mode it becomes larger and its gradients become steeper with increasing cesium pressure.

4. The ion current density Γ_i or J_i vs. x has the following features: (a) in the ignited mode positive ions are created in the plasma and diffuse toward the electrodes (i.e., the ion current is negative near the emitter and positive near the collector); (b) net ionization rates are largest for high currents; (c) in the ignited

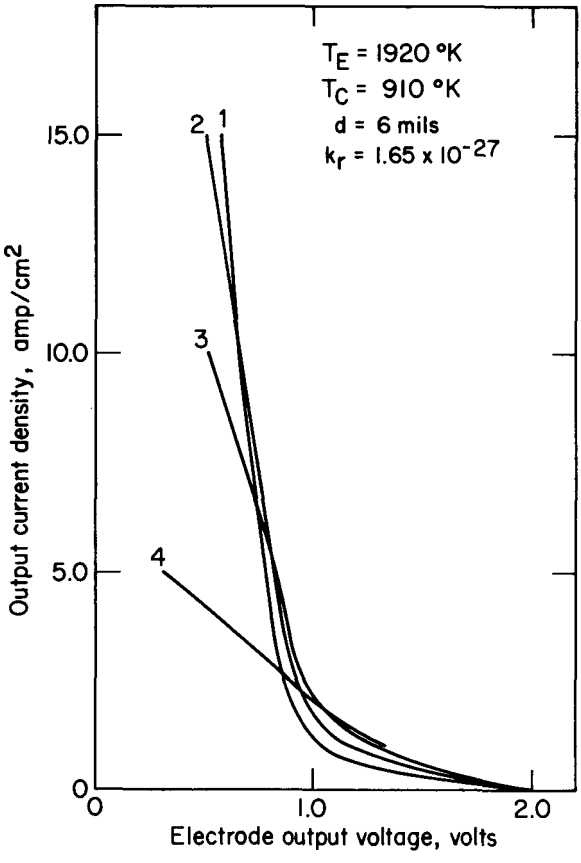


Fig. 12.3 Computed output current characteristics. (Computed according to the methods of Ref. 5.)

Curve	T_R, °K	ϕ_{E_0}, ev	ϕ_{C_0}, ev	pd, mil-torr	β_E
1	635	2.655	1.729	47.0	0.26
2	614	2.798	1.711	29.6	0.92
3	595	2.924	1.694	18.9	2.64
4	571	3.113	1.669	10.2	14.43

mode the ionization rate is largest near the emitter, and the peak ionization rate moves closer to the emitter as cesium pressure is decreased; (d) ion production rates are small or negative near the collector; and (e) in the unignited mode charge neutrality in the plasma is achieved because of thermionically emitted ions.

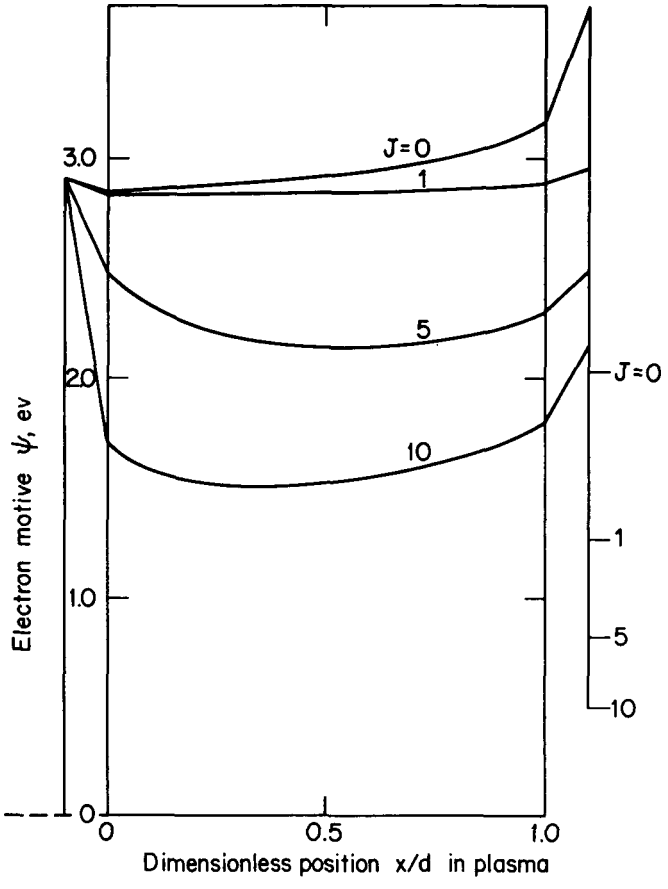


Fig. 12.4 Computed plasma motive ψ vs. x/d. (Computed according to the methods of Ref. 5.)

$T_E = 1920°K$	$d = 6$ mils	$pd = 18.9$ mil-torr
$T_C = 910°K$	$\phi_{E_0} = 2.924$ ev	$\beta_E = 2.69$
$T_R = 595°K$	$\phi_{C_0} = 1.694$ ev	$k_r = 1.65 \times 10^{-27}$

Other observations are given in Ref. 5.

The THRIVE computer program has also been used for a series of sensitivity studies of performance characteristics to the values of various input constants. The numerical results of these studies will not be reproduced here. Some general conclusions reached by Wilkins and coworkers are:

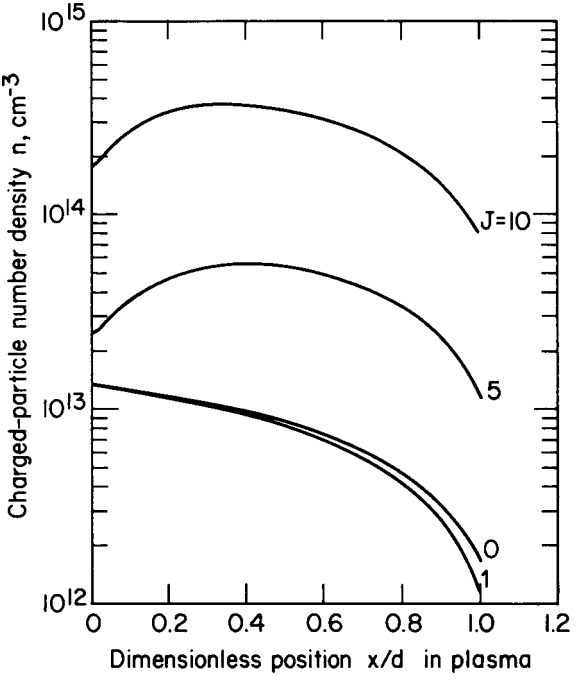


Fig. 12.5 Computed charged-particle number density n vs. x/d. (Computed according to the methods of Ref. 5.)

$T_E = 1920°K$	$d = 6$ mils	$pd = 18.9$ mil-torr
$T_C = 910°K$	$\phi_{E_0} = 2.924$ ev	$\beta_E = 2.69$
$T_R = 595°K$	$\phi_{C_0} = 1.694$ ev	$k_r = 1.65 \times 10^{-27}$

1. Variations of several hundredths of an electron volt in the emitter work function have a significant effect on the output current characteristics in the ignited mode but a relatively small effect in the unignited mode.

2. When collector emission is very small, changes in collector work function have a negligible effect on the transport processes.

3. Large variations of the coefficient k_r (Eq. 12.10) in the range $10^{-27} \leqslant k_r \leqslant 10^{-26}$ have a small effect on the output-current characteristics but an appreciable effect on the distribution T_e vs. x in the plasma.

4. Output-current characteristics and T_e vs. x distributions are sensitive to the values of σ_{ea} and σ_{ia}.

5. In the ignited mode output-current characteristics are sensitive to the estimate of the Schottky effect since this effect is equivalent to a change of electrode work functions.

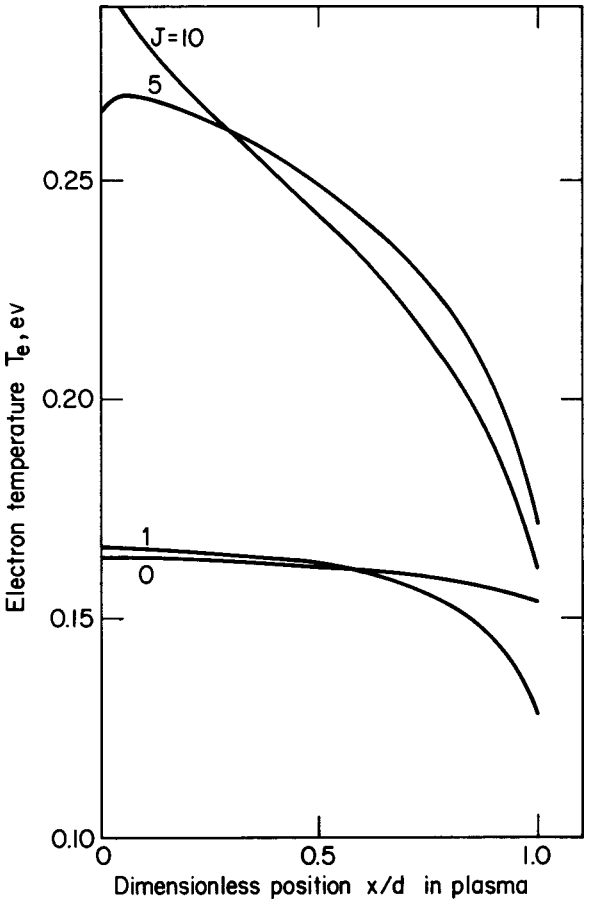


Fig. 12.6 Computed electron temperature T_e vs. x/d. (Computed according to the methods of Ref. 5.)

$T_E = 1920°K$	$d = 6$ mils	$pd = 18.9$ mil-torr
$T_C = 910°K$	$\phi_{E_0} = 2.924$ ev	$\beta_E = 2.69$
$T_R = 595°K$	$\phi_{C_0} = 1.694$ ev	$k_r = 1.65 \times 10^{-27}$

Wilkins, McCandless, and Derby[5] also report results that correspond to decelerating emitter sheaths. The validity of these results is questionable, however. A decelerating emitter sheath implies a double-valued sheath. In view of the complete lack of detailed analyses of double-valued sheaths, representing the effects of such a sheath by a barrier height and a Schottky effect, as was done in THRIVE, may be completely inadequate.

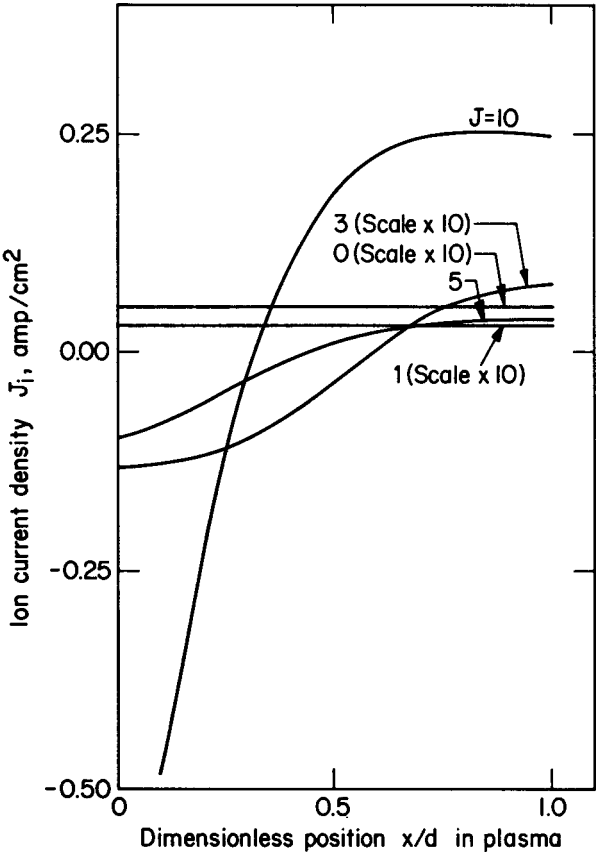


Fig. 12.7 Computed ion current density J_i vs. x/d. (Computed according to the methods of Ref. 5.)

$T_E = 1920°K$	$d = 6$ mils	$pd = 18.9$ mil-torr
$T_C = 910°K$	$\phi_{E_0} = 2.924$ ev	$\beta_E = 2.69$
$T_R = 595°K$	$\phi_{C_0} = 1.694$ ev	$k_r = 1.65 \times 10^{-27}$

In addition, the boundary conditions used in THRIVE are inaccurate even under conditions for which the gradients of temperature and motive at the sheath–plasma interface are negligible, namely, when

$$\left|\frac{\hat{\lambda}_e}{T_e}\frac{dT_e}{dx}\right|_\alpha \ll 1 \quad \text{and} \quad |a_e|_\alpha = \left|\frac{\hat{\lambda}_e eE}{2kT_e}\right|_\alpha \ll 1 \quad \text{for } \alpha = 1 \text{ and } 2$$

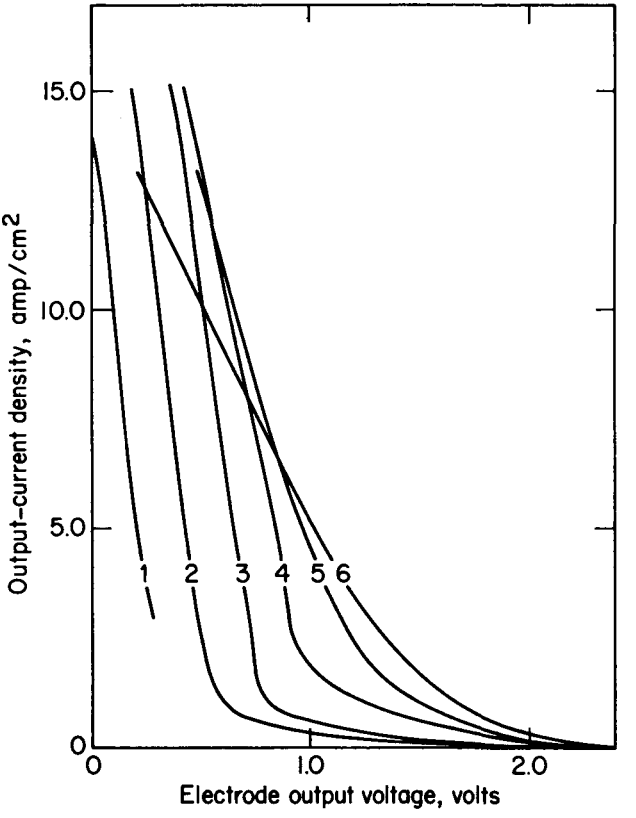


Fig. 12.8 Computed output-current characteristics. (Computed according to the methods of Ref. 5.)

Curve	T_E, °K	T_C, °K	T_R, °K	ϕ_{E0}, ev	ϕ_{C0}, ev	pd, mil-torr	β_E
1	1599	818	603	2.268	1.783	22.9	0.004
2	1715	840	605	2.436	1.767	24.0	0.03
3	1820	882	605	2.622	1.729	24.0	0.16
4	1920	910	605	2.856	1.703	24.0	1.50
5	2020	910	605	3.094	1.703	24.0	11.82
6	2120	910	605	3.350	1.703	24.0	92.60

For example, under these conditions and $(\Delta\gamma_{Ee})_1 > 1$, boundary condition 11.213 can be approximated by the relation

$$\{\Gamma_e = \Gamma_{ve}^E - [\Gamma_{re} \exp(-\Delta\gamma_{Ee})]\}_1$$

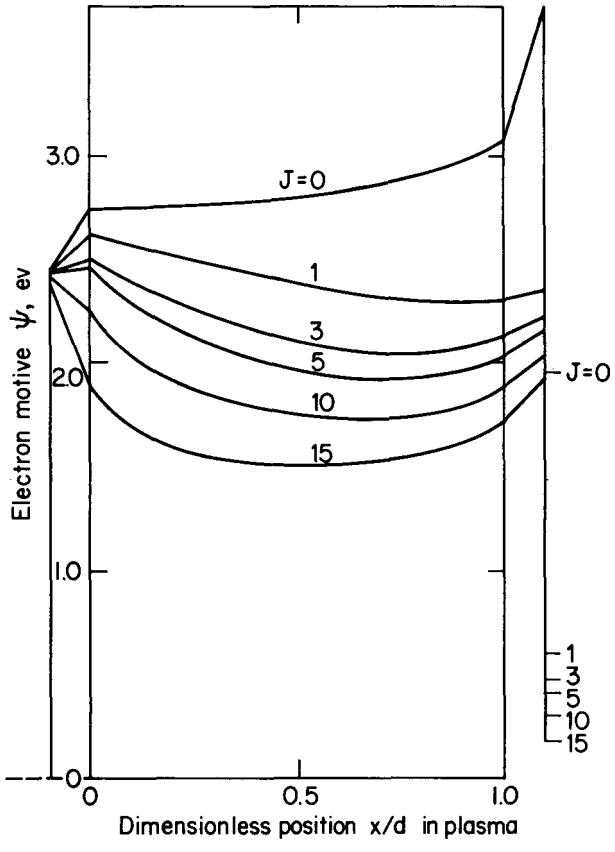


Fig. 12.9 Computed plasma motive ψ vs. x/d. (Computed according to the methods of Ref. 5.)

$T_E = 1715°K$	$d = 6$ mils	$pd = 24.0$ mil-torr
$T_C = 840°K$	$\phi_{E_0} = 2.436$ ev	$\beta_E = 0.03$
$T_R = 605°K$	$\phi_{C_0} = 1.767$ ev	$k_r = 1.65 \times 10^{-27}$

whereas boundary condition 11.167, used in THRIVE, is given by the approximate relation

$$\left[\Gamma_e = \Gamma_{ve}^{E} - \left(\Gamma_{re} - \frac{\Gamma_e}{2}\right) \exp\left(-\Delta\gamma_{Ee}\right)\right]_1$$

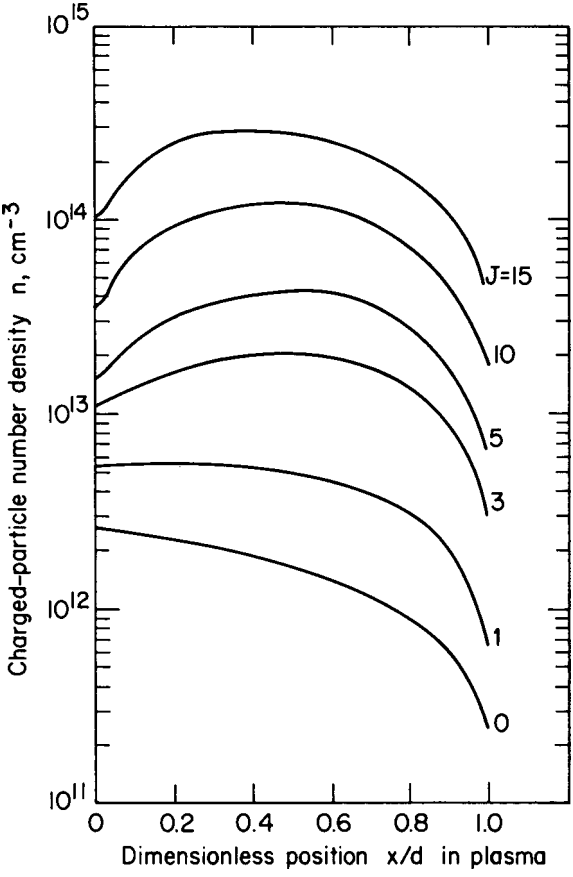


Fig. 12.10 Computed charged-particle density n vs. x/d. (Computed according to the methods of Ref. 5.)

$T_E = 1715°K$	$d = 6$ mils	$pd = 24.0$ mil-torr
$T_C = 840°K$	$\phi_{E_0} = 2.436$ ev	$\beta_E = 0.03$
$T_R = 605°K$	$\phi_{C_0} = 1.767$ ev	$k_r = 1.65 \times 10^{-27}$

Again, under the small gradient conditions and in the limit $\Delta\gamma_{Ee} \to 0$, boundary condition 11.215 reduces to the relation

$$[u_e = 4(kT_E - kT_e)\Gamma^E_{ve} + (2kT_e + \psi_E)\Gamma_e]_1$$

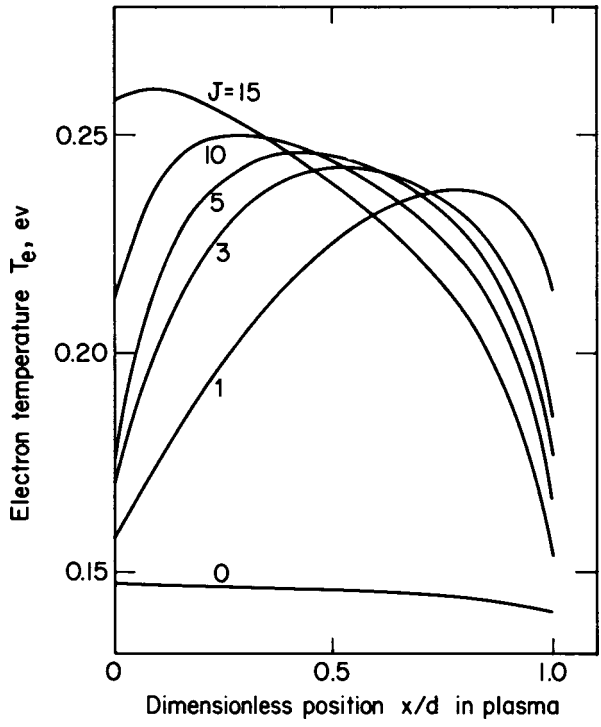


Fig. 12.11 Computed electron temperature T_e vs. x/d. (Computed according to the methods of Ref. 5.)

$T_E = 1715°K$	$d = 6$ mils	$pd = 24.0$ mil-torr
$T_C = 840°K$	$\phi_{E_{01}} = 2.436$ ev	$\beta_E = 0.03$
$T_R = 605°K$	$\phi_{C_0} = 1.767$ ev	$k_r = 1.65 \times 10^{-27}$

whereas boundary condition 11.169, used in THRIVE, is given by the relation

$$[u_e = 2(kT_E - kT_e)\Gamma_{ve}^E + (2kT_e + \psi_E)\Gamma_e]_1$$

Finally, as already indicated, THRIVE and other computer programs face numerical difficulties in the integration of the transport differential equations. We will illustrate some of these numerical difficulties by considering the explicit solutions of Eq. 12.23 for the special case of uniform electron and ion temperatures. Although approximate, such solutions are very informative.

We will consider a plasma ($n_e \approx n_i$) extending over an effective thickness, d_e, which consists of the actual thickness of the plasma ($x_2 - x_1 \approx d$) plus two

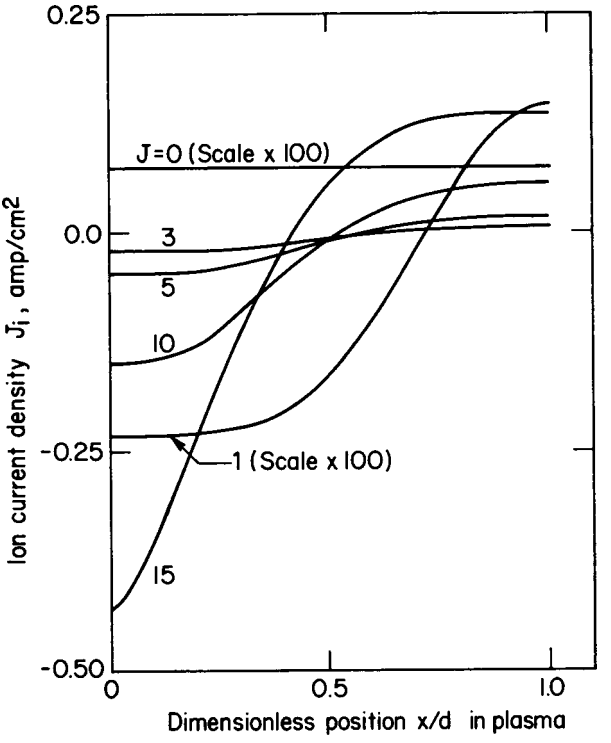


Fig. 12.12 Computed ion current density J_i vs. x/d. (Computed according to the methods of Ref. 5.)

$T_E = 1715°K$	$\phi_{E_0} = 2.436$ ev	$d = 6$ mils
$T_C = 840°K$	$\phi_{C_0} = 1.767$ ev	$\beta_E = 0.03$
$T_R = 605°K$	$pd = 24.0$ mil-torr	$k_r = 1.65 \times 10^{-27}$

extrapolation thicknesses (d_1 and d_2), as shown schematically in Fig. 12.13, i.e., $d_e = d_1 + d + d_2$.

For T_e^* and T_i^* constant, differentiating Eq. 12.23 with respect to x^* and using Eq. 12.20, we find the relation

$$y'' = -\delta^2 y\,(1 - y^2) \tag{12.27}$$

where

$$y = \frac{n^*}{n_{eq}^*} \quad \text{and} \quad \delta^2 = \frac{1}{\sqrt{\pi}\lambda_e^*\lambda_r^*}\,\frac{T_e^{*1/2}}{T_e^* + T_i^*}\,\frac{m_i\hat{\nu}_i}{m_e\hat{\nu}_e}\,n_{eq}^{*2}$$

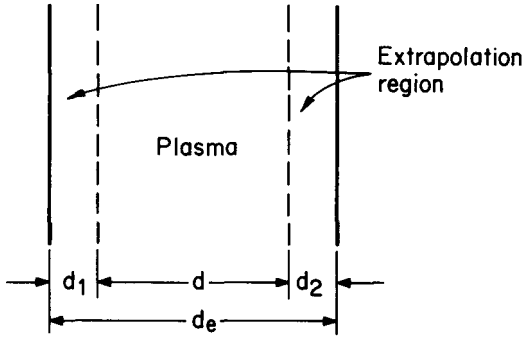


Fig. 12.13 Schematic of plasma with extrapolation regions.

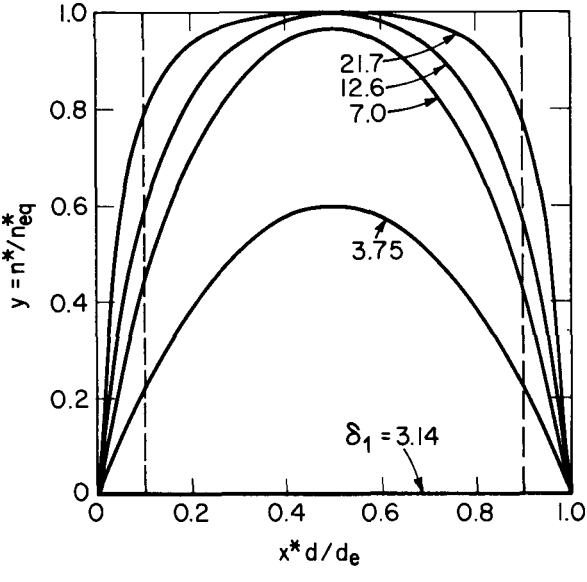


Fig. 12.14 Solutions for Eq. 12.27.

and the double prime denotes second derivative with respect to x^*. The solutions of Eq. 12.27 are Jacobi elliptic functions determined by the boundary conditions $y = 0$ at the two extrapolated boundaries of the plasma. A family of solutions y vs. $x^* d/d_e$ for different values of $\delta_1 = \delta(d_e/d)$ is shown in Fig. 12.14.

We can make several important observations about the solutions which are of general interest and are related to the problem of numerical difficulties:

1. For $\delta_1 < \pi$ the only solution is $y = 0$. Physically this implies that insufficient ions are produced by inelastic collisions and no plasma can be sustained.

2. The condition $\delta_1 = \pi$ can be interpreted as an ignition criterion defining the minimum electron temperature, the ignition temperature, necessary for the establishment of a self-sustained plasma.

3. For $\delta_1 > \pi$ nontrivial solutions exist. It is noteworthy that an increase of δ_1 by less than an order of magnitude above π results in solutions that are $y \approx 1 (n^* \approx n^*_{eq})$ virtually throughout the entire thickness of the plasma. Since δ_1 depends exponentially on T^*_e, the transition from the diffusion-dominated solution $y = 0$ to the ionization–recombination dominated solution $y \approx 1$ is stimulated by a small increase (of the order of a few hundred degrees Kelvin) in the electron temperature above the ignition temperature.

4. The values of the ignition temperature and n^*_{eq} depend on the cesium pressure, the plasma thickness, and the ionization–recombination characteristics.

One source of the numerical difficulties mentioned earlier can now be traced to the existence of solutions of the form $y \approx 1$ over a broad range of positions throughout the plasma. Reproducing such solutions by means of numerical methods is an extremely difficult task that has not been overcome in the various computer programs such as THRIVE.

Under certain conditions of operation the numerical difficulty can be reduced by using semianalytical methods. For example, Keck[14] has suggested using a parametric form for the electron production rate and integrating the transport equations to determine the electron number density, motive, and temperature distributions. These distributions are used to determine the electron production rate from the ionization-rate equation. Then the parameters in the initial parametric form are adjusted to make the assumed production rate match the calculated production rate as closely as possible. Again an iterative technique can be used.

In the limit $\lambda^*_e \Gamma^*_{re} \gg 1$, Eqs. 12.21 and 12.22 yield the zero-order solutions

$$T^{*(0)}_e = \text{constant} \quad \text{and} \quad \mu^{*(0)}_e = \text{constant}$$

These solutions do not imply that the bracketed quantities in the right-hand sides of Eqs. 12.21 and 12.22 are identically equal to zero but that the gradients of T^*_e and μ^*_e are very small because the magnitude of $\lambda^*_e \Gamma^*_{re}$ is very large. They are reasonable when the electron free path λ^*_e is much greater than the ion free

path λ_i^*, which is generally true for certain conditions of operation of converters.

The 0*th*-order value $T_e^{*(0)}$ can be used in Eqs. 12.20 and 12.23 to obtain the 0*th*-order solution $n^{*(0)}$ for the density. The density, together with $T_e^{*(0)}$ and $\mu_e^{*(0)}$, can then be used to evaluate the energy-flux density $u_e^{*(0)}$, the random-flux densities $\Gamma_{r\alpha}^{*(0)}$, the particle current densities $\Gamma_\alpha^{*(0)}$, and the motive $\psi^{*(0)}$. These results can then be substituted in Eqs. 12.21 and 12.22 to evaluate the first-order solutions $T_e^{*(1)}$ and $\mu_e^{*(1)}$, and the procedure can be repeated to obtain $n^{*(1)}$ etc.

The iterative procedure cited can be repeated as many times as necessary to achieve desired accuracy. Preliminary results indicate, however, that convergence is extremely fast and that, in general, the 0*th*-order solution $n^{*(0)}$ is sufficient for most practical problems.

References

1. J. V. Dugan, Three-Body Collisional Recombination of Cesium Seed Ions and Electrons in High-Density Plasmas with Argon Carrier Gas, Report NASA-TN-D-2004, National Aeronautics and Space Administration, October 1964.
2. L. P. Harris, Ionization and Recombination in Cesium-Seeded Plasmas Near Thermal Equilibrium, *J. Appl. Phys.*, **36**: 1543 (1965).
3. E. Hinnov and J. G. Hirschberg, Electron-Ion Recombination in Dense Plasmas, *Phys. Rev.*, **125**: 795 (1962).
4. D. R. Wilkins, SIMCON: A Digital Computer Program for Computing Thermionic Converter Performance Characteristics, USAEC Report GESR-2109, General Electric Company, Pleasanton, Calif., Feb. 1, 1968.
5. D. R. Wilkins, R. J. McCandless, and S. L. Derby, Theoretical Analysis of Thermionic Converter Volume Phenomena, Report AFAL-TR-G9-354, General Electric Company, Johnson City, N. Y., 1970.
6. C. Warner and L. K. Hansen, Basic Research in Thermionic Energy Conversion, Report AI-67-155, Atomics International, 1967; and Report AI-73-7, Atomics International, 1973.
7. F. G. Baksht, G. A. Djuzkev, V. B. Kaplan, I. L. Korobova, A. M. Martzinovskiy, B. Ya. Moizhes, G. A. Schohnasarova, and V. G. Yuriev, Theoretical and Experimental Investigation of Low-Voltage Arc in Thermionic Converter, in *Proceedings of the Second International Conference on Thermionic Electrical Power Generation, Stresa, Italy, May 27–31, 1968*, pp. 1147-1159.
8. D. Lieb and W. Bornhorst, Analysis of the Diffusion-Dominated Thermionic Converter Plasma, in *Proceedings of the Thermionic Conversion Specialists Conference, Palo Alto, Calif., Oct. 30–Nov. 1, 1967.*
9. R. H. Curry, D. W. Norcross, P. M. Stone, and K. J. Nygaard, An Experimental and Theoretical Study of the Non-Equilibrium Plasma in Thermionic Discharges, Report SRRC-CR-68-4, Sperry Rand Research Center, Sudbury, Mass., January 1968.
10. R. H. Bullis and W. L. Nighan, Investigation of Collision Probability of Electrons and Ions with Alkali Metals and Atoms, Report UACRL-E-920243-27, United Aircraft Corp., East Hartford, Conn., 1966.

11. J. W. Sheldon, Semiclassical Calculation of the Differential Scattering Cross Section with Charge Exchange: Cesium Ions in Cesium Vapor, Report NASA-TN-D-2484, National Aeronautics and Space Administration, September 1964.
12. J. M. Houston, Cross Section Values to Use in Analyzing the Cesium Thermionic Converter, in *Proceedings of the Thermionic Conversion Specialists Conference, Cleveland, Ohio, Oct. 26–28, 1964;* and in *Proceedings of the Twenty-Fourth Annual Conference on Physical Electronics, Cambridge, Mass., Mar. 25–27, 1964.*
13. P. M. Stone and J. R. Reitz, Elastic Scattering of Slow Electrons by Cesium Atoms, *Phys. Rev.*, **131**: 2101 (1963).
14. J. Keck, An Approximate Analytic Technique for Determining the Operating Characteristics of Thermionic Converters in the Ignited Mode, in *Proceedings of the Thermionic Conversion Specialists Conference, Miami Beach, Fla., Oct. 26–29, 1970.*

13

Experimental Techniques

13.1 General Remarks

This chapter describes the experimental techniques commonly used in the study of thermionic conversion. Some of the techniques are employed in many other fields, and for this reason they are discussed only briefly. On the other hand, techniques peculiar to thermionic conversion are discussed in detail.

The chapter includes discussions of techniques used in chemical and metallurgical characterization of converter materials, in the determination of electrode work functions, and in the measurement of input and output converter parameters, such as electrode temperatures and output currents and voltages.

13.2 Characterization of Materials

Characterizing materials used in the construction of a device consists in determining the features of composition and structure of the materials which are relevant to the function of the device. For example, characterizing a thermionic emitter requires determining the composition and crystallographic orientation of its surface, because these two features define its emission capabilities. The precision of the characterization required for a material is not the same for all its properties. For example, the presence of a minute quantity of an impurity in the bulk of an electrode can be unimportant in evaluating its structural integrity, but the presence of a minute quantity of oxygen on the surface of an electrode is very important in determining the emission properties of the surface.

In general, the characterization of materials used in thermionic converters involves the study of the composition and structure of (1) the bulk of each material, (2) solid–solid interfaces, and (3) solid–vapor interfaces, namely, surfaces.

A number of experimental techniques are employed to characterize materials used in thermionic converters. Some techniques are described briefly in the following several sections. We should note, however, that the problem of characterization in general, and of surfaces in particular, is very difficult and present results are mostly qualitative. Moreover, even if complete characterization were possible, rigorous relations between composition and structure and emission characteristics of a surface are not available. For these reasons, characterization of converter materials contributes primarily to the reproducibility of these materials.

Table 13.1 Techniques Used to Determine Bulk Composition of Materials[1]

Technique	Application	Sensitivity	Precision
Emission spectroscopy	Survey of impurities	0.1 to 100 ppm	5 to 10%
Spark-source mass spectrometry	Survey of impurities	0.01 to 0.1 ppm	Semiquantitative
Laser and micro-spark spectroscopy	Survey of impurities near surface	Variable	Semiquantitative
Vacuum-fusion mass spectrometry	Interstitial impurities (O_2, N_2, H_2)	0.07 ppm	20%
Ion-probe mass spectroscopy	Local survey of impurities on surface	Semiquantitative	
Electron-microprobe analysis	Homogeneity of major phase	0.1% over 1 to 5-μ scan	
Neutron activation	Trace of impurities	0.001 to 0.01 ppm	2 to 10%
X-ray fluorescence spectrometry	Major and minor constituents	20 to 200 ppm	0.1%
Mössbauer spectroscopy	Major phase and trace impurities	0.1% to 0.0001%	
Wet-chemistry titrimetry	Major and minor phases	$10^{-2}M$ in solid	0.1%
		$10^{-5}M$ in solid	0.1%
		10^{-6} to $\sim 10^{-7}M$ in solid	0.2 to 1%

13.3 Characterization of Bulk

13.3.1 Determining Bulk Composition. Table 13.1 is a partial list of experimental techniques that can be used to determine the composition of the bulk of materials of interest to thermionic conversion. Included in the table are the purpose, sensitivity, and precision of each technique. More detailed lists can be found in standard texts.[2,3]

Of the techniques listed, the most frequently used are *spark-source mass spectrometry* and *electron-microprobe analysis.* The principles of operation and procedures used in these two techniques are as follows:

1. *Spark-Source Mass Spectrometry.* The principles of operation of spark-source mass spectrometry consist in the ionization and electrostatic acceleration of atoms extracted from a sample under test, the mass separation of the accelerated ions by means of a magnetic field, and the collection of ions of a

given mass number on a collector. Ions are usually collected on photographic plates, and special tables are used to relate the intensities of exposures of the plates to the composition of the sample.

The sample, electric and magnetic fields, and photographic plates are in a highly evacuated chamber. Recent advances in ultrahigh vacuum techniques and in methods for cleaning samples have substantially improved the reliability and versatility of spark-source mass spectrometers.

The ultrahigh vacuum spectrometer operates as follows: Thin sample electrodes are cut from the material being tested. Two samples are chemically etched, mounted on holders of known composition and purity, and introduced into the spectrometer. The system is purged, a relatively inert gas is introduced, and a cathodic etch is performed by establishing a glow discharge between the two electrodes. The inert gas is removed by adsorption pumps, and then the pressure is reduced to 10^{-8} torr by ion and cryo pumps. Ionized atoms of the samples are formed by means of a spark discharge between the two electrodes. Usually a precursory spark is used to minimize surface contaminants.

The results of analysis of a tungsten sample used in the fabrication of emitters, given in Table 13.2, are similar to those reported in the literature except that, by using ultrahigh vacuum and in situ cleaning, we could detect interstitial elements such as carbon, nitrogen, and oxygen. Determining interstitial elements is important because of the strong effect trace amounts of these elements can have on structural and electronic properties of thermionic-converter materials.

The main limitations of the technique arise in selecting and preparing the sample. Materials, and therefore samples, are not homogeneous. For example, zone-refined materials often have variations in impurities related to axial position, and rolled or drawn materials have different composition in a direction perpendicular to the axis of rolling or drawing. We must recognize the nature of impurity variations to adequately determine the composition of a material without an excessive number of test samples.

Because preparing a sample electrode by machining can introduce nonrepresentative impurities, extensive sample cleaning should follow the electrode manufacturing process. If the manufacturing process for the sample is similar to that for the fabrication of converter electrodes, however, low-intensity sparks can be used to ionize the layers near the surface and thus to obtain information on surface impurities.

2. *Electron-Microprobe Analysis.* The principles of operation of electron-microprobe analysis consist in bombarding a sample by a high-intensity high-energy electron beam and observing the intensity and energy patterns of back-scattered electrons or X rays, or both. Since the electron beam can be highly localized, it can be used for scanning sample surfaces.

Table 13.2 Typical Mass-Spectrometric Analysis of a Tungsten Emitter*†

Element	Detection limit	Sample W–L monocrystal	Element	Detection limit	Sample W–L monocrystal
H	0.01	2.7	K	0.002	19
Li	0.001	1.0	Ca	0.02	2.8
B	0.01	0.19	V	0.03	0.53
C	0.01	230	Cr	0.03	15
N	0.01	17	Mn	0.03	0.78
O	0.01	110	Fe	0.05	4.9
F	0.01	100	Co	0.03	2.2
Na	0.001	200	Ni	0.05	19
Mg	0.01	5.4	Zn	0.1	0.3
Al	0.01	32	Ga	0.05	76
Si	0.02	20	Ge	0.07	4.3
P	0.02	6.1	As	0.1	7.7
S	0.02	0.54	Mo	0.3	120
Cl	0.02	1.1			

*Tantalum, zirconium, and niobium are not included. The mass spectrometer uses tantalum electrodes that introduce a spurious tantalum background signal. The high concentrations of tungsten and molybdenum interfere with the analysis of zirconium and niobium. Other impurities not listed are those with concentrations below their detection limit.

†Impurity concentrations in tungsten are given in parts per million by weight (ppmw). Spark intensity was 3×10^{-7} coulomb.

Microprobe analysis lacks the quantitative resolution of spark-source mass spectrometry, but the advantages of spatial resolution and direct readout of results (for a limited number of elements only) make it appropriate for studies of mass transport across interfacial regions, such as interdiffusion between a thermionic emitter material and the material of the supporting structure. Interdiffusion is an important process because it can result in contamination of the emitter surface or formation of voids in the interfacial region, or both. Void formations in rhenium–tantalum electrodes used in high-temperature converters have been so severe that they impede heat transfer and, in some cases, induce failure of the rhenium–tantalum bond.

Suitable samples are used to calibrate electron microprobes. For example, for analysis of the rhenium–tantalum electrode mentioned previously, calibration is achieved by means of various rhenium–tantalum alloys. We can avoid the need for extensive calibration by using a simple physical model for the sample which permits extrapolation of the results of a few key calibration points.[4] In addition, attempts have been made to predict the back-scattered patterns by means of the

relations that govern elastic and inelastic electron scattering and X-ray production and transmission.[5]

13.3.2 Determining Bulk Structure. Although many techniques are available for determining bulk structure, visual examination, combined with *optical photomicroscopy* and *X-ray diffraction*, has been used almost exclusively for the study of converter materials.

Optical photomicroscopy consists in photographic documentation of the visual examination often used to determine bulk structure. After a sample of the material is cross sectioned, polished, and etched, photomicrographs of various degrees of enlargement are made. The photomicrographs yield information on grain size, diffusion zones, integrity of structures, and, in conjunction with other experimental methods, local crystal orientation. A photomicrograph of a cross section of a chemically vapor-deposited tungsten electrode is shown in Fig. 13.1. For resolution higher than that possible with photomicrographs, we must use electron microscopy.

The *X-ray diffraction* technique is used in two ways. For monocrystalline materials or materials that consist of large-size grains, back-reflection Laue patterns are observed and correlated to the structure of the material. With conventional mechanical and photographic equipment, the angular orientation can be resolved to 1°, whereas, with more sophisticated apparatus and procedures, the uncertainty can be reduced to 0.01°.

Quantitative information for polycrystalline materials is obtained by means of X-ray rocking curves. The sample is set at the Bragg angle of an oriented lattice plane and then rocked, and the intensity of the diffracted beam is recorded. Typical arrangement and results[6] are shown in Figs. 13.2 and 13.3.

13.4 Characterization of Surfaces

Characterization of surfaces is of paramount importance to thermionic conversion because the performance characteristics of converters depend strongly on the composition and structure of electrode surfaces.

Table 13.3 is a partial list of experimental techniques typically used to determine the composition, structure or morphology, and orientation of surfaces. Each of the techniques is briefly discussed in the section indicated in the table.

As stated earlier, the problem of characterizing surfaces is very difficult, and presently available experimental techniques yield either ambiguous or qualitative results. For example, the only surface structure that has been studied in atomic detail is that of the needlelike tip used as the emitter in an ion microscope. Nevertheless, the techniques described in the next few sections yield helpful information about general features of electrode surfaces.

Fig. 13.1 Photomicrograph of chemically vapor-deposited tungsten electrode.

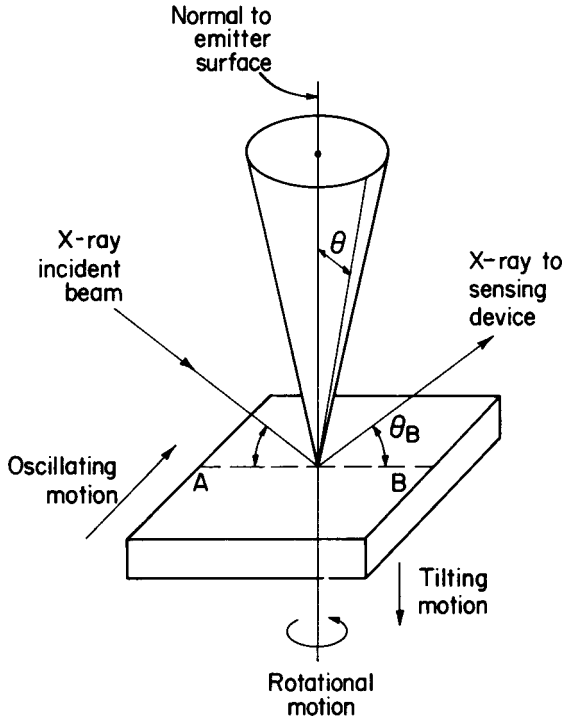


Fig. 13.2 Schematic of arrangement for obtaining spatial distribution of lattice direction of preferred orientation in a planar emitter. During X-ray exposure the emitter is oscillated perpendicular to the line AB, rotated in its own plane, and tilted about AB.

13.4.1 Mass Spectrometry. Variations of the mass-spectrometry approach discussed in Sec. 13.3.1 can be used to determine impurities on a surface. For example, when a sample is heated suddenly to a high temperature, impurity particles are emitted from its surface (thermal desorption) and can be analyzed by means of a mass spectrometer. Again, particle emission can be induced by bombarding the sample with energetic ions and electrons (ion and electron desorption), and mass spectrometry can be performed.

Often information on the bulk composition suggests particular surface impurities. A chemically active gas can then be used to react with the suspected impurity, and the resultant gaseous product (chemical desorption) can be mass analyzed. For example, we can use gaseous oxygen to determine the presence of carbon on a surface by observing the mass peaks corresponding to carbon monoxide and carbon dioxide. Frequently this chemical detection method is used in conjunction with surface preparation and cleaning.

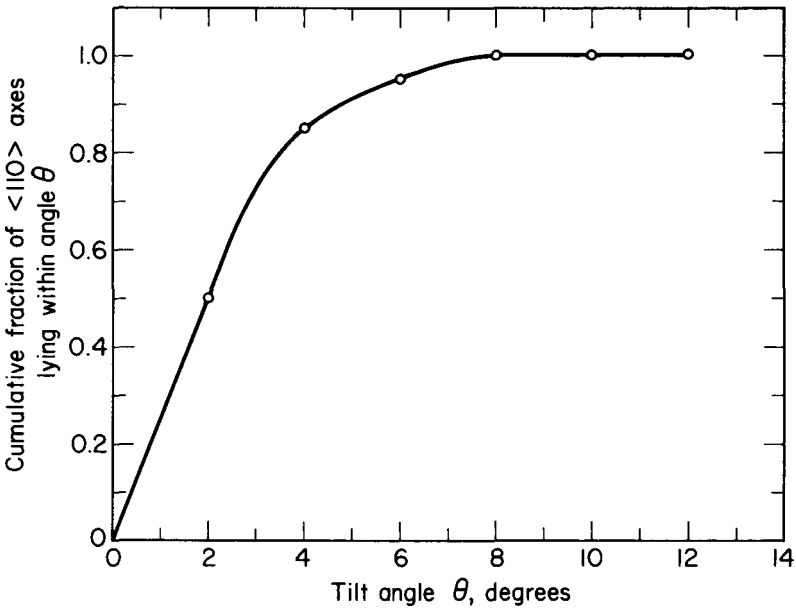


Fig. 13.3 Typical results of X-ray diffraction experiments.

Table 13.3 Techniques Used to Characterize Surfaces

Technique	Section
Surface Impurities	
Mass spectrometry	13.4.1
Thermal desorption	
Ion and electron desorption	
Chemical desorption	
Auger emission	13.4.2
Surface Morphology	
Optical and electron microscopy	13.4.3
Profilometry	13.4.4
Surface Orientation	
X-ray diffraction	13.4.5
Laue back reflection	
Rocking curves	
Low-energy electron diffraction (LEED)	13.4.6
Etch-pit analysis	13.4.7
Light-reflection goniometry	13.4.8

13.4.2 Auger Emission. In Auger emission a sample surface is bombarded with electrons in the energy range from 100 to 1000 ev. Inner-shell electrons of atoms at or near the surface are excited from a lower to a higher energy level, and each atom de-excites by emitting an electron. The energy of an emitted electron is related to the difference of energy of two inner-shell energy eigenstates of an atom and, therefore, is characteristic of that atom. Thus, in principle, the atomic composition of the surface can be determined.

In Auger emission the signal level is small with respect to noise, and therefore the technique requires the use of sophisticated instrumentation. Because of the small signal level, the sample is usually kept at or near room temperature. This, in turn, requires the use of ultrahigh vacuum so that the sample surface is maintained clean because the product of pressure and exposure time required for a monolayer of gas to form on a nonemitting surface is of the order of 10^{-6} torr-sec.

Although much progress has been made in determining the Auger emission spectrum of most elements of interest, the interpretation of spectra from surfaces with several elements is difficult. Only recently the technique was used to determine the oxygen coverage of a tungsten (1 1 0) surface.[7]

The interpretation of Auger emission from rough or multiple-layered surfaces is in the formative stage. Any structure or material that attenuates the energy of the Auger emitted electrons complicates the interpretation. In general, Auger emission should be considered as a guide for surface-preparation procedures rather than as a method for determining the composition of electrode surfaces.

13.4.3 Optical and Electron Microscopy. Optical and electron microscopy are universally used in studies of morphology of converter-material surfaces. These low-cost nondestructive methods yield information on flatness, grain size, faceting, imperfections, existence of impurities, work damage, and other features of surfaces. Modern apparatuses include stereographic photography and the use of comparators, such as small microspheres on the surface. Thus quantitative information on the surface topography can be obtained.

In optical microscopy direct visual observation is ordinarily used, and in electron microscopy the replication technique is employed. Replication is achieved by means of thermal, solvent, or pressure attachment of organic films to the surface of interest. The organic film surface is then removed, a thin metallic film is condensed on the embossed surface (shadow casting), and the surface is viewed in the electron microscope. The ease of sample preparation and examination makes electron microscopy an excellent complement to optical examination. Photographic records ranging from a view of the entire surface to a view of a portion of the surface with a resolution of 100 Å are readily available.

Judicious use of optical and electron microscopy to examine electrode surfaces before and after experiments has yielded valuable information about converters.

Numerous examples of the application of these forms of microscopy appear in the literature.

13.4.4 Profilometry. In profilometry a stylus-activated transducer is moved across a sample surface, and the vertical deflections are recorded. Usually the record consists of a graph of the deflection vs. position of the stylus on the sample surface. Sometimes the output of the stylus transducer is fed directly into computerized data processors that, in turn, analyze and classify the various surface irregularities.

Profilometry is nondestructive since the stylus exerts a very low pressure, e.g., 50 mg distributed over an area 200 μin. in diameter. The vertical resolution is typically 254 Å, with an accommodation of a vertical deflection of 16 mils for a high-precision profilometer.

Profilometry is an excellent technique for studying the local scale and roughness of surfaces and provides a valuable quantitative complement to microscopy.

13.4.5 X-Ray Diffraction. X-ray diffraction[2] (Sec. 13.3.2) is often used as the starting point in determining the crystallographic orientation of electrode surfaces. This technique yields only qualitative information about the surface, however. X rays penetrate several hundred angstroms below the surface, and therefore the diffraction patterns represent an integrated average over many atomic layers. Since the surface layer is only a small fraction of the diffracting field, it follows that the orientation of the surface atomic layer is not obtainable by means of X-ray diffraction.

13.4.6 Low-Energy Electron Diffraction. Low-energy electron diffraction (LEED) is widely used because it is one of the few techniques that yield information about the structure of the first several atomic layers in the vicinity of the surface. The principles of operation of LEED are as follows: A sample surface is bombarded with a beam of electrons 1 mm in diameter and 50 to 200 ev in energy. Because of its low energy, the beam is diffracted by the first few atomic layers near the surface of the sample. The back-reflected electron pattern is received either in a movable Faraday cage or on a fluorescent screen. The pattern is then related to the structure of the surface by the usual relations governing wave interference patterns generated by ordered structures.[8]

One of the major limitations of LEED is the extremely low pressure (about 10^{-11} torr) required in the test chamber. Higher pressures cause the surface to become contaminated during examination. Other limitations are: (1) The beam diameter is large relative to small sample grain sizes. (2) Amorphous deposits or surface structures cannot be detected. (3) The theory and experimental

interpretation of patterns are in the formative stage. (4) The sensitivity of the method to fractional surface coverage by additives or contaminants is not yet well established. As presently employed, LEED does not yield practical information for electrodes used in converters. The preparation and examination of small samples by means of LEED, however, have provided considerable insight to the care required in preparing atomically clean surfaces and surfaces oriented on an atomic scale.

13.4.7 Etch-Pit Analysis. Etch-pit analysis is an outgrowth of various methods of surface preparation. Electrode surfaces can be etched to expose particular crystallographic orientations by chemical action of an oxidizing agent in an aqueous solution, molten salt, or gas. For example, if a sample electrode is immersed in an electrolyte and a low voltage is applied, the current will cause a selective anodic dissolution. Also, heating the sample to a high temperature can induce thermal etching in vacuum or in an inert atmosphere. Finally, cathodic bombardment of the sample surface by a low-pressure gas discharge can cause the selective removal of atoms.

The theory of the effects of each etching procedure on a surface of given crystallographic orientation is not well understood. Only qualitative results are available. The literature contains extensive recipes describing how to expose various index planes of many of the electrode materials of interest to thermionic conversion and what to expect as a result of the exposures.

The empirical results that have been established can be used as a diagnostic tool in studies of surface orientation, and they are referred to as *etch-pit analysis.* Several types of information can be obtained. For example, the shape of the etch pit formed in response to a selective etchant can be easily related to surface orientation by simple geometry.[9,10] The misorientation between adjacent well-aligned grains can be determined if we know the value of lattice spacing and the number of etch pits in the grain boundary.[11] Skilled metallographers can use information on the number, size, and form of etch pits in a grain to establish the number and the nature of surface and bulk imperfections.

Inherently, selective etching used as a diagnostic method is a destructive test since etching changes the orientation of the surface. If the duration of the etch is short, however, the amount of change can be kept small.

13.4.8 Light-Reflection Goniometry. Light-reflection goniometry, which consists in observing the angle of reflection of light impinging on different facets of a surface, can be used to establish the orientation of the facets. It is an outgrowth of a light-reflection method introduced by Langmuir. Langmuir rotated a filament stuck in a cork in front of a lamp. As he observed a reflection, he would stick a pin in the cork. A few pins and a few twists yielded the

perpendicular directions of the flats. Langmuir's method has been updated recently by the use of a low-power helium laser as a light source and photographic plates for recording the reflected light.[12] The sample, laser, and photographic plates are arranged in a fashion almost identical to a Laue X ray. The pattern of the back-scattered light can be readily related to the orientation of the surface facets. In addition, the intensity of the back-scattered light can be qualitatively related to the distribution of the various facets. The precision is equivalent to Langmuir's pin-and-cork approach.

13.5 Analytical Relations for Electron-Emission Properties

13.5.1 General Remarks. Electrode properties of primary interest to thermionic conversion are those which determine the ability of the electrode surface to emit (or collect) particles. The most important of these is the electron work function or simply the work function.

A number of techniques for measuring work functions are discussed in Sec. 13.6. Most of the techniques are based on current-vs.-voltage measurements under either space-charge-limited or space-charge-neutralized conditions or in the presence of accelerating fields. Work function values are then deduced from these measurements by means of the analytical relations in Chaps. 2 to 4, Vol. I, and Chaps. 9 and 10. For convenience some of the relations between current-vs.-voltage measurements and work-function values are repeated in Secs. 13.5.2 to 13.5.5, together with some practical modifications.

13.5.2 Analysis of Electron Space-Charge-Limited Measurements. The bare work function of an electrode can be determined in a device that operates like a vacuum diode, with proper instrumentation for measuring temperature, current, and voltage. A schematic of such a device with planar electrodes is shown in Fig. 13.4. The *guard ring* around the collector is used to maintain a uniform electrostatic field between emitter and collector so that emission and collection surface areas are equal. Guard rings are very important in thermionic experimentation. For this reason a separate section of this chapter (Sec. 13.9.2) is devoted to their design and use.

Information about electrode properties can be extracted from the output-current characteristics of the vacuum diode by means of graphic techniques. Most of these techniques were developed by Nottingham.[13]

The graph of the natural logarithm of current vs. voltage in the retarding range is a straight line. The slope of the line is related to the emitter temperature T by the relation

$$T = \frac{11{,}606\,(V_2 - V_1)}{\ln\,(J_1/J_2)}\ ^\circ\text{K} \qquad \text{for } V \text{ in volts} \tag{13.1}$$

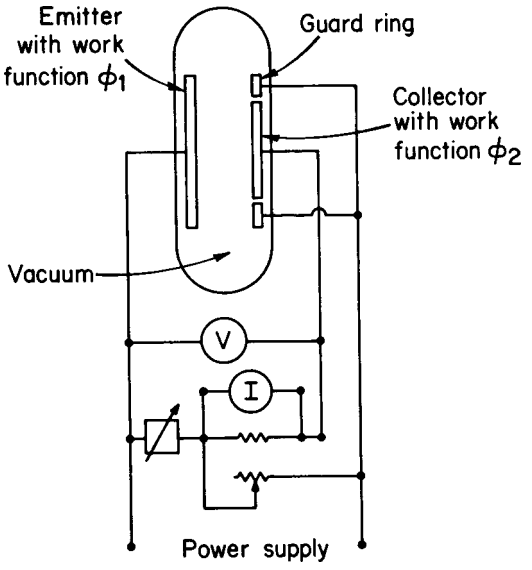


Fig. 13.4 Schematic of experimental setup for bare-work-function measurements.

where V_j and J_j are the voltage and current at point j on the line (the Boltzmann line of the characteristic). Thus the emitter temperature T can be determined from the characteristic and compared with the directly measured value.

In the graph of ln J vs. eV/kT and in the retarding range, data taken at any temperature should fall on the same straight line, as shown in Fig. 13.5. Moreover, the onset of the space-charge-limited range, the critical point, for each temperature can be established as the point of departure of each current characteristic from the straight-line part of the graph. Thus the values of the critical current J_R and the critical voltage V_R can be determined. These two values define the work function of the collector by means of the relation given later in this section.

The critical current J_R can also be used in a graph of ln (J/J_R) vs. eV/kT, like that shown in Fig. 13.6. Nottingham observed that at the saturation point the value ln (J_S/J_R) vs. eV_S/kT is described by a universal relation independent of the interelectrode spacing d of the diode. This observation follows directly from Eqs. 10.47, 10.49, and 10.50. A graph of the universal relation (master curve), ln (J_S/J_R) vs. eV_S/kT, is superimposed on Fig. 13.6 so that the linear parts of the experimental data, ln (J/J_R), and the master curve, ln (J_S/J_R), coincide (i.e., so that the points of departure of the two curves from linearity coincide). When this is done, we can easily verify that the intersection of the curvilinear portions of the experimental data and the master curve determines the saturation point,

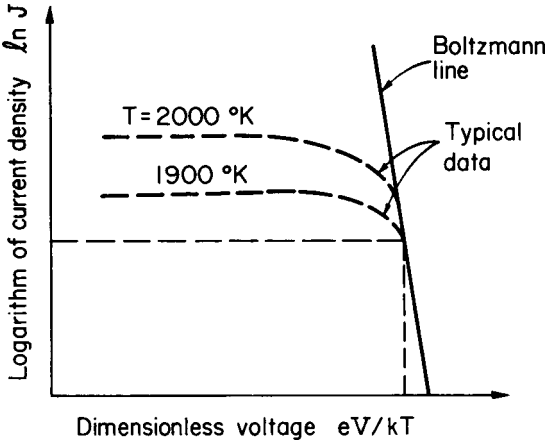


Fig. 13.5 Typical graph of ln J vs. eV/kT.

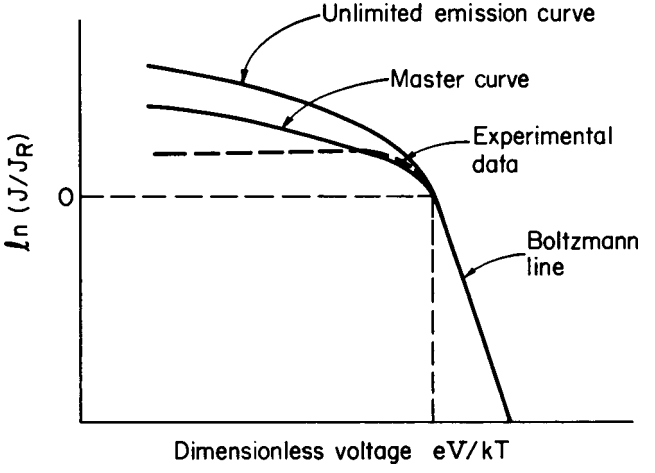


Fig. 13.6 Typical graph of ln (J/J_R) vs. eV/kT.

namely, J_S and V_S. The saturation-current density then determines the emitter work function. Additional details of this form of analysis appear in Ref. 14.

Often experimental data do not follow the analytical predictions. A variety of causes can contribute to the discrepancy; e.g., the electrodes can be patchy, or the instrumentation can be inaccurate. Patch effects are discussed in Chap. 9. Errors in instrumentation can be partially detected by comparing the emitter

temperature deduced from the slope of the Boltzmann line with a measured value and by comparing the measured interelectrode spacing with the value determined from J_R by means of Eq. 10.47.

Nottingham used the unlimited-emission curve shown in Fig. 13.6 to determine the accuracy of experimental data. This curve, derived for an emitter that is an unlimited source of electrons, corresponds to the upper envelope of all electron space-charge-limited characteristics. Experimental data rising above this curve, after experimental and analytical informations are superimposed, are questionable since they do not meet the conditions of analysis. Space-charge-limited characteristics should tend asymptotically to the unlimited-emission curve at high output voltages and should change monotonically into the saturation conditions at low output voltages.

13.5.3 Analysis of Electron-Emission Measurements in the Accelerating Range. The electron saturation current and, therefore, the emitter work function can also be determined from an analysis of the accelerating mode. For an applied voltage V_B moderately greater than the saturation-point voltage V_S, the current J is influenced by the Schottky effect (Sec. 9.2.6) and is given by the relation

$$\ln J = \ln J_S + \frac{11{,}606\,(3.8 \times 10^{-4}\,E^{1/2})}{T} \qquad \text{for } E \text{ in volts/cm and } T \text{ in } ^\circ\text{K} \qquad (13.2)$$

where E is the strength of the applied electrostatic field at the emitter surface.

As a first approximation, E is given by the relation

$$E = \frac{|V_B| + V_0}{d} \qquad (13.3)$$

where V_0 is the contact potential between emitter and collector and d is the interelectrode spacing. The quantity $|V_B| + V_0$ is equal to the difference between the electrostatic potentials just outside the emitter and just outside the collector.

Using Eq. 13.3 in Eq. 13.2 indicates that a graph of the logarithm of the current density J vs. $11{,}606\ (|V_B| + V_0)^{1/2}/T$ should be a straight line, the Schottky plot. The slope of this straight line is related to the interelectrode spacing d by means of the relations

$$d = \frac{2.72 \times 10^{-8}}{s^2} \text{ cm} \qquad (13.4)$$

and

$$s = \frac{T}{11{,}606} \frac{\log (J_1/J_2)}{(|V_{B1}| + V_0)^{1/2} - (|V_{B2}| + V_0)^{1/2}} \qquad (13.5)$$

If the measured spacing and the spacing calculated by Eqs. 13.4 and 13.5 are equal, then the Schottky plot can be extrapolated to the zero-voltage line to yield the saturation-current density J_S.

The use of Eqs. 13.2 to 13.5 implies an iterative procedure since, in general, the value of V_0 is not known. On the other hand, in the accelerating range the value of the applied voltage $|V_B|$ is much greater than that of the contact potential, and therefore V_0 can be neglected.

Experimental errors associated with measurements in the accelerating range are similar to those mentioned in connection with measurements in the space-charge-limited range. In addition, the high values of the applied voltage required in the accelerating range introduce the problem of collector heating. For example, if the accelerating voltage is of the order of several hundred volts and the current density is significant, the energy flux at the collector is of the order of kilowatts per square centimeter. This problem can be avoided by a reduction of the interelectrode spacing and by use of low-duty-cycle fast-voltage sweeps that take advantage of the thermal inertia of the electrodes to keep both emitter and collector at relatively constant temperatures.

13.5.4 Determining Work Functions. Measuring the saturation current density† J_S by means of either of the methods discussed in Secs. 13.5.2 and 13.5.3 yields the work function ϕ_E of an emitter surface through the relation (Eq. 9.12)

$$\phi_E = \frac{T}{11{,}606} \ln \frac{120(1 - \bar{r})T^2}{J_S} \text{ ev} \qquad \text{for } J_S \text{ in amp/cm}^2 \text{ and } T \text{ in } ^\circ\text{K} \tag{13.6}$$

provided that the average reflection coefficient $\bar{r}$ at temperature T is known.

If the work function ϕ_E is a linear function of temperature with a temperature coefficient α, namely,

$$\phi_E = \phi^* + \alpha T \qquad \text{for all } T > 0 \tag{13.7}$$

and $\bar{r}$ is independent of temperature, a graph of experimental results in the form of $\log (J_S/T^2)$ vs. $1/T$ yields a straight line. This line is called the Richardson plot. The slope of the Richardson plot is proportional to ϕ^*, the emitter work function in the limit of 0°K, and the intercept of the graph with the zero inverse-temperature axis is equal to the logarithm of $120\,(1 - \bar{r}) \exp(-\alpha/k)$. Because of the significant extrapolation to $1/T$ equal to zero, a large number of data points is required in the emitter temperature range available for

†The saturation current density is obtained from the measured current and the area of the emitter surface.

experimentation. Unfortunately, in practice the temperature range is limited. A lower temperature limit exists because the signal-to-noise ratio decreases exponentially with decreasing temperature. An upper temperature limit exists because of considerations of structural integrity of materials, excessive currents, and possible changes of crystallographic orientations of the emitter surface.

Because of the many ambiguities about the values of the average reflection coefficient, the temperature coefficient, and the nonuniformity of surfaces, in many applications a work function ϕ_E is defined by means of the relation

$$\phi_E = \frac{T}{11{,}606} \ln \frac{120T^2}{J_S} \quad \text{ev} \qquad \text{for } J_S \text{ in amp/cm}^2 \text{ and } T \text{ in } ^\circ\text{K} \tag{13.8}$$

where J_S is the experimentally obtained saturation current density. A work function computed from experimental data by means of Eq. 13.8 is often called an *effective work function.*

Measuring the critical electron current density J_R and the output voltage V_R yields the collector work function ϕ_C by means of the relation

$$\phi_C = V_R + \frac{T}{11{,}606} \ln \frac{120(1-\bar{r})T^2}{J_R} \quad \text{ev} \tag{13.9}$$

where T and $\bar{r}$ are the temperature and average reflection coefficient of the emitter, respectively.

13.5.5 Determining Work Functions of Cesiated Emitter Surfaces. When cesium is present, the output-current characteristics of a diode-type electrode arrangement differ substantially from those of the vacuum diode. Nevertheless, under certain conditions of operation (Chap. 4, Vol. I), an output-current characteristic exhibits an electron saturation mode. Measured values of the electron-saturation-current density can be used in Eq. 13.8 to determine the work function of the cesiated emitter surface.

Under other conditions of operation, an output-current characteristic exhibits an ion saturation mode in the deep retarding range. Measured values of the ion-saturation-current density $\tilde{J}_S$ and the cesium pressure p_{Cs} can be used in the Saha–Langmuir equation (Eq. 9.150) to yield the work function of the cesiated emitter surface. Specifically, ϕ is given by the relation

$$\phi = V_i + \frac{T}{11{,}606} \ln \frac{2\tilde{J}_S(2\pi m_{Cs} kT)^{1/2}}{ep_{Cs}} \quad \text{ev} \tag{13.10}$$

where V_i is the first ionization potential of cesium.

13.6 Measurements of Electron Emission and Collection Properties

Table 13.4 gives a partial list of methods that can be used for electron-emission, collection, and work-function measurements relevant to thermionic conversion.

Table 13.4 Methods Used in Electron-Emission, Collection, and Work-Function Measurements

Method	Section
Diode arrangement	13.6.1
Triode arrangement	13.6.2
Emission microscopy	13.6.3
Plasma anode device	13.6.4
Electron beam collector	13.6.5
Field-emission retarding potential device	13.6.6
Kelvin contact potential detector	13.6.7

Each of these methods is discussed briefly in the section indicated in the table. Other methods are available but have not been used in the study of electrodes suitable for converters.

13.6.1 Diode Arrangement. The diode arrangement consists in the use of two electrodes in the form of a planar diode to measure either average emission properties of an emitter or average collection properties of a collector. It can be used both in vacuum and in the presence of cesium.

When the two electrodes are in vacuum (Fig. 13.4), the discussion in Secs. 13.5.2 to 13.5.4 is directly applicable, and therefore only a few additional comments are in order here.

The high bare work functions of interest to thermionic conversion require high emitter temperatures so that easily observable currents can be achieved and an accurate Richardson plot can be made. This requirement has resulted in the development of the energetic-electron-bombardment heating technique for the high-temperature emitter. The wide use of electron-bombardment heating and of highly refractory metals is the main difference in vacuum diode designs encountered in thermionic conversion as compared with those prevailing in electronics.

When the vacuum diode is used for collector measurements, the work function ϕ_C is deduced from experimental results either at the critical point (Eq. 13.9) or on the values of the contact potential V_0 and the emitter work function ϕ_E or, in general, on the location of the Boltzmann line. Specifically, the contact potential V_0 is given by the intercept of the Boltzmann line and the saturation current line, and the emitter work function ϕ_E by the saturation current density. It follows that ϕ_C is given by the relation

$$\phi_C = \phi_E - V_0 \tag{13.11}$$

In general, the close proximity of the high-temperature emitter can result in contamination of the collector surface. The contamination is due to evaporation of materials from the emitter surface and condensation of these materials on the low-temperature collector. This contamination is not necessarily undesirable in collector measurements if it simulates the contamination actually present in a practical converter.

The two electrodes of the diode arrangement can also be used for measurements in the presence of cesium in the interelectrode space. The results of such measurements form the basis for evaluating performance characteristics of practical converters having electrodes of approximately the same composition and structure but operating at cesium pressures higher than those used to measure work functions.

In addition to the difficulty resulting from its being chemically very active, cesium introduces other difficulties: (1) It ionizes readily, and the resulting large ion currents can interfere with electron emission and collection measurements. (2) It can partially cover surfaces of structural components and lower their work functions so that they become active and spurious sources of electron currents. (3) It can coat refractory insulators and thus increase leakage currents. (4) Cesium positive ions and neutral atoms can scatter emitted electrons and thus impede current transmission from emitter to collector through the interelectrode space. (5) Cesium ion densities can be affected by applied or space-charge fields and thus alter arrival rates at the electrodes under steady-state conditions in ways that are difficult to predict theoretically. Despite all these difficulties, the cesium diode arrangement is in wide use as a tool for rating and selecting high-performance emitters and collectors for practical converters.

A typical diode used for cesiated work-function measurements is shown in Fig. 13.7. Guard rings are used to suppress stray currents, to establish the emission and collection areas, and to reduce leakage currents. Variable spacing is often employed as a means of either evaluating or reducing electron scattering in the interelectrode space.

Emitter work functions are determined in two ways, depending on the conditions of operation of the cesium diode.

When the emitter ion-richness ratio is much greater than unity and the interelectrode spacing is so small that collisions in the interelectrode space are negligible, the output-current characteristics exhibit an electron saturation mode. Thus the emitter work function is determined from the saturation current density in the same way as in the case of vacuum diode measurements.

When the emitter ion-richness ratio is smaller than unity and the interelectrode spacing is so small that collisions in the interelectrode space are negligible, the diode is operated in the deep decelerating range where the output-current characteristic exhibits an ion saturation mode. Thus the emitter work function is

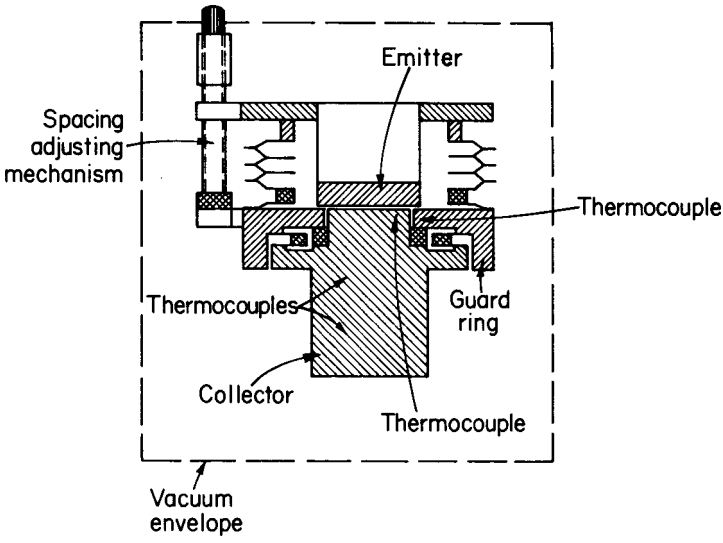


Fig. 13.7 Typical cesium-gas-filled diode for emitter and collector studies.

determined from the measured ion-saturation-current density by means of the Saha–Langmuir equation.

Finally, under conditions of operation which result in output-current characteristics that do not exhibit easily identifiable saturation modes, emitter-work-function estimates can be obtained by other means (see Ref. 15).

13.6.2 Triode Arrangement. The triode arrangement consists in the use of a collimating magnetic field and an auxiliary electrode (accelerating for emitter electrons) between the two electrodes of a diode arrangement to overcome space-charge effects and to establish a local current density of the surface of the emitter. Thus work-function values and their variations along a nonuniform surface can be studied.

A schematic of the triode used in the absence of cesium (vacuum triode) is shown in Fig. 13.8. The auxiliary accelerating electrode has a small aperture of known area. The applied magnetic field is approximately normal to the surface of the emitter and forces electrons originating from the emitter to spiral toward the collector via the aperture. If the Larmor radius of gyration of the electron is appreciably less than the radius of the aperture (i.e., if the magnetic field is sufficiently strong), electrons reaching the collector originate from an emitter area equal to the small area of the aperture. Since the aperture can be located at different positions along the surface of the emitter, emission properties of

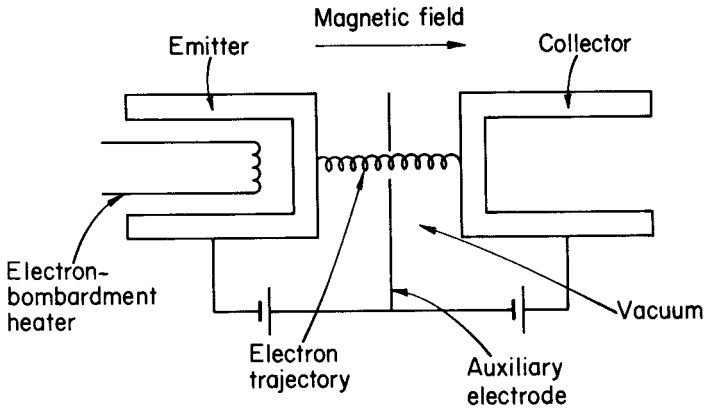


Fig. 13.8 Schematic of triode used in work-function studies in vacuum.

different electrode areas can be investigated. The position of the aperture can be controlled either electrically or mechanically.

In emitter studies the collector is biased so that all current going through the aperture reaches it, and the accelerating electrode is biased so that the current–voltage characteristic between emitter and collector is in the accelerating mode. The zero-field steady-state condition is established by means of a Schottky plot. The work function of the sampled area of the emitter is determined by means of the relations given in Sec. 13.5.

A schematic of the triode used for emission measurements in the presence of cesium[16] is shown in Fig. 13.9. In this device the accelerating electrode is stationary, and the direction of the applied magnetic field with respect to the normal to the emitter surface can be altered. Thus *scanning* of various small areas of the emitter can be achieved. The accelerating electrode is biased until saturation current is established. For this to be possible, the restrictions discussed in Sec. 13.6.1 must be satisfied. In addition, the bias voltage should be below the breakdown voltage of the cesium vapor.

The triode arrangement is not suitable for collector studies by means of the retarding field method, because the accelerating electrode can alter appreciably the distribution of velocities of the electrons.

13.6.3 Emission Microscopy. Emission microscopy consists in the projection of electrons emitted from a small area onto a fluorescent screen by means of appropriate electrostatic lenses.[17] It is particularly suited to the study of emission properties of nonuniform surfaces in vacuum, but it cannot be used for cesiated surfaces.

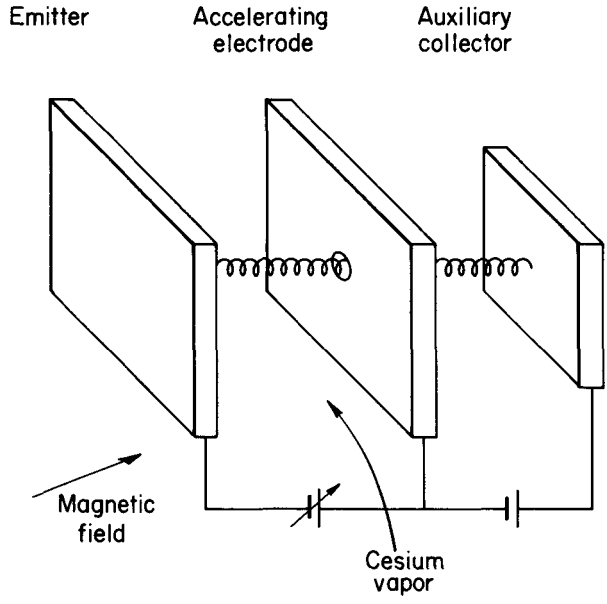


Fig. 13.9 Schematic of scanning triode used in cesiated-work-function studies.

The projection can easily be made 10^3 times as large as the emission area. Lower magnifications are usually employed, however, because the characteristic sizes of nonuniformities of surfaces are relatively large. The current density can be measured by a densitometer trace of the image on the fluorescent screen, a movable Faraday cup ahead of the screen, or the use of a sampling aperture in the screen. The aperture acts as a Faraday cage, and sampling is accomplished by electrostatic, electromagnetic, or mechanical motion of the image with respect to the sampling hole. The observed current is related to a local work function by means of the expressions given previously.

13.6.4 Plasma Anode (Collector) Device. The plasma anode device consists of one or more emitting surfaces immersed in the plasma of a gas discharge. The plasma acts as a collector for the emitters and as a source of positive ions for the neutralization of electron space charge. Thus the saturation current of each emitter is measured. This current and the knowledge of the area of the emitter yield the emitter work function.

Work functions of cesiated surfaces can be determined in a cesium-gas discharge. The results are expressed as functions of the emitter temperature and the cesium-reservoir temperature.

The plasma anode method with cesium has been employed extensively by Houston and Webster[18] to screen various metals for use as electrodes in converters. In their experiments the emitters were in the form of filaments and, therefore, were not representative of electrode surfaces of practical converters. The method is not restricted to filaments only, however.

13.6.5 Electron Beam Collector. The electron beam collector consists of a beam of electrons used to measure the motive barrier just outside a collector and the work function of a collector. The results are independent of the distribution of velocities of the electrons and can be obtained with large interelectrode spacings.

The beam of electrons is directed either on a reference surface with a known work function ϕ_{ref} or on a surface with an unknown work function ϕ. The beam can be deflected from target to target mechanically,[19] electrostatically, or magnetically. Each target is biased with respect to the source of electrons.

When the beam impinges on the reference surface, the bias voltage is adjusted so that collection takes place in the retarding mode. In this mode a particular operating point is selected corresponding to a current J_{ref}, bias voltage V_{ref}, and barrier ψ_{ref} for the incoming electrons. The barrier ψ_{ref} satisfies the relation

$$\psi_{ref} = \phi_{ref} + eV_{ref} \tag{13.12}$$

When the beam impinges on the surface under test, the bias voltage V_t is adjusted so that the collected current is equal to J_{ref}. Thus the barrier is again ψ_{ref}, and therefore the unknown work function ϕ is given by the relation

$$\phi + eV_t = \phi_{ref} + eV_{ref} \tag{13.13a}$$

or

$$\phi = \phi_{ref} + e(V_{ref} - V_t) \tag{13.13b}$$

An elegant extension to the electron beam method, a *retarding beam scanner,* has been developed by Haas and co-workers.[20]

13.6.6 Field-Emission Retarding-Potential Device.[21] The field-emission retarding-potential device uses electrons originating from the Fermi level of a cold emitter to detect the work function of a collector. It is based on the observations that such electrons are practically monochromatic and that the bias voltage between emitter and collector at which the collected current goes abruptly to zero is equal to the collector work function.

In practice the source of electrons is a cold field-emission tip in a suitably designed system of electrostatic lenses. Thus a well-collimated and nearly monochromatic beam of electrons is achieved.

Sometimes it is convenient to allow the field-emission tip to assume ambient temperature. Then the extracted electrons are still close to monochromatic, and only a minor correction of the zero-current bias voltage is necessary. Since the beam size is usually 1 mm in diameter, patches smaller than 1 mm can distort the results.

13.6.7 Kelvin Contact-Potential Detector. The Kelvin contact-potential detector uses two electrodes as plates of a condenser to establish the average contact potential between the electrodes by means of a null measurement. The method can be used only at very low temperatures so that no particles are emitted from the electrode surfaces.

One plate of the condenser is vibrated by means of a solenoidal drive, such as a loudspeaker coil, and the bias voltage between the two plates is adjusted so that no displacement current flows in the external resistive circuit.[22] When this is achieved, the value of the bias voltage is equal to the contact potential between the two electrode surfaces. Thus the work function of one electrode can be determined if the work function of the other is known.

13.7 Temperature Measurements by Means of Thermocouples

Temperature measurements in thermionic conversion are made by means of *thermocouples* and *optical pyrometry*. Thermocouples are discussed here, and optical pyrometry is discussed in Sec. 13.8. In general, temperature measurements are difficult.

13.7.1 Principle of Operation. The theory[23,24] and practice[25] of temperature measurement by means of thermocouples are well known and are reviewed here only briefly. Measurements are based on the *Seebeck effect*, i.e., the existence of an electromotive force (emf) in a circuit consisting of two different metallic conductors if the junctions of the conductors are at different temperatures. The emf results in a current flow unless it is balanced by a voltage interposed in the circuit so that the current is equal to zero. When this is done, the voltage is a function of the conductor materials and the junction temperatures only, and therefore it is a measure of the temperature of one junction if that of the other is known. The voltage is often called the *Seebeck emf*.

In a Seebeck emf measurement, the current through the thermocouple circuit must be kept as close to zero as possible. Nonzero current can affect the result and its interpretation by virtue of any one of the following effects: (1) ohmic voltage drops; (2) *Peltier effects* (heating or cooling of the junctions as a result of the current flowing through them); and (3) *Thomson effects* (heating or

cooling of the conductors by virtue of current flowing along a temperature gradient).

The following corollaries of the theory of thermoelectricity provide useful guidelines to the experimenter.

The rule of intermediate metals: If a circuit consisting of any number of dissimilar metals is at uniform temperature, the algebraic sum of the Seebeck emf's produced by thermoelectric effects is equal to zero.

The rule of intermediate temperatures: If Seebeck emf's V_1 and V_2 are generated in a circuit composed of two dissimilar homogeneous metals when the junctions are at temperatures T_1 and T_2 and T_2 and T_3, respectively, then, when the junctions are at T_1 and T_3, the Seebeck emf will be $V_1 + V_2$.

The rule of the homogeneous circuit: Temperature gradients alone cannot cause the flow of an electric current in a closed circuit consisting of a single homogeneous metal.

13.7.2 **Practical Considerations.** Despite their conceptual simplicity, in practice, thermocouple measurements are subject to many sources of error. The most common of these sources are: (1) heterogeneities of wires; (2) mislocated junctions; (3) stray signals; (4) temperature uncertainty at the cold junction; and (5) thermocouple instrumentation problems. In addition to these general sources of error, others exist that are peculiar to thermocouple applications in thermionic conversion. A detailed description of the various sources of error and their consequences is given in the following paragraphs.

(a) Heterogeneities of Wires. Metallic wires as received from a manufacturer are not always homogeneous. Composition can vary along the length of the wire and often varies significantly from spool to spool from the same manufacturer. Moreover, spools of allegedly identical material can vary greatly from one manufacturer to another. It is always advisable, therefore, to calibrate any given thermocouple against known temperatures.

Kinks or strains can introduce local heterogeneities and, if present in a temperature gradient, spurious emf's. Accordingly, wires that are swaged to form insulated jacketed thermocouples should always be calibrated after fabrication.

High vacuum and high temperature cause the selective removal of the more volatile metallic constituents and, therefore, result in changes in composition of the wire. These effects are particularly prevalent and difficult to avoid in thermionic conversion. For this reason recalibration of thermocouples from time to time is recommended.

High-temperature operation of junctions in a zone of severe temperature gradient, such as regions near the emitter, can accelerate interdiffusion of the elements comprising the junction, thereby causing heterogeneous composition. Consequently, thermocouples are rarely used to measure emitter temperatures

greater than 1800°K. Moreover, even when used at emitter temperatures smaller than 1800°K, they should be frequently calibrated.

False junctions can occur as a result of swaging small-diameter sheathed thermocouples. If the false junction is at a temperature different from that of the main junction, shunt losses and spurious emf's arise. False junctions also result from sharp bends of swaged wires. Bends at high temperatures are troublesome because they often cause intermittent false junctions. Conductive soldering fluxes used to join wires and their sheaths and insulators, which become conductive in high-temperature zones, also produce false junctions.

(b) Mislocated Junctions. The end-closure method used in sheathed thermocouples can cause the location of the thermocouple junction to be somewhat in doubt. Uncertainties of ¼ in. in the junction location are common.

Demountable connectors in the thermocouple termination in vacuum systems are sometimes used in regions of thermal gradients. The thermal resistance and temperature of the contacts can vary so that a mislocated and spurious junction is created.

Mislocated junctions can result from the use of solders of abnormally high or abnormally low Seebeck emf. This is especially important in converter testing where solders selected for weldability are used to seal the vacuum feedthroughs in the test apparatus. The vacuum feedthrough can be a zone of significant temperature gradient. Then, if the Seebeck emf of the soldered feedthrough is not matched in both legs of the vacuum penetration, appreciable errors result. For this reason the Seebeck emf of the solder must be equal to that of the thermocouple wire if a thermal gradient exists across the vacuum penetration. Rarely is this requirement met in thermionic experimentation. Illustrations of faulty feedthrough installations are shown in Fig. 13.10.

Errors can arise in a feedthrough design where the penetration and change of materials is in a zone presumed to be isothermal. In one experiment the vacuum-seal zone was short (Fig. 13.11), and thus nearly isothermal, but the wire contacted the feedthrough intermittently past the seal zone and errors of about 30°K resulted.

(c) Stray Signals. Bare thermocouple wires in ionized gases can act as plasma probes. In thermionic experimentation, effects of ionized gases on thermocouple readings have been noted in: (1) nuclear reactors, where gases near the thermocouple have been ionized by radiation (principally gamma radiation and secondary electrons); (2) systems evacuated by ion pumps; and (3) vacuum systems using electron-bombardment heating.

Electron bombardment used for heating can introduce spurious signals by direct electron impact, which affects the thermocouple indication either through an electron current or through electron heating near a junction. The presence of

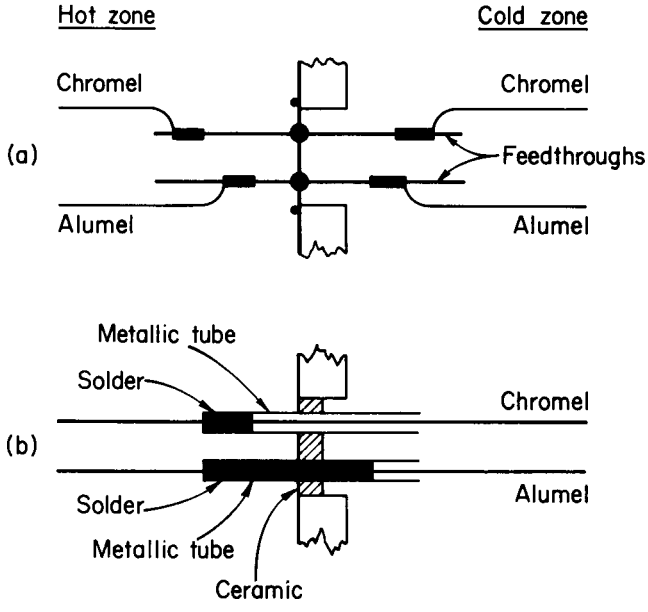


Fig. 13.10 Faulty thermocouple penetrations between hot and cold zone. (a) Unbalanced series thermoelectric elements. (b) Unbalanced emf produced by solder.

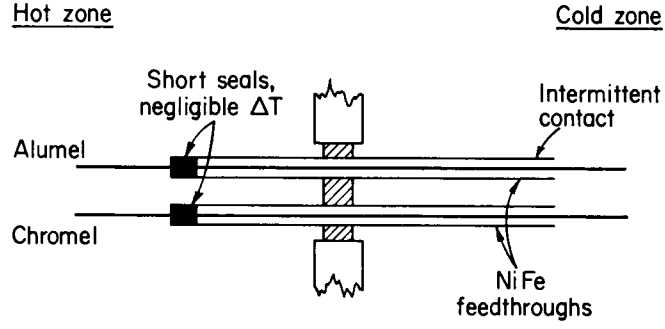


Fig. 13.11 Schematic of thermocouple error caused by intermittent contact.

a spurious emf due to electron bombardment can be investigated by observing the change in emf when electron bombardment is suddenly turned off.

Thermocouples that are grounded (i.e., electrically connected with the test element) can be subject to errors caused by currents flowing in extraneous ground loops. The "ground-loop" error, usually related to errors in instrumentation, is discussed in more detail later.

Finally, thermocouple junctions in contact with converter components carrying large currents can develop spurious signals related to voltage drops across the junction area.

(d) Cold-Junction Temperature Uncertainty. Poorly designed cold-junction environments, such as unstirred ice-water baths, can be nonuniform in temperature by several degrees. Thermostatically controlled commercially available reference junctions will provide reproducible cold-junction temperatures within ±0.10°K when carefully maintained. All too often the bulky reference junctions are tucked away at inaccessible parts of the instrument racks and are unmonitored. Thermostat failures throughout the life of the test facility are possible and, if undetected, can be expensive in rejected data and lost time. Hum and contact noise of thermostatically controlled cold junctions can also interfere with the accuracy of thermocouple readings unless the indicating instruments used in the thermocouple circuit are designed for existing conditions.

(e) Thermocouple Instrumentation Problems. Many serious errors in thermocouple measurements arise from poorly designed instrumentation, namely, instrumentation in which leakage currents are caused by the thermocouple zone's being at a potential different from ground potential.

Seebeck emf's of thermocouples in general use are in the range from 15 to 80 μvolts/°K. This signal can be considerably less than the potential differences between many parts of a thermionic apparatus nominally at ground potential. High currents and significant Seebeck emf's are the main contributors to voltage variations in a structure that, from superficial examination, would be presumed to be at ground potential.

Single-ended (grounded-input-terminal) amplifiers are susceptible to ground-loop leakage currents even with allegedly isolated thermocouples. The use of single-ended, or floating-thermocouple, recorders in conjunction with transient testing is conceptually acceptable if the readings are made after the transient test and the converter is allowed to return to a steady-state low-current condition. But unfortunately, since the time for the recorders to recover from the transient test can be long, errors can arise.

Differential input amplifiers with a high common-mode rejection eliminate many of the problems cited, but these amplifiers are not infallible in correcting signal errors. High common-mode rejection capabilities of amplifying circuits are usually related to low-frequency symmetric wave forms, and symmetry of noise is not often present in thermionic experiments. Since thermocouple wires usually cannot be twisted without causing strain-induced heterogeneity, unequal alternating-current pickups can exist in the separate legs of the thermocouple circuit.

13.7.3 Recommended Procedures. Some procedures for avoiding thermocouple problems are recommended in Sec. 13.7.2. Additional recommendations are offered in this section.

The high level of temperature required in the emitter can cause changes in thermocouple calibration with time and large temperature gradients along the thermocouple wires. Calibrations can be stabilized by preconditioning for 50 to 100 hr at temperatures in excess of the conditions of operation.

Since small-diameter wires are more sensitive to high temperatures than large wires, sheath diameters of less than 60 mils should be avoided. For example, a thermocouple made of an alloy of 3% rhenium in tungsten and an alloy of 25% rhenium in tungsten in a hard-fired high-purity BeO insulation with a tantalum sheath of high purity may be quite stable. The hot junction should be immersed to a reasonable depth, and compensating leads should be avoided.

Where thermocouple wires penetrate vacuum enclosures, they should be electrically shielded or enclosed whenever possible, changes of composition should be avoided, and solders used in construction of demountable connectors should have low thermal emf. Soldered joints should be free of flux, and welded joints should be spot-welded and uncontaminated with electrode material.

In regions of intense temperature gradients, intimate contact must be provided between the thermocouple and the part whose temperature is to be measured, and the thermocouple stem should be perpendicular to the direction of the

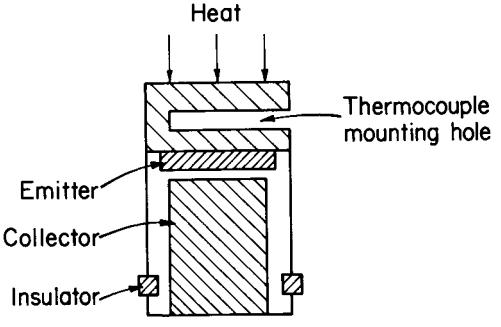


Fig. 13.12 Schematic of thermocouple mounting in a region of large temperature gradients.

temperature gradient (Fig. 13.12). If the temperature of the instrumented location is sensitive to heat leakage, fine thermocouple wires should be used.

All thermocouple junctions should be located where the voltage gradient is small. Wires should be sheathed and insulated with due regard for the reduced electrical resistance of oxide insulation at high temperatures.

In the interest of safety, if at all possible, sheathed or shielded thermocouples should be used in a test apparatus. High voltages used in electron bombardment heating can inadvertently reach an unshielded wire. High-voltage electron-bombardment filaments and wires sometimes sag from overheating and touch thermocouple leads and circuits designed for low voltage only. If cesium is suddenly released or if a leak occurs in the vacuum enclosure, plasma conduction can transfer high voltages to thermocouple leads and endanger both instruments and operators.

13.8 Optical Pyrometry

13.8.1 General Remarks. The operation of pyrometers is based on the balance of intensities of radiation from a target and a reference filament. When the two intensities are equal, optical images of the target and the filament are indistinguishable. If the emissivity of the target is well defined, then the temperature of the target can be related to that of the calibrated filament. The emissivity of the target is usually defined by monitoring a cavity on it which acts as a source of blackbody radiation.

Manually balanced disappearing-filament pyrometers are extensively used to determine emitter temperatures either directly or indirectly. Almost all high-temperature sensors used in thermionics are calibrated by the use of this simple low-cost pyrometer. The theory and reduction to practice of optical pyrometry is adequately treated in the literature.[26,27]

In contemporary pyrometer designs the human eye is replaced by a photosensitive detector. Although their precision, accuracy, and response are over an order of magnitude better than those of manually balanced pyrometers,[28] automatic photoelectric pyrometers have not been used in thermionic experimentation to any great extent.

Errors encountered in the use of manually operated pyrometers are seldom caused by the limitations of the pyrometer itself. They frequently result from vignetting, from limited optical resolution, from uncertainty about the target emittance (cavity design), and from poor calibration procedures. They are discussed in more detail in Secs. 13.8.2 to 13.8.5, with emphasis on experiments with converters.[29]

13.8.2 Vignetting. Vignetting is the nonuniform radiance of the target at the image plane (for the visual optical pyrometer, the image plane coincides with the plane of the filament). Vignetting results from obstruction of rays from the target and can be caused by improper window size, misaligned optics, wires, shutter elements, and defects or local deposits on the viewing window. The reduced target radiance shifts the filament brightness for contrast reversal away from its correct value. In manual optical pyrometry a skilled operator can detect

its existence because vignetting broadens the range of filament currents (temperatures) over which contrast reversal occurs.

The existence of vignetting caused by misaligned optics can be established by simple geometric considerations illustrated in Fig. 13.13. Although the entrance pupil diameter d depends on the brand of pyrometer used,* to have some margin for misalignment, we can assume that it is the diameter of the objective lens. The

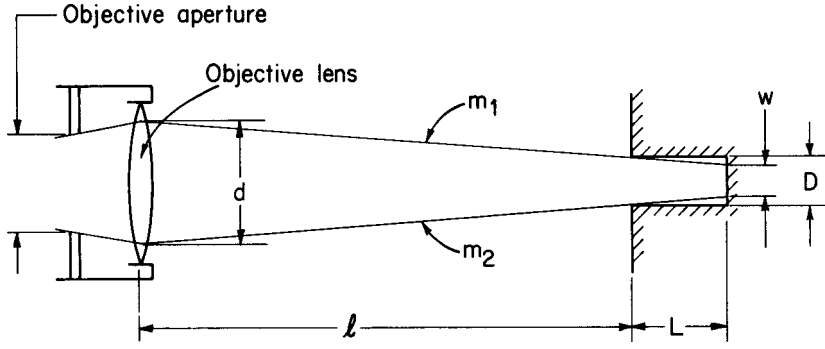


Fig. 13.13 Schematic diagram of pyrometer focused on back wall of cavity. Symbols are: diameter of cavity entrance, D; entrance pupil diameter, d; distance from objective lens to cavity entrance, l; depth of cavity, L; rays, m_1 and m_2; and target diameter, w.

diameter w is the minimum allowable target size. As explained in Sec. 13.8.3, w should be about 3 to 5 times the diameter of the filament as viewed in the image plane. Obstructions in the solid cone bounded by the rays m_1 and m_2 will cause vignetting. Since

$$\frac{D-w}{L}=\frac{d-D}{l} \tag{13.14}$$

the minimum blackbody-cavity diameter D free of vignetting is given by the relation

$$D \approx w+\frac{L}{l}d \tag{13.15}$$

where d = entrance pupil diameter
L = cavity length
l = distance from the objective lens to the cavity entrance

*In a commercial pyrometer the pupil diameter is 1.4 cm, and the objective lens diameter is 3 cm.

Unless the optical alignment of the pyrometer is precise, with protection from vibration and provision against relative motion of the target caused by thermal expansion, the diameter of the cavity and the positions of obstructions must be kept well above the limit indicated.

13.8.3 Optical Resolution. Because converters usually contain intense temperature gradients in parts of small dimensions, the cavity used as a target for temperature measurement must, in turn, be small. Errors in measurement result when the size of the cavity aperture approaches the limit of optical resolution of the pyrometer and its operator. The minimum error-free target size that can be used for good-quality disappearing-filament optical pyrometers is discussed in Ref. 28. There, it is observed that the magnitude of the error changes rapidly with changes in the temperature of the field surrounding a small target. For example, if the field around a target has an apparent temperature 100°K different from that of a target one-half the diameter of the minimum error-free target, a 1% error can exist. The error increases rapidly with increased difference of temperature between the surrounding field and the target. For L/D equal to zero, the minimum error-free target diameter can be approximated by

$$D \approx \frac{l}{d} 10^{-3} \text{ cm} \tag{13.16}$$

For L/D different from zero, the minimum error-free target diameter vs. l is larger than that given by Eq. 13.16.

Since experimental installations preclude error-free optical pyrometer measurements, compromises must be made. Vignetting errors can usually be reduced by increasing the distance between the target and the pyrometer, along with the target size. If the apparent temperature adjacent to the target differs significantly from that of the target, no compromise on target size can be tolerated. An installation that resulted in an error of several hundred degrees Kelvin is shown in Fig. 13.14.

13.8.4 Blackbody Cavity. Target cavities are used to establish the temperature of high-temperature parts of thermionic converters in optical pyrometry. In principle, polished metallic surfaces could be used if the observed brightness temperature is corrected for the surface emissivity so as to yield the true temperature. The difficulties of meeting requirements of surface finish and of eliminating spurious radiation in the field of view make a polished surface a poor choice, however. Using cavities that approach blackbody radiating characteristics to varying degrees reduces this difficulty. Unfortunately, the size of the thermionic converter and the intense temperature gradients in the converter parts make it difficult to find a good location for a cavity.

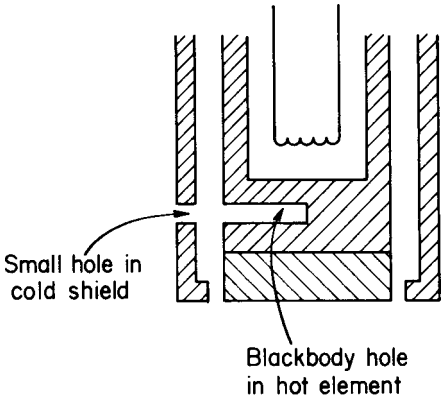


Fig. 13.14 Schematic of installation with two temperatures that differ substantially and with poorly selected target size, which results in erroneous pyrometric temperature measurements.

A good cavity should have: (1) a target diameter, w, significantly larger than that specified in the preceding section; (2) an emittance as near 1.0 as practical because it is difficult to establish predictable cavity emittances at values substantially less than 1.0 (at a source temperature of 2000°K and wavelength of 6500 Å, the wavelength of most optical pyrometers, an indicated temperature of 1991°K will result if the cavity emissivity is 0.95); (3) uniform radiance over the target area; (4) isothermal walls; and (5) emittance independent of changes caused by alterations in the radiant properties of the cavity walls. These five qualifications are not easy to obtain in a converter environment and, along with the other constraints of converter design, tend to be mutually exclusive.

In principle, the emittance of a cavity can be determined analytically if the properties of the surfaces bounding the cavity are known. The greatest causes of uncertainty in the results of such an analysis are the radiation properties of the walls. Are they diffuse radiators? Are they reflective? What is the surface roughness inside a 20- to 30-mil diameter cavity? For want of better knowledge, the walls are often assumed to be diffuse radiators[30,31] since this assumption makes the analysis easier and corresponds to one limiting case. This analysis indicates that long cavities should be preferred over short cavities, especially when the cavity wall is a material that is subject to thermal polishing. When thermal polishing occurs, the cavity shifts from diffuse radiation to reflective radiation. Then the emissivity of a cavity with an L/D of 5 changes enough to alter the indicated temperature by more than 100°K. The temperatures and time durations encountered in thermionic experimentation are sufficient to cause such a shift in tantalum cavities.

An additional test for cavity emissivity and possible change in the radiation characteristics of the wall can be provided by means of a simple fabrication procedure. Since most of the blackbody cavities are now machined by electrical-discharge machining (EDM), a small annular surface adjacent to the cavity opening is also machined at the same spark setting to establish a surface equivalent to the cavity surface. Next to the annulus a small target area is polished. If the relative brightness temperatures of the several surfaces do not change with time (with due regard to equivalence of spurious radiation), then the probability that surface rearrangement of the cavity walls has occurred is slight.

Of course, there are always interesting cases that do not follow the expected patterns. For example, a molybdenum cavity was prepared by the approved EDM method. Although the target area indicated the uniformity expected of a diffuse radiator, a change occurred in the cavity emittance in relation to the emittance of a polished surface adjacent to the cavity. This suggested a change in spectral radiation. Subsequent microscopic examination of the cavity showed that a large number of small filaments or whiskers had grown in the cavity. The whiskers near the opening of the cavity radiated to the external surroundings and lowered the apparent cavity temperature significantly. The whiskers yielded diffuse but low-temperature radiation. Because their diameters were only about 10^5 Å (0.001 cm), they were very difficult to detect. After a long thermal soak, the whiskers were clipped and did not return, and the tests proceeded smoothly. Local stress or impurities or a particle trapped in the cavity can induce whisker formation.

Since large values of L/D for the cavity usually require compromises in the optical arrangement, tungsten is often employed as the wall material for the cavity because metals as reflective as tantalum tend to alter cavity emittance through thermal polishing. Tungsten is quite stable in a high vacuum and at the usual emitter temperatures. A tungsten cavity can be made in any emitter material by drilling a hole and inserting a liner made from roughened tungsten foil. Alternatively, a cavity with a rough surface can be formed by EDM in a small tungsten plug inserted with good thermal contact in the drilled hole. Experimentally determined emittances for tungsten cavities formed by EDM are shown in Fig. 13.15. An L/D of 5 or greater yields cavity emissivities of 0.95 or more.

One of the most difficult applications of optical pyrometry is the measurement of emitter temperature of a cylindrical converter. In the past such measurements were made with arrangements that were entirely unreliable. A reliable arrangement that can be used in experimental cylindrical converters was employed by Houston and Webster.[18] They viewed the emitter through holes in the collector, as shown in Fig. 13.16. Because of the limited emitter thickness,

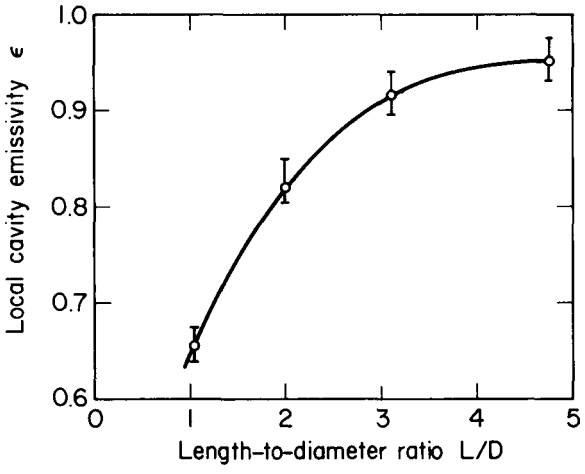


Fig. 13.15 Experimentally determined emissivity for tungsten. Cylindrical cavities were drilled by electrical discharge method. Temperature range is from 1500 to 2200°K.

the diameter of the cavity must be small to attain a suitable value of L/D. In this arrangement the cavity is not isothermal, and, moreover, radiation from the cavity can be partially absorbed by cesium, especially if the cesium pressure is high and the path through the cesium is long. Houston and Webster recognized the cesium absorption problem and used the work of Lapp and Harris[32] to derive pyrometer corrections as functions of cesium pressure and path length. When these and other corrections are made, useful results are obtained. This arrangement cannot be used in practical converters, however.

13.8.5 Calibration of Optical Pyrometers. Calibration procedures in optical pyrometry are well established.[26,27] Thermionic conversion laboratories often use a pyrometer calibrated at the National Bureau of Standards (NBS) or an NBS-calibrated tungsten strip lamp as secondary standards.

Photoelectric pyrometers now appearing in test laboratories are a valuable and convenient addition to the calibration procedures. If these are calibrated at NBS and used in conjunction with a strip lamp, as shown in Fig. 13.17, the uncertainties associated with the strip-lamp temperature can be reduced by an order of magnitude. The uncertainties of the conventional-filament optical pyrometers still remain, unfortunately. Reviewing some of the apparatus and procedures used in the calibration facility, we note that the primary factors are: (1) a calibrated tungsten strip lamp; (2) a stable filament power supply; (3) a precision shunt and precision voltmeter; (4) a lamp enclosure; (5) a calibrated photoelectric pyrometer (traceable to NBS standards); (6) an optical bench; and

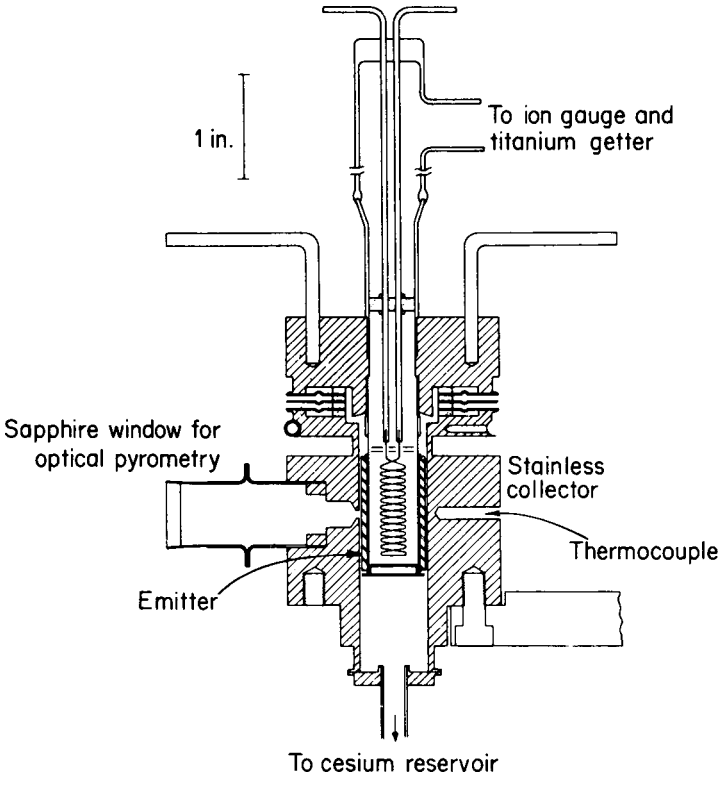


Fig. 13.16 Arrangement for pyrometric emitter-temperature measurements in a cylindrical converter.

(7) the optical pyrometer to be calibrated. Two standards, the lamp and the pyrometer, are used merely to cross-check on the malfunctions of either.

One reference lamp in frequent use is a gas-filled, notched, flat, tungsten-filament lamp model 30 A/6V/T24 manufactured by the General Electric Corporation. The lamp can be conveniently heated by a commercial highly stabilized power supply. A constant current-stabilization feature is desirable. The lamp should be mounted in a nonreflective housing, with the vertical alignment and convective cooling being the same as those used in the NBS calibration. The lamp current should be allowed to stabilize for 10 min for each 100°K temperature change in the range from 1300 to 2000°K. The filament temperature should not be allowed to exceed 2573°K, and the lamp should not be used over 200 hr without some form of recalibration.

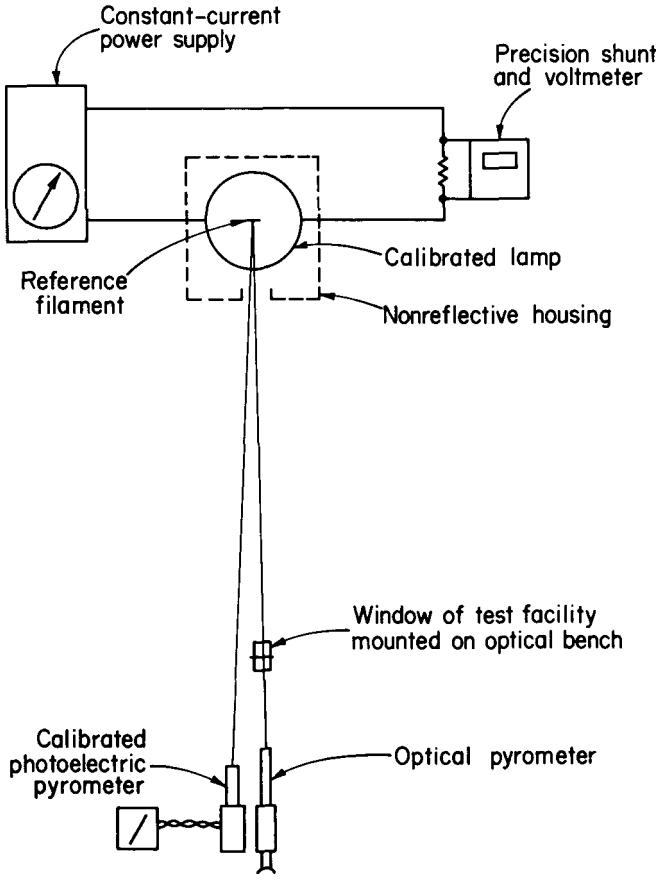


Fig. 13.17 Schematic of optical-pyrometer calibration facility.

The calibrated reference lamp is used as a target, and, when available, the calibrated photoelectric pyrometer can be used to cross-check calibrations. The alignment must correspond to the NBS procedures. Then, with constant current flowing through the lamp, the photoelectric pyrometer should be moved slightly so that the optical pyrometer can be moved into the normal position of incidence required for calibration. The photoelectric pyrometer reading must not change after the alignment procedures if the lamp is at constant current. Usually the windows used in the thermionic experiment are introduced into the field of view of the optical pyrometer, but they must not interfere with the radiation supplied to the photoelectric pyrometer. The calibration then can

proceed. In almost all cases, a calibration by this procedure introduces uncertainties that are trivial to the other aspects of optical pyrometry.

Eye fatigue of the investigator is a major problem in optical pyrometry. Since spurious radiation is a prime cause of fatigue, unwanted radiation in the apparatus and around the pyrometer should be eliminated. A dark cloth placed over the head of the observer and body of the pyrometer, reminiscent of the photographers of the early 1900s, does improve temperature measurements. If the target is sufficiently large, a slight motion of the filament with respect to it sometimes improves accuracy. The pyrometers using circular filaments often exhibit a brightness temperature change at the edge of the filament since emissivity is angle dependent. Slightly defocusing the optics of the eye piece will provide a uniform yet adequately defined filament.

13.9 Voltage and Current Measurements

Voltage and current measurements are quite conventional except for certain precautions dictated by the nature of experimental converters. These precautions, which are related to electrical connections, guard rings, and transient tests, are discussed in Secs. 13.9.1 to 13.9.4.

13.9.1 Electrical Connections. Wires used to measure electrode output voltage are often made of the same material as the electrodes to avoid contamination of the electrodes.

Since most electrode materials have a low Seebeck emf, thermoelectric corrections to the measured voltage are small and are usually neglected. These corrections should be made for accurate results, however. They are the sum of emf's in the emitter wire from the voltmeter at room temperature to the emitter at heat-source temperature minus the sum of emf's in the collector wire from the voltmeter at room temperature to the collector at its temperature. The largest emf, which occurs in the emitter leg, can be about 20 mv.

We can combine the functions of wires for voltage measurements with those of thermocouple wires if we use care. Although a properly functioning high-impedance voltmeter will not interfere with the thermocouple output, it is customary to disconnect the voltmeter at the time of the thermocouple measurement for the sake of improved experimental reliability.

Connections for current measurements are conventional except when transient tests are involved. Then, if feasible, the current lead, vacuum feedthrough, and shunt connections should be designed to be as close as possible to a coaxial configuration. Thus the magnetic fluxes caused by forward and return current flows almost cancel each other, and their inductive effects on response time are minimized.

13.9.2 Guard Rings. A collector guard ring is used in experimental converters, as shown in Fig. 13.18. If properly designed, it serves the purposes of reducing leakage currents from the side surfaces of the collector to the converter envelope and of defining of an emitter–collector area in which the particle flows are not affected by fringe fields. For the latter purpose the guard ring must be maintained electrically isolated from and at the same potential as the collector.

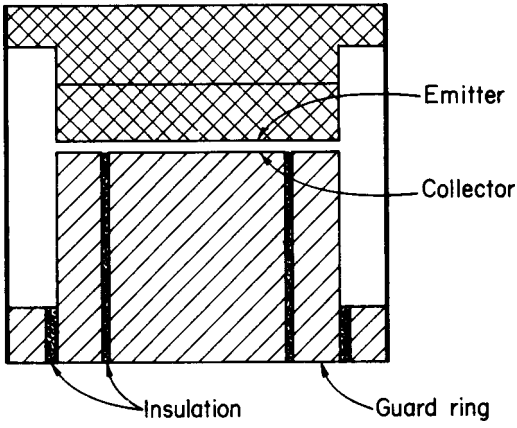


Fig. 13.18 Schematic of collector guard ring in an experimental converter.

The use of collector guard rings under steady-state conditions is straightforward. In transient tests, however, it is somewhat more difficult to maintain the collector and the guard ring at the same potential. The difficulty can be partially overcome by tying the guard-ring lead downstream of the measuring shunt. An arbitrary amount of resistance can be introduced into the guard-ring circuit to help balance the resistive drops in the two current collection paths. Since an equal balance can rarely be established because of the fringe fields at the outside edge of the guard ring, the simpler expedient of using low-resistance shunts and a high-gain amplifier is often selected.

Some experimenters use active elements to balance the potential between the guard ring and the collector. The circuit used is similar to that found in highly regulated power supplies where the voltage regulation is maintained by fast-response solid-state power transistors. With minor modifications the power supplies can be adapted to the voltage-follower purpose.

Emitter guard rings are rarely used because they usually create more problems than they cure. For this reason they will not be discussed here.

13.9.3 Transient Tests. Electrical characteristics of converters depend on several parameters, each of which can vary over a wide range of values. Studying

them experimentally requires amounts of data that can be conveniently obtained only by means of transient test methods.

Usually the voltage is varied rapidly, and the current and voltage are recorded. The time for the recording of a current–voltage relation is selected largely by experience, backed by first-order analysis. An attempt is made to select a time interval sufficiently small so that the temperatures of the converter parts are nearly constant and yet sufficiently long so that transient effects in the instrumentation, interelectrode plasma transport, and relaxation processes are minimized. Accordingly, a voltage excursion of the order of 2 volts is accomplished over a period of about 5 to 10 msec; usually thermal time constants of electrodes are of the order of a fraction of a second, whereas ion transit times in the interelectrode space based on thermal velocities are about 1 to 10 μsec.

The voltage excursion can be applied in a variety of ways, some of which employ transformer coupling to reduce input noise and improve instrument response time. Often the current–voltage relation is sampled only during specified discrete intervals, but the voltage excursion is varied as a periodic function of time and is recorded on the screen of an x,y oscilloscope. Of course, for the full current–voltage relation to be sampled, the sampling period must be out of step with the excursion period. The sampling technique makes it possible to use slow-writing-speed instruments such as a pen-and-ink recorder. The slow-speed recorder has the advantage of providing an accurate and convenient graph of current vs. voltage. This advantage is obtained at the expense of time required for the acquisition of data.

Solid-state high-current wave generators are now displacing many of the transformer drive systems. An emitter–follower power-transistor amplifier easily converts the standard low-current wave (or function) generator to a high-current supply. Single or multiple variable-amplitude variable-frequency, variable-form pulses of the wave generator can then be transformed into high-current counterparts directly on the thermionic converter. Changes in rise time of the current in a sweep or of dwell time (duty cycle) are easily achieved in such a flexible sweep circuit.

13.9.4 Instruments for Transient Tests. Oscilloscopes used in transient tests must be of high quality. The current should be measured through a differentially coupled amplifier to avoid ground-loop problems associated with single-ended inputs. A differential amplifier is preferred for the voltage measurement also when the voltage leads are attached directly to the emitter and collector, rather than to the collector and ground.

The differential amplifier in the oscilloscope should be calibrated and adjusted frequently because of the tendency of the independent sections of the amplifier to drift with respect to each other.

A commonly ignored feature of amplifiers is the hysteresis in recovery time which can occur if the input is overdriven. The exponential-like rise of current with voltage or the discontinuous current surge after ignition in a converter has been observed to saturate the amplifier. The recovery from saturation is generally long compared with other amplifier response times and can be a major source of error. A convenient rule is to disregard data taken after the x,y trace departs from the fluorescent screen.

The instrumentation used in data samplers is usually an integral part of the sampling procedure and cannot be discussed in a general fashion. Nevertheless, attention should be given to interpreting amplifier specifications. The sampling devices often employ a capacitance storage with a finite charge capacity. Depletion of the charge (and, therefore, of the voltage) is usually avoided by use of high-impedance amplifiers for voltage sampling. Some of the amplifier circuits tend to be potentiometric in their design and have the characteristics of a high-input impedance when the input and output are properly proportioned. If, at the time it is switched into the circuit, the amplifier has a prior history of quite different input and output voltages, it will exhibit the low-input impedance until internal balance is achieved. Modern sampling circuits, called sample-and-hold amplifiers, use active elements to restore and maintain charge on the sampling capacitor. In general, however, the characteristics of any sampler must be carefully reviewed, and an effort should be made to minimize voltage changes of the sampling device during its nonsampling interval or during a sequential sampling step.

When current is measured, a signal imbalance can arise from inappropriate connection of the guard or shield of the input to the amplifier used to amplify the emf developed across the shunt. Usually the shunt is at collector potential and is biased negatively with respect to the grounded emitter. Then, as a voltage pulse is applied across the converter, the voltage in both leads from the shunt varies rapidly with time. If the shield of the amplifier leads to the shunt is inadvertently grounded, a phase shift in the current signal often occurs. The problem can be avoided by connecting the shield to a point immediately downstream of the current shunt and to the chassis ground of the input stage of the amplifier and then decoupling the chassis ground of the output stage. Most differential amplifiers are provided with this chassis-separation feature.

13.9.5 Computer-Controlled Data Acquisition. The large amount of data involved in the study of electrical characteristics suggests the use of computerized procedures. Computer systems have been designed for data acquisition, processing, and presentation.

The flow chart of one design is shown in Fig. 13.19. The principal elements of the design are:

1. *Test apparatus,* which consists of the converter, high-current pulse generator, conventional *x,y* oscilloscope, pen-and-ink recorder, and computer cell and control panel.

2. *Test-facility data center,* which consists of the precision switch gear for multifacility operation, differential input–output amplifier bank, sample-and-hold amplifiers, multiplexer (high-speed scanner), analog-to-digital converter, 8K-word high-speed computer, various outputs, typewriter, paper punch, digital-to-analog converter, and translator and transmission line that conditions and transmits the data in a format similar to that used in the central computation center.

3. *Central computation center,* which consists of the magnetic-tape record used as a central data file plus the usual hardware of large computation installations.

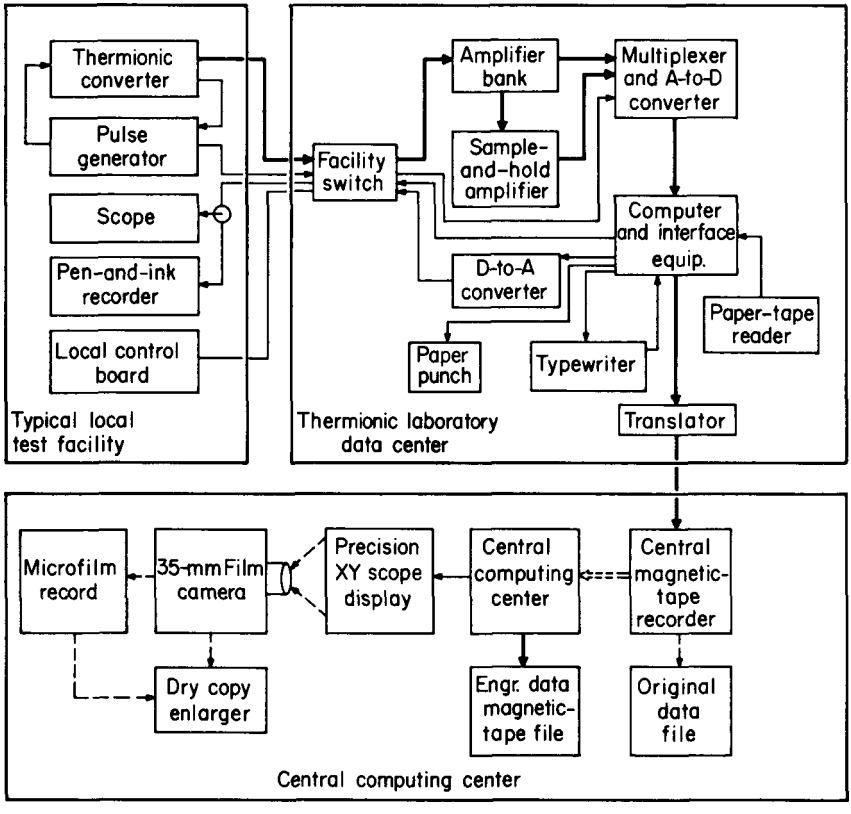


Fig. 13.19 Typical data-processing system.

The small high-speed digital computer controls and directs the flow of data through the various input and output devices shown in Fig. 13.19.

The flexible command structure of the computer permits a classification of input data. For example, input data are available as either time synchronized (high frequency) or infrequently sampled (low frequency) or in forms suitable for control functions and bookkeeping entries.

The accuracy of the system is limited by the differential amplifiers at high frequency (−3 db at 30,000 Hz) and by the sample-and-hold amplifiers at low frequency (±1 part in 8000). The remaining parts of the system have an accuracy of ±1 part in 8000. Playback lines provide for verification of accuracy.

The principal limitation in the speed of recording the data is in the conversion of the input signal from analog to digital (A to D) form. Conversion to a 13-bit digital number plus sign requires 30 μsec. The effective sampling rate is further reduced by the time required to store data in the core. The computer has a 0.9-μsec cycle time, but four cycles are required to address and store the time-synchronized data.

A specific data-recording path, with an illustration of some of the available data-playback options, can be traced with the aid of Fig. 13.19. Data and control information goes from the local facility through a facility switch. At the facility switch it is divided into a high-frequency set that enters the multiplexer directly. Control information enters the computer through appropriate interface equipment. On command from the converter control network and by activation of various pieces of equipment at the local test facility, all test data are generated, transferred, and stored in the computer core. Data are available at this step for review or further transfer. The data can be reviewed by returning two analog signals from the digital-to-analog converter to the local facility to provide the inputs for x,y graphs of current and voltage. Individual or arbitrary combinations of up to 12 graphs can be displayed in normal or expanded current (output from low- or high-gain amplifiers) as a function of voltage. The oscilloscope or pen-and-ink recorder can be used. The oscilloscope is usually used to screen data for appropriate scaling of signals and to make rough checks on the equipment, whereas the pen-and-ink recorder is used to check reproducibility.

Additional options exist at the local computer center. All or part of the data and the bookkeeping information in the core can be typed out in well-indexed tabular form, or the complete set of the data stored can be punched out on a paper tape.

An advantage of computer-controlled thermionic-converter experiments is that data acquisition can be scheduled and controlled to meet the natural thermal characteristics of the converter. The need for expensive temperature controllers can be eliminated by letting the computer search for the appropriate

data-recording intervals during simple ramped temperature excursions. Other advantages are the short time required for accurate data acquisition, an effective data file from both entry and retrieval points of view, and the ability to edit at the test station to minimize excessive and erroneous data generation.

References

1. Characterization of Materials, National Research Council, National Technical Information Service No. AD-649 941, 1967.
2. Methods of Experimental Physics, in *Solid State Physics*, Vol. 6, Part A, Academic Press, Inc., New York, 1958.
3. G. L. Clark (Ed.), *The Encyclopedia of Spectroscopy*, Reinhold Publishing Corp., New York, 1961.
4. R. Costaing, in *Advances in Electronics and Electron Physics*, L. Marton (Ed.), Vol. 13, Academic Press, Inc., New York, 1960.
5. D. B. Brown, D. B. Wittry, and D. F. Kyser, Prediction of X-Ray Production and Electron Scattering in Electron-Probe Analysis Using a Transport Equation, *J. Appl. Phys.*, 40(4): 1627-1636 (March 1969).
6. L. Yang, R. G. Hudson, T. Tagami, and J. W. R. Creagh, Preferred Crystal Orientation Evaluation and Vacuum Work Function Measurements on Chemically Vapor-Deposited Tungsten Emitters, in *Proceedings of the Thermionic Conversion Specialists Conference, Boston, Mass., Oct. 21–24, 1968*, pp. 31-40, Institute of Electrical and Electronics Engineers, New York.
7. R. G. Musket and J. Ferrante, Auger Electron Microscopy Study of Oxygen Adsorption on W(110), *J. Vac. Sci. Technol.*, 7(1): 14-17 (1970).
8. J. W. May, Electron Diffraction and Surface Chemistry, *Ind. Eng. Chem.*, 57(7): 18-39 (July 1965).
9. J. R. Thompson, J. C. Danko, T. L. Gregory, and H. F. Webster, Surface Characterization Studies on CVD Tungsten, in *Proceedings of the Thermionic Conversion Specialists Conference, Boston, Mass., Oct. 21–24, 1968*, pp. 23-30, Institute of Electrical and Electronics Engineers, New York.
10. J. R. Smith and A. L. Smith, A Simple Empirical Formulation of Electron Emission from Cesiated Metal Surfaces, in *Proceedings of the Thermionic Conversion Specialists Conference, Carmel, Calif., Oct. 21–23, 1969*, pp. 53-58, Institute of Electrical and Electronics Engineers, New York.
11. J. B. Newkirk and J. H. Wernick (Eds.), *Direct Observation of Imperfections in Crystals*, Wiley-Interscience, Inc., New York, 1962.
12. H. F. Webster and K. H. McCoy, Evaluation of Surface Facet Angles by a Light Reflection Treatment, in *Proceedings of the Thermionic Conversion Specialists Conference, Carmel, Calif., Oct. 21–23, 1969*, pp. 42-45, Institute of Electrical and Electronics Engineers, New York.
13. W. B. Nottingham, *Thermionic Emission*, Report TR-321, Research Laboratory of Electronics, Massachusetts Institute of Technology, 1956.
14. W. B. Nottingham and R. Breitwieser, Theoretical Background for Thermionic Conversion Including Space-Charge Theory, Schottky Theory, and the Isothermal Diode Sheath Theory, Report NASA-TN-D-3324, National Aeronautics and Space Administration, March 1966.

15. R. J. McCandless and D. R. Wilkins, Interpretive Studies of Thermionic Converter Performance, in *Proceedings of the Thermionic Conversion Specialists Conference, Miami Beach, Fla., Oct. 26–29, 1970,* pp. 227-232, Institute of Electrical and Electronics Engineers, New York.
16. L. van Someren et al., Evaluation of Thermionic Emitter Surfaces, *Adv. Energy Convers.*, 7(4): 201 (1968).
17. D. L. Jackson and A. E. Campbell, Effective Work Function Determinations of Rhenium and Molybdenum, in *Proceedings of the Thermionic Conversion Specialists Conference, Boston, Mass., Oct. 21–24, 1968,* pp. 1-9, Institute of Electrical and Electronics Engineers, New York.
18. J. M. Houston and H. F. Webster, Thermionic Converter Research Studies. Final Report, General Electric Company, Report AFCRL-64-409, 1964.
19. D. L. Fehrs and R. E. Stickney, Contact Potential Measurements of the Adsorption of Cs, O_2 and H_2 on Ta, *Surface Sci.*, 8: 267-287 (1967).
20. R. E. Thomas, G. A. Haas, and F. H. Harrig, Principles and Instrumentation for the Measurement of Work Function Variations by Electron-Beam Scanning, Report NRL-6336, Naval Research Laboratory, 1965.
21. L. W. Swanson and R. W. Strayer, Final Report, Single Crystal Surface Work Function and Desorption Studies, National Aeronautics and Space Administration, Grant NGR 38-010-001, October 1970.
22. P. P. Craig and V. Radeka, Stress Dependence of Contact Potential: The ac Kelvin Method, *Rev. Sci. Instrum.*, 41(2): 258-264 (February 1970).
23. P. W. Bridgman, Thermoelectric Phenomena, in *The Thermodynamics of Electrical Phenomena in Metals,* pp. 39-69, The Macmillan Company, New York, 1934.
24. General Principles of Thermoelectric Thermometry, *Temperature, Its Measurement and Control in Science and Industry,* A. I. Dahl (Ed.), Vol. 3, Part 2, pp. 3-32, Reinhold Publishing Corp., New York, 1962.
25. Temperature Measurement Thermocouples, ASA Standard C96 1-1964, Instrument Society of America, 1964.
26. Basic Concepts, Standards and Methods, *Temperature, Its Measurement and Control in Science and Industry,* F. G. Brickwedde (Ed.), Vol. 3, Part 1, Reinhold Publishing Corp., New York, 1962.
27. H. J. Kostkowski and R. D. Lee, Theory and Methods of Optical Pyrometry, Monograph No. 41, National Bureau of Standards, 1962.
28. R. D. Lee, The NBS Photoelectric Pyrometer and Its Use in Realizing the International Practical Temperature Scale Above 1063°C, *Metrologia,* 2: 150 (1966).
29. J. R. Branstetter, Some Practical Aspects of Surface Temperature Measurement by Optical and Ratio Pyrometers, personal communication to R. Breitwieser, September 1966.
30. E. M. Sparrow, L. U. Albers, and E. R. G. Eckert, Thermal Radiation Characteristics of Cylindrical Enclosures, *J. Heat Transfer,* 84(1): 73-81 (1962).
31. J. C. De Vos, Evaluation of the Quality of a Blackbody, *Physics,* 20: 669-689 (1954).
32. M. Lapp and L. P. Harris, Influence of Alkali-Vapor Atmospheres on Pyrometrically Determined Temperatures: Cesium and Potassium, *J. Appl. Phys.*, 34(12): 3622-3625 (1963).

14

Experimental Studies and Correlations of Characteristics

14.1 General Remarks

The output of a converter is fully specified by the output-current density J and the electrode output voltage V. For a cesium diode converter with given electrode surfaces, the values of J (or V) depend on the values of V (or J) and those of the emitter temperature T_E, collector temperature T_C, cesium-reservoir pressure p or temperature T_R, and interelectrode spacing d. It follows that, for a given pair of electrode surfaces, any one of the six quantities J, V, T_E, T_C, p, and d can be considered a dependent variable when the remaining five are regarded as independent parameters. The relation between a dependent variable and the corresponding independent parameters is called a *characteristic of the electrode pair* or simply a *characteristic*.

The values of certain distinguishing features of a set of characteristics can depend either on fewer than five independent parameters or on a number of combinations of the independent parameters which is less than five. For example, the cesiated work function that determines the emitter-saturation-current density is independent of the value of d. The relation between the values of a distinguishing feature of a characteristic and fewer than five independent parameters will be called a *correlation*.

Experimental characteristics and correlations contain a wealth of information about the thermionic-conversion process and can be used for studies of electrode properties, plasma properties, and performance optimization. In addition, they can be used as control and diagnostic tools.

The purposes of this chapter are: (1) to present a more or less comprehensive collection of experimental characteristics that illustrate the interdependences among converter variables; (2) to discuss some correlations that have been established; (3) to indicate how the characteristics and correlations can be used to gain insight into the thermionic-conversion process; and (4) to describe the use of the data in design, control, and diagnostics of practical converters. These topics are discussed in Secs. 14.2 to 14.4. Unless otherwise stated, the data presented in this chapter are obtained from a research converter with a (1 1 0) preferred-orientation chloride tungsten emitter and a polycrystalline molybdenum collector.

14.2 Characteristics

For a given pair of electrode surfaces, the most readily measurable characteristics are the output-current characteristics, namely, values of J vs. V (or V_T) with T_E, T_C, p, and d as parameters.

The shape of the $J-V$ curves depends strongly on the values of converter parameters; its general features are shown schematically in Fig. 14.1. The

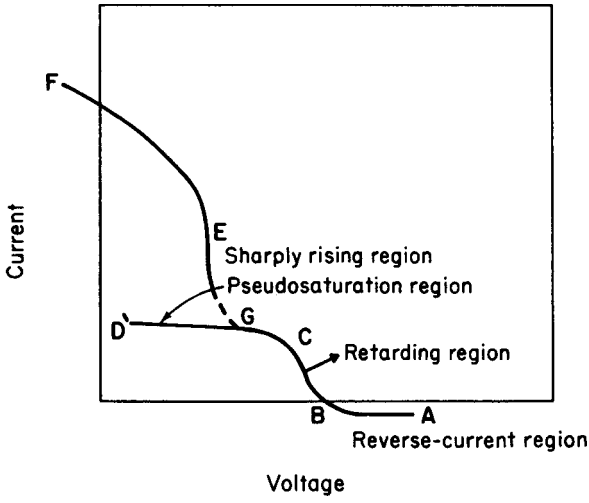


Fig. 14.1 Schematic of $J-V$ characteristics.

unignited mode AD and the ignited mode GF can be thought of as consisting of several regions. For example, the unignited mode consists of a reverse-current region AB, a retarding region BC, and a pseudosaturation region CD. Again, the ignited mode consists of a sharply rising region GE and a quasisaturation region EF. All these regions cannot always be precisely defined. In general, only some of these regions are visually noticeable for a given set of converter parameters, and the exact shape of the $J-V$ curve in the vicinity of the ignition point D and the deignition point G depends not only on the dynamic (as opposed to static) parameters of the converter but also on those of the external load.

The value of pd is an approximate measure of the number of collisions in the interelectrode space, namely, of the ratio of d/λ of the interelectrode spacing d and the mean free path λ for particle collisions. For this reason the conditions of operation of converters can be conveniently divided into two categories, low-pd conditions for which collisions are negligible and the high-pd conditions for which collisions are important.

14.2.1 J–V Characteristics for Low Values of pd (pd < 1 mil-torr). Although interparticle collisions are negligible for low values of *pd*, the *J–V* characteristics are not ideal, because of space-charge effects. Depending on whether the emitter is ion rich ($\beta > 1$) or electron rich ($\beta < 1$), either the electron current or the ion current will be space-charge limited.

(a) Ion-Rich Conditions. A typical *J–V* characteristic for low-*pd* and ion-rich conditions is shown in Fig. 14.2. Electron-saturation current is observed in the

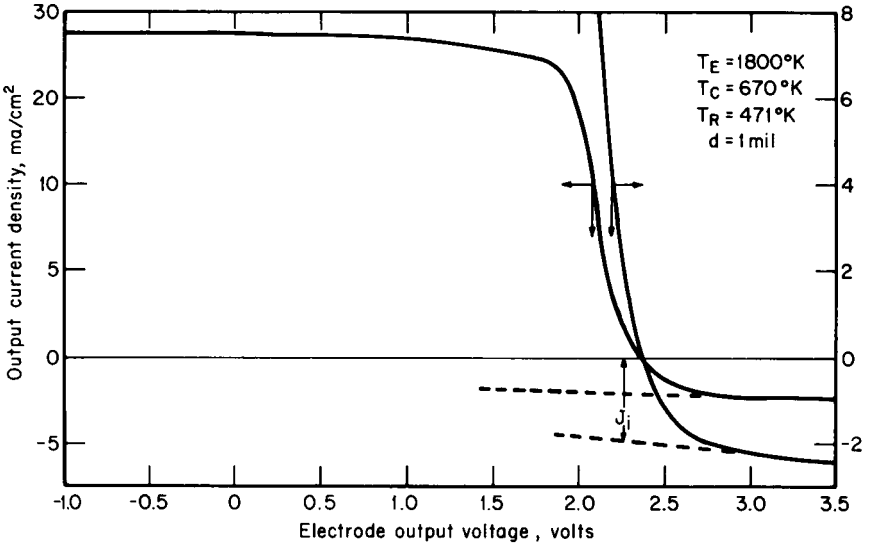


Fig. 14.2 Typical *J–V* characteristic for low-*pd* and ion-rich emitter conditions.

electrode-output-voltage range from −1 to +1 volt because the relatively large numbers of cesium ions generated at the emitter neutralize the electron space charge and because the infrequent particle collisions do not cause appreciable electron backscattering. The saturation current density can be used in the Richardson equation to calculate the cesiated-emitter work function.

The electrode-output-voltage range in excess of 2.5 volts in Fig. 14.2 represents the reverse-current region. For the set of conditions for which the data were obtained (relatively high emitter temperature and low cesium-reservoir and collector temperatures), the reverse current is composed primarily of ions generated at the emitter. Its saturation current density, given by the Saha–Langmuir equation, should be 100 ma/cm^2. The observed ion current density is only 2 to 3 ma/cm^2, however; this indicates the existence of ion space charge in

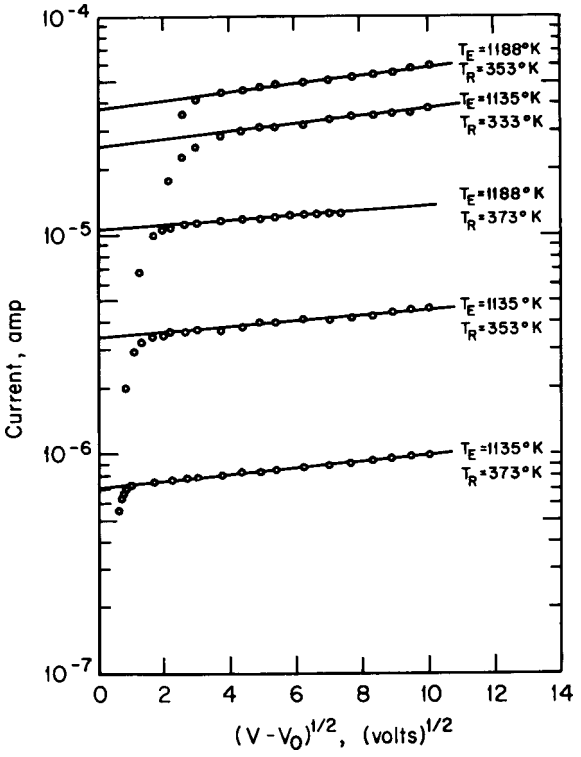


Fig. 14.3 Typical Schottky plots of cesium-ion currents for a (1 0 0) tungsten surface. V is the applied voltage, and V_0 is the contact potential.

the vicinity of the emitter. Ion saturation current can be measured by applying several hundred volts between the electrodes and using Schottky plots. A number of such plots[1] is shown in Fig. 14.3. The ion saturation current can then be used in the Saha–Langmuir equation to determine the cesiated-emitter work function.

The electrode-output-voltage range from 2 to 2.5 volts in Fig. 14.2 corresponds to the retarding range for electrons. Here the relation between the electron current density J_e from emitter to collector and the electrode output voltage is expected to follow the equation for the Boltzmann line. Since, in the absence of back emission, the output current density J is given by the expression

$$J = J_e - J_i \tag{14.1}$$

the ion current density J_i must be added to the observed J to obtain J_e. In the range of 2 to 2.5 volts, J_i can be estimated by extrapolating the reverse-current

characteristic as shown by the dashed lines in Fig. 14.2. A semilogarithmic graph of J_e vs. V, obtained by adding J_i to the observed J, is shown in Fig. 14.4. We see from this graph that a straight line with slope $1/kT_E$ is observed and that it extends over several decades in current. We might be tempted to conclude that we can use Boltzmann lines under low-*pd* ion-rich conditions to measure collector work functions. When this is done, however, the resulting collector work functions appear to depend on the emitter temperature or the ion current.

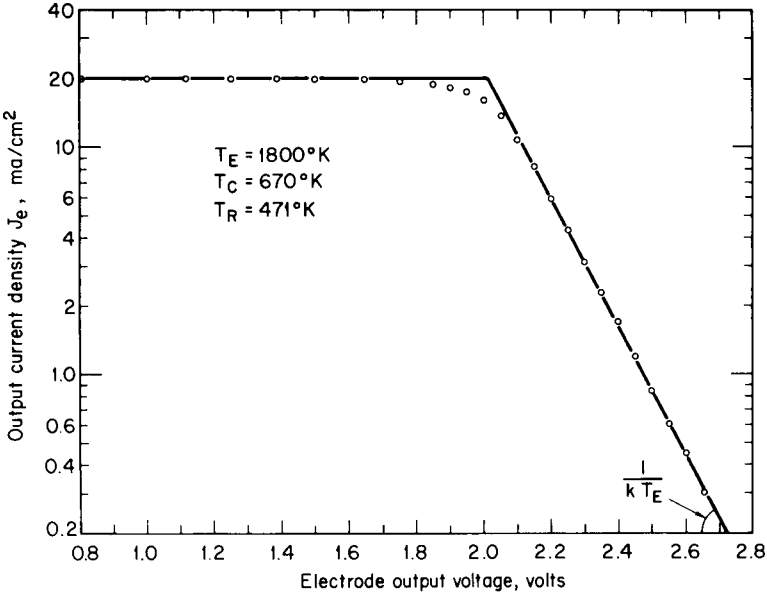


Fig. 14.4 Boltzmann line from retarding region of data in Fig. 14.2.

This dependence indicates that the method used to account for ion current is not sufficiently accurate and, therefore, that the collector work function results are not reliable.

The effect of cesium-reservoir temperature on $J-V$ characteristics under ion-rich low-*pd* conditions[2] is shown in Fig. 14.5. As the cesium temperature increases, the cesium coverage of the emitter surface increases, the emitter work function decreases, and the electron current increases and ion current decreases in accordance with the Richardson and Saha–Langmuir equations, respectively.

The dependence of $J-V$ characteristics on emitter temperature[2] is shown in Fig. 14.6. As the temperature increases, the cesium coverage of the emitter surface decreases, resulting in lower electron current and higher ion current. For

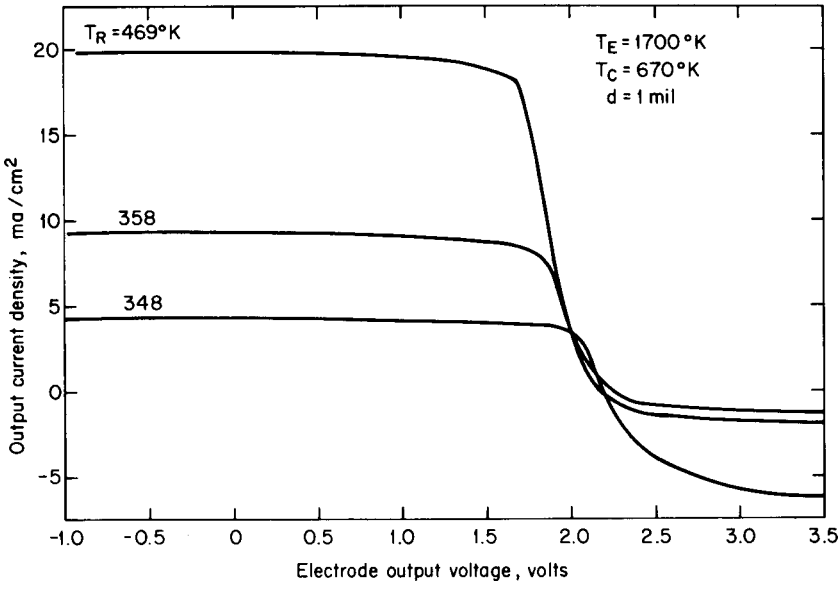


Fig. 14.5 Dependence of $J-V$ characteristics on cesium temperature for $\beta > 1$ and $pd <$ 0.1 mil-torr.

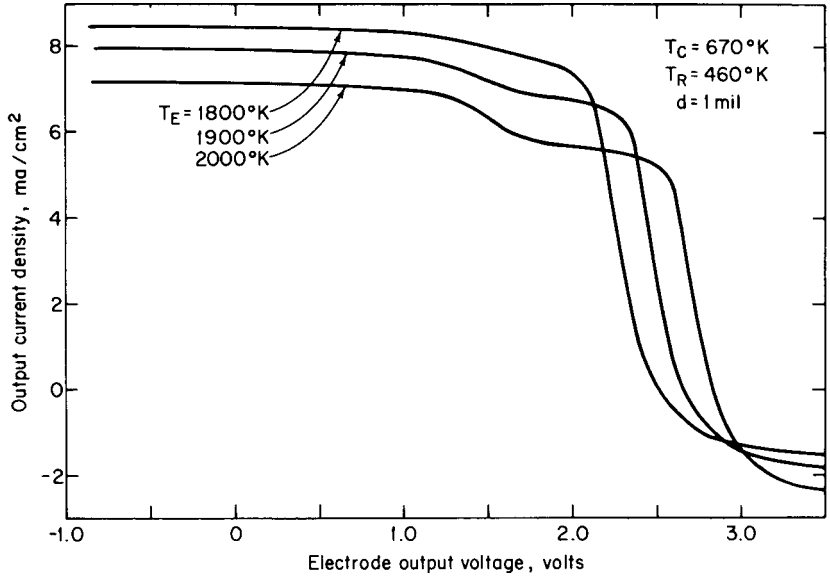


Fig. 14.6 Dependence of $J-V$ characteristics on emitter temperature for $\beta > 1$ and $pd <$ 0.1 mil-torr.

a given current in the retarding region, increase in the emitter temperature also results in an increase in output voltage.

(b) Electron-Rich Conditions. Typical $J-V$ characteristics for a cesium pressure of 0.34 torr and several values of interelectrode spacing[3] are shown in Fig. 14.7. The emitter temperature is low so that ion generation at the emitter surface is negligible and the converter behaves almost like a vacuum diode. The collector temperature is close to the emitter temperature so that electron emission from the two electrodes is comparable in magnitude. Since, for the interelectrode spacing of 0.5 mil, current density is less than about 7 ma/cm^2 and the critical current density J_R, calculated from Eq. 10.47, is approximately 150 ma/cm^2, the $J-V$ characteristic is not space-charge limited. For the spacing of 5 mils, however, space charge becomes effective, particularly in the voltage range from 0 to -2 volts.

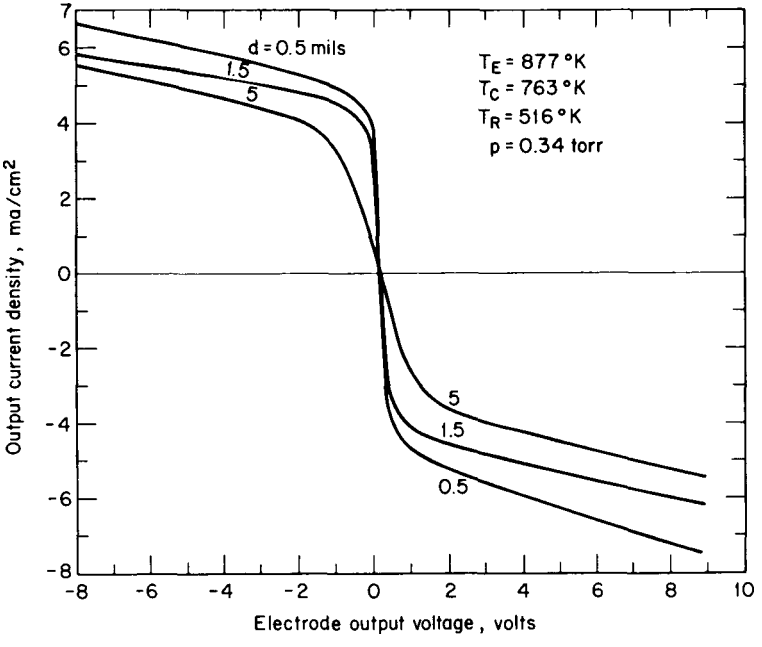


Fig. 14.7 Typical $J-V$ characteristics for $\beta < 1$, $pd < 1$ mil-torr, and interelectrode spacing as a parameter.

For low-pd electron-rich conditions, the region of $J-V$ characteristics around the open-circuit voltage ($J \approx 0$) provides a great deal of information. Typical data are shown in Fig. 14.8a. In the range from 1.5 to 2 volts the characteristic can be described by the equation for the Boltzmann line. A graph of $\ln J$ vs. V in

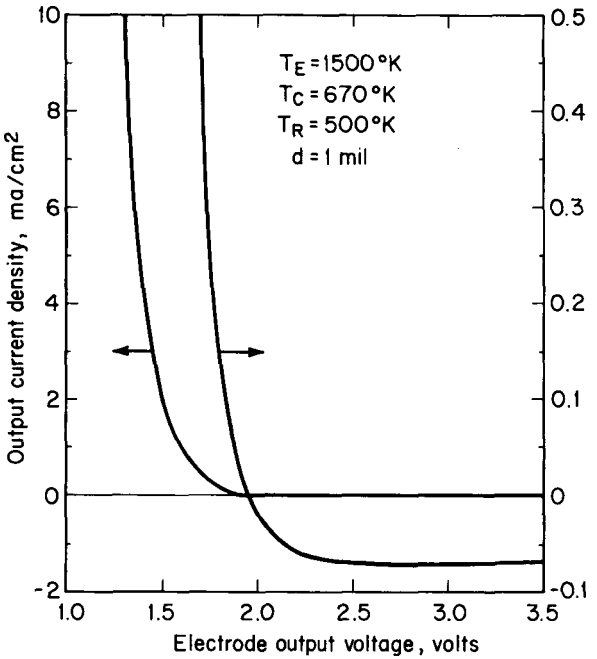


Fig. 14.8a Typical $J-V$ data around open-circuit voltage for $\beta < 1$ and $pd < 1$ mil-torr.

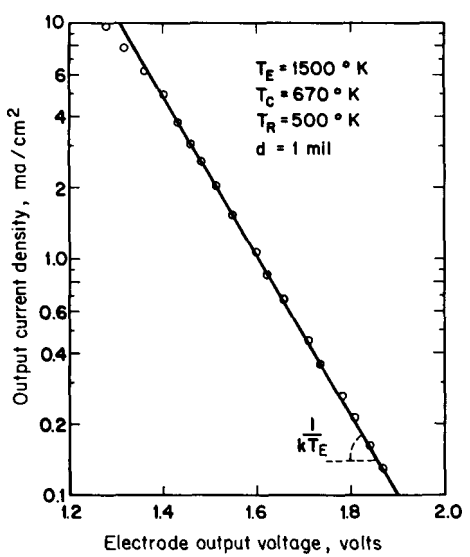


Fig. 14.8b Semilogarithmic graph of data in Fig. 14.8a.

this range gives a straight line with the slope of $1/kT_E$ as shown in Fig. 14.8b. Here the ion and back-emission currents are negligible, and therefore collector-work-function determination using the Boltzmann line is free of the errors discussed earlier.

The dependence of the Boltzmann line on cesium-reservoir temperature is shown in Fig. 14.9. Variations in cesium-reservoir temperature affect the collector work function and displace the Boltzmann line horizontally.

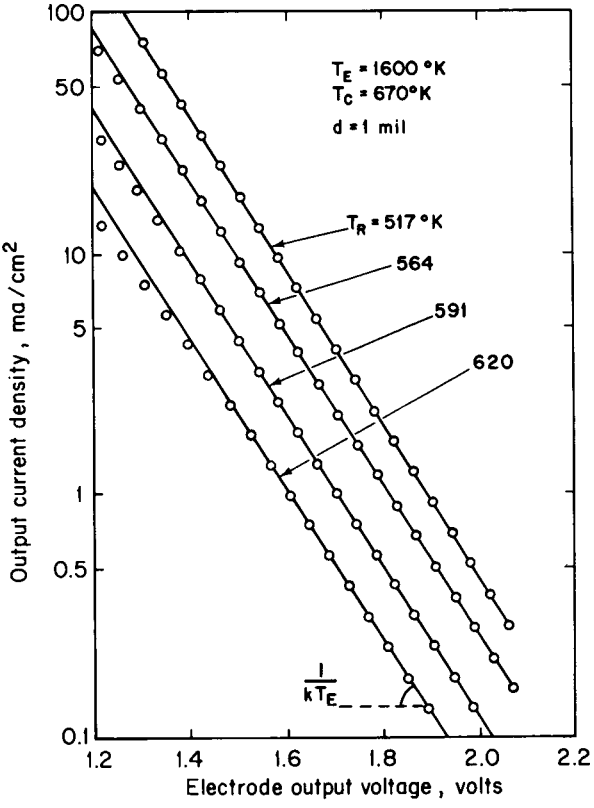


Fig. 14.9 Effect of cesium-reservoir temperature on Boltzmann line for $\beta < 1$ and $pd \approx 1$ mil-torr.

The collector-work-function value calculated from the data of Fig. 14.8a is 1.8 ev. From the Richardson equation the collector saturation current J_{CS} is found to be only 1.6 μa/cm^2. To a good approximation, therefore, the reverse current of 70 μa/cm^2 can be attributed entirely to the current of ions generated at the emitter surface. Furthermore, because of the electron-rich conditions, the ion current is not space-charge limited and can be used to determine the emitter

work function by means of the Saha–Langmuir equation. Section 14.3.1 shows that emitter-work-function values found by this method are in good agreement with the values calculated from measurement of electron saturation current from the emitter.

The dependence of ion current on cesium-reservoir temperature[4] is shown in Fig. 14.10. As the cesium-reservoir temperature increases, the emitter work function and the ion current decrease in accord with the Saha–Langmuir equation.

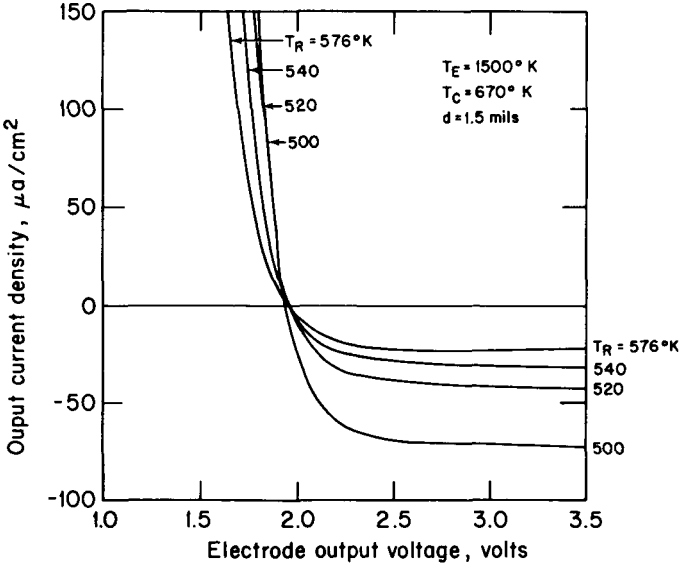


Fig. 14.10 Effect of cesium-reservoir temperature on ion current for $\beta < 1$ and $pd \approx 1$ mil-torr.

14.2.2 J–V Characteristics for High Values of pd (pd > 1 mil-torr). At high cesium pressures and high interelectrode spacings, particle collisions in the interelectrode space cause the $J-V$ characteristics to deviate considerably from ideal.

(a) Ion-Rich Conditions. A typical $J-V$ characteristic[4] for $pd = 40$ mil-torr and ion-richness ratio $\beta = 10$ is shown in Fig. 14.11. Superimposed on the same graph is the corresponding ideal characteristic. We see from this figure that the current is considerably less than ideal for practically all voltages.

The dependence of $J-V$ characteristics on spacing is shown in Figs. 14.12 and 14.13. The transition from low to high pd values can be observed[5] clearly in Fig.

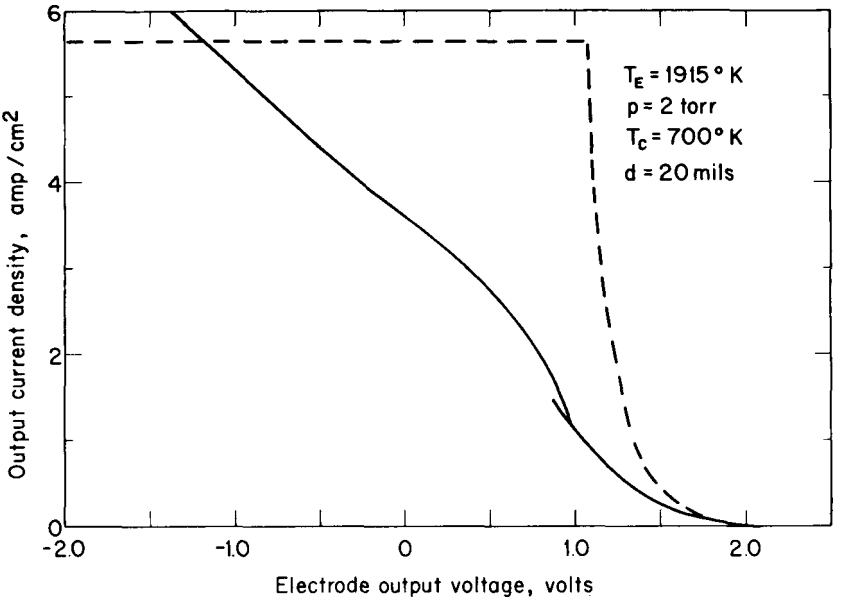


Fig. 14.11 J–V characteristic for $\beta = 10$ and $pd = 40$ mil-torr.

14.12. For pd = 0.5 mil the characteristic is nearly ideal. As the spacing increases, the current decreases until, at a spacing of 2.5 mils, ignition occurs, and there is a corresponding sharp rise in current and a drop in voltage. For certain voltages the current for the ignited mode exceeds the ideal current. The excess current is probably caused by the Schottky effect and by ion current from the interelectrode space to the emitter. The pseudosaturation region is observed for spacings greater than 0.5 mil. For the higher cesium pressure of Fig. 14.13, the pseudosaturation region appears only for spacings greater than 10 mils.

(b) Electron-Rich Conditions. At high-pd electron-rich conditions, the converter performance is considerably poorer than ideal because of (1) the insufficient number of ions generated at the emitter surface for space-charge neutralization and (2) the numerous collisions in the interelectrode space. Nevertheless, this region is of practical interest because relatively high power densities can be obtained at practical interelectrode spacings. Since detailed discussion of the dependence in this region of output power on converter variables is given in Sec. 14.4, only a brief discussion is presented here.

The dependence of J–V characteristics[5] on interelectrode spacing is shown in Figs. 14.14 and 14.15. As the spacing increases, the unignited-mode and

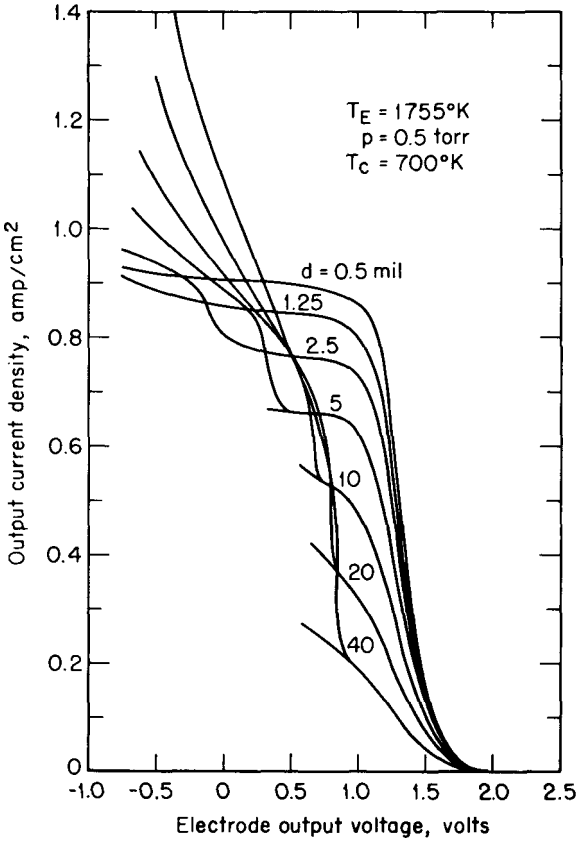


Fig. 14.12 $J-V$ characteristics for $\beta = 10$, $p = 0.5$ torr, and various interelectrode spacings.

ignited-mode currents decrease. The sharply rising regions of the $J-V$ characteristics show an initial increase and a subsequent decrease in voltage with increasing spacing.

The dependence of $J-V$ characteristics on emitter temperature[2] is shown in Fig. 14.16. As the emitter temperature increases, the emitter work function increases, the electron saturation current decreases, and the electrode output voltage increases. The increase in emitter work function enhances the generation of ions; this, in turn, reduces the electron space charge and increases the unignited-mode currents, as shown[6] in Fig. 14.17.

The dependence of $J-V$ characteristics on collector temperature[7] is shown in Fig. 14.18. As the collector temperature increases in the range from 630 to 970°K, the output voltage increases because of a decrease in collector work

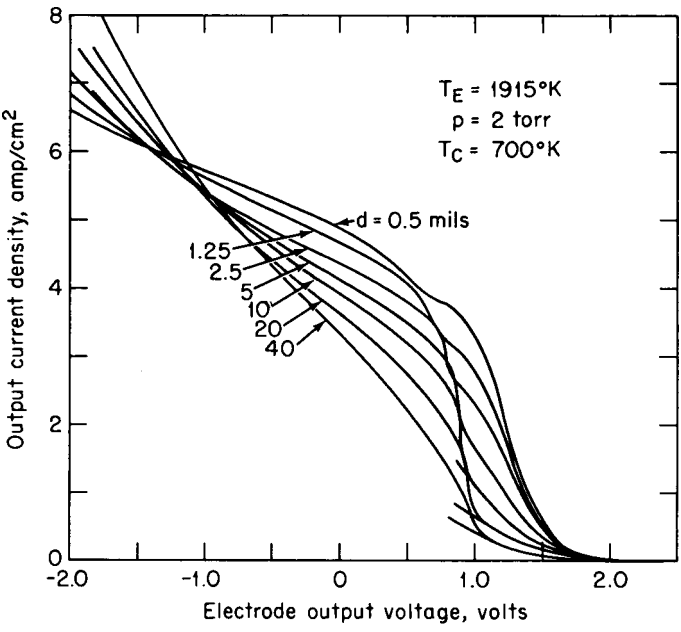


Fig. 14.13 $J-V$ characteristics for $\beta = 10$, $p = 2$ torr, and various interelectrode spacings.

function. Further increase in collector temperature results in a reduction in voltage in the sharply rising region of the characteristics, however, because electrons are emitted from the collector.

The dependence of $J-V$ characteristics on cesium-reservoir temperature[5] is shown in Fig. 14.19. The trends are self-evident.

The $J-V$ characteristics are more sensitive to changes in cesium-reservoir temperature than to changes in emitter temperature. As shown in Fig. 14.19, a 15°K increase in cesium-reservoir temperature increases the current at zero voltage from 9 to 15 amp/cm^2. As shown in Fig. 14.16, however, a 200°K change in emitter temperature is required to cause the same variation in current density.

14.3 Correlations

14.3.1 Emitter-Work-Function Correlations. The emitter work function significantly influences electron and ion emissions from the emitter surface and is directly related to the output voltage of the converter. Therefore accurate values of emitter work functions are necessary in studies of thermionic converters.

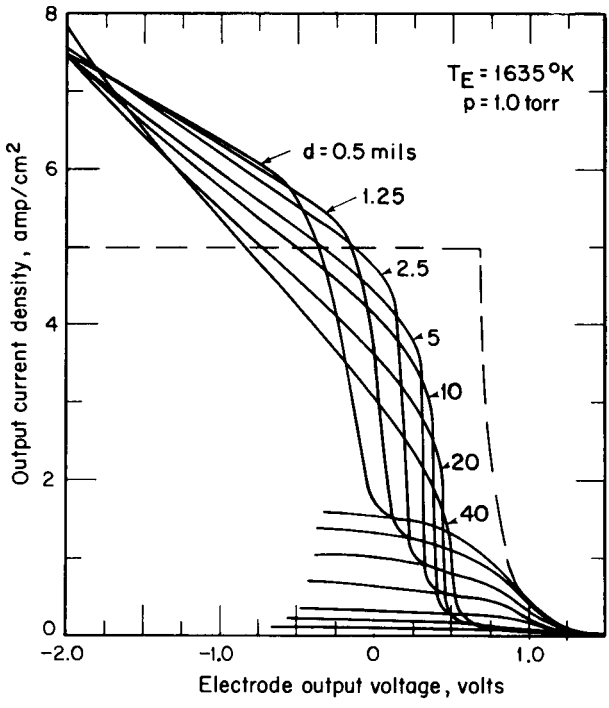


Fig. 14.14 J–V characteristics for $\beta = 0.1$, $p = 1$ torr, and various interelectrode spacings.

Accurate work-function values can be obtained only for certain ranges of converter variables. Outside these ranges, work functions are evaluated by means of correlations.

Figure 14.20 shows typical emitter-work-function values[8] vs. inverse cesium-reservoir temperature. They are obtained from measurements of electron saturation currents under ion-rich low-*pd* conditions and various emitter temperatures. The parameter T_E/T_R has been suggested for correlation of work functions.[9] A graph of emitter work function vs. this parameter (Fig. 14.21) shows that the data fall within a band of ± 0.050 ev. Such a band can result in a 100% error in calculating electron and ion emissions from the emitter surface. Thermodynamic analysis of surface films, discussed in Chap. 9, suggests another correlation based on the parameter $T_E \ln (p/p^0)$, where p is the cesium pressure and p^0 is a constant. A graph of emitter work function vs. $T_E \ln (p/p^0)$ (Fig. 14.22) shows that the data fall within a band of ± 0.020 ev.

Under ion-rich low-*pd* conditions, emitter work functions obtained from electron saturation currents are greater than about 3 ev. The corresponding

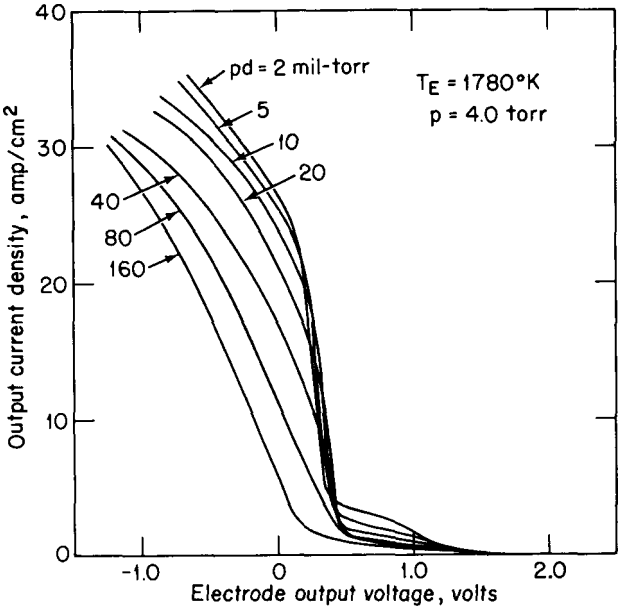


Fig. 14.15 $J-V$ characteristics for $\beta = 0.1$, $p = 4$ torr, and various interelectrode spacings.

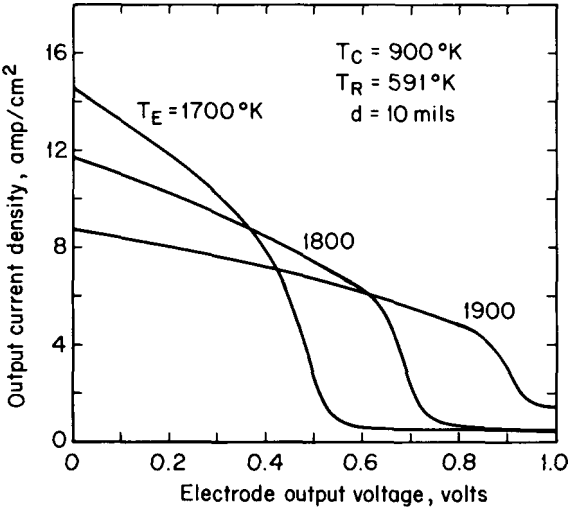


Fig. 14.16 Effect of emitter temperature on $J-V$ characteristics.

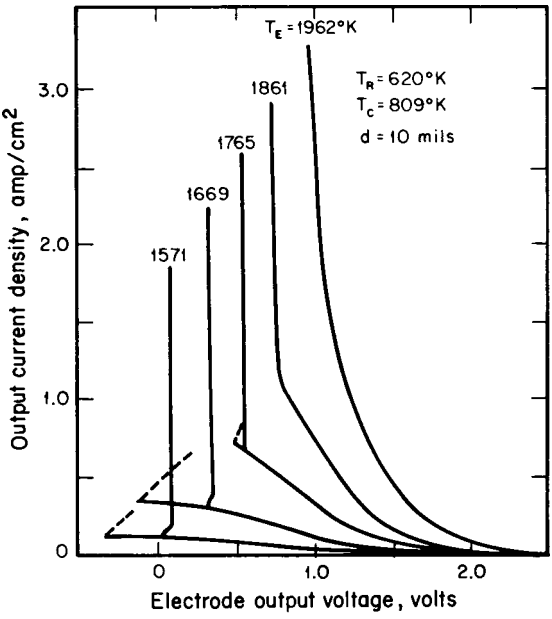


Fig. 14.17 Effect of emitter temperature on unignited-mode currents of a tungsten emitter prepared by chemical vapor deposition and a niobium collector.

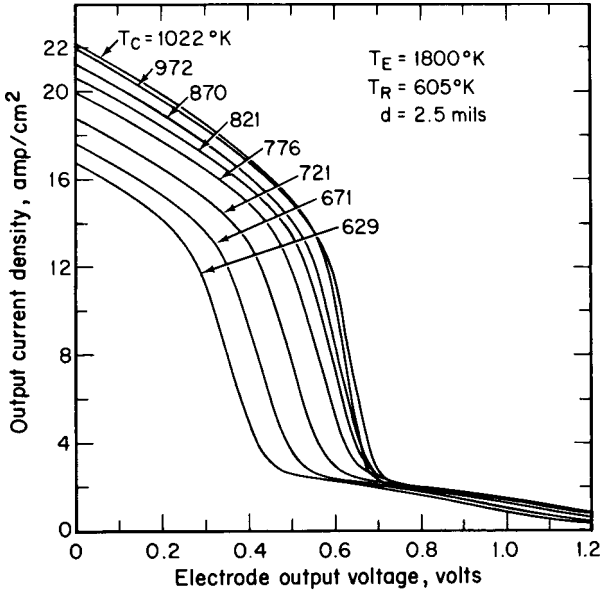


Fig. 14.18 Effect of collector temperature on $J-V$ characteristics.

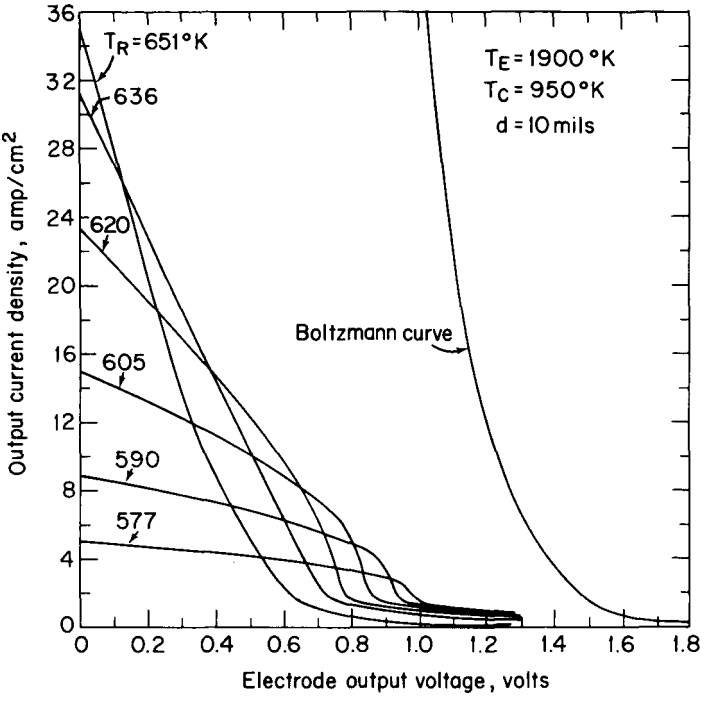


Fig. 14.19 Typical $J-V$ characteristics for $\beta < 1$, $pd > 1$ mil-torr, and various cesium-reservoir temperatures.

correlations can be extrapolated to values between 2 and 3 ev, and the extrapolations can be compared with experimental data obtained by measuring ion currents according to the procedure described in Sec. 14.2.1. In Figs. 14.23 and 14.24, the dashed lines show the extrapolation of work-function data from electron-current measurements, and the circles show the experimental data from ion-current measurements.

14.3.2 Collector-Work-Function Correlations. The parameters discussed for correlations of emitter work functions can also be used for correlations of collector work functions. Because collector-work-function measurements are more susceptible to error, however, the simpler parameter T_C/T_R is adequate for this purpose. Typical data[6] plotted vs. this parameter are shown in Fig. 14.25. We can appreciate the difficulties associated with collector-work-function measurements if we consider experimental results on the effects of collector temperature on converter performance.

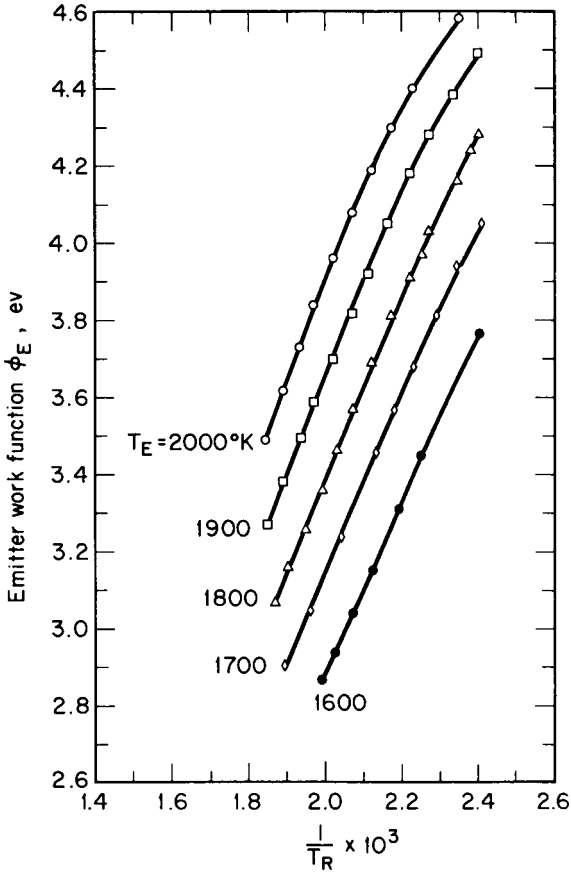


Fig. 14.20 Emitter work function vs. $1/T_R$, calculated from measurements of electron saturation current under ion-rich and low-*pd* conditions.

14.3.3 Effect of Collector Temperature. Collector temperature affects converter performance in two ways: (1) it changes the collector work function and, therefore, the electrode voltage according to the equation

$$e\,\Delta V = -\Delta\phi_C \tag{14.2}$$

and (2) it influences the collector emission according to the Richardson equation.

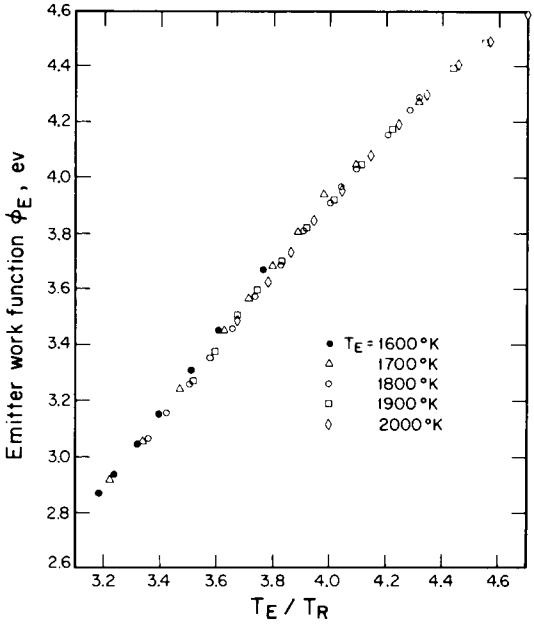


Fig. 14.21 Correlation of emitter work function with T_E/T_R.

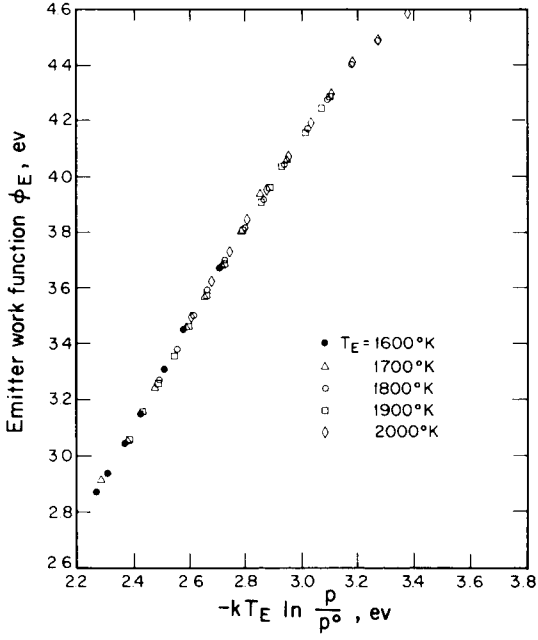


Fig. 14.22 Correlation of emitter work function with $kT_E \ln (p/p^0)$ for $p^0 = 3 \times 10^6$ torr.

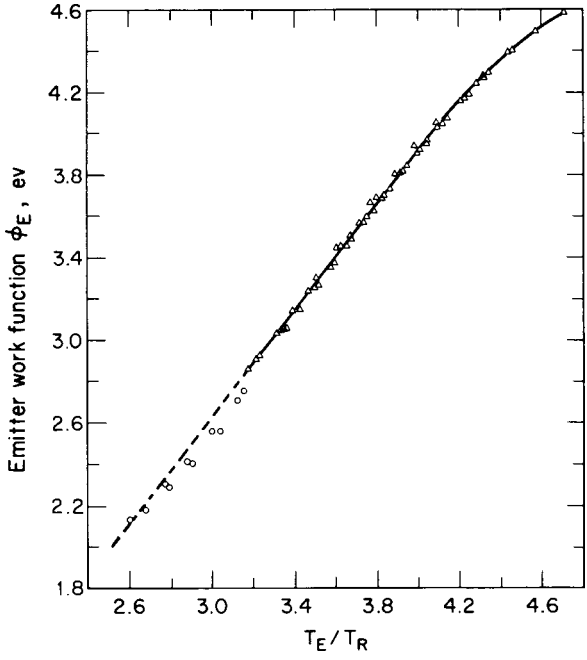


Fig. 14.23 Emitter work functions predicted by extrapolation of the T_E/T_R graph (- - - -) and deduced from ion-current measurements (○). The T_E/T_R graph is determined by electron-current measurements (△).

For low collector temperatures, collector emission is negligible; therefore collector temperature affects only the electrode output voltage according to Eq. 14.2. For high collector temperatures, back emission is no longer negligible, and an increase in collector temperature results in an increase in reverse current; this tends to reduce converter output power.

On the basis of most of the present converter analyses,[10,11] back emission becomes important only when it is comparable to the electron current moving from emitter to collector. Experimental results, however, show that back emission becomes important at values that are much smaller than those of the electron current moving from emitter to collector.

The effect of collector temperature on $J-V$ characteristics[7] is shown in Fig. 14.26. For a given current, increasing the collector temperature in the range 630 to 775°K results in a monotonic increase of the electrode output voltage. Further increase in collector temperature still results in further increase in the electrode output voltage but at a lower rate. When the collector temperature

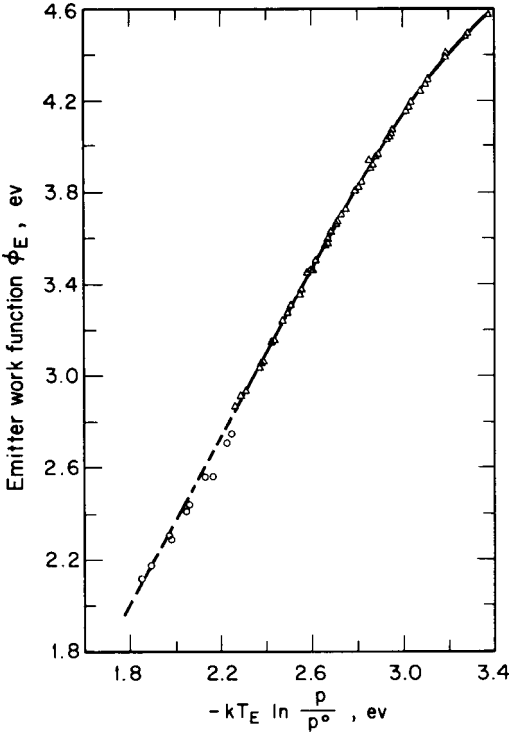


Fig. 14.24 Emitter work functions predicted by extrapolation of the $kT_E \ln (p/p^0)$ graph (- - - -) and deduced from ion-current measurements (○) for $p^0 = 3 \times 10^6$ torr. The $kT_E \ln (p/p^0)$ graph is determined by electron-current measurements (△).

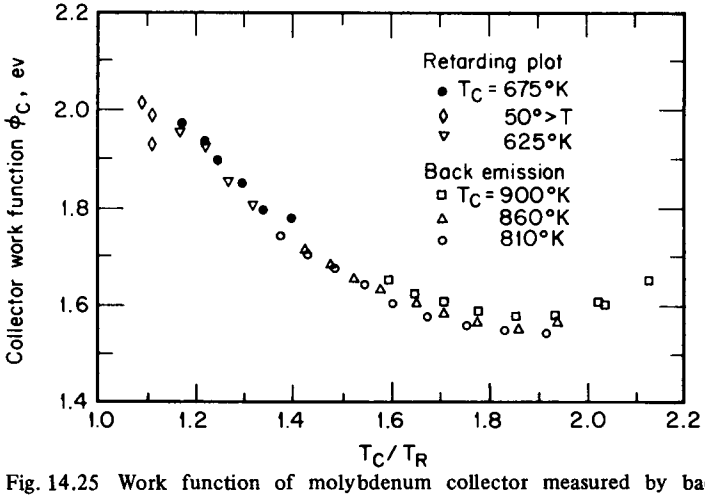


Fig. 14.25 Work function of molybdenum collector measured by back emission and retarding plot techniques.

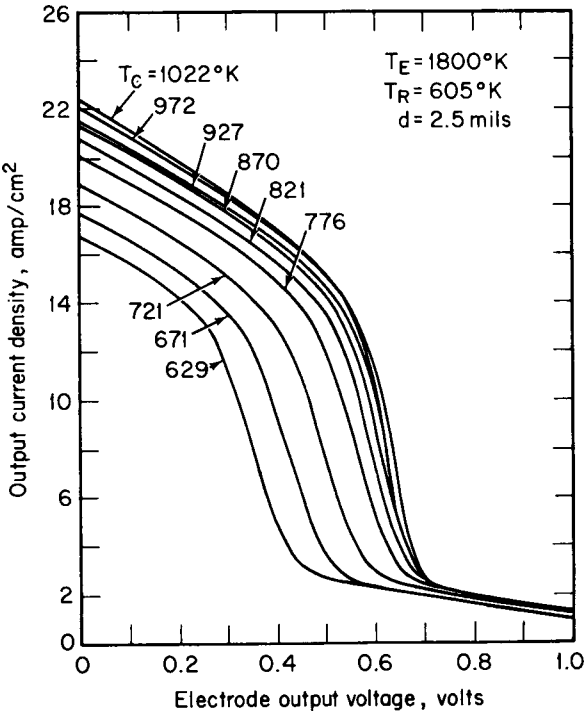


Fig. 14.26 Typical dependence of $J-V$ characteristics on collector temperature for a molybdenum collector.

exceeds 970°K, an increase in collector temperature results in a decrease in electrode output voltage.

For an output current density of 12 amp/cm^2, a graph of the change ΔV vs. T_C/T_R (for T_R = 605°K) is shown in Fig. 14.27. By virtue of Eq. 14.2 and if back emission were negligible, the shape of the curve in Fig. 14.27 should be identical to the shape of ϕ_C vs. T_C/T_R in Fig. 14.25. These two curves can be superimposed on the same graph (Fig. 14.28) by adding to ΔV the value of ϕ_C/e at T_C/T_R = 1 since then the collector temperature is relatively low, the collector work function is relatively high, and back emission is indeed negligible. We see from Fig. 14.28 that the two curves coincide for T_C/T_R in the range from 1 to 1.3 and that they differ by an amount V^* for T_C/T_R greater than 1.3. Data obtained at other cesium-reservoir temperatures show the same behavior (Fig. 14.29). The discrepancy between these two types of measurements ($\Delta\phi_C$ and ΔV) cannot be explained by the effects of back emission. It occurs even when back emission is of the order of 10^{-4} to 10^{-3} amp/cm^2 and the emitter current is several orders of magnitude larger.

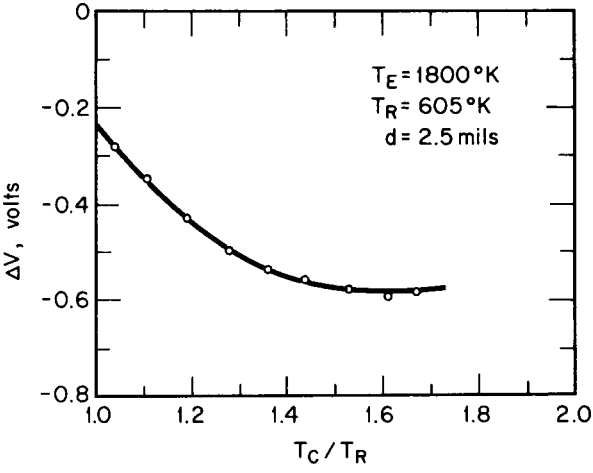


Fig. 14.27 Cross plot of Fig. 14.26 showing the dependence of output voltage on collector temperature at a current density of 12 amp/cm^2.

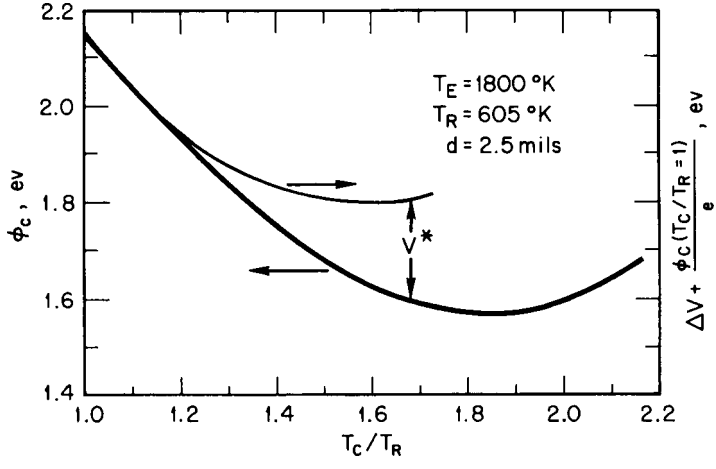


Fig. 14.28 Superposition of the curves in Figs. 14.25 and 14.27.

Present analyses cannot explain the discrepancy. One possibility is that it is caused by a double-valued sheath formed at the collector–plasma interface and that the effective collector work function that limits converter performance is the sum of the collector work function and the height of the double-valued sheath. If this is true, it follows that we must understand the criteria for

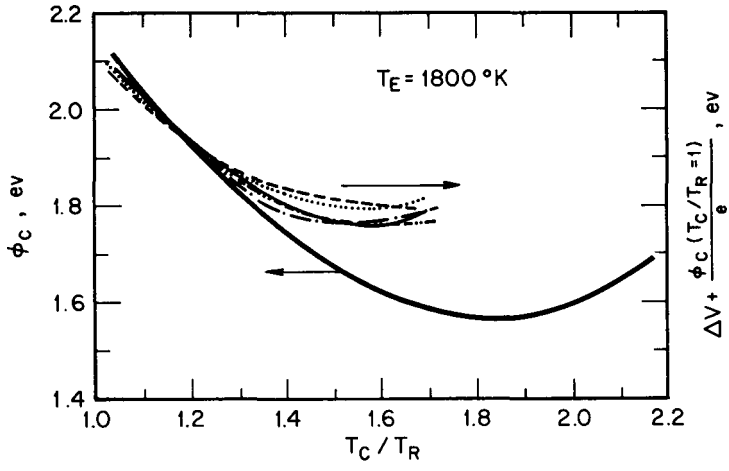


Fig. 14.29 Comparison of the changes in collector work function with changes in electrode output voltage for fixed output current and for several cesium-reservoir temperatures.

formation of double-valued sheaths to determine the figure of merit for collectors.

14.3.4 Voltage-Loss Correlations. A family of $J-V$ characteristics for several values of interelectrode spacing is shown in Fig. 14.30, and the corresponding cross graph of electrode output voltage vs. interelectrode spacing at a current density of 3 amp/cm^2 is shown in Fig. 14.31 (solid line). Superimposed on the graph in Fig. 14.31 is also the voltage loss V_d, which represents the difference between the actual and the ideal electrode output voltages of the converter vs. interelectrode spacing for a current density of 3 amp/cm^2. The maximum in the electrode output voltage can be qualitatively explained by the competition of two processes, ion generation and elastic scattering. At small spacings, ion generation is inadequate and hence limits the flow of electrons. At large spacings, ion generation is plentiful, but elastic scattering collisions are so frequent that they predominate and limit the flow of electrons.

Several measures can be used to describe deviations of a $J-V$ characteristic from the corresponding ideal characteristic. Among the most practical is the *voltage loss* V_{dm}, which is defined as the difference between the ideal electrode output voltage and the electrode output voltage of the converter at the point of maximum electrode output power. Since the point of maximum power occurs at the point of tangency of the $J-V$ characteristics to the cesium-optimized current characteristic, it can be readily identified from experimental data. Then V_{dm} can be determined provided that the collector work function is known so that

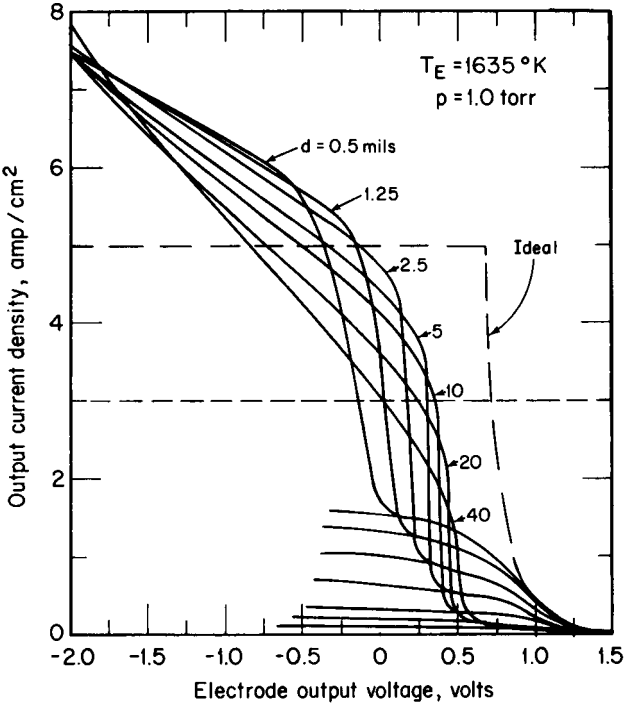


Fig. 14.30 $J-V$ characteristics for various interelectrode spacings.

the Boltzmann line of the ideal characteristic can be established. In contrast to other measures of deviation from ideality, the value of V_{dm} does not require knowledge of the emitter work function.

The dependence of the voltage loss V_{dm} on interelectrode spacing d is shown in Fig. 14.32 for several values of the cesium pressure. The same data are shown in Fig. 14.33 in the form of V_{dm} vs. pd, and additional data are shown in Figs. 14.34 and 14.35. The graphs of Figs. 14.33 to 14.35 indicate some degree of correlation between V_{dm} and pd although a persistent dependence on p is also evident. This latter dependence is better illustrated by the graphs in Fig. 14.36, in which V_{dm} vs. p is shown for various values of pd. The dependence on pressure, in addition to that on pd, is of the order of 0.1 volt; this is a significant fraction of the output voltage of thermionic converters.

The voltage loss V_{dm} represents the sum total of losses due to a number of causes, such as plasma losses, space-charge losses, and possibly losses due to a double-valued sheath at the collector–plasma interface. The minimum value of V_{dm} observed in experimental data is about 0.5 volt. If a double-valued sheath

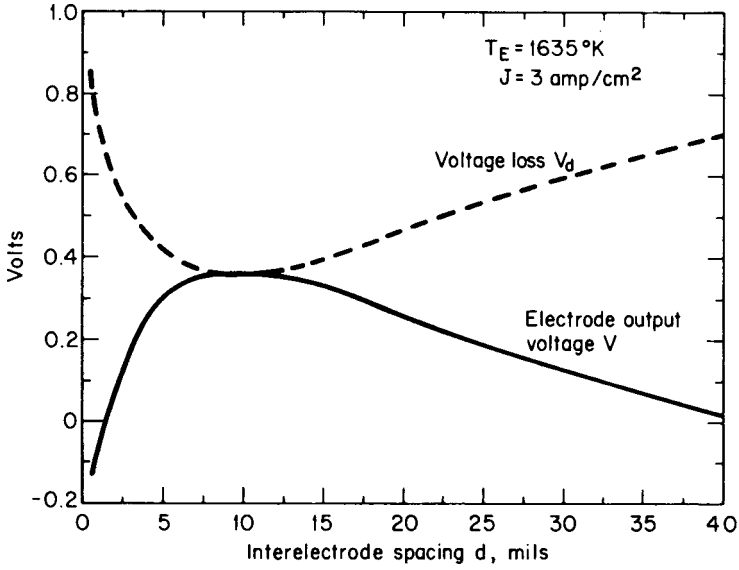


Fig. 14.31 Graphs of electrode output voltage and voltage loss V_d vs. interelectrode spacing for $J = 3$ amp/cm² and data from Fig. 14.30.

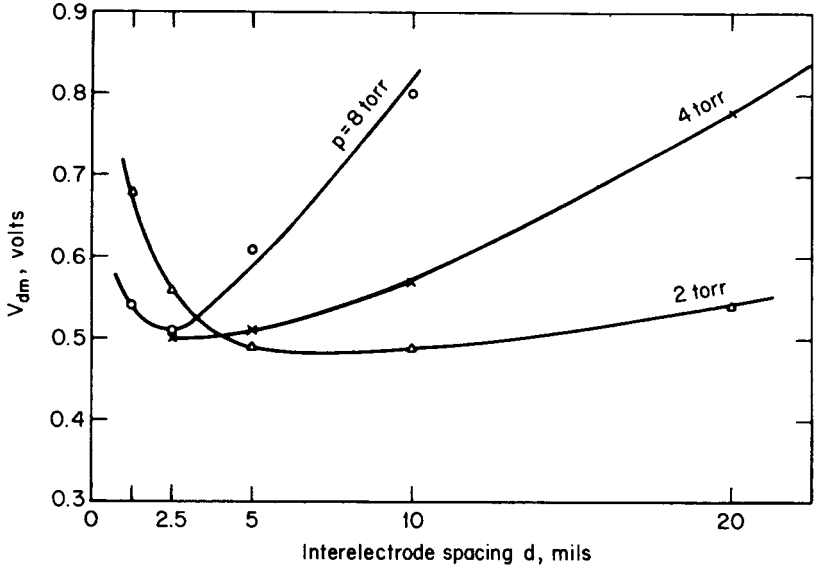


Fig. 14.32 Dependence of voltage loss V_{dm} on interelectrode spacing for several cesium pressures and $T_C \approx 1000°$K.

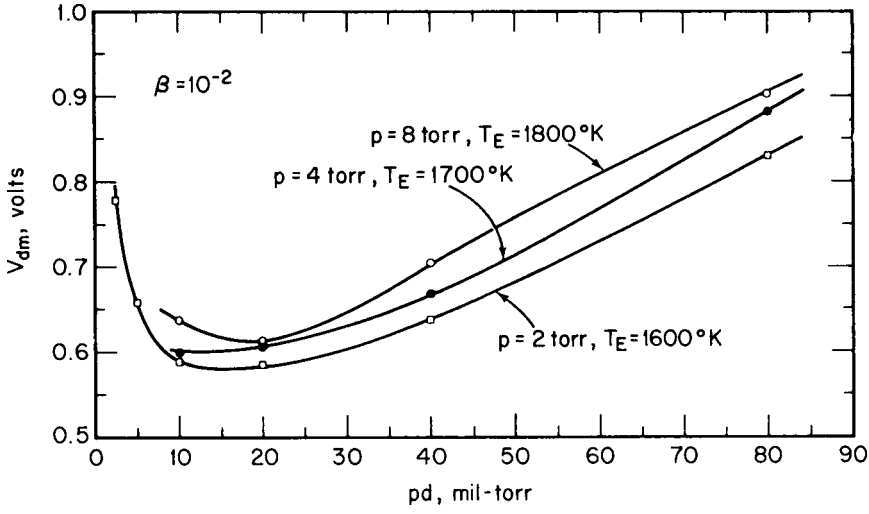


Fig. 14.33 Dependence of voltage loss V_{dm} on pd for $\beta = 10^{-2}$ and $T_C \approx 1000°K$.

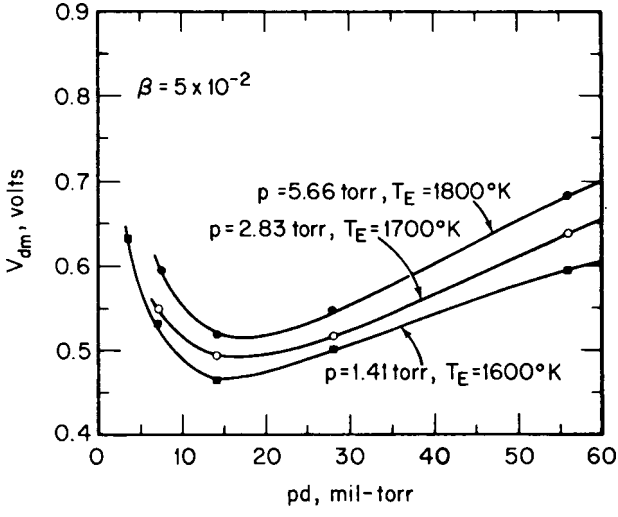


Fig. 14.34 Dependence of voltage loss V_{dm} on pd for $\beta = 5 \times 10^{-2}$ and $T_C \approx 1000°K$.

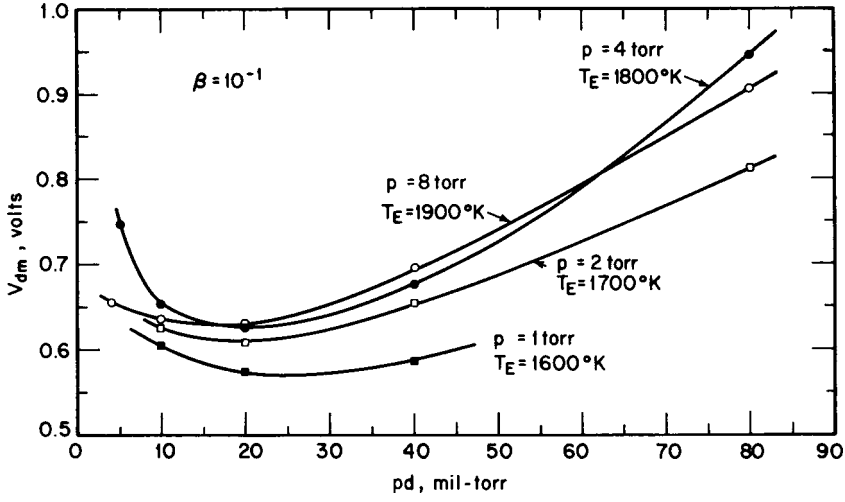


Fig. 14.35 Dependence of voltage loss V_{dm} on pd for $\beta = 10^{-1}$ and $T_C \approx 1000°K$.

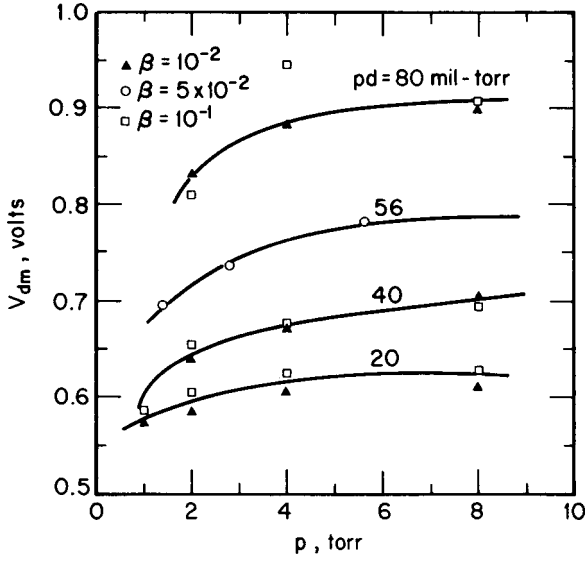


Fig. 14.36 Dependence of V_{dm} on cesium pressure and ion richness for $T_C \approx 1000°K$.

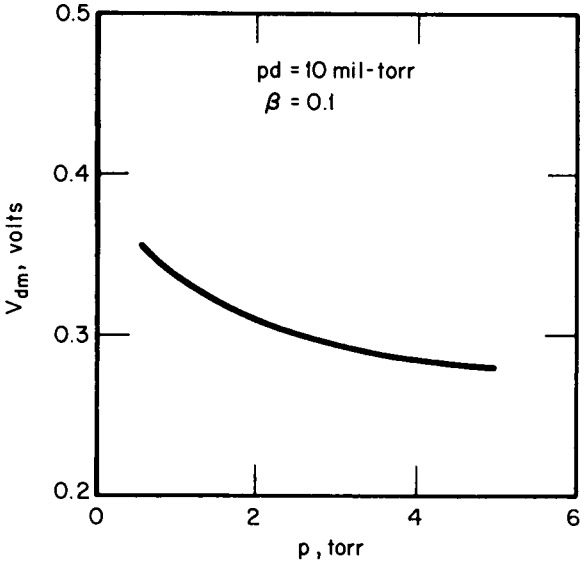


Fig. 14.37 Voltage loss V_{dm} vs. p for conditions for which back emission is negligible.

forms at the collector–plasma interface, its contribution to V_{dm} would be about 0.2 volt (Sec. 14.3.3), and therefore losses associated with the interelectrode plasma would account for the remaining 0.3 volt of the minimum V_{dm}.

The voltage-loss data presented in Figs. 14.33 to 14.35 were taken at optimum collector temperatures (900 to 1000°K) for which collector work functions are low and collector emissions are high. Hence a double-valued sheath may be present at the collector–plasma interface.

To investigate further the degree to which the voltage loss is due to collector effects rather than to plasma effects, we have examined data on V_{dm} vs. p taken under conditions for which back emission is negligible (Fig. 14.37). Two differences are observed by comparing Fig. 14.37 with the curves in Fig. 14.36: (1) The minimum voltage loss at low collector temperatures is about 0.3 volt, namely, about 0.2 volt less than the minimum observed at high collector temperatures. (2) The voltage loss decreases as pressure increases at low collector temperatures, whereas the voltage loss increases with pressure at high collector temperatures. These two differences give credence to the speculation that a double-valued sheath may be forming at the collector–plasma interface.

In converters the cesium pressure is adjusted so that the desired emitter work function can be achieved, and the interelectrode spacing is selected so that manufacturing restrictions can be accommodated. The product of practical values

of these two parameters is generally larger than 20 mil-torr, i.e., larger than the value required for minimum voltage loss (Figs. 14.33 to 14.35). Therefore it is desirable to find means for reducing the cesium pressure required to achieve a given emitter work function. The most promising techniques for reducing the required cesium pressure are those of improving the crystallographic orientation of surfaces and using additives such as oxygen.

14.3.5 Correlation of J–V Characteristics with Parameters β and pd. The basic phenomena in thermionic conversion can be conveniently divided into two classes, emission phenomena and transport phenomena. Two parameters that may be significant in characterizing these phenomena are β, the ion-richness ratio of the emitter surface, and pd, the product of cesium pressure and interelectrode spacing. In this section we examine the correlation of $J-V$ characteristics with these parameters under conditions of low and high collector temperature.

Figure 14.38 shows a number of $J-V$ characteristics for $\beta < 10$ and $pd < 100$ mil-torr along with the corresponding ideal characteristics. Currents are normalized with respect to the emitter-electron-saturation current. At high-β low-pd conditions (e.g., $\beta = 10$ and $pd = 0.5$ mil-torr), nearly ideal characteristics are observed since scattering losses are negligible and, until ignition occurs, sufficient ions are present for space-charge neutralization. As β is decreased and pd is increased, the unignited-mode characteristics deviate more and more from ideal. On the other hand, the behavior of the ignited-mode characteristic is much richer in structure than that of the unignited mode. For example, at $\beta = 0.1$ and low values of pd, the output current can exceed the electron saturation current significantly, probably because of Schottky and ion-current effects. As pd increases, however, the output current decreases and becomes considerably smaller than ideal for all practical values of electrode voltage. In addition, the slope of the ignited-mode characteristic increases. Finally, the voltage loss V_{dm} decreases for $0.5 < pd < 10$ mil-torr, reaches a minimum for $10 < pd < 20$ mil-torr, and increases for $pd > 20$.

Figure 14.39 shows $J-V$ characteristics for $\beta < 10$, $pd < 100$ mil-torr, and several values of interelectrode spacing. We see from this figure that, although β and pd are useful parameters for data correlations, they do not specify uniquely the characteristics. For example, for $\beta = 1$, $pd = 5$ mil-torr, and small interelectrode spacings, the output current is smaller than ideal and the pseudosaturation region is not observed. On the other hand, for $\beta = 1$, $pd = 5$ mil-torr, and large interelectrode spacings, the output current can exceed the electron saturation current and the pseudosaturation region is observed. Again, for a given value of pd, the voltage loss decreases as d increases and reaches a minimum value of 0.3 volt for $10 < pd < 20$ mil-torr.

Figure 14.40 shows $J-V$ characteristics for $\beta < 0.5$, $pd < 100$ mil-torr, and a collector temperature $T_C = 1000°K$. The pseudosaturation region is not included

in these data. For $10 < pd < 40$ mil-torr, the output current is smaller than ideal throughout the power quadrant. In addition, the minimum value of the voltage loss V_{dm} is about 0.5 volt as compared with 0.3 volt observed under low-collector temperature conditions.

In summary, the parameters β and pd give an adequate description of $J-V$ characteristics in the unignited mode. In the ignited mode, however, significant variations are observed as interelectrode spacing is varied and β and pd are held constant. The variations are observed in the pseudosaturation region, in the comparison of output current with electron saturation current, and in the voltage loss.

For β and pd constant and d varying, other parameters, such as emitter temperature and saturation current, are indirectly varied. Therefore an apparent dependence on interelectrode spacing may in fact be caused by emitter-temperature or current-level changes. This is particularly true for the preignition and the high-output-current regions. On the other hand, the dependence of voltage loss on interelectrode spacing (or pressure) at constant pd does not appear to be caused by emitter temperature and current level.

The data in Fig. 14.36 show that the voltage loss is independent of β. Once ignition occurs, the number of ions generated at the emitter is negligible compared with that in the interelectrode space. It is not surprising, therefore, that β is not a controlling factor in the ignited mode.

A useful overall correlating parameter of $J-V$ characteristics is the quantity $pd/\sqrt{\beta}$. The lines of short dashes in Fig. 14.38 correspond to approximately constant values of $pd/\sqrt{\beta}$ and show the similarity in the general features of $J-V$ characteristics crossing each line.

14.4 Use of Electrical Characteristics in Design and Diagnostics

14.4.1 Optimization of Converter Variables. For most applications the practical ranges of converter variables are:

$$T_E = 1500 \text{ to } 2100^\circ\text{K} \qquad T_R = 500 \text{ to } 700^\circ\text{K}$$

$$T_C = 800 \text{ to } 1200^\circ\text{K} \qquad d = 1 \text{ to } 20 \text{ mils}$$

At a given emitter temperature, maximum output power and maximum efficiency are achieved at optimum values of T_C, T_R, and d. Generally, the higher the emitter temperature, the higher is the performance of the converter. The materials problems become more severe at the higher emitter temperatures, however, and a compromise must be made.

The emitter temperature is often dictated by the application, and then it is interesting to determine the optimized performance of the converter for a given

(This section continues on page 554.)

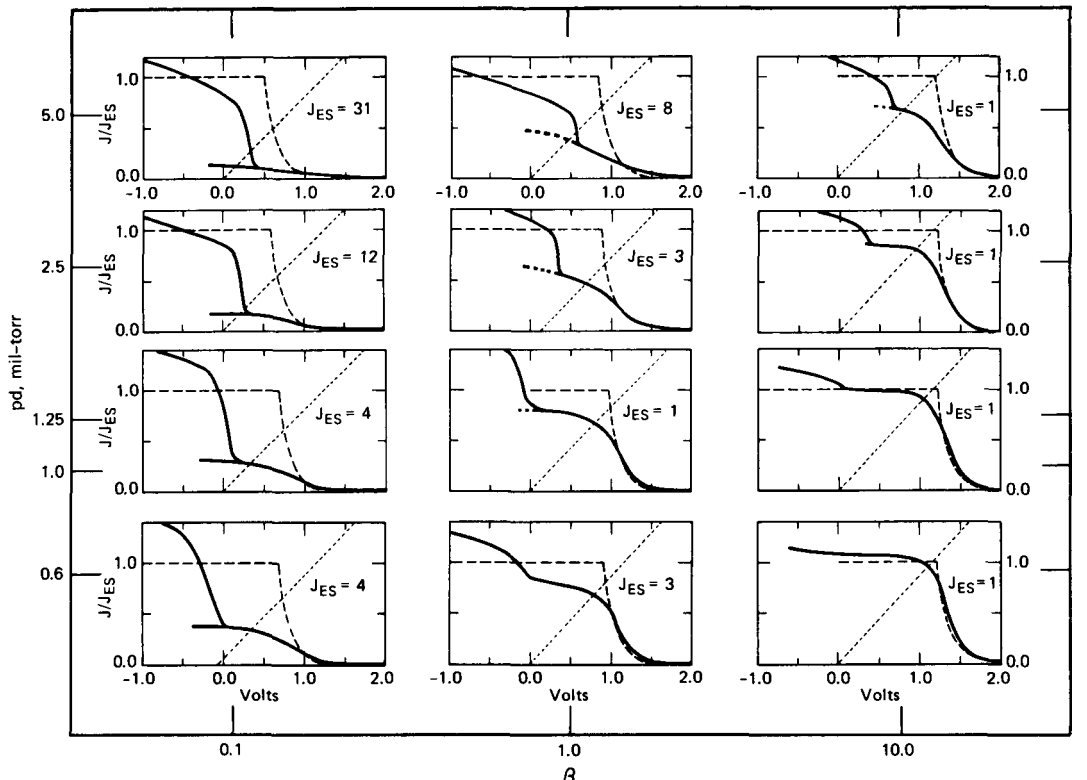


Fig. 14.38a $J-V$ characteristics for various values of β and pd at T_C = 700°K. Output current densities normalized with respect to emitter-electron-saturation-current density J_{ES} (amp/cm²). ——, measured characteristic. – – – –, ideal characteristic. – – – –, locus of constant $(pd/\sqrt{\beta})$ values.

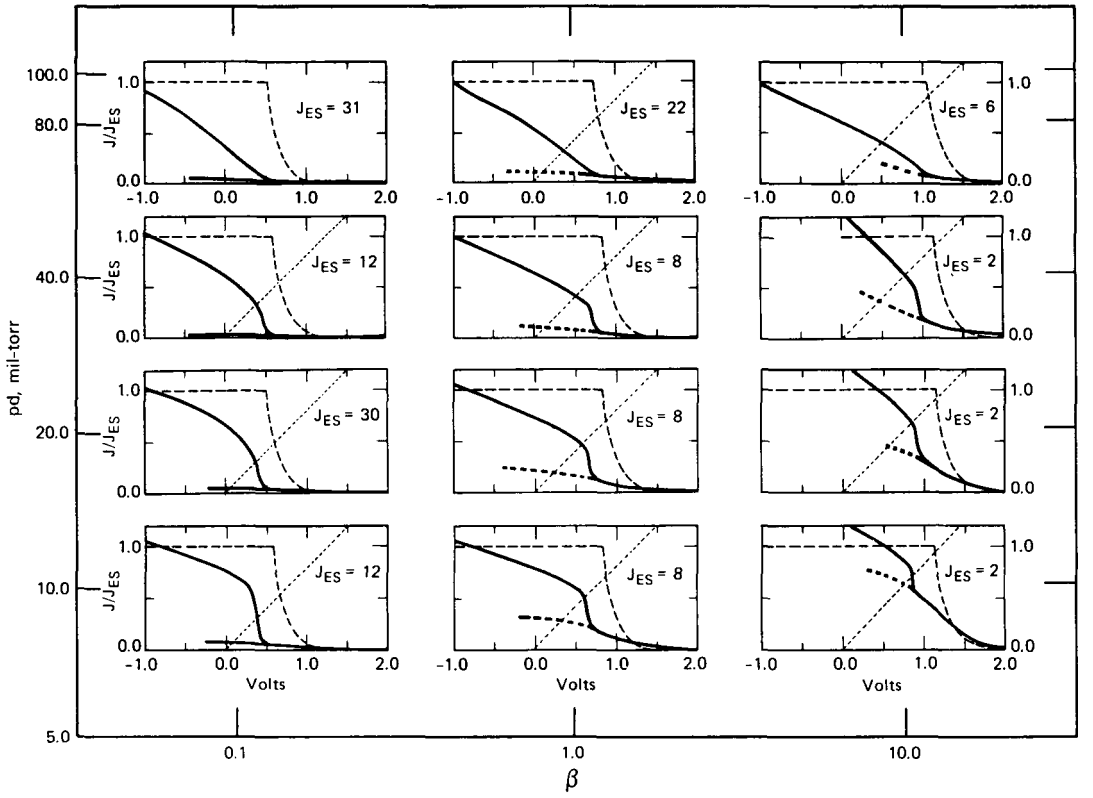


Fig. 14.38b J–V characteristics for various values of β and pd at T_C = 700°K. Output current densities normalized with respect to emitter-electron-saturation-current density J_{ES} (amp/cm²). ——, measured characteristic. – – – –, ideal characteristic. – – – –, locus of constant ($pd/\sqrt{\beta}$) values.

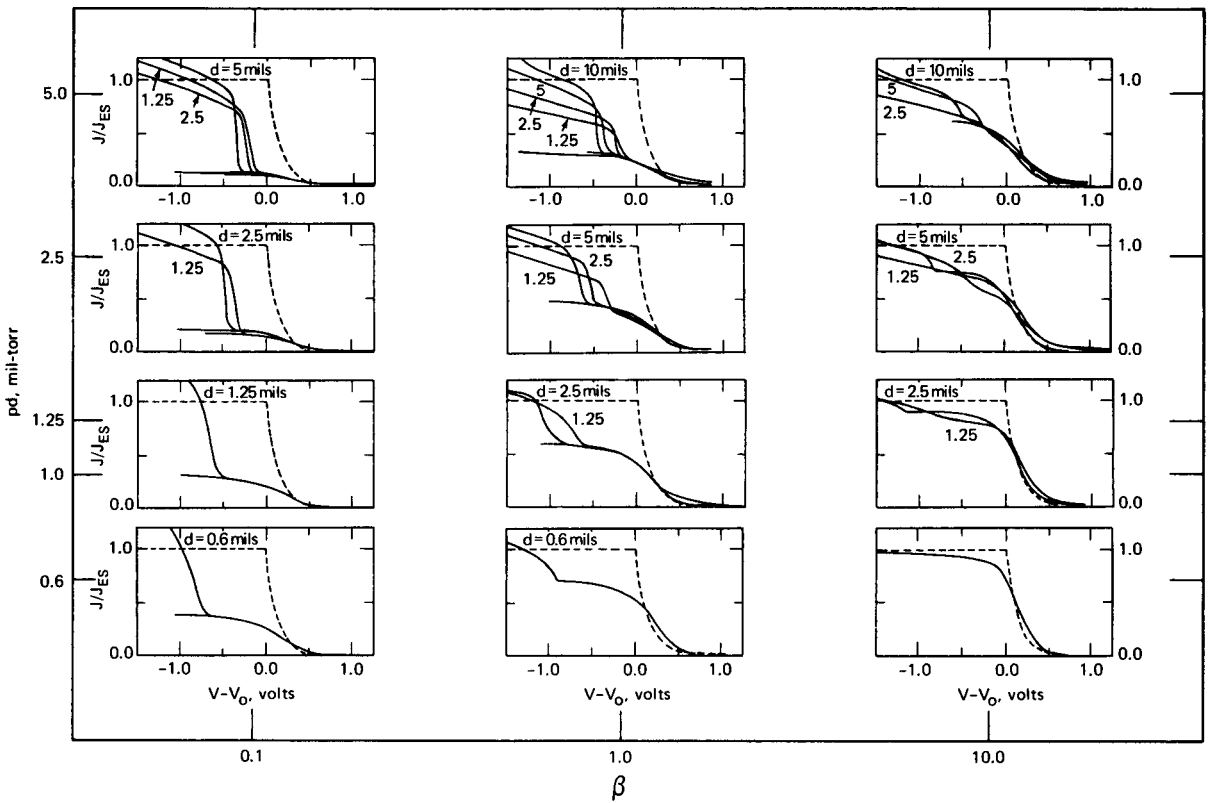


Fig. 14.39a $J-V$ characteristics for various values of β, pd, and d at $T_C = 700°K$. Output current densities normalized with respect to emitter-electron-saturation-current density J_{ES} (amp/cm^2). ——, measured characteristic. – – – –, ideal characteristic.

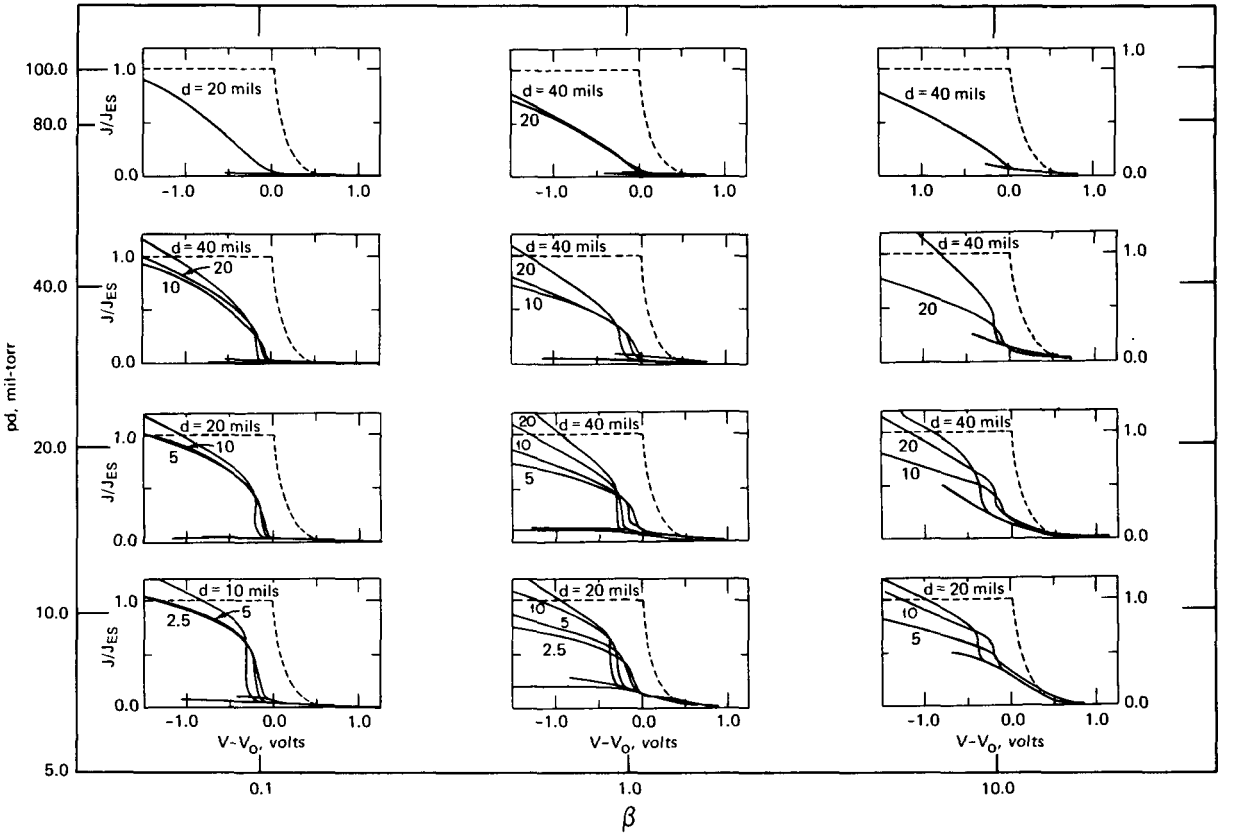


Fig. 14.39b J–V characteristics for various values of β, pd, and d at $T_C = 700°K$. Output current densities normalized with respect to emitter-electron-saturation-current density J_{ES} (amp/cm^2). ——, measured characteristic. – – – –, ideal characteristic.

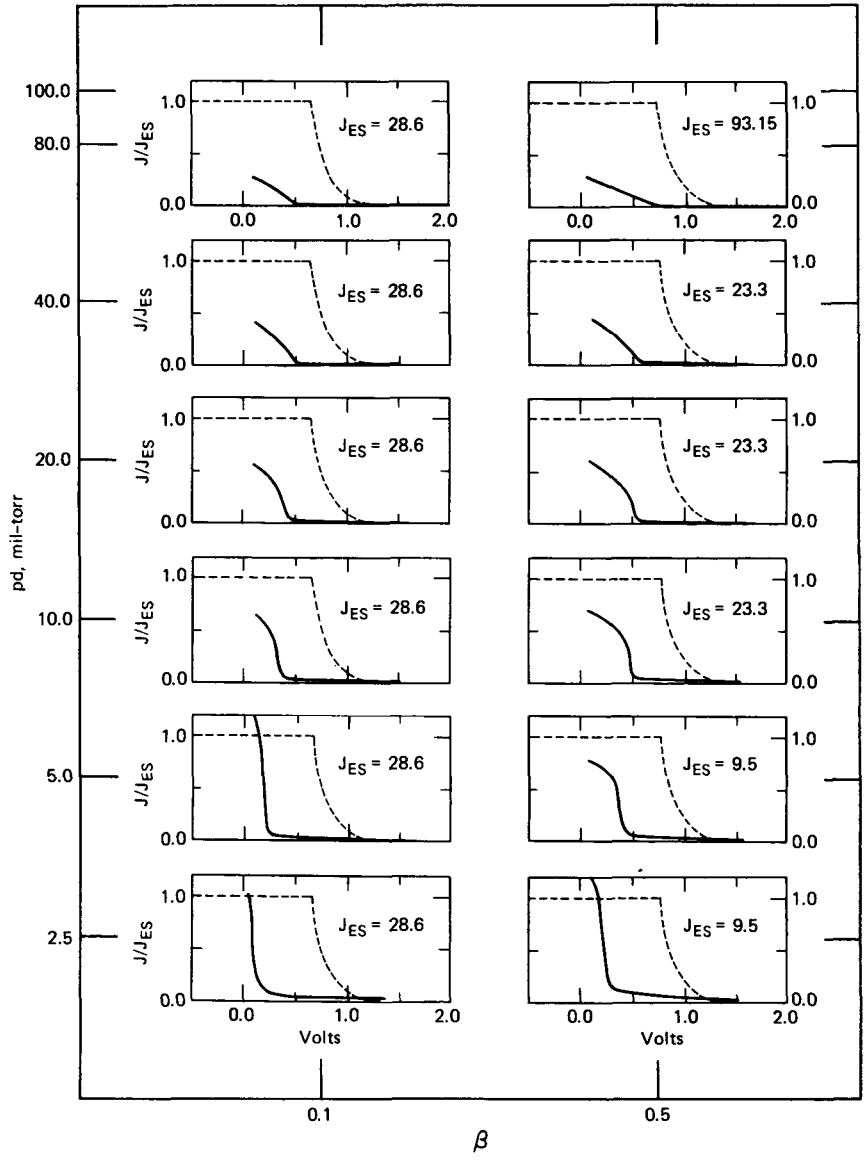


Fig. 14.40a $J-V$ characteristics for various values of β and pd at $T_C = 1000°K$. Output current densities normalized with respect to emitter-electron-saturation-current density J_{ES} (amp/cm^2). ——, measured characteristic. - - - -, ideal characteristic.

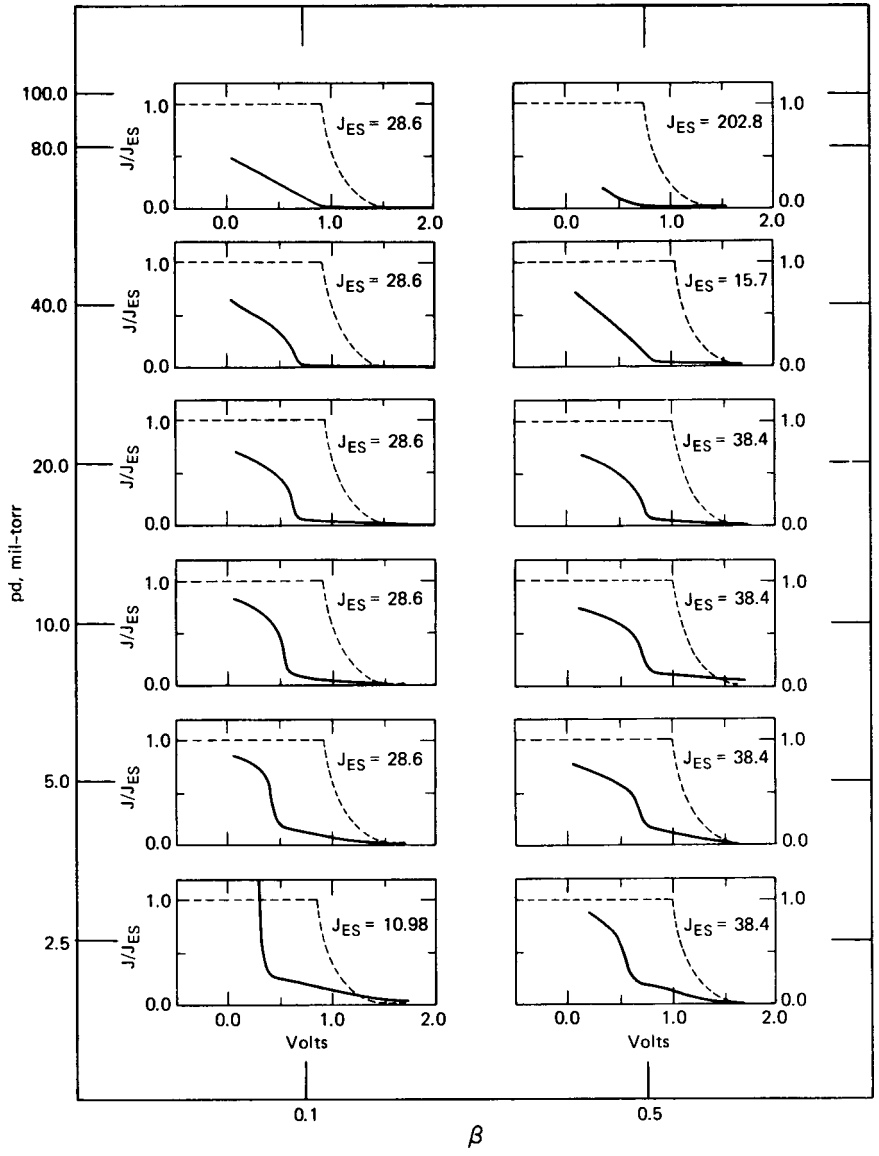


Fig. 14.40b J–V characteristics for various values of β and pd at T_C = 1000°K. Output current densities normalized with respect to emitter-electron-saturation-current density J_{ES} (amp/cm^2). ——, measured characteristic. - - - -, ideal characteristic.

value of emitter temperature. To this end, we (1) adjust the cesium-reservoir temperature to obtain optimum emitter work function; (2) adjust the collector temperature to obtain optimum collector work function; and (3) adjust the interelectrode spacing to optimize the transport phenomena, subject to manufacturing considerations and restrictions inherent to the application.

As an example, we will discuss the optimization of performance of a pair of electrodes for an emitter temperature of 1800°K. First, the collector temperature and the interelectrode spacing are set at some moderate values, e.g., T_C = 900°K and d = 10 mils. Then the $J-V$ characteristics are recorded for several cesium-reservoir temperatures. The data[2] in Fig. 14.41 show that the $J-V$ characteristics corresponding to the various cesium-reservoir temperatures have

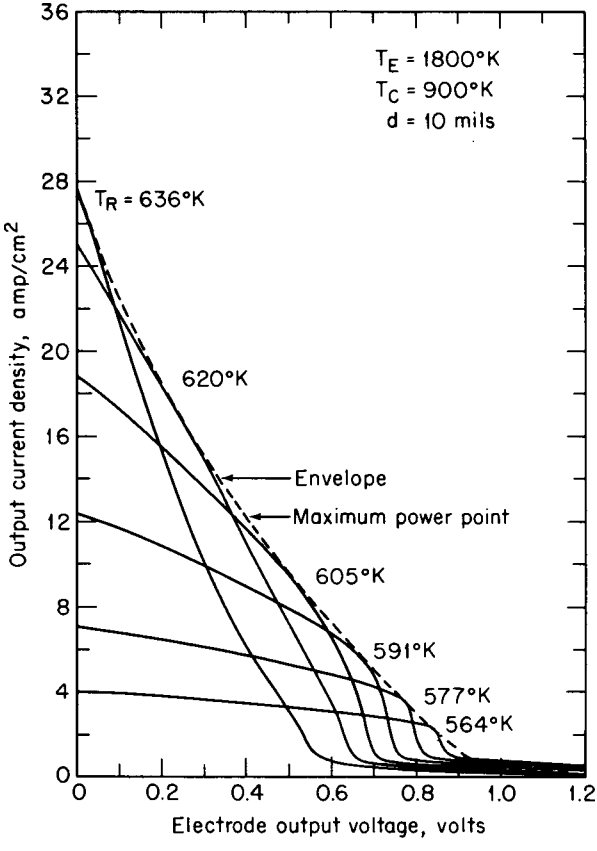


Fig. 14.41 Optimization of cesium-reservoir temperature at T_C = 900°K and d = 10 mils.

an envelope that is tangent to each $J-V$ curve. The envelope represents the cesium-optimized current characteristic for the given values of T_E, T_C, and d.

Second, the collector-temperature optimization is performed. For a cesium-reservoir temperature near the maximum power point of the envelope (T_R = 605°K in Fig. 14.41) and for T_E = 1800°K, d = 10 mils, and T_R = 605°K, the $J-V$ characteristics are recorded for collector temperatures in the range from 700 to 1050°K. Figure 14.42 shows that the highest output power is achieved at

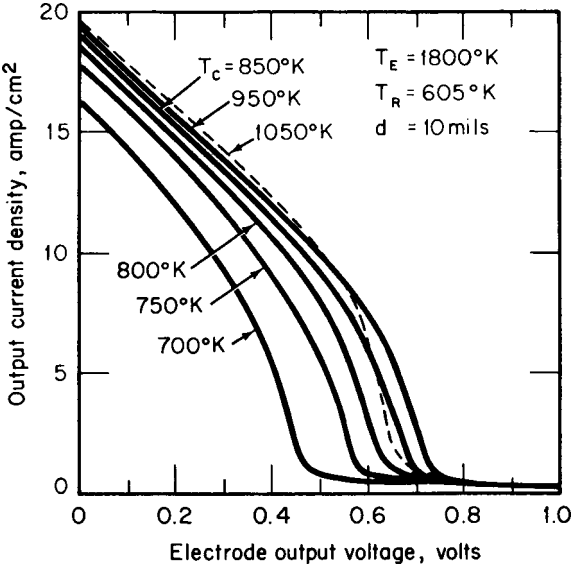


Fig. 14.42 Optimization of collector temperature at T_R = 605°K and d = 10 mils.

T_C = 950°K. Because around optimum performance the output characteristics are not sensitive to the value of the collector temperature, the value T_C = 950°K turns out to be near optimum. Nevertheless, to establish the optimum collector temperature precisely, we have obtained the cesium-optimized current characteristics for T_E = 1800°K, d = 10 mils, and T_C = 900, 950, and 1000°K. Comparing these three optimized characteristics (Fig. 14.43) shows that the collector temperature of 950°K yields the highest output power throughout the range of practical interest. Hence the curve at T_C = 950°K (Fig. 14.43) represents the cesium- and collector-temperature-optimized characteristics at T_E = 1800°K and d = 10 mils.

Third, the interelectrode spacing optimization is performed. For T_E = 1800°K and T_C = 950°K, we obtain cesium-optimized current characteristics for several

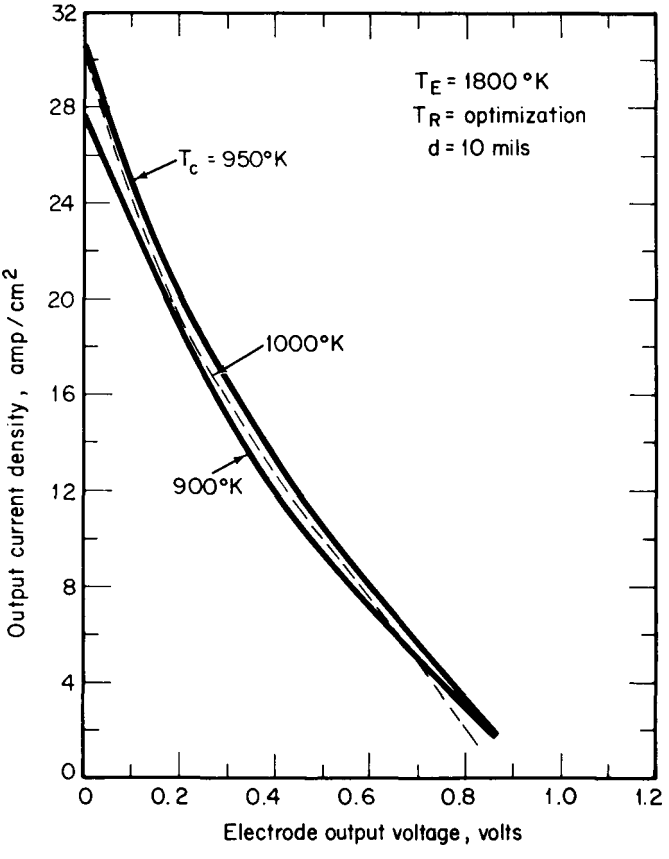


Fig. 14.43 *Cesium-optimized characteristics for various collector temperatures.*

interelectrode spacings. These characteristics are shown in Fig. 14.44. Finally, we can establish the fully optimized characteristic with respect to T_R, T_C, and d for T_E = 1800°K by drawing the envelope of the curves in Fig. 14.44.

This optimization procedure can be repeated for emitter temperatures of 1600, 1700, 1900, and 2000°K. The resulting fully optimized characteristics[12] of the pair of electrodes in question are shown in Fig. 14.45.

14.4.2 Performance Comparisons. This section presents performance data from several converters[1,3] (pairs of electrodes). The performance data were obtained by inserting the electrodes in the same type of variable-spacing research converter and by using the same processing, outgassing, cesiating, and temperature-measurement techniques. The pairs of electrodes studied are:

Converter FW–Nb (Ref. 14)
Emitter: (1 0 0) preferred-orientation fluoride tungsten
Collector: Ground bulk niobium

Converter ClW–Mo (Ref. 2)
Emitter: (1 1 0) preferred-orientation chloride tungsten; bare work function = 4.78 ev
Collector: Ground bulk molybdenum

*Converter ClW–Mo**(Ref. 13)
Emitter: (1 1 0) preferred-orientation chloride tungsten electropolished and heat treated for 20 hr at 2300°K; bare work function = 4.98 ev
Collector: Sublimed molybdenum on a niobium substrate

Converter ClW–Nb (Ref. 13)
Emitter: Same as converter ClW–Mo*
Collector: Ground bulk niobium

Converter (1 1 0) W–Nb (Ref. 15)
Emitter: Nominal (1 1 0) single-crystal tungsten; bare work function = 5.13 ev
Collector: Ground bulk niobium

Converter (1 1 0) W–Mo–O_2
Emitter: Nominal (1 1 0) single-crystal tungsten
Collector: Ground bulk molybdenum
Additive: Oxygen

Figure 14.46 compares the performance of three groups of these converters at an interelectrode spacing of 10 mils and a current density of 10 amp/cm^2. The lowest curve represents the performance of fluoride tungsten emitters.[14] The middle group shows the performance of (1 1 0) tungsten emitters[2,13,15] and represents the present state of the art. The top curve shows performance with oxygen additives and indicates what can be expected in the future.

For the middle group the difference in performance can be expressed as about 0.1-volt difference in output voltage. The emitter with the bare work function of 4.78 ev has the lowest output power; as the bare work function increases, the output power increases.

Although niobium and molybdenum collectors are used, the variations in performance can be explained solely by the variations in emitter bare work function. For example, there is no difference in the performance of the two emitters with similar bare work functions (4.98 ev) although the collectors were niobium and (1 1 0) molybdenum, respectively.

Figures 14.47 and 14.48 show the power-output comparison as a function of interelectrode spacing for emitter temperatures of 1800 and 1900°K. The order of the curves remains the same throughout the interelectrode-spacing range.

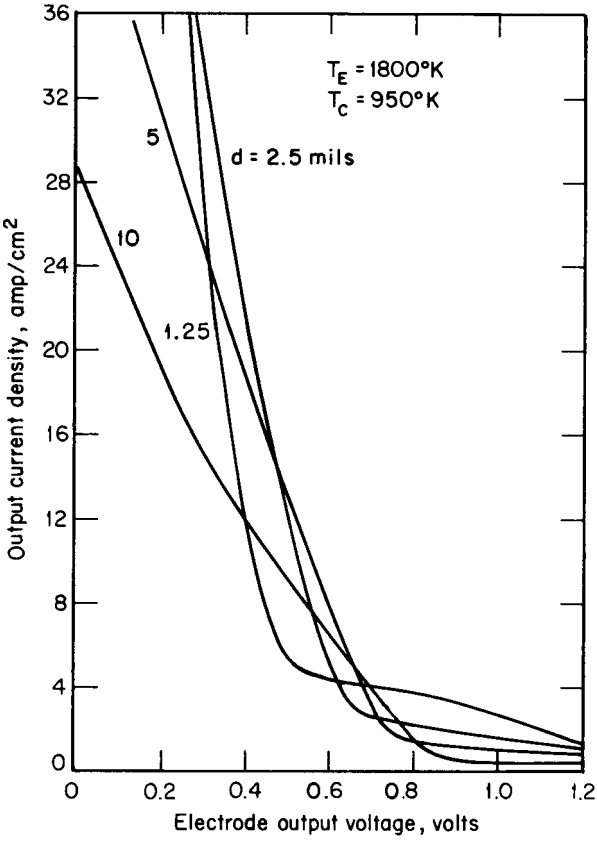


Fig. 14.44 Cesium-optimized characteristics for various interelectrode spacings and T_E = 1800°K.

The black dots on the graphs represent data from a fixed-spacing converter[16] with a different design. This converter was processed and outgassed in a manner similar to that used for the variable-spacing devices. Its emitter is chloride tungsten with a high (1 1 0) preferred orientation and a bare work function of 4.92 ev, and its power output is in close agreement with the CIW emitters having similar work functions.

An additional consistency can be observed in this group of converters. Table 14.1 shows the cesiated emission of the emitters for several cesium-reservoir and emitter temperatures. The higher the bare work function, the higher is the observed cesiated emission (and therefore the higher the output power is). The qualitative correlation of bare work function and cesiated emission observed

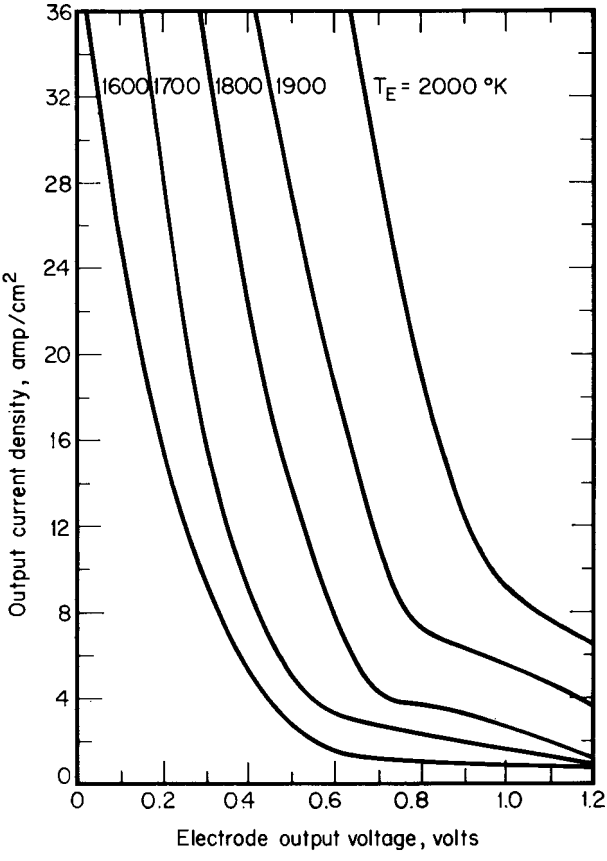


Fig. 14.45 Fully optimized characteristics.

with these groups of (1 1 0) preferred-orientation tungsten emitters, however, may not hold for other types of emitters. McCandless and Wilkins[17] have shown that the bare-work-function correlation does not hold for the single-crystal (1 1 0), (1 1 1), and (1 1 2) tungsten (see also Chap. 9).

14.4.3 Converter Diagnostics. Because of practical difficulties certain parameters of converters may not be accessible to direct measurement by standard techniques. For example, the emitter temperature of converters designed for installation in the core of a nuclear reactor is not accessible to direct measurement.

A number of techniques have been developed in which values of parameters are inferred from certain features of $J-V$ characteristics. These techniques are

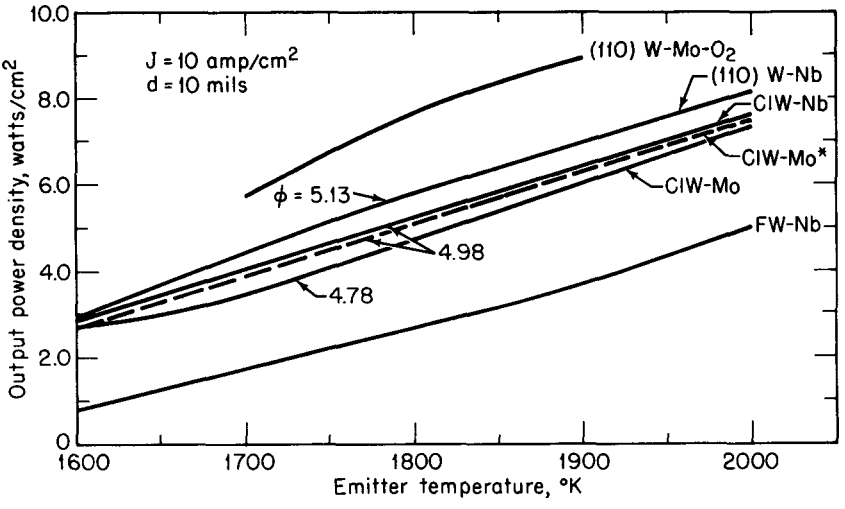


Fig. 14.46 Comparison of performance of several electrode materials at an interelectrode spacing of 10 mils.

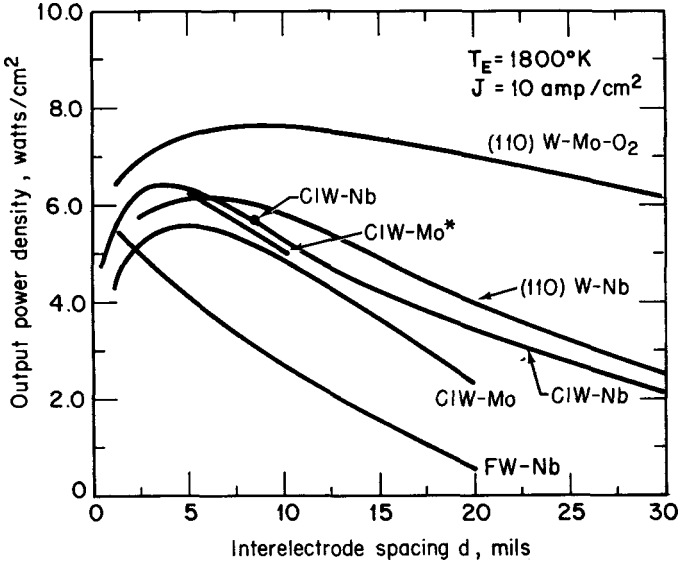


Fig. 14.47 Comparison of performance of several electrode materials at an emitter temperature of 1800°K.

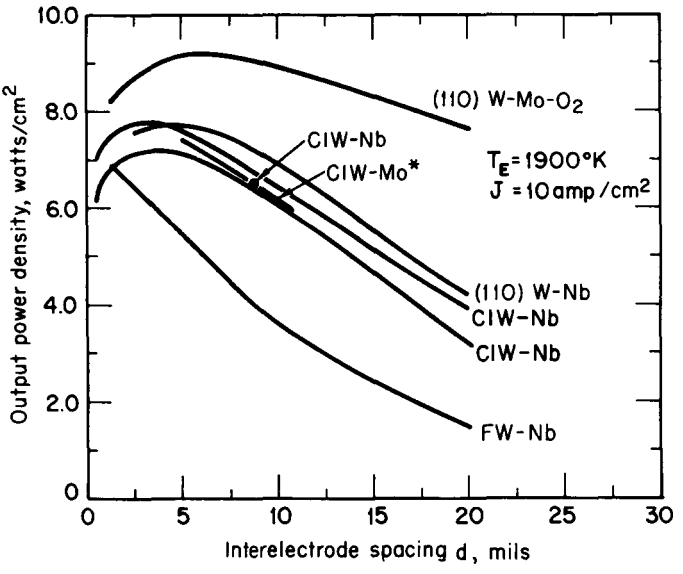


Fig. 14.48 Comparison of performance of several electrode materials at an emitter temperature of 1900°K.

Table 14.1 Comparison of Cesiated Emission and Bare Work Function

Converter	Bare work function, ev	Short-circuit density, amp/cm² T_E = 1600°K T_R = 551°K	T_E = 1800°K T_R = 564°K	T_E = 2000°K T_R = 591°K
(1 1 0) W–Mo–O_2	5.13	12.8	7.6	10.6
ClW–Nb	4.98	7.6	4.8	7.0
ClW–Mo	4.78	6.0	3.4	6.7

referred to as converter diagnostics. In addition, converter diagnostics can be used to infer the causes of abnormal changes in the performance of a converter.

(a) Interelectrode Spacing. A measure of the interelectrode spacing can be inferred from the behavior of the cesium-optimized current characteristics. For example, a decrease in interelectrode spacing can result in an increase of the slope of the cesium-optimized characteristic (Fig. 14.49). In addition, for small current values that represent the unignited mode, the smaller the spacing, the

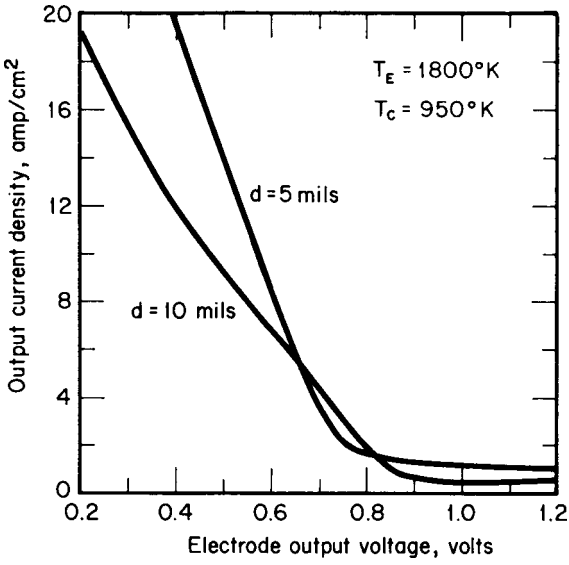


Fig. 14.49 Effect of interelectrode spacing on the cesium-optimized current characteristics.

larger is the output current. The cesium-reservoir temperatures corresponding to a given current level are approximately the same.

Fixed-spacing converters are generally designed to achieve the desired interelectrode spacing when the electrodes are operating at design temperatures. It is possible that the interelectrode spacing will not reach its desired value, however, because of warping of the emitter assembly or peeling of the emitter surface from the back plate. Then converter output characteristics would be different from those expected, and the need to determine the interelectrode spacing after assembly is apparent. One technique is to deduce the spacing from measurements of external dimensions of the converter. These measurements can uncover possible warping of the emitter structure but not variations due to peeling of the emitter surface. Another technique is X-ray radiography. These techniques cannot be carried out under operating conditions, however. Then the quantity of heat conducted through the cesium vapor from emitter to collector can be used to deduce the interelectrode spacing.[18] The relation between heat conducted through the cesium vapor and d is given by Eq. 4.55 and Fig. 4.53, Vol. I, Chap. 4.

The experimental procedure for determining the interelectrode spacing is as follows: The heat-input-rate density q'_{in} required to keep the emitter at constant temperature T_E is recorded as a function of cesium pressure under open-circuit conditions. For cesium pressures less than 1 torr, cesium conduction is

negligible, and the measured value of the heat-input-rate density $(q'_{in})_0$ represents the sum of all energy-rate densities to and from the emitter except for conduction q'_v through the cesium vapor. It follows that q'_v is given by the relation

$$q'_v = q'_{in} - (q_{in})_0 \tag{14.3}$$

Knowing q'_v, we can determine d from Fig. 4.53, Vol. I, Chap. 4.

(b) Emitter. Changes in the work function of an emitter (changes in emission characteristics) can occur because of prolonged operation at high temperature or because of exposure to gases. An increase in the cesiated emitter work function results in effects similar to those caused by an increase in the interelectrode spacing, except that the cesium temperature for a given output current increases instead of staying approximately constant.

(c) Emitter Temperature. An increase in emitter temperature results in: (1) increase in output voltage; (2) decrease in the slope of the cesium-optimized current characteristics (Fig. 14.50); (3) higher cesium temperatures for a given output current; and (4) significantly larger output currents in the unignited mode.

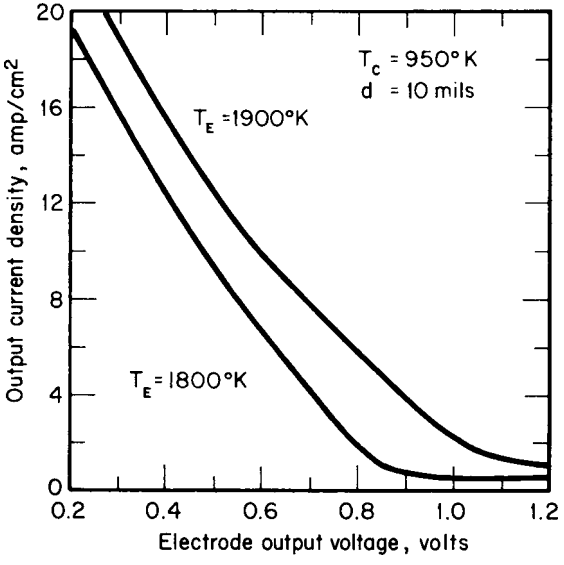


Fig. 14.50 Effect of emitter temperature on cesium-optimized current characteristics.

Measuring the pseudosaturation current density in the unignited mode has been suggested[6,19,20] as a means of determining the emitter temperature. Theoretical arguments[19-21] show that under electron-rich conditions this current density is a function of emitter temperature, interelectrode spacing, and cesium pressure only and is independent of electrode work functions.

Typical data showing the effects of spacing and emitter temperature[6] are shown in Figs. 14.51 and 14.52. Because the pseudosaturation current density is not always clearly observed, the data are correlated by using the value of output current density immediately prior to ignition, the so-called ignition current density J_{ig}.

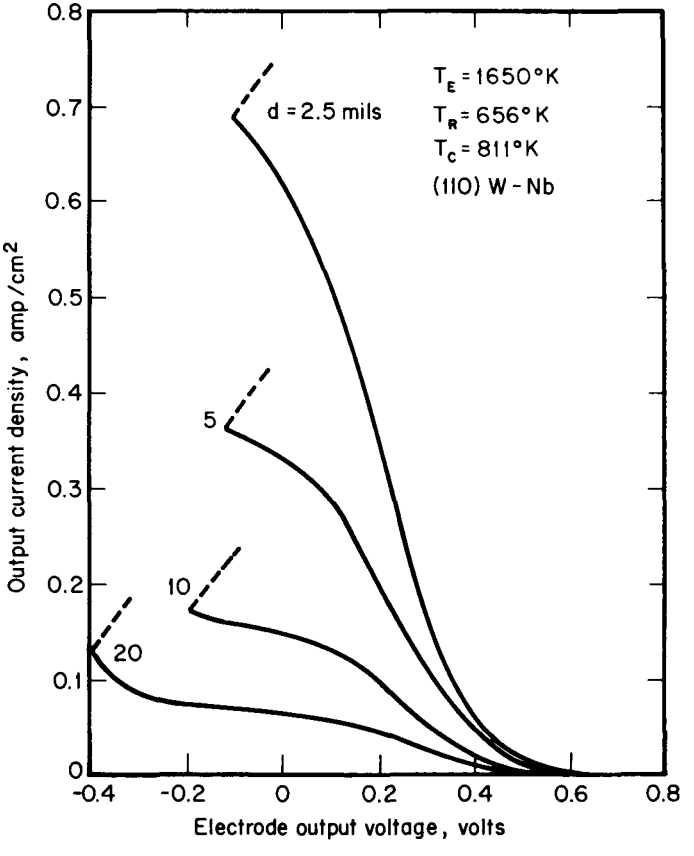


Fig. 14.51 Effect of interelectrode spacing on unignited mode.

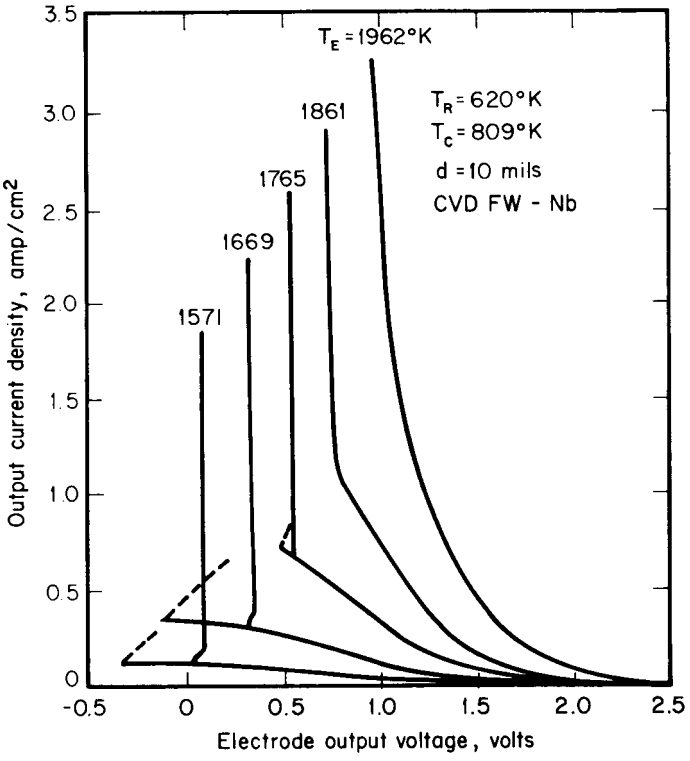


Fig. 14.52 Effect of emitter temperature on unignited mode.

A graph of the inverse ignition current density vs. interelectrode spacing at a reservoir temperature of 620°K is shown in Fig. 14.53. A linear relation is observed in accord with theory.[19,20]

Graphs of ignition current density vs. collector temperature for interelectrode spacings of 10 and 20 mils are shown in Fig. 14.54. We see from this figure that collector temperature and, therefore, collector work function do not affect ignition current.

Graphs of ignition current density vs. emitter temperature for several values of interelectrode spacing and a cesium-reservoir temperature of 620°K are shown in Fig. 14.55. Except for the smallest spacing, the ignition current depends linearly on emitter temperature; this dependence is in contrast with the theoretical expressions that show an exponential relation.[19-21] The discrepancy is probably due to the fact that the theory describes the relation between pseudosaturation current and T_E and not that between ignition current and T_E. At high emitter

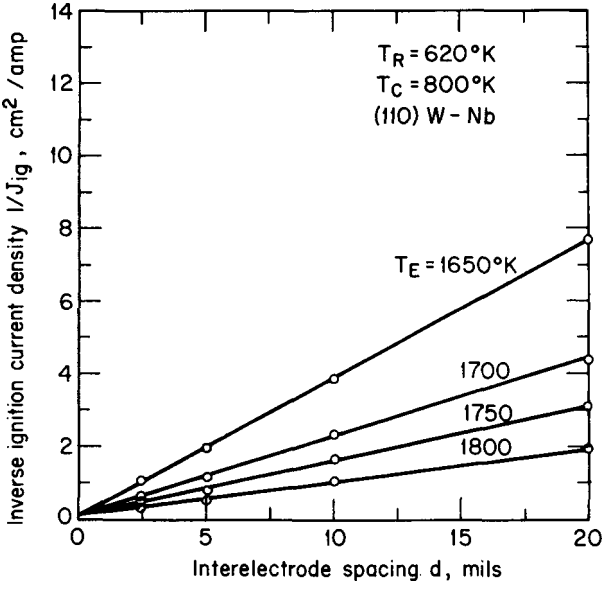


Fig. 14.53 Dependence of ignition current density on interelectrode spacing.

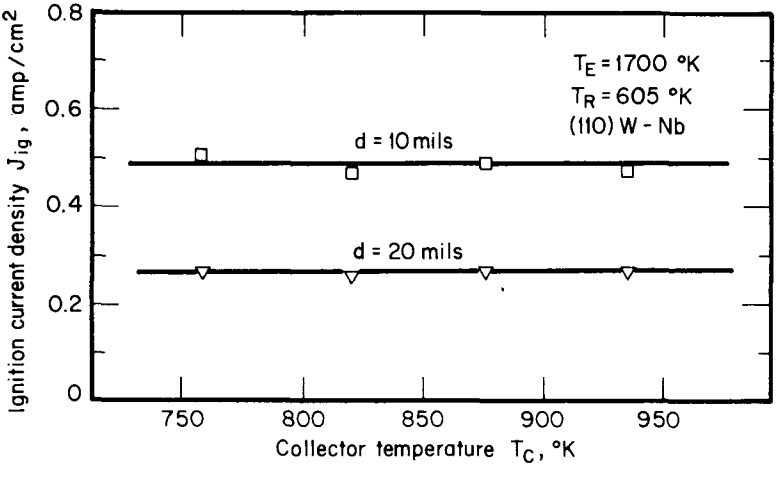


Fig. 14.54 Data illustrating that ignition current density is not dependent on collector temperature.

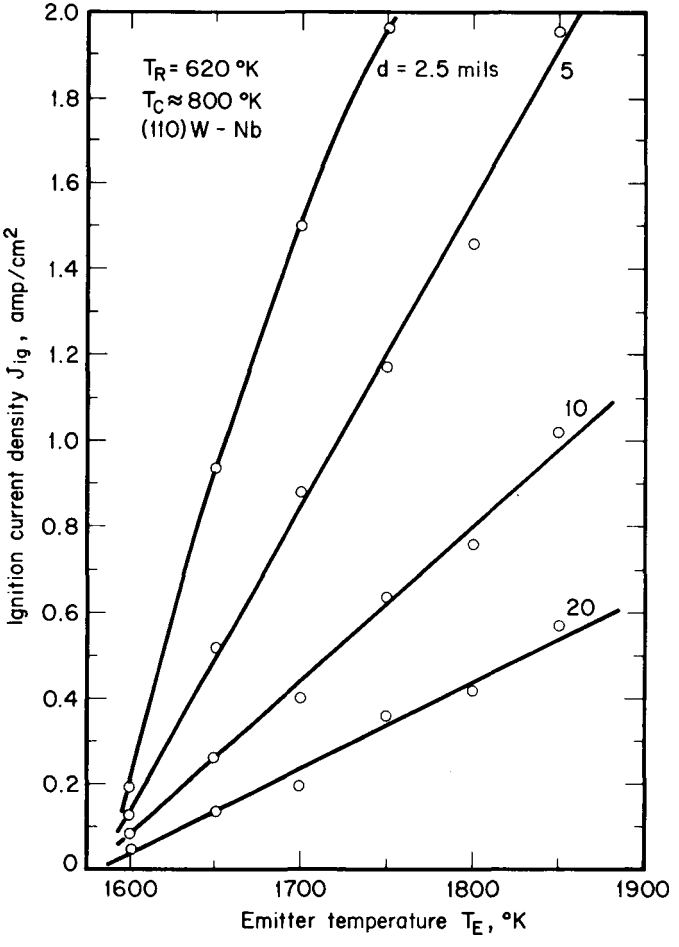


Fig. 14.55 Summary of ignition current data at a reservoir temperature of 620°K.

temperatures, however, the pseudosaturation current is not observed (Fig. 14.52).

Graphs of ignition current density vs. emitter temperature for interelectrode spacings of 10 and 20 mils and emitters with considerably different work functions are shown in Fig. 14.56. We see from this figure that the ignition current density under electron-rich conditions is independent of emitter work function in accordance with theory and that the data are in agreement within 25°K in emitter temperature.

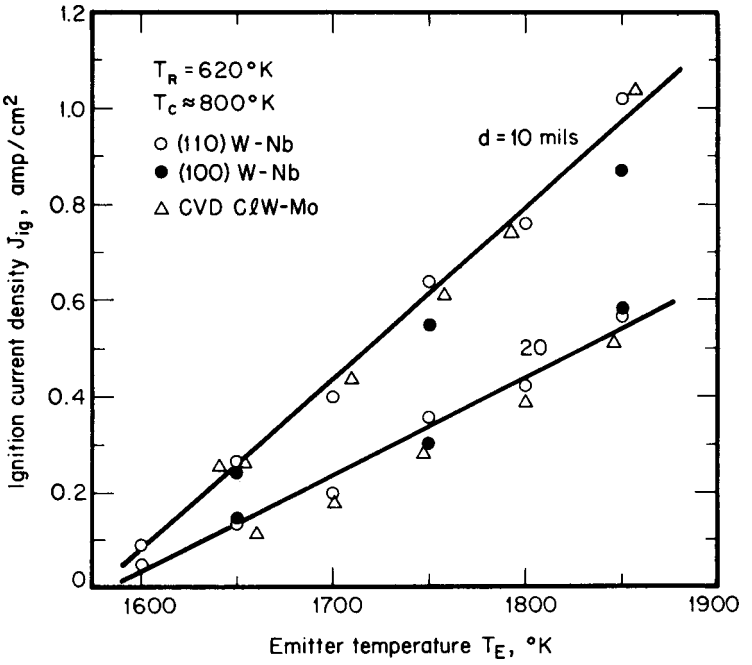


Fig. 14.56 Ignition current density vs. emitter temperature for three emitters with different work functions.

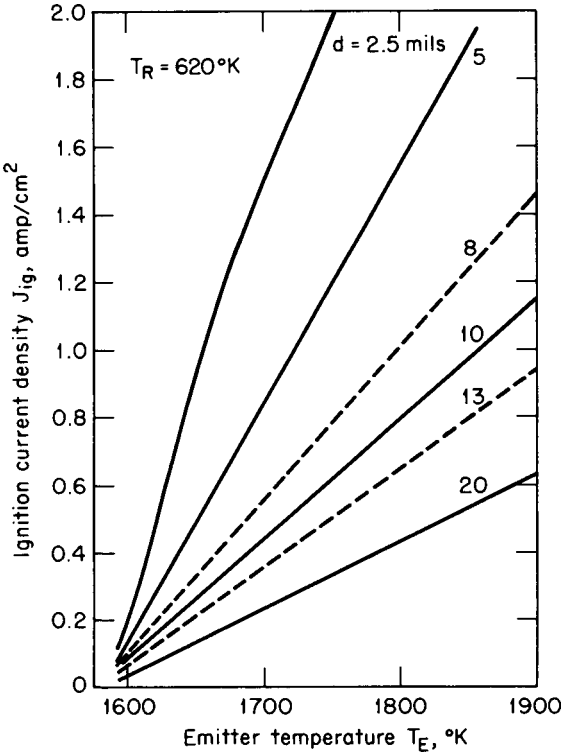


Fig. 14.57 Effect of interelectrode spacing on correlation of ignition current density with emitter temperature.

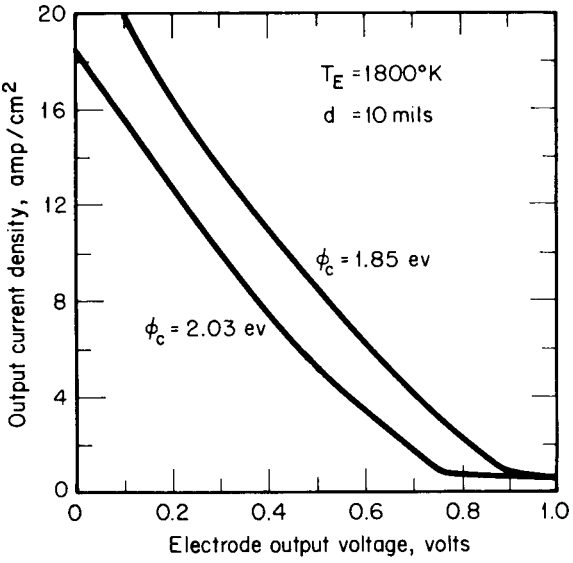


Fig. 14.58 Effect of collector work function on cesium-optimized characteristics.

Additional graphs are shown in Fig. 14.57 to illustrate the error that can occur in emitter-temperature determination if the interelectrode spacing is not known accurately. An error of 2 mils in the value of an interelectrode spacing of 10 mils results in an error of 50°K in emitter temperature.

(d) Collector. An increase in collector work function results in translation of the $J-V$ characteristics or the cesium-optimized current characteristic (Fig. 14.58). The amount of translation along the voltage ordinate corresponds approximately to the change in collector work function.

References

1. D. Koenig and T. Pigford, Work-Function Depression and Surface Ionization for Planar (1 0 0) Tungsten Exposed to Cesium, in *Proceedings of Twenty-Sixth Annual Conference on Physical Electronics, Cambridge, Mass., Mar. 21, 1966,* pp. 233-243, Massachusetts Institute of Technology, 1967.
2. F. Rufeh, D. Lieb, and L. van Someren, Applied Thermionic Research, Report TE-4092/3-108-69, Thermo Electron Corp., 1969.
3. F. Rufeh, B. Gunther, and D. Lieb, Collector Work Function Measurements, in *Proceedings of the Thermionic Conversion Specialists Conference, Miami Beach, Fla., Oct. 26–29, 1970,* pp. 233-242, Institute of Electrical and Electronics Engineers, New York, 1970.

4. F. Rufeh and D. Lieb, The Output Characteristics of a Vapor-Deposited Tungsten Emitter and a Molybdenum Collector, Report TEE-3874-2, Thermo Electron Corp., June 1969.
5. F. Rufeh and D. Lieb, Emission Characteristics of a Duplex Vapor-Deposited Tungsten Emitter, in *Proceedings of the Thermionic Conversion Specialists Conference, Boston, Mass., Oct. 21–24, 1968,* pp. 141-150, Institute of Electrical and Electronics Engineers, New York, 1968.
6. B. Gunther, Emitter Temperature Dependence of Lower Mode Current, in *Proceedings of the Thermionic Conversion Specialists Conference, Miami Beach, Fla., Oct. 26–29, 1970,* pp. 446-452, Institute of Electrical and Electronics Engineers, New York, 1970.
7. F. Rufeh and D. Lieb, The Dependence of the Volt-Ampere Characteristics on Collector Temperature, in *Proceedings of the Thermionic Conversion Specialists Conference, Carmel, Calif., Oct. 21–23, 1969,* pp. 237-246, Institute of Electrical and Electronics Engineers, New York, 1969.
8. N. Rasor and C. Warner, Correlation of Emission Processes for Adsorbed Alkali Films on Metal Surfaces, *J. Appl. Phys.*, **35**(9): 2589-2600 (1964).
9. G. Hatsopoulos and F. Rufeh, Thermodynamic Correlation of Work Function, *Energy Convers.*, **10**: 135-138 (1970).
10. D. Lieb and W. Bornhorst, Analysis of the Diffusion Dominated Thermionic Converter Plasma, in *Proceedings of the Thermionic Conversion Specialists Conference, Palo Alto, Calif., Oct. 30–Nov. 1, 1967,* pp. 191-197, Institute of Electrical and Electronics Engineers, New York, 1967.
11. D. R. Wilkins and E. P. Gyftopoulos, Thermionic Converters Operating in the Ignited Mode. Part I: Theoretical Output Current Characteristics, *J. Appl. Phys.*, **37**(7): 2888-2891 (1966).
12. S. Kitrilakis and J. Weinstein, Second Annual Technical Summary Report for the Thermionic Emitter Materials Research Program, Report 27-64, Thermo Electron Corp., 1964.
13. D. Lieb and F. Rufeh, Thermionic Performance of CVD Tungsten Emitters with Several Collector Materials, in *Proceedings of the Thermionic Conversion Specialists Conference, Miami Beach, Fla., Oct. 26–29, 1970,* pp. 471-480, Institute of Electrical and Electronics Engineers, New York, 1970.
14. F. Rufeh and D. Lieb, Thermionic Performance of Fluoride CVD Tungsten–Niobium Converter, in *Proceedings of the Thermionic Conversion Specialists Conference, Miami Beach, Fla., Oct. 26–29, 1970,* pp. 462-470, Institute of Electrical and Electronics Engineers, New York, 1970.
15. D. Lieb, A. Donaker, and F. Rufeh, Performance of a Thermionic Converter with a Nominal Single-Crystal (1 1 0) Tungsten Emitter and a Niobium Collector, in *Proceedings of the Thermionic Conversion Specialists Conference, Carmel, Calif., Oct. 21–23, 1969,* pp. 66-75, Institute of Electrical and Electronics Engineers, New York, 1969.
16. C. Wang and J. Ward, Performance of Chloride and Fluoride Vapor-Deposited Tungsten Emitters in Thermionic Converters, in *Proceedings of the Thermionic Conversion Specialists Conference, Boston, Mass., Oct. 21–24, 1968,* pp. 134-140, Institute of Electrical and Electronics Engineers, New York, 1968.
17. R. G. McCandless and D. R. Wilkins, Theory of Thermionic Converter Surface Phenomena, in *Proceedings of the Thermionic Conversion Specialists Conference, Miami Beach, Fla., Oct. 26–29, 1970,* pp. 210-216, Institute of Electrical and Electronics Engineers, New York, 1970.

18. S. Kitrilakis and M. Meeker, Experimental Determination of the Heat Conduction of Cesium Gas, *Advan. Energy Convers.*, 3: 59-68 (1963).
19. D. R. Wilkins and E. P. Gyftopoulos, Theory of Thermionic Converter Extinguished-Mode Operation with Applications to Converter Diagnostics, in *Proceedings of Twenty-Sixth Annual Conference on Physical Electronics, Cambridge, Mass., Mar. 21, 1966*, pp. 40-52, Massachusetts Institute of Technology, 1967.
20. D. R. Wilkins, Massachusetts Institute of Technology, personal communication.
21. C. Warner and L. K. Hansen, Transport Effects in the Unignited Mode of Thermionic Diodes, Report on the Twenty-Third Annual Conference on Physical Electronics, pp. 400-412, Massachusetts Institute of Technology, 1963.

15

Metallurgy of Electrode Materials

15.1 General Remarks

This chapter describes some basic elements of physical metallurgy which are pertinent to the study and preparation of electrode surfaces. Some terms used in describing perfect crystals are defined, and imperfections found in real metals, properties of refractory-metal electrodes, and metallurgical properties and methods of preparation of electrode surfaces are discussed.

15.2 Structure of Perfect Crystals

A perfect crystal of metal consists of a regular array or lattice of positive-ion cores permeated by a cloud or "gas" of electrons. The electron cloud consists of the valence electrons only, and in pure metals all the ion cores are identical. The electrons in a perfect crystal of a metal are dispersed throughout the volume of the crystal; i.e., they do not have properties associated with any particular location in the crystal. In ionic and covalent crystals, the valence electrons are not completely dispersed. The difference between metal crystals and either ionic or covalent crystals is not hard and fast (see also Chap. 5, Sec. 5.10.8). Nevertheless, it is helpful in clarifying what is meant by a perfect crystal of a metal.

At temperature T the ion cores in a perfect crystal occupy equilibrium positions.* The equilibrium positions correspond to the minimum Gibbs free energy per atom at temperature T. Deviations of ion cores from their equilibrium positions lead to a nonequilibrium condition that can be described by a Gibbs free energy larger than that of the equilibrium condition.

The ion cores in metal crystals are found experimentally to arrange themselves in patterns that can be approximated by the patterns of a set of closely packed hard spheres. In general, the crystal can be divided into cells. One such division is discussed in Sec. 5.10.1. Another division consists of equal-size, contiguous parallelepipeds defined by lines intersecting at positions occupied by ion cores; i.e., the intersections are lattice points. Each intersection has identical surroundings. The parallelepipeds, called *unit cells*, are the smallest repeating units in the crystal. In general, each unit cell contains more than one ion core or

*Each equilibrium position is described by an expectation value.

atom.* The directions and lengths of the edges of the unit cell define a coordinate system for the crystal.

The simplest and most regular crystal lattice is the *cubic*, which has one atom at each corner of the cube. This lattice has large interstices at the center of each cube, however, and for this reason very few materials crystallize in this crystal form.

Two other crystal lattices related to the cubic lattice are the *body-centered cubic* and *face-centered cubic*. In the body-centered cubic (bcc) lattice, in addition to the atoms at each corner of the cube, an atom occupies the center of each cube. Many refractory metals (e.g., niobium, tantalum, molybdenum, tungsten, and the alkali metals) exhibit this crystal structure. In the *face-centered cubic* (fcc) lattice, in addition to the atoms at each corner of the cube, atoms occupy the center of each face of the cube. Metals such as gold and aluminum exhibit this crystal structure. Each of the cubic crystal lattices can be characterized by three sets of equivalent orthogonal axes. The density of packing in the fcc lattice is larger than in the bcc lattice and that in the bcc lattice is larger than in the cubic lattice.

Another important crystal lattice is the *hexagonal close-packed* (hcp). This lattice consists of layers in each of which atoms are packed together in a triangular net. An example of a metal having this structure is rhenium. Unlike the cubic lattices, the hcp lattice has one unique direction perpendicular to the layers, called the *c-axis*, and three sets of equivalent directions perpendicular to the unique direction, called the *a-axes*. A unit cell of the hcp lattice is shown in Fig. 15.1. It is seen from this figure that the cell is determined by the two lengths a and c. If the crystal consists of hard spheres, the ratio c/a is equal to $(8/3)^{1/2}$. Metals with hcp structures have values of c/a different from the hard-sphere value. The value of c/a for each metal is a unique function of temperature.

A corollary of the presence of distinct directions in the hcp lattice is that many properties of the metal crystal can depend on direction. For example, electrical conductivity, thermal conductivity, and thermal expansion, as well as the corresponding temperature coefficients along the c-axis, have values different from those along the a-axes. For this reason metals with hcp structures are said to be *anisotropic*. Metals with cubic structures are, in general, *isotropic*.†

15.3 Directions, Planes, and Orientations in Crystals

In a unit cell three edges with a common lattice point define *three axes in the*

*An atom means an ion core plus a number of electrons equal to the number of positive charges on the core.

†The elastic properties of cubic metals can vary with direction.

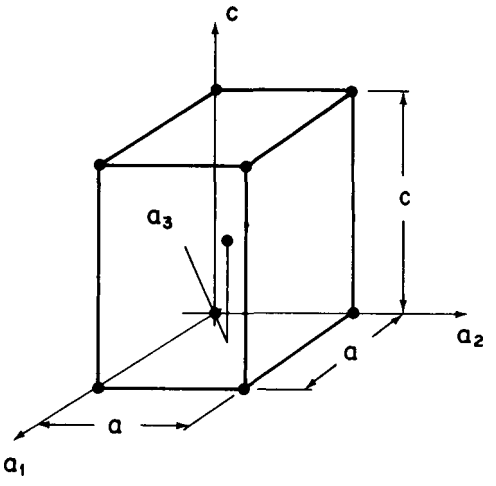


Fig. 15.1 Unit cell of a hexagonal close-packed lattice.

crystal. These three axes are used to describe *directions*, *planes*, and *orientations* in the crystal. In cubic crystals the three axes are mutually orthogonal.

A *direction* with respect to the three axes is defined as a vector from the origin to another lattice point of the crystal. Any direction can always be represented as a vector sum of displacements along the three axes, each displacement being an integer multiple of the side of the unit cell associated with the axis. For this reason directions can be denoted by the three integer multipliers placed inside broken brackets, i.e., ⟨ ⟩. For example, in the unit cell at the origin, the direction of the vector from the origin to the corner farthest distant from it is the ⟨1 1 1⟩ direction.

A (rational) *plane* passing through three lattice points, one on each axis of the crystal, can be determined by three *Miller indices*. The Miller indices of a plane are defined as follows:

We shall assume that a, b, and c are the lengths of the sides of the unit cell and that n_1a, n_2b, and n_3c are the intercepts of the plane with the three axes x, y, and z of the crystal, respectively, where n_1, n_2, and n_3 are integers different from zero and infinity. Moreover, we shall assume that r, s, and t are the highest common factors of n_1 and n_2, n_2 and n_3, and n_3 and n_1, respectively. Then the Miller indices h for x, k for y, and l for z are given by the relations

$$h = \frac{n_2 n_3}{rst} \qquad k = \frac{n_1 n_3}{rst} \qquad l = \frac{n_1 n_2}{rst} \tag{15.1}$$

The plane is denoted by placing the three indices in parentheses.

For a plane parallel to the z-axis, n_3 is equal to infinity and the three Miller indices are given by the relations

$$h = \frac{n_2}{t} \quad k = \frac{n_1}{t} \quad l = 0 \quad \text{for } n_3 = \infty \tag{15.2}$$

The cross-hatched plane shown in Fig. 15.2 intercepts the x- and y-axes at unit displacements from the origin and is parallel to the z-axis. Therefore its Miller indices are 1, 1, and 0, and the plane can be denoted by (1 1 0).

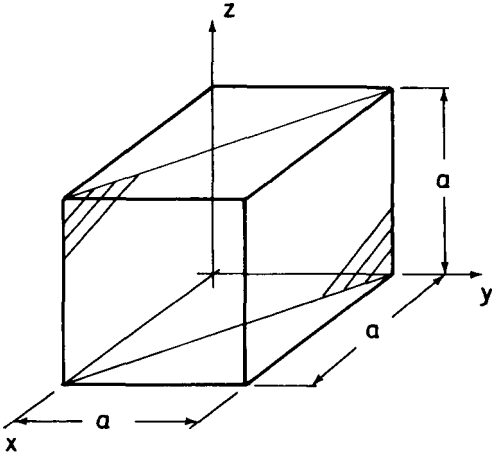


Fig. 15.2 Unit cell of a cubic lattice.

The planes corresponding to (1 1 1), (−1 −1 −1), (1 1 −1), etc., are subsumed to be included in the {1 1 1} class of planes. Examples of three principal classes of planes, {1 1 1}, {1 1 0}, and {1 0 0}, are shown in Fig. 15.3.

Indices for planes in hexagonal lattices are more complicated, and detailed discussions are given in standard texts.[1] For present purposes it suffices to say that for hexagonal structures four *Miller–Bravais* indices are used.

The *orientation* of a grain, crystallite, or homogeneous region of a crystal with respect to a reference grain of the same crystal is defined as the angle between two identical directions, one direction drawn in the grain and the other in the reference grain. Orientation can also be defined in terms of the angle between two identical planes, one in each of the two grains.

15.4 Structure of Real Metals

Perfect crystals are idealized pictures of the crystals of real metals. Deviations

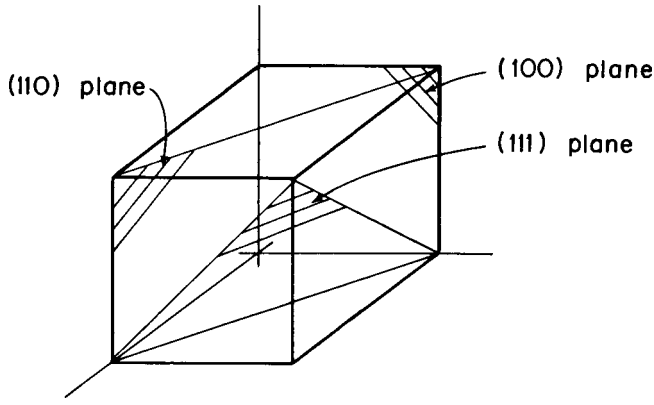


Fig. 15.3 The (1 1 1), (1 1 0), and (1 0 0) planes of cubic lattices.

from ideality occur because of a number of defects such as *impurities, vacancies, grain boundaries,* and *dislocations.* Each of these defects is discussed in this section.

15.4.1 Impurities. Atoms of other elements can be present in the crystal of a metal. The atoms that are unintentionally present are called impurities.* When present in a lattice, small atoms, such as atoms of hydrogen, carbon, nitrogen, or oxygen, are squeezed in between the regularly arrayed metal atoms and are called *interstitial impurities.* On the other hand, impurities with atom sizes comparable with those of the metal atoms appear on sites normally occupied by the metal atoms and are called *substitutional impurities.* Metals with either interstitial or substitutional impurities, or both, are called *solid solutions.* Impurities can cause changes in the physical and mechanical properties of the metal. Such changes depend on the concentration and the nature of the impurities involved.

Usually, as the concentration of impurities is increased, the metal lattice is distorted and strained. At certain high concentrations impurities will form a *precipitate* of a second phase. The second phase has a distinctly different chemical composition and frequently a different crystal structure from the *matrix* of the metal.† In general, the presence of the second phase results in substantial changes in physical and mechanical properties. For example, most alloying for increased strength depends on the precipitation of a second phase.

*Impurities should be differentiated from alloying elements added intentionally to a metal to modify its properties.

†The matrix of a metal is defined as the real crystal composed of metal atoms and defects.

The concentration of atoms of element B in the matrix of element A necessary to cause a second phase to precipitate is called the *solubility limit of B in A*. Its value varies widely (from less than 1 part per million to 50%) between metal pairs. For a given pair the solubility limit varies with temperature in the manner indicated by the phase or equilibrium diagram for the pair.

In a sample of a refractory metal of high purity (by current standards), one interstitial atom for every 10^3 to 10^4 matrix atoms, i.e., one interstitial atom per cube 10 to 20 atoms long, is present. The interstitial atoms can be distributed randomly throughout the matrix or precipitated locally, depending on the past thermal and mechanical history of the sample. As discussed later, each distribution has different effects on the mechanical properties of the metal.

In the sample of a refractory metal just cited, substitutional impurity concentrations are frequently higher than interstitial impurity concentrations. A large fraction of the substitutional impurities, however, consists of atoms from the same group in the periodic table as the metal, i.e., atoms with chemical properties similar to those of the atoms of the metal. For this reason substitutional impurities have relatively little effect on most properties of the metal.

15.4.2 Vacancies. Certain lattice sites of the crystal of a metal are not occupied by ion cores. These sites are called vacancies. The equilibrium concentration of vacancies increases exponentially with temperature. At temperatures T smaller than about one-half the melting temperature T_m on the Kelvin scale, the equilibrium concentration of vacancies is very small. For this reason specimens can be prepared in which there are virtually no vacancies and their effects are negligible.

Vacancies are critically involved in elevated-temperature deformation processes such as creep. In addition, concentrations above the equilibrium concentration can be "frozen in" by quenching a sample from a high to a low temperature. At the low temperature the frozen vacancies are not mobile and can affect the physical and mechanical properties of the sample.

15.4.3 Grain Boundaries. Most samples of a metal of a given lattice structure consist of a large number of small crystals each of which is called a *grain*. Each grain can be characterized by an orientation with respect to an external reference direction, and adjacent grains have different orientations. A grain boundary between adjacent grains is defined as the transition zone in which the change in orientation from one grain to the other occurs. A grain boundary has a thickness equal to a few atomic layers and consists of atoms occupying compromise positions between those dictated by the two adjacent grains. It follows that the atoms of a grain boundary do not occupy equilibrium positions,

and therefore their Gibbs free energy is higher than that of atoms in the bulk of a grain. The difference in the two energies depends on the degree of misorientation between adjacent grains. Because of their higher free energies, grain boundaries are thermodynamically unstable or metastable. Accordingly, sufficient heating of the sample results either in cancellation of boundaries or in movement of boundaries to a free surface. This process, called *grain growth*, proceeds at a diminishing rate and usually leads to a metastable network of boundaries.

The linear dimensions of grains range from those barely visible with the naked eye to a few millimeters. A sample containing many grains is said to be polycrystalline. Distinct grains are not usually visible on finished-metal surfaces, although they are visible on fracture surfaces and on surfaces that have been etched.

Except under conditions of creep failure (plastic flow below the yield stress at high temperature), grain boundaries are not zones of weakness. In fact, a fine-grained material with a large concentration of boundaries is usually stronger than a coarse-grained material.

15.4.4 Dislocations. A dislocation is a line *defect* which, as it moves through a perfect crystal, leaves behind it material that has been displaced.[2] A dislocation can be explained by the following mechanical process. If we cut a crystal composed of layers of atoms in half along a plane perpendicular to the layers, remove one layer from the middle of one half, and reassemble the two halves by connecting the layers as well as possible, one layer is incomplete and terminates in the middle of the crystal. A section of the crystal, perpendicular to the layers and to the plane of the initial cut, appears as shown in Fig. 15.4a. The edge of the incomplete layer is called an *edge-dislocation line*.

The section of the crystal in Fig. 15.4a can be sheared by a force in the plane of the figure and normal to the layers of the crystal. The shearing breaks and re-forms bonds between atoms along lines in the vicinity of the edge dislocation (Fig. 15.4a). Thus the edge of the incomplete layer moves upward. The process can be continued until the edge-dislocation line has reached the upper surface of the crystal (Fig. 15.4b). The term *slip* is used to describe plastic deformation by the movement of dislocations across a plane called the *slip plane*.

Once the dislocation has reached the surface, further shear would be very difficult because it would entail breaking all the bonds in the plane of shear simultaneously, instead of breaking just those to one row of atoms. An analogous situation arises in moving a carpet on a floor. Sliding the whole carpet at once is difficult, but a rumple made at one edge can be propagated quite easily. When the rumple reaches the opposite edge, the whole carpet has been moved.

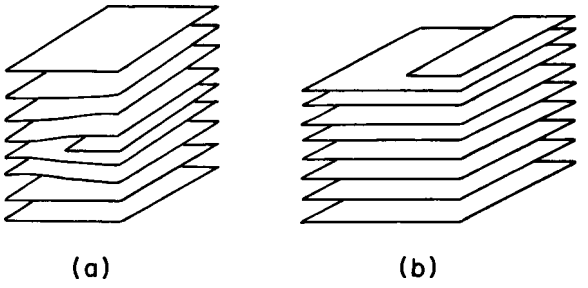


Fig. 15.4 (a) Schematic representation of edge-dislocation line. (b) Position of edge-dislocation line after shearing.

Dislocations of various shapes are present in all ordinary samples of metals. They are very important because they are responsible for the ductility that makes metals so useful.

Movement and proliferation of dislocations results in the plastic deformation of metals. Movement of a dislocation is inhibited by other dislocations. This inhibition is the basis of the phenomenon of *work hardening*. Grain boundaries can sometimes be considered as arrays of dislocations, and they always interfere with dislocation movement. For this reason fine-grained specimens are stronger than coarse-grained specimens. Strengthening mechanisms in alloys are those which interfere with dislocation movement. Dislocation-line densities are typically in the range 10^5 to 10^8 cm^{-2}.

Unlike metal crystals, ionic crystals and ceramics either cannot accommodate dislocations readily or do not allow motion of dislocations, or both. As a result, these materials are brittle and fail without plastic flow.

15.5 Refractory Metals for Electrodes

The materials used in converters are subjected to severe environmental conditions, including high temperatures and chemical attack by cesium. Experience has shown that the most useful electrode materials are among the refractory metals, specifically, niobium and tantalum from Group V, molybdenum and tungsten from Group VI, and rhenium from Group VII. In addition to strength and corrosion resistance, these metals have reasonably low evaporation rates even when used at emitter temperatures. On the other hand, they are difficult to fabricate into the required shapes and are susceptible to attack by residual gases in the converter environment. The following discussion presents some properties of refractory metals relevant to their use as electrodes in a converter.

Metals in Groups V and VI differ in their tolerance for interstitial impurities. For example, the amounts of hydrogen, carbon, nitrogen, and oxygen in niobium and tantalum which lead to precipitation of a second phase are relatively much larger than the corresponding amounts in molybdenum and tungsten. The second phases in niobium and tantalum are extremely stable and do not occur until 10^{-4} to 10^{-3} parts by weight of interstitial impurities have dissolved. The second phases in molybdenum and tungsten are not stable.

By virtue of the preceding remarks, niobium and tantalum components can be fabricated by the conventional techniques used for ductile materials. For example, rolling can be used to form strips, drawing to make wires and tubes, and electron beam welding and brazing to make complicated shapes. Because of their unique tendency to form hydrides and disintegrate, metals from Group V cannot be processed in hydrogen. For this reason extraction from ores is carried out by way of the halides. Finally, because of the stability of the compounds of niobium and tantalum with hydrogen, carbon, nitrogen, and oxygen, these metals are useful as getters in vacuum systems.

When certain very small amounts of carbon, nitrogen, and oxygen are exceeded, precipitates form in molybdenum and tungsten. The precipitates tend to segregate at grain boundaries, thus causing the material to be weak and brittle. Molybdenum and tungsten are extracted from ores by way of the oxides, which are reduced to metal powders with hydrogen. A powder can be either remelted in vacuum to yield an ingot or pressed to form shapes that can be hot-rolled to form bars or strips with a fine grain size. Any precipitate present is dispersed over so much grain-boundary area that its embrittling effect is small. Accordingly, thin sheet and fine wire, which have the finest grain sizes, can be readily worked into intricate shapes such as lamp filaments. The fine grain structure is unstable, however, and operation at incandescent temperatures (even in vacuum) causes a coarser grain structure to form. A fixed small amount of oxide or carbide precipitate can then form a continuous film at the grain boundaries and again give a weak, brittle material. In summary, the mechanical properties of a given piece of molybdenum or tungsten depend strongly on its past thermomechanical history. In addition, parts fabricated from powder cannot attain the ductility of wire or foil, because the powder always absorbs oxygen during handling.

Because of the fabrication difficulties cited, an alternative method, called *pyrolytic deposition* or *chemical vapor deposition* (CVD), has been developed for fabricating strong and complicated shapes of molybdenum, tungsten, and other metals. This method can be illustrated by the following example. A volatile tungsten halide is reduced with hydrogen in a surface-catalyzed reaction on a hot mandrel. Metallic tungsten deposited on the mandrel can be built up to a thickness of about 1 cm, thus replicating exactly the shape of the mandrel. Intricate shapes can be made with useful strength.

Small amounts of halide remaining in the tungsten produced by pyrolytic deposition have a profound effect on its properties. This effect has been studied extensively for fluorine. In the absence of fluorine, grain boundaries move easily, and the grains can lose their columnar (oriented) form and become large and equiaxed (not elongated in any particular direction). When large amounts of residual fluorine are present initially in solution, fluorine diffuses to the grain boundaries and forms large bubbles, which embrittle the material and cause gross swelling. On the other hand, when the concentration of residual fluorine is about 20 parts per million, the bubbles that form are microscopic in size and do not cause swelling. But the microscopic fluorine bubbles interfere with the movement of grain boundaries so much so that tungsten retains a typical columnar grain size at high temperatures. The factors controlling the amount of residual fluorine are not known. Apparently they are determined by the conditions of temperature and gas pressure used during deposition.

Mechanical-working processes, such as rolling and drawing, and pyrolytic deposition result in polycrystalline materials with *preferred orientations*, i.e., crystals having grains with a more or less common direction rather than grains with randomly distributed directions.* For example, wrought bars (or wires) of molybdenum and tungsten tend to have many grains with ⟨1 1 0⟩ directions nearly parallel to the bar axis.† Similarly, grains in sheets tend to have the ⟨1 1 0⟩ directions of the grains parallel to the rolling direction and the (1 0 0) planes parallel to the sheet surface.

Preferred orientation is a bulk and not a surface property of a material. This property is studied experimentally by means of interactions of X rays with the bulk of a crystal. Accordingly experimental data on preferred orientation do not necessarily reveal anything about the surface planes although the bulk structure is related to the surface structure.

The preferred orientation developed by a pyrolytically deposited material is particularly strong but is simpler than that occurring in sheet materials. The term "simpler" means that a particular type of crystal plane tends to be preferred parallel to a surface of the sample but no particular directions in the plane of the deposit are preferred. The reduction of tungsten hexafluoride results in grains that grow along ⟨1 0 0⟩ directions away from the mandrel surface, i.e., grains having (1 0 0) planes parallel to the surface. The reduction of tungsten

*In a sample with randomly distributed grains, many grains with different orientations occur with equal frequency. For example, a section through such a sample discloses a set of grains whose ⟨1 1 0⟩ directions point in all directions in space with equal frequency. On the other hand, a section of a sample with a preferred orientation discloses that the ⟨1 1 0⟩ directions of all the grains are parallel.

†In wrought materials this occurs because the material deforms more easily by slip along certain crystal planes.

hexachloride results in grains that have (1 1 0) planes parallel to the surface. Presumably the difference in results is caused by the crystallographic dependence of the adsorption of species active in the reduction reaction.

Rhenium from Group VII has an extraordinary combination of properties. It is less volatile than any metal except tungsten and has good tolerance for interstitial impurities. It is ductile, but it work-hardens very rapidly and can be made harder than glass. It has a hexagonal crystal structure, strongly anisotropic thermal expansion, and a particularly high bare work function for unmodified, low-index, crystal surfaces. It is also rare, expensive, and dense. Fabrication is achieved either by pressing, sintering, and cold-working powder prepared by reduction of the soluble salt ammonium perrhenate or by pyrolytic deposition from the chloride. Both fabrication processes produce pieces having strong preferred orientations with basal (0 0 0 1) planes parallel to sheet and deposit surfaces.

Each of the refractory metals discussed in this section is available in the form of very pure, small-diameter, single-crystal rods prepared by float-zone refining. Slices from these rods are useful for experimental purposes but are too expensive for practical applications.

15.6 Electrode Surfaces

This section discusses some factors that influence the properties of refractory-metal surfaces. For a comprehensive discussion of metal surfaces in general, the reader is referred to the literature.[3]

15.6.1 Surface Free Energy. Atoms at or near the surface of a metal are not subjected to the same force fields as those in the bulk of the metal. As a result, a metal, and for that matter any solid, is not homogeneous near its bounding surface. The heterogeneity near a surface manifests itself by variations in intensive properties such as atom density and energy density per unit volume along a direction normal to the surface, and most of the variations of these properties occur over a distance of the order of one or two lattice constants.

Because of the thinness of the heterogeneous region, a sample of a metal can be analyzed by means of two thermodynamic systems: (1) a *homogeneous bulk system B* having the size of the sample and the intensive properties of the bulk of the sample; and (2) a *homogeneous surface system S* having a thickness of one or two lattice constants and extensive properties such that the superposition of S and B yields all the extensive properties of the metal sample. Thus the value of any extensive property of S, such as mass and energy, must be equal to its value in the sample minus its value in the homogeneous bulk system.

This technique for analyzing samples, which was first developed by Gibbs, has been used extensively to study surface effects on both solids and liquids. For

liquids the phenomenon of heterogeneity is called *capillarity*, and sometimes this term is also used for this phenomenon in solids.

The energy associated with the homogeneous surface system S is called *surface energy*. Because of the thinness of S, a surface-energy density per unit area rather than per unit volume is ordinarily defined for the surface system. The surface-energy density can be used to determine the *surface-Gibbs-free-energy density* and the *surface Gibbs free energy*. The surface Gibbs free energy is one of the most important quantities in the study of stability of the surface structure of a metal sample. Hereafter we refer to this quantity simply as the *surface free energy*. The surface free energy is always positive.

Calculating the surface free energy of a sample of a metal from its characteristics is difficult. Estimates have been made by using the number of bonds broken in the process of forming the surface by cleavage. Correlating measurements yields that, for metals with strong interatomic bonds (i.e., metals with high melting temperatures), surface-free-energy values are relatively larger than for metals with weak bonds. Refractory-metal surfaces have surface free energies of the order of 3000 ergs/cm^2, or 2 ev/atom.

The surface free energy depends on the crystallographic orientation of the surface. For a given metal it can vary by ±20% from one orientation to another. It is found experimentally that the surface free energy of a plane is lower the higher the mass density per unit area of the plane. It is noteworthy that planes with high mass densities also tend to have high bare work functions.

The surface free energy is also affected by curvature. An estimate of the difference, $\delta\mu$, between the surface free energy (μ_1) per atom of a curved region of the surface of an isotropic solid and (μ_2) of a flat region of the same surface is given by the Gibbs–Thomson relation:

$$\delta\mu = \gamma\Omega\kappa \tag{15.3}$$

where γ is the surface-free-energy density, Ω is the atomic volume, and κ is the radius of curvature of the curved region. For a typical refractory metal, γ is approximately 3000 ergs/cm^2, Ω is approximately 3×10^{-23}cm^3, and κ is approximately 10^{-6}cm. Therefore $\delta\mu$ is of the same order as kT at relatively high temperatures. It follows that curvature can contribute significantly to surface-free-energy differences or total potential differences* that cause migration of atoms and eventually cause changes in the curvature and orientation of the surface of the metal.

Finally, surface free energy is affected, usually increased, by *mechanical damage*. If a surface is ground or lapped, enough damage is done so that the

*As shown in Chap. 7, the Gibbs free energy per atom is also equal to the total potential of the atoms. For this reason the symbol μ is used in both discussions.

surface loses its regular crystal structure and appears amorphous, or at least randomly distributed. Most of this damage anneals out with rising temperature before the metal reaches a temperature where surface changes can occur. Nevertheless, the amount of surface free energy due to mechanical damage is so large that even 1% of it, distributed over all the atoms on the surface, would be a large amount compared with kT.

15.6.2 Effects of Heating on Metal Surfaces. For stable equilibrium of a system held at temperature T, the Gibbs free energy must be at a minimum. A corollary of this condition is that the total potential of a given component of the system must be uniform throughout the system.

If the Gibbs free energy of the system is greater than the minimum value corresponding to T, spontaneous changes of the state of the system will occur unless they are prohibited by constraints imposed on the system. Similarly, if the total potential of a component is greater in one region than in another, atoms will flow from the region having greater total potential to the region having smaller total potential unless this migration is prohibited by constraints imposed on the system.

Usually a piece of a metal at temperature T is not in a stable equilibrium state. It is in a state that changes and tends to approach the stable equilibrium state corresponding to T. Indicative of the nonequilibrium state are the different surface free energies that the metal can have as a result of various orientations, curvatures, and degrees of mechanical damage.

At relatively low temperatures the changes leading from the nonequilibrium to the stable equilibrium state of the metal can occur very slowly because of the existence of some potential barrier that inhibits the migration of atoms. At high temperatures, however, kT becomes comparable to the potential barrier, and usually atoms can migrate more rapidly and bring the metal to its stable equilibrium state. Such phenomena are observed experimentally at surfaces.

Increasing the temperature of a flat, single-crystal surface can cause *faceting*, i.e., local rearrangement of the atoms on the surface into two other surfaces with a larger total area and a smaller total surface free energy (Fig. 15.5). Faceting can occur when the areas A_i and surface-free-energy densities γ_i before and after faceting satisfy the relation (Fig. 15.5)

$$A_0\gamma_0 > A_1\gamma_1 + A_2\gamma_2 \tag{15.4}$$

The change is likely to happen on a single-crystal surface having a high surface-free-energy orientation.

Increasing the temperature of a flat polycrystalline surface causes another kind of change. For example, consider the surfaces of two adjacent grains, I and II. A

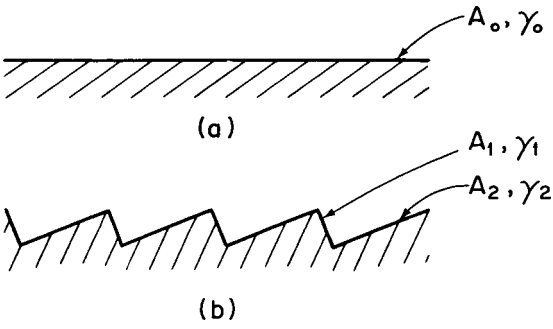


Fig. 15.5 Cross sections of a single-crystal surface. (a) Before faceting. (b) After faceting.

cross section of the grains by a plane normal to the surface and the grain boundary is shown schematically in Fig. 15.6a. If γ_1 is the surface-free-energy density of I, γ_2 that of II, and γ_g that of the grain boundary,* then it can be shown that, for mutual stable equilibrium of grains I and II, γ_1, γ_2, and γ_g must add vectorially to zero as shown in Fig. 15.6b; namely, the surface is initially distorted on a microscopic scale and a groove is formed at the grain-boundary edge at the surface. The sloping surfaces of the groove meet the original grain surfaces and form corners or high curvature regions. As a result, the surface-free-energy densities at the corners are altered, and further migrations of atoms occur; this leads to rounding off of the corners. The net effect of all these competing processes is that the groove grows at a steadily decreasing rate. The effect can be visible eventually to the naked eye.

If the surface is roughened by mechanical damage (abrasion, grinding, sandblasting, etc.), the surface free energy due to this damage can enhance the changes occurring when the surface temperature is increased.

The mechanisms by which changes occur on a surface as either a function of time or a function of temperature, or both, are discussed by Mullins.[4] In general, it is concluded that surface and volume diffusion are the principal mechanisms, with surface diffusion being predominant. In addition, treating a surface for a long time at a low temperature was found to be equivalent to treating it for a short time at a high temperature.

In general, changes leading to a stable equilibrium state result in rounding and smoothing of edges and corners. The significance of the effects obviously depends on how many edges and corners are present initially. At converter operating temperatures a finely etched polycrystalline surface can change

*The free-energy density of the grain boundary is defined in a manner completely analogous to that used for surfaces.

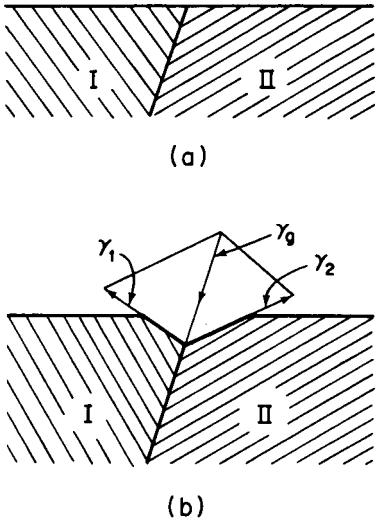


Fig. 15.6 Cross sections of a polycrystalline surface (two grains). (a) Before grain-boundary grooving. (b) After grain-boundary grooving.

enormously on a scale visible with a light microscope, whereas an etched, (1 1 0), single-crystal surface may not change at all.

15.7 Metallurgical Preparation of Electrode Surfaces

The primary objective of metallurgical preparation of an electrode surface for a practical thermionic converter is that of achieving a stable surface with a high bare work function. Stability is required so that the performance characteristics will remain constant over the lifetime of the converter. A high bare work function is required because, in general, it results in better performance in the presence of cesium under practical conditions of operation.

The techniques used for preparation are *mechanical polishing, thermal treatment without mass transfer to or from the surface, mechanical damage and heat treatment, chemical and electrochemical etching*, and *thermal treatment with mass transfer*. In this section each of these techniques is discussed briefly, and the rationale for their use is given.

For convenience suppose that a polycrystalline, randomly oriented piece of metal with a cubic structure must be prepared as an electrode with a flat surface. The simplest way to do this is to use machining and mechanical polishing techniques. The resulting smooth and shiny surface contains some residual mechanical damage and consists of a large number of grain surfaces with orientations representative of the randomly oriented bulk of the metal. If the

metal is not treated further, the mechanical damage at operating temperatures usually is not sufficiently strong to produce significant changes. On the other hand, at operating temperatures the grain boundaries will move and grain-boundary grooving will occur. A grain boundary appears as a groove and as it moves leaves traces of the groove within a single grain. Thus heating the surface results in the formation of a network of grooves, mostly at the grain boundaries, but the orientations of the surface remain unaffected. Since the widths of the grooves are considerably smaller than the sizes of the grains, the grooves do not have a significant effect on the emission properties of the surface. It follows that machining and polishing do not produce surfaces with features different from those dictated by the polycrystalline nature of the bulk of the metal.

These results can be partially altered by means of intensive heat treatment of the surface, which can result in faceting and in an increase of the bare work function of the surface. A surface of randomly oriented grains has a distribution of surface-free-energy densities. Under intensive heat treatment, regions with high surface free energy sometimes tend to lower their energies by rearranging atoms on their surfaces, namely, by diffusion of atoms from one position on the surface to another (Sec. 15.6.2). Because of the general tendency of a region of low surface free energy to have a high bare work function, faceting is accompanied by an increase in the bare work function of the surface. Faceting has been observed to occur frequently in metals heated above about one half their melting point in degrees Kelvin. The width of the facets lies between a few hundred Angstrom units and a millimeter.

Faceting by means of heat treatment is effective only on surfaces having orientations close to a low surface-free-energy orientation. An activation energy is necessary to convert a surface of (1 1 0) to a (1 1 0) surface plus another plane. This energy is larger the larger the misorientation from (1 1 0), and the larger the energy the higher is the temperature at which the heat treatment must be performed. Extreme heating causes evaporation and eventually melting, however. For example, an attempt to promote faceting on a surface 20° from (1 1 0) is likely to be unsuccessful because the temperature necessary to activate the process will also lead to evaporation, which causes smoothing of the surface and counteracts faceting. Thus thermal faceting can be achieved only on surfaces in certain restricted ranges of orientations, namely, near local minima in the graph of surface free energy vs. orientation.

The activation energy necessary for faceting can be provided by a combination of severe mechanical damage and heat treatment. The mechanical damage can be introduced either by sandblasting or by abrasion with fine alumina particles on a glass plate instead of the customary polishing cloth. The effectiveness of this technique was shown in one special case of a tungsten surface oriented close to (1 1 0). When it was heat treated, the initially rough surface developed many

small flat regions of exact (1 1 0) surface bounded by curves and connected by an undulating background that might have been faceted on a fine scale (Fig. 15.7). Thus a rough surface that initially had *no* (1 1 0) regions developed a significant fraction (40 to 60%) of stable (1 1 0) surface. If the original surface had been electropolished so as to be completely stress-free, no changes would have been detectable after heating. The mechanical-damage plus heat-treatment technique is not practical, because it requires monocrystalline specimens.

Another class of preparatory methods relies on the modification of a machined surface until it is no longer flat on a fine scale but is flat on a gross scale. As already stated, the surface of a polycrystalline sample consists of regions having low surface-free-energy and high work-function values and regions having high surface-free-energy and low work-function values. This suggests that material be removed from the surface selectively so that more low surface-free-energy planes are exposed. Certain electrochemical processes do just that. In a suitable electrolyte the rate of material removal (dissolution) is determined by the surface free energy. Low surface-free-energy planes are attacked by the solution less rapidly than high surface-free-energy planes. For this reason, after a period of attack, low surface-free-energy planes come to occupy a larger fraction of the total surface. Depending on the reactants involved, such a process, commonly called etching, can occur with or without the application of an external electric

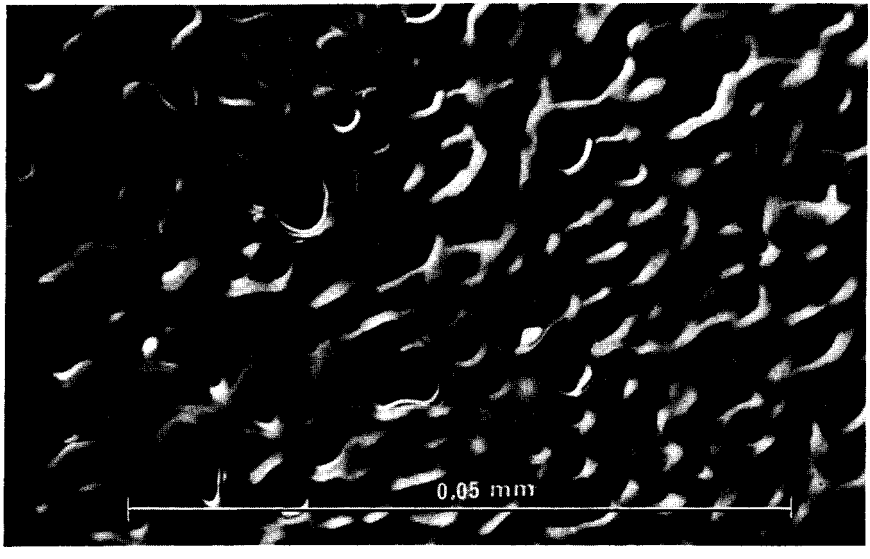


Fig. 15.7 Surface of a tungsten crystal oriented close to (1 1 0), after abrasion and annealing.

field. For example, tungsten can be treated with alkaline potassium ferricyanide or made anodic in dilute sodium hydroxide. In either case (1 1 0) surfaces are removed less rapidly than others; this leads to a desirable increase in the value of the bare work function. Similarly, rhenium can be electroetched in a perchloric acid–alcohol bath with a suitable electric field. This bath can also be used for electropolishing, but then the field and, therefore, the applied potential is higher than that used for electroetching.

The etching procedures cited always attack grain boundaries and dislocations. Accordingly the resulting surface texture depends on the initial grain size and on the duration of treatment. The effects of the attack are clearly visible to the naked eye. A rough surface bounded by regular crystal planes and containing many sharp edges and corners is formed. In addition to the increase in the value of the bare work function, the total area of the surface is considerably greater than that of the original plane surface.

The facts that the etched surface is formed at room temperature and that it contains many edges and corners indicate that such a surface may be unstable at high temperatures. Although a high value of the bare work function can be obtained, this value can diminish considerably when the surface is used at practical temperatures because of the rounding and smoothing processes discussed earlier. It follows that etching techniques are practical for preparing low-temperature electrodes only.

A number of electrochemical conditions can be found under which the process of *electropolishing* occurs. In this process the removal of the material of the surface occurs most rapidly at protrusions. As a result the surface undergoes smoothing on a microscopic scale and develops a highly reflecting appearance featureless under the microscope. Like etched surfaces and unlike mechanically polished ones, the electropolished surface is completely free of stress. Its stability under heating depends only on its orientation. Electropolishing is useful in preparing particular single-crystal planes for research studies.

The final technique of surface preparation consists in the use of heat treatment of and mass transfer to a surface. This technique is discussed in connection with cubic structures only. If a large flux of atoms impinges on a surface, the atoms tend to arrange themselves to form layers of high-atom-density, low surface-free-energy planes, rather than other planes. For example, if atoms of tungsten from a very high temperature source are condensed on a hot tungsten surface, they form a new surface consisting almost entirely of (1 1 0) planes *regardless* of the orientation of the substrate (Fig. 15.8). Moreover, the new surface is not parallel to the surface of the substrate. For example, a pyramid or roof of four (1 1 0) planes will develop on a (1 0 0) grain of the substrate. A rhombohedral, truncated pyramid bounded by five (1 1 0) planes will develop on a (1 1 0) grain, and so on. The formation of such a surface requires a certain high

Fig. 15.8 Surface of a tungsten crystal exposed to a condensing flux of tungsten atoms.

substrate temperature and is facilitated by very high incident fluxes and by mechanical damage in the substrate. A range of incident fluxes exists for which damaged (scratched or sandblasted) substrate surfaces lead to the new type of surface. On polished surfaces, however, the impinging beam condenses uniformly and gives a replica of the substrate. Because of the high temperatures involved, surfaces formed by heat treatment and mass transfer are stable at converter operating temperatures. This technique can be used to modify an arbitrary surface and produce a surface consisting of (1 1 0) planes.

References

1. C. S. Barrett and T. B. Massalaski, *Structure of Metals*, 3rd ed., p. 12, McGraw-Hill Book Company, Inc., New York, 1966.
2. Ref. 1, p. 384.
3. W. D. Robertson and N. A. Gjostein, *Metal Surfaces—Structure, Energetics and Kinetics*, American Society for Metals, Metals Park, Ohio, 1963.
4. W. W. Mullins, Flattening of a Nearly Plane Solid Surface Due to Capillarity, *J. Appl. Phys.*, 30: 77 (1959); see also Ref. 3.

16

Design and Fabrication of Practical Converters

16.1 General Remarks

The design of a practical converter is guided primarily by the need to couple it to a given heat source in a way that minimizes loss of heat. The designer must evaluate possible configurations in view of such objectives as small temperature differences in heat reception and rejection, light weight, rapid warm-up, structural strength, and simplicity of fabrication.

To aid the designer in devising an adequate variety of possible system configurations prior to design evaluation, we identify in this chapter the functional requirements that must be satisfied by the converter structure. Typical design procedures are described in the light of limitations imposed by available materials and techniques of fabrication, and broad guidelines are presented for the fabrication of prototypes.

16.2 Functional Requirements on Converters

The principles of operation of a converter suggest a number of functional requirements that must be satisfied by the various structural parts of the converter. These requirements, which can be classified as general and specific, are discussed in Secs. 16.2.1 and 16.2.2.

16.2.1 General Requirements. General requirements resulting from considerations of heat transfer, optimization of performance, delivery of electrical power, and maintenance of cesium vapor in the interelectrode space are discussed in the following paragraphs.

(a) Heat Transfer. In a converter heat flows from a high-temperature source to the emitter and from the collector to the heat sink or coolant. Adequate structures must be provided so that the heat flows to the emitter with the least possible loss and at the rate required to maintain the selected emitter temperature and operating conditions. The temperature drops between the source and the emitter and between the collector and the heat sink usually must be kept as small as possible to ensure efficient and reliable operation.

(b) Interelectrode Spacing. Performance characteristics of a converter depend on the interelectrode spacing. The electrodes must be shaped and positioned so that at operating temperatures the interelectrode spacing is uniform throughout the

interelectrode space and is relatively insensitive to expected electrode displacements caused by grain-boundary movements, grain growth, and creep. This provision ensures that the converter performance will be uniform across the electrode surfaces and that electrode shorting will be avoided.

(c) Delivery of Electrical Power. The electrical connections to the electrodes of the converter usually involve a substantial heat flow from emitter to collector. In addition to its effect on efficiency, this flow can substantially affect the temperature distribution over the emitter surface. Special attention must be given to lead design to optimize the efficiency and maintain temperature uniformity. The lead optimization procedure is more complicated than that given in Vol. I because of the effect of radiation heat transfer on the temperature distribution of the lead. A more realistic design procedure is discussed in Sec. 16.2.2.

(d) Cesium Envelope. The cesium vapor produced by the reservoir must be contained by a hermetic envelope, which must include an insulating seal to prevent shorting of the electrodes and must allow the vapor to reach and fill the interelectrode space. Moreover, care must be taken to ensure that no area of the envelope is at a temperature lower than that of the reservoir. This is necessary so that the cesium-vapor pressure can be controlled by the temperature of the reservoir.

In addition the envelope usually provides mechanical support for the converter with respect to the structures that surround it.

16.2.2 Specific Requirements. Specific requirements are dictated by the heat source coupled to the converter and by the intended application. These requirements are discussed in the following paragraphs.

(a) Compatibility with Heat Source. The heat source has important implications for both the geometry and the physcial properties of the materials of the converter. The converter geometry must be such that the emitter is preferentially exposed to the heat flux delivered by the heat source. The design requirements are affected by the nature of the heat source. In some instances, only the emitter is affected. For example, in converters heated by combustible fuels, compatibility of the emitter with the combustion environment is the prime requirement. On the other hand, in converters heated by nuclear fuels in a nuclear reactor, not only must the emitter be metallurgically compatible with the reactor fuel but also the converter design is influenced by criticality requirements for the reactor. Finally, high-temperature materials that are good heat conductors and good electrical insulators must be used if several converters connected electrically in series are heated by conduction from a common heat source.

(b) Compatibility with Heat Sink. The heat sink is an important boundary condition of the design of the thermal path for heat rejection. Relatively high-temperature materials that are good heat conductors and good electrical insulators must be used when the collectors of series-connected converters are cooled by a common heat sink.

(c) Converter Envelope. In addition to the requirements discussed in Sec. 16.2.1 (d), the design of the envelope of a converter is influenced by considerations such as weight, life, and resistance to shock, vibration, and thermal cycling. The envelope is one of the key components of the converter and often is the component most amenable to design innovations.

(d) Controls. In general, the heat balance at the emitter depends on the conditions of operation of the converter. Variations of electrical load and heat input can result in deterioration of performance characteristics and loss of structural integrity of components. The procedures used for start-up and control of the converter must be carefully considered. The variations cited can be avoided by controlling either the output of the heat source or the electrical load, or both. In addition, performance characteristics are sensitive to cesium-vapor pressure. This implies that some form of control of the cesium-vapor pressure or cesium-reservoir temperature must be provided.

16.3 Component Design

The functional requirements discussed in Sec. 16.2 can be satisfied by the components of a converter in a variety of ways. In the best designs each component is made to perform several functions without duplicating those of other components.

Designing a component that performs a number of required functions presents several problems that are generally independent of converter geometry. In this section the discussion of these problems is based on the typical arrangements of components of plane-electrode and cylindrical-electrode converters shown schematically in Figs. 16.1 and 16.2.

16.3.1 Emitter Electrode. The emitter electrode receives heat from the heat source and transfers it primarily to the emitter surface. Considerations of low-cost operation, temperature drop, and weight indicate that the emitter electrode should be as thin as is consistent with strength to withstand atmospheric pressure loads and with uniformity of temperature at the emitter surface. Typical values for the thickness of this component are 0.060 to 0.100 in., the larger value being associated with plane geometries.

In plane-electrode converters with small interelectrode spacing and large electrode-surface areas, heat transfer normal to the surface causes a camber

through differential thermal expansions. The magnitude of the camber can be large compared with the interelectrode spacing. For example, for circular plates and small deformations, the deflection Δd of the center of the plate with respect to its rim is given by the relation

$$\Delta d = D^2 \frac{\alpha q}{8K} \tag{16.1}$$

where D = diameter
α = coefficient of linear thermal expansion
q = heat-flux density
K = thermal conductivity of the plate

The displacement Δd is independent of plate thickness and occurs in a direction opposite to the direction of heat transfer. For a tantalum emitter electrode having a diameter of 2 in. and conducting 100 watts/cm^2, Eq. 16.1 yields that the value of Δd is equal to 1.2 mils. Such a displacement is significant in converters with an interelectrode spacing of a few mils.

16.3.2 Emitter Support. The emitter support (Fig. 16.1) is a thin component designed to locate the emitter with respect to the collector while accommodating the difference in the thermal expansions of the high-temperature emitter and the low-temperature insulating-seal structure. In many designs it also serves as part of the converter envelope and as a conductor of electrical current from the emitter to the positive terminal. In other designs the supporting function is shared by several components, as indicated in Fig. 16.2. In plane-electrode converters the expansion of the emitter support can be used to achieve a desired interelectrode spacing. At room temperature the electrodes are in full contact (zero spacing). At operating temperatures the electrodes are separated because of the difference in thermal expansion of the emitter support and the collector.

Attaching the emitter support to the emitter electrode involves two design problems: (1) The abrupt change in material cross section at the point of attachment can cause cumulative plastic deformation on thermal cycling; this deformation eventually leads to cracking. (2) The emitter support provides a path for heat loss and creates nonisothermal conditions over the emitter surface.

The stresses at the joint between the emitter electrode and its support can be effectively reduced either by using similar cross sections for these two elements or by providing a gradual change in cross section.

The effect of thermal conduction on the temperature distribution over the emitter surface cannot be completely avoided. In some designs, however, this may not be entirely detrimental to the performance of the converter since the power-producing capability of the hot portions of the emitter support can be

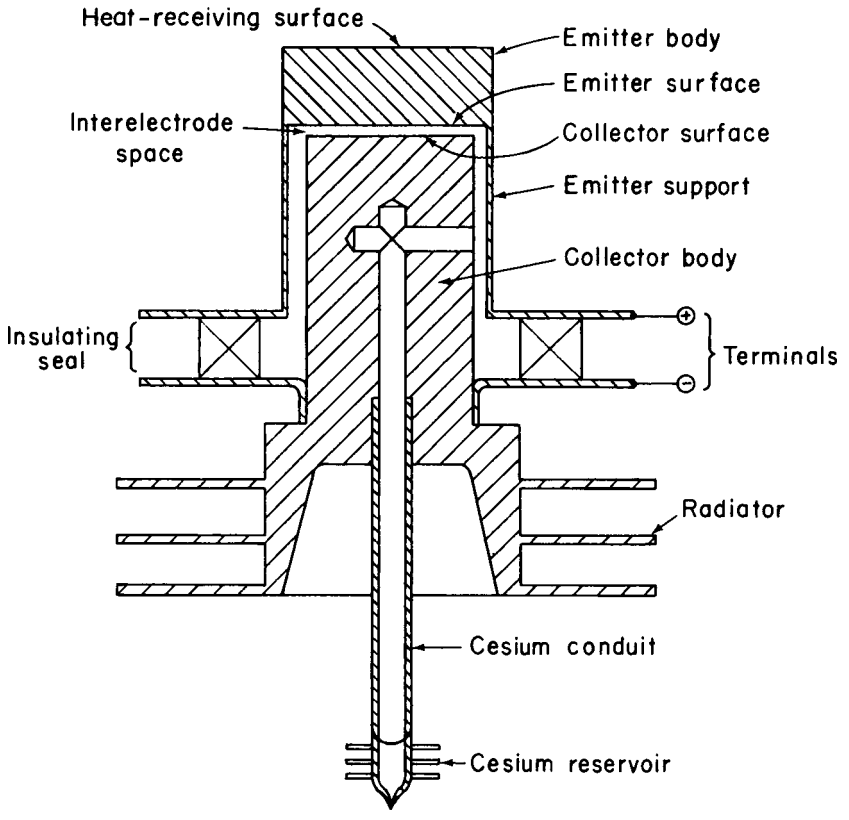


Fig. 16.1 Basic converter structure. Planar geometry.

significant and, thus, can partially or totally offset the loss in performance resulting from the nonisothermal condition of the emitter surface.

Ideally the emitter-support dimensions should be optimized so that a decrease in cross section reduces the required heat input by an amount equal to the heat equivalent of the increase in electrical-power dissipation caused by the Joule effect.

The heat equivalent Q_j of the Joule effect is given approximately by the relation

$$Q_j = \left(\frac{1}{\eta_e} - \frac{1}{2}\right) I^2 R_e \tag{16.2}$$

where η_e is the efficiency of the converter based on the net heat input to the emitter, I is the output current, and R_e is the electrical resistance of the emitter

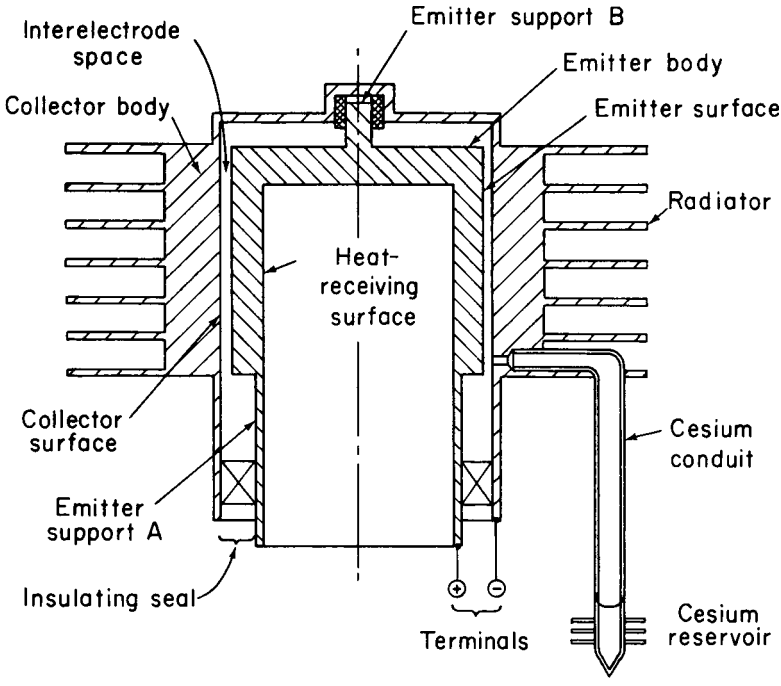


Fig. 16.2 Basic converter structure. Cylindrical geometry.

support. The term $(\frac{1}{2})I^2R_e$ represents approximately the amount of Joule heat transmitted to the emitter. This approximation is adequate for all practical purposes.

The heat Q_k conducted away by the emitter support can usually be expressed in the form

$$Q_k = \frac{\Delta T_s}{R_t} \tag{16.3}$$

where ΔT_s and R_t are the temperature difference across the emitter support and its thermal resistance, respectively. In calculating R_t we must account for the radiation-heat-transfer-enhanced temperature gradients in the high temperature portions of the emitter support. Estimating the magnitude of this effect by numerical analysis, we find that the heat conducted away from the emitter can be greater than that calculated by the laws of heat conduction alone.[1]

The calculation of optimum lead geometry is affected considerably by the ratio of the correct heat loss to heat loss calculated without radiation effects. Figure 16.3 illustrates the effects of a ratio of 1.5.

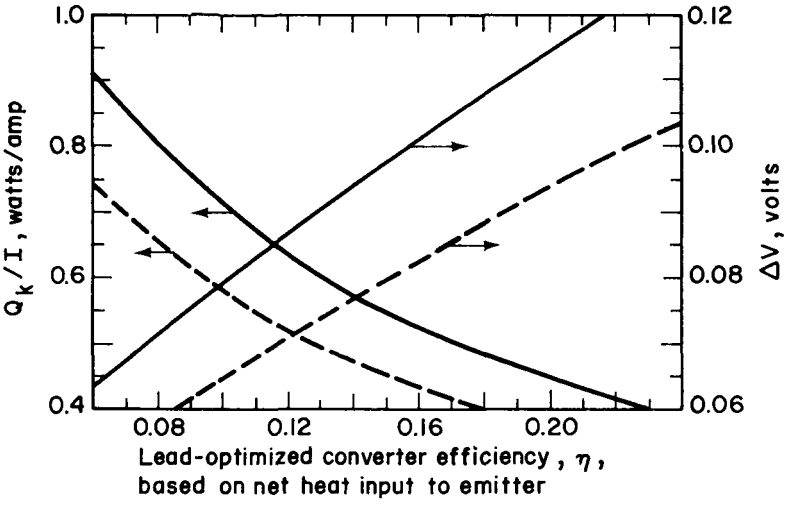


Fig. 16.3 Optimum emitter support losses. - - - -, losses of ideal optimized lead (no radiation effects). —, losses of optimized lead when radiation increases conduction loss by 5% (see text).

In small converters the optimum emitter-support configuration may be impossible to attain because of practical limitations on the fabrication of very thin cylindrical walls. The minimum practical wall thickness for a cylindrical emitter support is of the order of 0.003 in. or ½% of the diameter, whichever is greater. In general, emitter supports are made cylindrical so that they can withstand exposure to atmospheric pressure.

16.3.3 Insulating Seal. The functions of the insulating seal (Figs. 16.1 and 16.2) are (1) to provide electrical insulation between the emitter and the collector, (2) to form part of the converter envelope, (3) to provide mechanical support for the emitter electrode with respect to the collector electrode, and (4) to conduct heat from the emitter to the collector. Because these functions are practically the same as those of insulating seals used in the metal–ceramic tube industry, respective designs of this component exhibit substantial similarities. Some typical configurations are shown in Fig. 16.4. The insulating material is usually a ceramic. Because of the high notch sensitivity of ceramics, all corners where chips and stress-concentrating cracks are likely to form must be rounded. Moreover, tensile loads must be kept to a minimum by avoiding large temperature gradients and by using metal flanges of relatively thin cross section with a coefficient of thermal expansion which closely matches that of the ceramic.

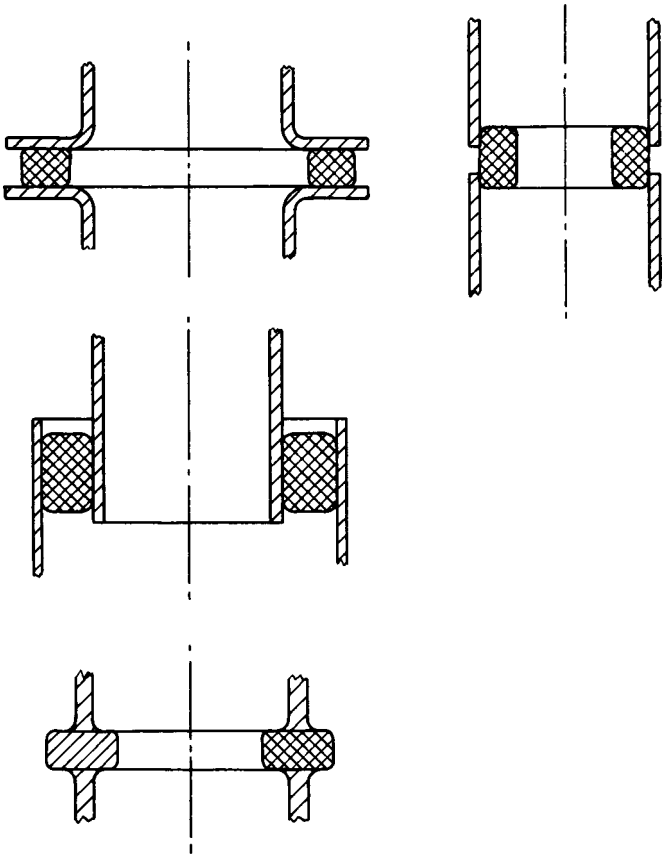


Fig. 16.4 Typical insulating-seal designs.

16.3.4 Cesium Reservoir. In most converter designs a temperature-controlled liquid-cesium reservoir is in vapor communication with the converter envelope. This reservoir is connected to the envelope by means of a conduit, the connection being made most often at a region where the temperature of the envelope is near that of the collector. The conduit provides a heat-transfer resistance that allows the reservoir to operate at temperatures lower than those prevailing in the converter. In many converters the conduit is also used before the envelope is sealed to exhaust the residual gases of the converter. Because of this function, the tube must have dimensions that will ensure a relatively free exhaust of the gases produced during final outgassing.

In some converter designs the liquid-cesium reservoir is replaced by a cesium-adsorption reservoir, i.e., cesium that either is adsorbed in a matrix of or

is chemically combined with another substance. Cesium-adsorption reservoirs are designed to establish the required cesium-vapor pressure at a temperature between the temperatures of the emitter and the collector. They are often thermally coupled to both the emitter and the collector. When this is done, the reservoir can maintain the temperature required for optimum operation even when the emitter temperature and the electrical load vary.

The design of the collector structure is discussed in Chap. 17.

16.4 Materials

The high temperatures of operation and the need to contain cesium vapor restrict the choice of converter materials. Although further restrictions can be imposed by the particular environment in which the converter operates, this section is concerned only with material problems common to all converters.

16.4.1 Thermophysical Properties. Since the converter is an electrical device operating over a wide range of high temperatures, only materials with suitable thermal and electrical properties can be used in its construction. Values of thermal conductivity, electrical resistivity, and thermal-expansion coefficients for metals used in converters are given in Figs. 16.5 to 16.9. Values of the

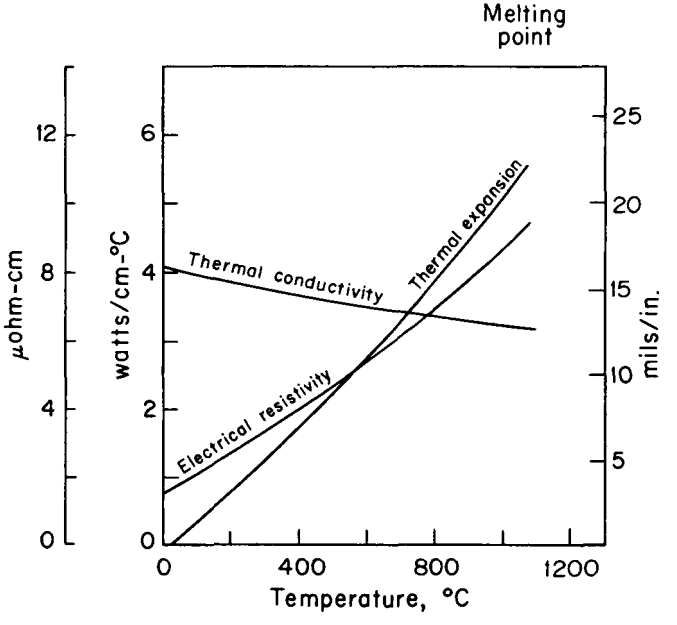


Fig. 16.5 Selected properties of OFHC copper.[2]

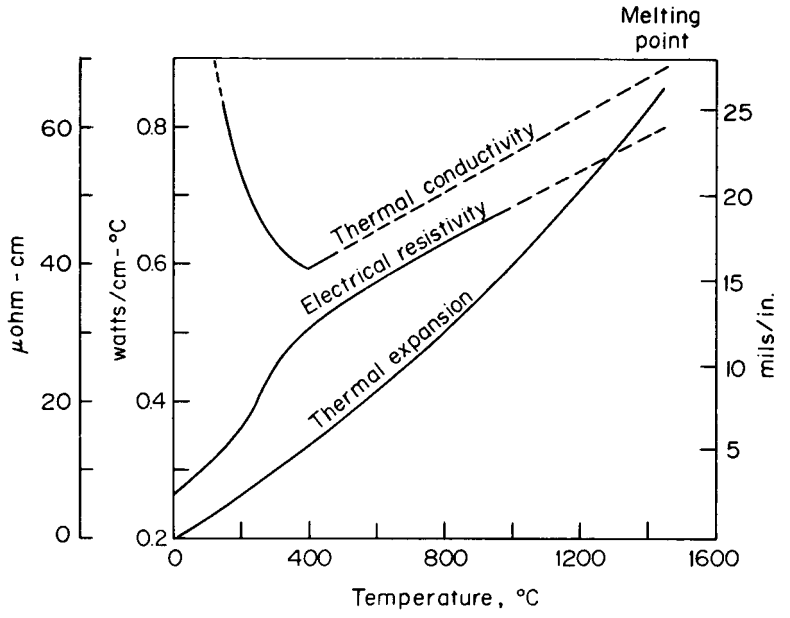


Fig. 16.6 Selected properties of high-purity nickel.[2,3]

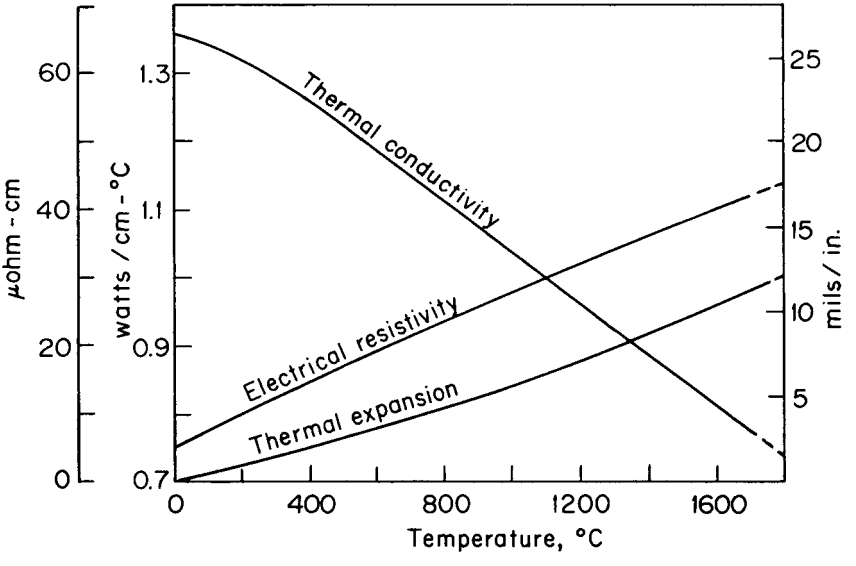


Fig. 16.7 Selected properties of niobium.

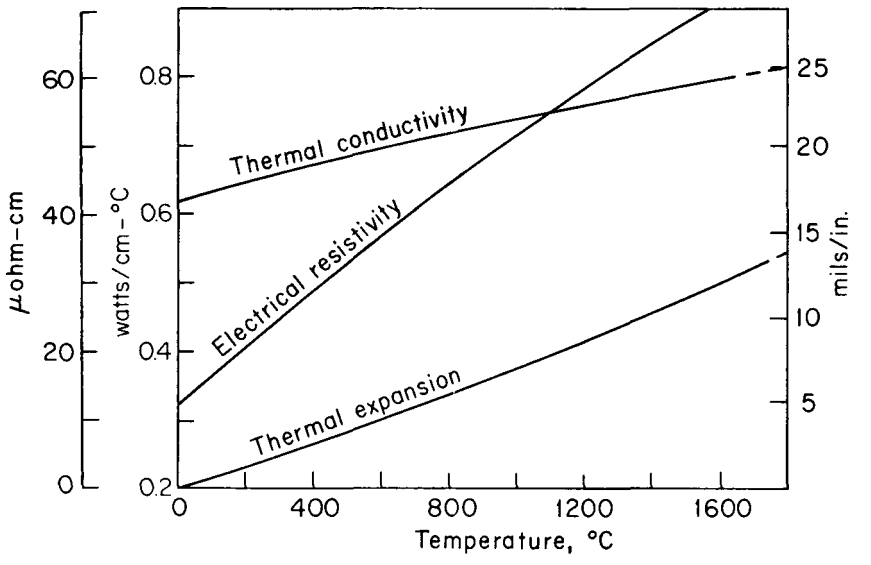


Fig. 16.8 Selected properties of molybdenum.[1]

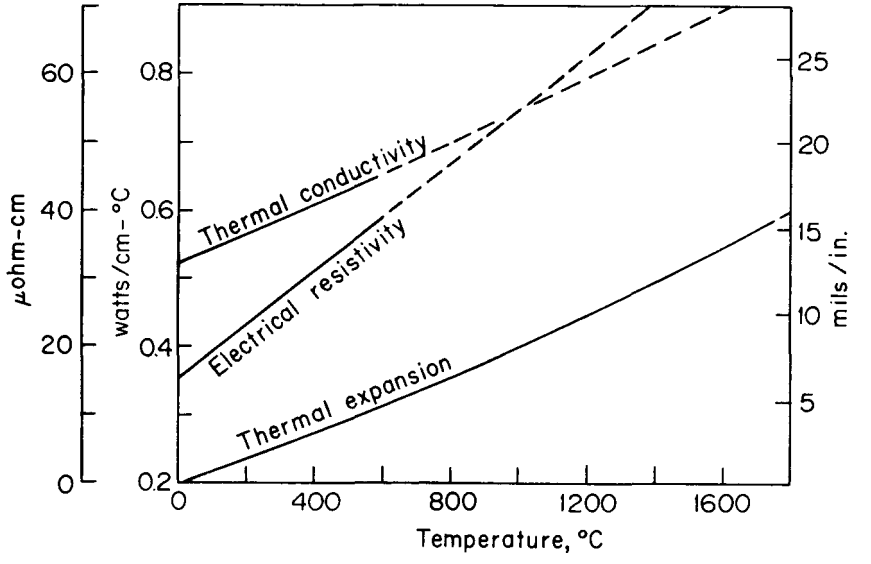


Fig. 16.9 Selected properties of tantalum.[2]

properties of commercial metals can show considerable variations from the values given in the figures because of differences in grain structure, composition, and method of fabrication. Values of thermal emissivities for many materials as functions of temperature, surface condition, and radiation wavelength are given in Ref. 4. Evaporation rates of a number of metals are shown in Fig. 16.10.

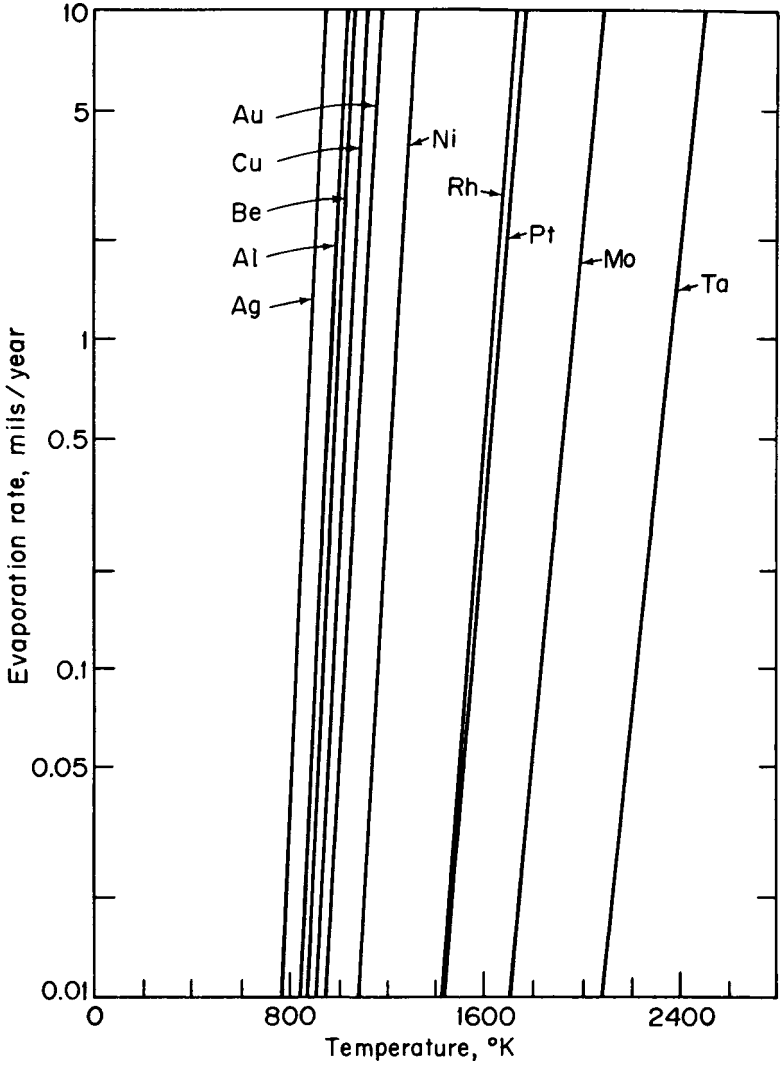


Fig. 16.10 Rate of evaporation of metals.

Metals with high evaporation rates must be excluded from converters because they can cause changes in dimensions, depletion of brazes, and deposition of conducting layers on ceramic insulators.

16.4.2 Mechanical Considerations. A converter is a relatively compact and rigid structure made up of many materials that are subjected to high and different temperature levels, often under cyclic conditions. The thermal stresses occurring during steady-state operation can lead to fracture of joints, and those occurring during warm-up and cool-down operations can lead to an accumulation of plastic deformation, which, in turn, can result in torn and fractured materials. Moreover, static loads at high temperatures can produce creep of components to such an extent that the interelectrode spacing of the converter is significantly altered. For these reasons converter materials should be lightweight, ductile, strong, and similar to each other in thermal expansion at all temperatures of operation.

Special attention must also be given to the joining processes. Materials used in joints must satisfy requirements similar to those for materials of components.

Heating to a high temperature can change the properties of some materials considerably. For example, the transition temperatures of tungsten and molybdenum undergo an appreciable shift when recrystallized. As a result these materials can become embrittled after the first cycle at high temperature although originally they are ductile at room temperature. Because of this tendency to embrittlement, tungsten and molybdenum are seldom used in structures that are apt to encounter any type of impact loading.

16.4.3 Chemical Considerations. Although the subject is much broader, this discussion is devoted solely to considerations of chemical attack on converter components by cesium and gaseous impurities.

Several workers have investigated the compatibility of cesium in both liquid and vapor phases with possible converter materials. Slight variations in the reported results are usually due to variations in experimental procedures. For example, the cesium-vapor velocity in some experiments may be appreciable; this leads to erosion rather than corrosion. In other experiments gases other than cesium which may be present may be responsible for the effect observed. Data on cesium corrosion are summarized in Ref. 5. The most important conclusion that emerges from these data is that all refractory metals, such as tungsten, molybdenum, tantalum, rhenium, niobium, and hafnium, have excellent resistance to attack by cesium vapor at elevated temperatures (about 1800°K). On the other hand, iridium, rhodium, vanadium, palladium, and aluminum oxide have good resistance only up to about 1600°K; titanium and zirconium and all alloys of nickel, iron, chromium, and cobalt are good up to about

1300°K; and, finally, copper and platinum are good up to about 900°K but can have limited resistance up to about 1300°K.

Materials that *do not* resist attack by cesium are silver, gold, and silicon dioxide. The attack on silver and gold is due to the formation of alloys with low melting points similar to the amalgams that these metals form with mercury. The attack is much less rapid on silver or silver-bearing alloys than on gold or gold-containing alloys, and the silver–copper eutectic can be used as a braze filler to handle cesium vapor below 700°K for short periods of time. The attack by cesium on silicon dioxide in ceramics is attributed to the formation of a cesium–aluminum silicate.[5] It is observed in ceramic bodies containing about 1% or more of silica.

The problem of attack by gaseous impurities can be very complicated, but it occurs rather infrequently in converters. Usually gaseous impurities are present in a converter in quite small amounts. For this reason their effects tend to be severe only when the chemical reaction at hand involves a cycle in which the deleterious gaseous compound is continuously reconstituted. A classic example is the tungsten–water-vapor cycle. Trace amounts of water vapor can cause transport of material from incandescent filaments through the vacuum to a glass envelope at rates far exceeding those produced by evaporation alone.[3,5] In converters the water-vapor cycle has been observed only with molybdenum emitters.

In converters with tantalum emitters, rapid transport of tantalum from the emitter to the collector has been observed. A layer of powder 1 mil thick was formed on the collector in 10 hr. The cause of the attack has not been positively identified. Chemical analyses of the collector deposit showed rather high amounts of silicon. Silicon is often found in cesium as an impurity, particularly when cesium has been distilled in glass ampuls. Silicon can also be produced by cesium attack on ceramic seals containing silica.

Evaluating the possibility of a chemical vapor cycle requires calculation of net change in the free energy of formation of the compounds presumed to form from the species present at each location in the converter and at the temperature of the location. For the tantalum–silicon cycle, a favorable free-energy balance appears possible for the reactions

$$2\tfrac{1}{2}\,SiO_2 + 2Ta \underset{2000^\circ K}{\overset{1000^\circ K}{\rightleftharpoons}} 2\tfrac{1}{2}\,Si + Ta_2O_5 \qquad (16.4)$$

16.4.4 Metallurgical Considerations. Metallurgical considerations play an important role in the design and fabrication of converters. This remark is amply supported by the brief discussions that follow.

An important consequence of the requirement of long operating life is that the converter envelope must be fabricated to be relatively free of gaseous

contaminants. All materials used in fabricating converter envelopes must be cleaned chemically to remove surface contaminants and then heated (fired) in a selected atmosphere to promote diffusion and evaporation of most of the dissolved gases or gas-forming chemical elements. The processing of components is expedited and the contamination of processing equipment is reduced when vacuum-tube-grade materials are used. These materials have a minimum content of impurities such as carbon, manganese, phosphorus, sulfur, and silicon, which are undesirable because they can cause embrittlement and porosity of welds, poor wetting of braze materials, and blistering of electroplates.

For emitters the tolerance to impurities is even smaller than for converter envelopes. For example, substances such as iron and niobium can cause severe degradation of electrical output. For this reason emitters must be fired at high temperatures for several hours before they are assembled in a converter structure. This firing is important also because it promotes the formation of a stable emitter surface structure. Such a structure is essential to the construction of a converter with stable performance characteristics.

Components operating at an appreciable level of mechanical stress should have a fine grain structure so that loss of strength produced by massive precipitation of impurities at grain boundaries can be prevented (see also Chap. 15).

Tensile and creep strength at high temperature can be enhanced by adding alloying elements that induce strengthening through solid solution or dispersion. Examples are the molybdenum alloys with small amounts of titanium, carbon, and zirconium. Also, small additions of elements such as titanium and zirconium can prove effective in controlling the residual gas evolution from high-temperature components, but such elements must not be allowed to migrate to electrode surfaces at an appreciable rate because they can be detrimental to the performance characteristics of the converter.

Most joining processes in a converter require considerable attention. At a joint made by brazing or welding, the materials involved are deliberately alloyed to form an intimate metallurgical bond. Sometimes this alloying results in intermetallic compounds which are extremely hard and brittle and which have thermal-expansion characteristics such that the joints obtained are susceptible to cracking. Because of this possibility, proposed joint-fabrication techniques should be tested on samples subjected to various thermal cycles and stress loads.

Embrittlement by the formation of intermetallic compounds can also result from diffusion of materials at a joint interface. This process, which occurs in intermetallic-compound-forming materials that are joined by pressure bonding and operated at high temperature, is a frequent cause of failure of emitter structures.

Another diffusion phenomenon that can cause failure is the Kirkendall effect. Here considerably different mutual diffusion rates of joined materials can cause porosity in pressure-bonded and brazed joints.

One of the most complex areas of metallurgy in converters is that of ceramic-to-metal bonding at the insulating seals. Bonds between a ceramic and a metal are usually achieved either directly by pressure bonding or pyrolytic deposition or indirectly by such processes as active alloy brazing or metallizing and brazing. The great variety of chemical compounds and their mixtures which can be formed at the interface areas is often the source of difficulty in perfecting or using a bonding process. This is particularly applicable when composition, purity, cleaning procedures, furnace atmospheres, and thermal treatments are not carefully controlled.

Metallizing and brazing is one of the most frequently used processes for ceramic-to-metal bonding. Metallizing mixtures usually consist of powders of either molybdenum and titanium or molybdenum and manganese. Calcium oxide is often added to either the metallizing mix or the ceramic-body composition to promote calcia–alumina reactions. The metallizing mix is first applied as a paint, with a suitable organic vehicle, to the ceramic surface. Sintering is then carried out in a wet hydrogen atmosphere at about 2000°K. After this step the metallized surface can be plated and brazed as if it were a conventional metal surface.

16.5 Fabrication Processes

16.5.1 Material-Forming Processes. There are three broad categories of material-forming processes: *machining, forming,* and *deposition.*

Machining is a process that removes metal by mechanical, electrical, or chemical means to produce a part of a desired form. For example, metal cutting (milling, grinding, stamping, etc.), electrical discharge machining (EDM), electrochemical machining (ECM), and photoetching are machining processes.

Metal cutting can be difficult with both soft materials, such as annealed tantalum and nickel, and hard materials, such as molybdenum and tungsten. Soft materials tend to tear and gall, and hard materials cause rapid wear and breakage of tools. For this reason cutting a part with accurate dimensions and a good surface finish can be time-consuming and expensive. For soft metals the difficulties can be reduced by cutting in the unannealed condition, with the possible disadvantage of loss in dimensional accuracy when the part is subsequently heat treated. For hard metals favorable cutting conditions can be obtained by heating the metal above the brittle–ductile transition temperature. In some instances, however, rather elaborate set-ups are required to heat the metal adequately.

Electrical discharge machining (EDM) requires expensive equipment. The expense is often justified, however, because of the high cost of machining refractory metals with conventional machine tools. In addition to facilitating the machining of very hard materials, such as rhenium and tungsten, EDM permits us

to shape a piece of metal without applying tool forces; thus delicate shapes can be produced with minimum risk of breakage. However, the electrical sparking used in the process produces high stresses at a microscopic level. The stresses can induce such damages as microcrack patterns or checking on the finished product. High-dimensional accuracy is obtained, but, because of wear, frequent changes of the machining electrode and electrical adjustments are required. Generally, the EDM process is economical for cutting rhenium and for drilling small deep holes and cutting especially complicated shapes in any material.

The electrode wear problems of EDM are avoided in the electrodeplating or electrochemical (ECM) process. The ECM process is costly, however, and does not attain the accuracy of EDM. Because of these and other practical limitations, the ECM process is not used in fabricating converter components.

Forming is a process employing dies and patterns to change the shape of a piece of metal. For example, spinning, forging, powder metallurgy, and casting are forming processes.

Powder metallurgy is a relatively expensive process because it requires densification at high pressures and temperatures. It is used mostly in fabricating either rare materials or composite structures such as cermets or uranium oxide. Parts with high purity, a fine grain structure, and good mechanical strength can be obtained.

Deposition is a process employing atomic streams impinging on a substrate to build up a desired form. For example, pyrolytic deposition, electroforming, electroplating, and vacuum deposition are deposition processes.

Pyrolytic deposition is carried out by decomposition of vaporized metal compounds on contact with heated surfaces. The deposits so obtained usually exhibit a high degree of grain orientation and can be exceptionally pure. Adherence to the substrate is relatively poor unless great care is exercised to prevent surface contamination and chemical reactions with the substrate. Deposition conditions and the heating methods suitable for each specific metal deposit and substrate are usually determined by experiment.

16.5.2 Cleaning and Handling Processes. Cleaning machined parts involves a series of operations designed first to remove all surface contaminants by chemical cleaning and then to remove the bulk of the dissolved gases by heating in a selected atmosphere. Surface contamination usually results from lubrication during machining, contact with other metals during fixturing and inspection, use of abrasives, and hand contact.

Since the exact composition of the contaminants is unknown, the procedure generally followed in chemical cleaning begins with a degreasing treatment, such as condensation of trichloroethylene vapors, to remove the bulk of the lubricants. Foreign particles, such as metal chips and threads of rags used to wipe the part, are then removed by immersion in an ultrasonically agitated degreasing

solution, such as freon, trichloroethylene, or a detergent. Finally, a very thin layer of the material itself is removed by reacting it with a selected chemical solution; this is followed by a thorough rinsing and drying process. For many parts of a converter, removal of material during chemical cleaning must be reflected in the specification of mechanical dimensions and tolerances. Moreover, the chemical solution used should not react at appreciably different rates with material of different grain orientations. Because this requirement is difficult to meet, an electropolishing treatment is sometimes used in place of or in addition to chemical action by a reagent.

The effectiveness of surface-cleaning treatments can be evaluated with a water-wetting test, which consists in immersing the part in distilled water, withdrawing it, and observing whether water clings as a film or as segregated droplets. With only a few atoms of foreign substances present, the surface tension is sufficiently altered to cause a residual film of water to break up into drops.

Dissolved gases (and sometimes nongaseous impurities) are removed by heating the parts in an atmosphere selected to promote continuous withdrawal of the undesirable elements from the exposed surfaces. Parts should always be fired at a temperature exceeding their design operating temperature. The duration of the firing must take into account the kind and amount of gases and impurities that may be present initially, the thickness of the part at various sections, and the rates of diffusion attainable at the selected firing temperature. Firing in a very pure hydrogen atmosphere at as high a temperature as possible is recommended[6] provided hydrogen produces no harmful effects on the properties of the metal.

After they are fired in hydrogen, parts should be fired in vacuum. For some metals vacuum firing alone is necessary or desirable. Nickel must be fired in wet hydrogen because the principal impurity to be removed is carbon. Carbon reacts with the water vapor in the hydrogen and produces carbon monoxide (water–gas reaction). In this method the rate of outgassing, although still slow, is a hundred times faster than in vacuum.

After the outgassing treatment the *clean part* must be handled with the utmost care. It should be stored in a dry atmosphere at 400 to 500°K to avoid condensation of moisture from the air, and it must be protected from dust precipitation. Because a clean part is difficult to preserve for more than a few days, parts should be processed relatively close to the time of assembly.

16.5.3 Joining Processes. The assembly of a converter is based to a large extent on leak-tight joining of parts. Joining is accomplished by various forms of welding, bonding, or brazing. The many types of joining processes are discussed extensively in the literature and in the manuals prepared by metal suppliers. Our discussion is restricted to only a few processes and materials important to converter joints.

Although many early converter models were assembled entirely by brazing, welding has become the most widely used metal joining technique. Where welding can be used, it offers the advantage of rapid assembly, minimum heating of parts not related to the joint, and the formation of joints that are nearly as strong as the metals joined. Also, converter parts can be assembled by welding in practically any desired sequence without much concern for the effects of weld heat on brazes or other materials.

Welding is usually accomplished with an electron gun in vacuum or with a pointed electrode in an inert atmosphere. In vacuum the process is called *electron-beam welding* and is used in preference to *inert-gas welding* when the materials to be joined are highly sensitive to impurities, when a high degree of penetration is desired, or when the configuration of the parts is such that inert-gas welding would lead to problems in controlling the location and degree of melting. Because it does not affect the vacuum in a closed envelope, electron-beam welding is especially suited to the fabrication of heat pipes and containers for pressure bonding of parts. Electron-beam welding is discussed extensively in Ref. 7.

Inert-gas welding is usually performed with a thoriated tungsten electrode in argon. Argon is used because it is relatively inexpensive and can sustain an arc at low welding currents (below 60 amp). High-quality welds can be performed on impurity-susceptible materials, such as niobium. The welding is carried out in a leak-tight chamber which is vacuum-purged two or three times before the argon is admitted. In inert-gas welding the weld quality is more operator dependent than in electron-beam welding because, for a good weld, the distribution of arc power between the metals being joined and the arc length should be maintained constant. This requirement is difficult to achieve because arc power distribution and arc length cannot be measured conveniently. Also, inert-gas welding requires that the parts being joined be of similar cross section. For example, it is difficult, if not impossible, to weld a thin tube to a heavy plug.

In both electron-beam and inert-gas welding, weld quality is highly dependent on the configuration of the area of the joint. Parts should fit with a clearance of no more than 10% of the width of the material to be melted, and the material with the higher thermal-expansion coefficient should be used on the inside of a joint. Materials must be very clean and well outgassed before welding to ensure sound and nonporous welds.

Nickel and its alloys (e.g., Kovar and Driver Harris) yield porous welds when welded in an atmosphere that is not oxygen free because, since it is impossible to remove all the carbon impurity from nickel, carbon monoxide is released at the joint area during welding.[5]

Nearly any material can be welded to itself. Certain combinations of materials cannot be welded, however, because brittle intermetallics form and cause weld

cracks on cooling. For example, niobium cannot be welded to nickel-bearing alloys because a eutectic of $NbNi_3$ and NbNi forms at 1448°K. Again, the crack sensitivity of high-nickel-content ferrous alloys prevents welding of nickel to either Kovar or Driver Harris. Some weld configurations are less susceptible to crack development. For example, stainless steel can be welded to either Kovar or Driver Harris when the joint is a cylindrical shell that accommodates expansion differences with minimal stress.

Bonding is a process resulting in a leak-tight joint when high pressure and, usually, heat are applied. The principal forms of bonding are *cold welding* and *isostatic pressure bonding.* Cold welding is used extensively as a technique for sealing off converter envelopes by pinching a copper conduit with a pair of hydraulically activated, hard steel rollers.

Isostatic pressure bonding is used when a leak-tight joint is desired over extended surfaces that must transport significant amounts of heat. Joining is done in a high-pressure chamber at high temperature. Typical values of pressure and temperature are 700 atm at 1700°K.

Brazing involves the joining of metals with a filler metal. The filler melts at temperatures generally above 1000°K and flows into the joint by capillary action. The filler material must have good wetting characteristics, a suitable melting point, and a low evaporation rate at the temperature of operation of the joint. In addition, it must not form undesirable alloys with the metals to be joined. Often, the filler material must also be resistant to attack by cesium. Brazing techniques are discussed extensively in Ref. 3.

In converter fabrication the use of the brazing process is hindered by the tendency of the surfaces of some materials to form tenacious oxides that prevent wetting by the filler metal. One such material is niobium, which is used extensively in fabricating envelope seals because it matches the thermal expansion of most useful ceramics. The difficulty can be reduced by abrading the surface of the niobium just before brazing to remove the layer of oxide mechanically. Palladium and vanadium can be used as fillers. When palladium is used, substantial melting of niobium must be allowed for at the joint because niobium and palladium form extensive solutions near the melting point of palladium (1825°K).

Another difficulty with the brazing process is the tendency of some materials to form brittle intermetallics. One such material is rhenium, which is used for electrodes because of its good thermionic-emission characteristics and its strength at high temperatures. Rhenium forms brittle intermetallics with practically every braze material of interest. The only known satisfactory braze material at the high temperatures of operation of rhenium is vanadium, which has a melting point of about 2120°K. Because the formation of brittle intermetallics is a problem also in rhenium welding, rhenium is usually pressure bonded or welded only to itself.

Parts to be brazed are usually heated in a furnace or by radio-frequency induction in a bucket. For some assemblies the brazing must be performed with localized heating. Then either an RF concentrator or electron-bombardment heating can be used.

A useful variation of brazing is *fuse brazing*,[8] in which the brazing is done by melting one of the metals to be joined and using it as the braze filler material. Fuse brazing is used particularly for joining copper conduits to envelopes of converters. In this application the copper conduit is thinned at its end to facilitate localized melting and fitted around a smaller diameter connecting tube of high-purity nickel or nickel-plated tantalum or niobium. The fuse braze is performed in a period of seconds with an RF concentrator.

16.5.4 Converter Outgassing Processes. Outgassing of an assembled converter is the process of removing the gases that are adsorbed on all parts inside the envelope of the converter. Outgassing of converters is accomplished by the techniques used to outgas magnetrons and similar metal–ceramic devices. All parts are heated above their design operating temperatures for several hours while the device is connected to a vacuum pump.

When the envelope of the converter cannot be exposed to the atmosphere at its design temperature, the outgassing is performed in vacuum with a system such as that shown in Fig. 16.11. This figure shows that two vacuum pumps are used, a main pump and an outgassing or auxiliary pump. Initially, the main pump is used to evacuate the auxiliary pump to its required start-up level, and then the valve between the two pumps is shut off and the auxiliary pump is turned on.

Usually the auxiliary unit is an ion pump having a pumping speed of about 5 liters/sec. This type of pump has the advantage that the pressure can be measured electrically at the pump chamber. This measurement gives an indication of the degree of completion of outgassing of the converter.

It is noteworthy that the pressure at the ion pump chamber is much smaller than that in the converter envelope because of the flow resistance of the evacuation line. For a typical copper conduit 10 in. long and ¼ in. in inside diameter, the conductance is 0.2 liter/sec. It follows that, at low pressures and for an effective pumping speed of 2 liters/sec, the pump pressure is of the order of 10 times smaller than the pressure in the converter.

The pumping speed decreases at low pressures because part of the available pumping capacity is consumed by the gas load due to self-outgassing of the ion pump. The time required for converter outgassing can be estimated by plotting the ion pump pressure vs. time and extrapolating the graph to the time when the pump pressure will become 10^{-8} torr. If the ultimate pressure of the ion pump is larger than 10^{-8} torr, the time for outgassing is taken equal to the time corresponding to twice the ultimate pressure.

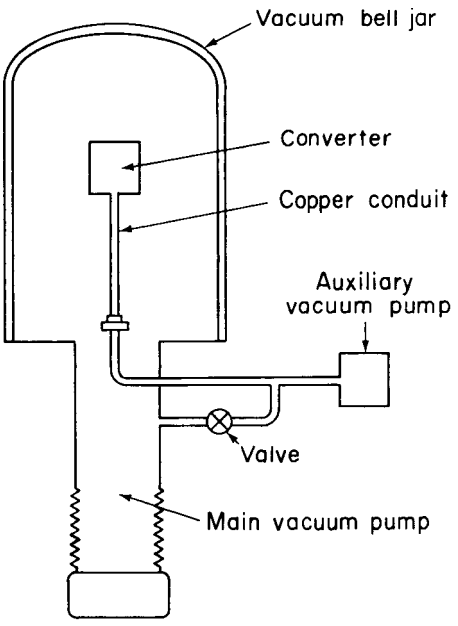


Fig. 16.11 Vacuum system for converter outgassing.

The final steps of converter outgassing are usually outgassing of the cesium-charging ampul and crimping off of the converter from the outgassing station. Although cesium ampuls are sealed in vacuum by most manufacturers, they contain small traces of gas. The ampul-outgassing procedure consists in rupturing the ampul while the converter is still connected to the pumping station. The ampul is ruptured by distorting the walls of the enclosure housing it; the outgassing lasts only a few seconds. The converter is severed from the auxiliary pumping line by crimping off the evacuation conduit. During this operation, as during rupture of the ampul, atmospheric pressure surrounds the converter, and the auxiliary pump can be used as an excellent detector of any leaks that could have developed during outgassing or gone undetected during assembly.

16.5.5 Cesium-Charging Processes. The required amount of cesium is usually introduced into the envelope of a converter by distillation from a container. The container is a small glass or metal ampul housed in a copper conduit so that it can be easily ruptured for outgassing, easily heated for distillation, and easily crimped off when the charging of the converter is completed. Two distillation methods are shown in Fig. 16.12. In the first distillation occurs directly between

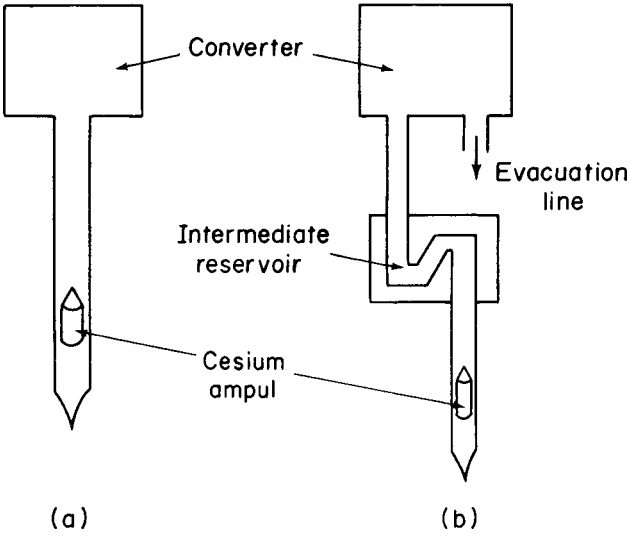


Fig. 16.12 Appendages for cesium distillation.

the ampul and the converter, and the same copper conduit is used both for outgassing and ampul housing. When this is done, the ampul should not block the outgassing line.

In the second distillation method, cesium is distilled to an intermediate reservoir and from there to the converter. When this is done, the converter envelope can be pumped continuously during the first distillation; this method allows for removal of some of the gases released by cesium.

The distillation of cesium must be accomplished at a temperature high enough to achieve rapid mass transfer and yet not high enough to endanger the converter and cause a chemical reaction between glass and cesium if a glass ampul is used. The distillation temperature is usually between 500 and 600°K.

In most converters the distillation is not as high as anticipated from evaporation data. In practice, the distillation temperature is monitored by a thermocouple on the conduit containing the cesium. A delay of unknown duration occurs before the cesium reaches conduit temperature and starts evaporating at the rate dictated by this temperature.

After cesium distillation is completed, the appendage used for charging is crimped off. Since cesium is solid at room temperature, solid cesium could be present at the region of the crimp-off. If this is suspected, the crimp-off area must be heated above the melting point of cesium (301.7°K) to ensure leak-tight bonding of the copper conduit at the crimp-off.

References

1. P. Brosens and I. Shai, The Calculation of Emitter Support Conduction Loss, *IEEE Thermionic Conversion Specialists Conference, Boston, Oct. 21–24, 1968,* pp. 272-276, Institute of Electrical and Electronics Engineers, 1968.
2. A. Goldsmith et al., *Handbook of Thermophysical Properties of Solid Materials. Volume I. Elements,* Pergamon Press, Inc., New York, 1961.
3. W. Kohl, *Materials and Techniques for Electron Tubes,* Reinhold Publishing Corporation, New York, 1960.
4. G. Gubareff, J. Jansson, and R. Torborg, Thermal Radiation Properties Survey, Minneapolis-Honeywell Regulator Co., Minnesota, 1960.
5. General Electric Company, Research and Development Program of Thermionic Conversion of Heat to Electricity. Volume I. Final Report on Contract NO-bs-77578, June 1964.
6. S. Dushman, *Scientific Foundations of Vacuum Technique,* John Wiley & Sons, Inc., New York, 1962.
7. R. Bakish and S. White, *Handbook of Electron Beam Welding,* John Wiley & Sons, Inc., New York, 1964.
8. The International Nickel Co., Inc., Fusion Welding of Nickel and High-Nickel Alloys, Technical Bulletin T-2, New York.

17

Thermionic Power System Engineering

17.1 General Remarks

The types of thermionic power generators that have received attention are solar thermionic generators, radioisotopic thermionic generators, nuclear reactor thermionic generators, chemical thermionic generators, and topping thermionic generators.

Engineering a thermionic power generator as part of an overall system involves successive iterations and compromises to bridge the gap between the characteristics of the generator and the requirements of the overall system.[1] In this process the engineer–designer cooperates with the designer of the overall system. The overall-system designer evaluates the discrepancies between realizable and desired characteristics, determines the acceptability of a proposed generator, and decides on the need to modify the design and revise the specified requirements. The interrelations between various technical procedures and engineering judgments leading to a final generator design can be represented by a feedback block diagram as shown schematically in Fig. 17.1.

Typically, design specifications include a description of the heat source and heat sink communicating with the generator and the engineering design information required to modify the source and sink if modification is necessary and permissible. The specifications also include the desired performance in terms of power, power per unit weight and volume, efficiency, control and start-up, reliability, and life. Finally, they include safety considerations, acceptance procedures, and cost limitations. The relative importance of each of these specifications depends on the type of thermionic generator in question and its intended application. For example, designers of nuclear reactor and radioisotopic thermionic generators must be more concerned with safety than must designers of solar and chemical thermionic generators, and designers of generators for applications in space must be more concerned with weight than with cost.

This chapter describes the general configuration for each type of thermionic generator, outlines problems arising at the interface of heat source and converter, and presents typical examples of hardware that has been developed.

17.2 Solar Thermionic Generators

Solar thermionic generators attracted attention at a time when solar photovoltaic power systems appeared to suffer seriously from exposure to high

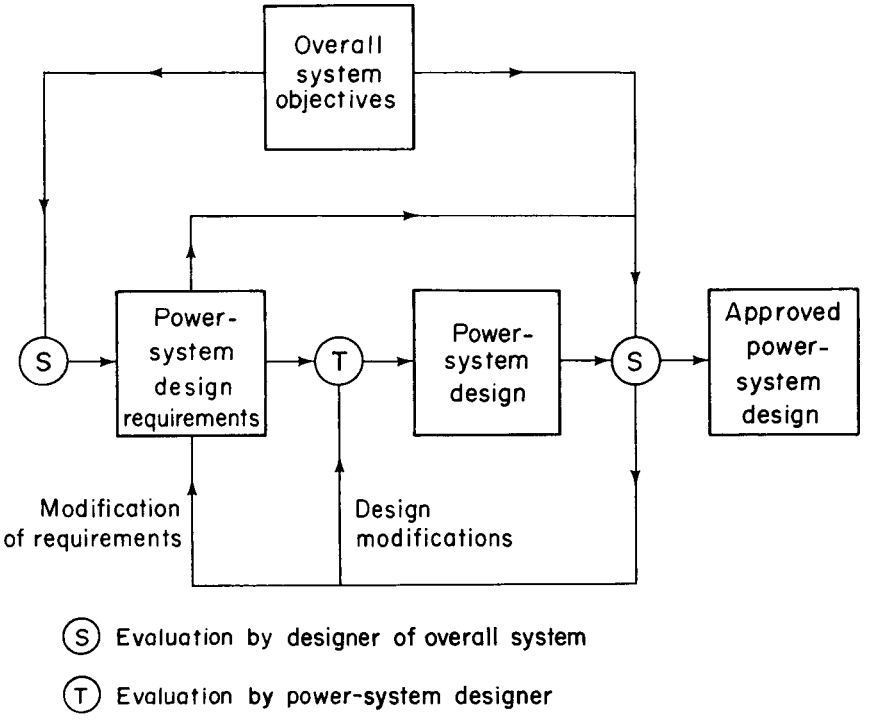


Fig. 17.1 Iterative power-system design process.

temperatures, radiation, and high-energy particle fields. They are the simplest of all thermionic systems in that heat transfer from the heat source, the sun, is achieved directly through the vacuum of outer space. Unlike those in all other thermionic generators, the high-temperature components of the converters do not contact other materials or atmospheres. Because the solar heat source is easily coupled to converters, the technology of solar thermionics advanced rapidly and yielded much of the information that was later used in designing other systems.

17.2.1 General Configuration. The typical components of a solar thermionic generator are shown in Fig. 17.2. Solar energy is collected by a *parabolic mirror* (concentrator) at a rate adjusted by means of a *solar-flux control.* In this manner a relatively fixed amount of heat flux is focused inside a *cavity* formed by adjacent faces of converters. The output power of the converters is conducted to suitable power-conditioning equipment.

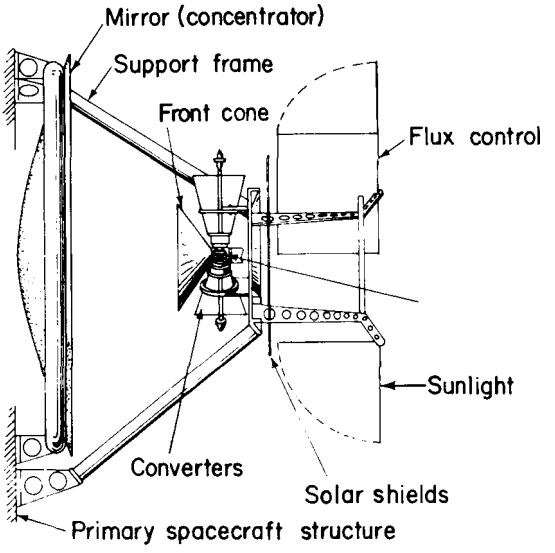


Fig. 17.2 Solar thermionic generator.

Because the required cavity-wall temperature is of the order of 2000°K, reradiation losses through the *cavity aperture* are high. These losses can be reduced only by reducing the aperture size. The smallest possible value of the aperture size is determined by the expected errors in the shape and orientation (*boresight error*) of the mirror and by the magnitudes of the focal length of and the distance of the sun from the mirror. This size, in turn, defines the largest possible number of converters of practical dimensions which can be assembled around the cavity.[2,3]

Analyzing the effects of the inaccuracies in shape and orientation of the mirror indicates that, for acceptable performance at thermionic temperatures, the sum of the mirror *standard-deviation-slope error* plus the boresight error should not exceed ½ of the collimation angle of the sun. Only mirrors of high accuracy can satisfy this requirement. Experience shows that, when fabricated by means of electroforming techniques, mirrors 5 and 9 ft in diameter tend to have standard-deviation-slope errors about 2 and 9 minutes of arc, respectively.[4,5]

The smallest practical generator output power is determined by the output power of the smallest converter that can be built without disproportionate heat loss. The power rating of the smallest practical converter has been found to be about 25 watts, and, since at least four converters are required to form a reasonable solar cavity, the corresponding minimum generator output power is about 100 watts.

The required minimum-size mirror for the smallest generator can be calculated as a function of its distance from the sun. For the data cited and by virtue of the known solar-energy flux [130 watts/sq ft at the mean sun-to-earth distance of 93 million miles, or 1 Astronomical Unit (AU)], it follows that the mirror must be at least 1 ft in diameter for each 30 million miles of distance from the sun.

When the orientation error is 4 minutes of arc, the preceding considerations indicate that the orbit of Mars is the greatest distance from the sun at which solar thermionic generators can be operated efficiently.

The closest possible distance from the sun at which the generator can operate is determined chiefly by the highest allowable temperature for either the mirror or the solar-flux control flaps, or both. At a small distance from the sun, the solar flux reflected by the mirror is so intense that any exposed material must withstand very high temperatures. The equilibrium temperature reached by a uniform flat plate exposed to the sun as a function of distance from the sun, surface properties, and angle of incidence of radiation is shown in Fig. 17.3. For untreated refractory-metal surfaces, a typical value of the ratio of solar absorptivity α over total emissivity* ϵ is unity. For a distance of 0.1 AU, the equilibrium temperature of the surface will be about 1250°K.

17.2.2 **Solar-Cavity Design.** Early designs were based on the assumptions that the reflections of solar radiation in the cavity would ensure a uniform input-heat flux to the converters and that the cavity would reradiate blackbody radiation through its aperture. Analysis of the absorption and reflection of solar energy and of reradiation of energy by the cavity showed, however, that these assumptions can lead to substantial errors in generator design. Because of the high absorptivity (about 0.45) of most refractory metals at the wavelengths of the solar spectrum, the cavity surfaces exposed directly to solar energy are apt to overheat. Moreover, the input-heat-flux density required for converter operation is considerably higher than the heat-flux density corresponding to a blackbody at the average temperature of the cavity walls. Consequently the heat reradiated through the aperture tends to be larger than the heat of a blackbody at cavity temperature.

To calculate properly the distribution of thermal fluxes, we must have an analytic description of the reflective characteristics of each point of the cavity surface as a function of angle of incidence and wavelength. We also need detailed knowledge of the intensity and direction of the flux produced by the mirror at the cavity aperture and of the temperature at each point of the cavity surface. Because the calculation is complicated, we have obtained approximate results by

*Note that ϵ stands for the sum of front-side and rear-side total hemispheric emissivities.

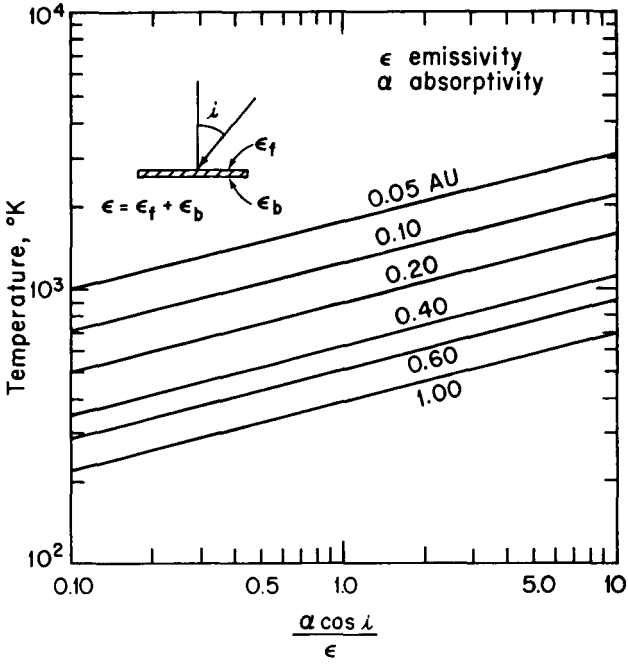


Fig. 17.3 Solar exposure temperature of uniform flat plates.

assuming that (1) the solar-energy flux enters the cavity as though it has radiated uniformly from a point on the optical axis in front of the aperture such that the angle subtended by the aperture from this point is equal to the rim angle of the mirror, (2) the cavity surfaces behave like perfectly diffuse reflectors and radiators, and (3) the cavity surfaces are at the desired temperature of the emitters of the converters. By virtue of these assumptions and by detailed considerations of the cavity-surface view factors and of the differences in radiative properties at solar-thermal-flux wavelengths, we can determine whether the net heat absorbed by every cavity-surface element is enough to maintain the element at its assumed operating temperature.[2]

In a practical generator the level at which this thermal balance occurs can be adjusted by varying the radiative properties of the cavity walls by such techniques as machining of fine grooves or changing the operating point of the converters.[2]

The efficiency of the generator can be described in terms of the efficiencies of each of the three major components shown schematically in Fig. 17.4. Use of the energy rates shown in the figure yields the relations

$$\text{Mirror efficiency} = \frac{Q_2}{Q_0} = \eta_m$$

$$\text{Cavity-absorber efficiency} = \frac{Q_2 - Q_3}{Q_2} = \eta_{ca}$$

$$\text{Heat-utilization factor} = \frac{Q_5 - Q_6}{Q_2 - Q_3} = \eta_h$$

$$\text{Converter ideal efficiency} = \frac{P}{Q_5 - Q_6} = \eta_i$$

$$\text{Generator efficiency} = \eta_m \times \eta_{ca} \times \eta_h \times \eta_i = \frac{P}{Q_0} = \eta_g$$

$$\text{Conversion-system efficiency} = \eta_h \times \eta_i = \frac{P}{Q_2 - Q_3} = \eta_s$$

Typical values of η_m, η_{ca}, η_h, and η_i are

$\eta_m = 0.85$ $\eta_{ca} = 0.90$ $\eta_h = 0.75$ $\eta_i = 0.20$

and therefore

$\eta_g = 0.11$ and $\eta_s = 0.15$

The advantage of the preceding definitions lies in the introduction of the concept of the heat-utilization factor η_h which represents essentially the effect of all the heat losses that can be affected by the design.

Solar-energy-flux-control requirements strongly depend on the range of solar distances for which the generator is designed. At any given value of input-heat flux, control of cavity temperature can be achieved by varying the converter load conditions. This point is illustrated by the discussion that follows.

Typical current vs. voltage characteristics, shown in Fig. 17.5, are obtained from the converter shown in Fig. 17.6 under conditions of optimum cesium-reservoir temperature but free-floating collector temperature* for each operating point. The emitter temperature vs. input-heat flux is shown in Fig.

*The collector is allowed to reach the equilibrium temperature dictated by the inherent heat-transfer characteristics of the converter. The collector temperature corresponding to the curves in Fig. 17.5 is optimum at output currents of about 35 amp.

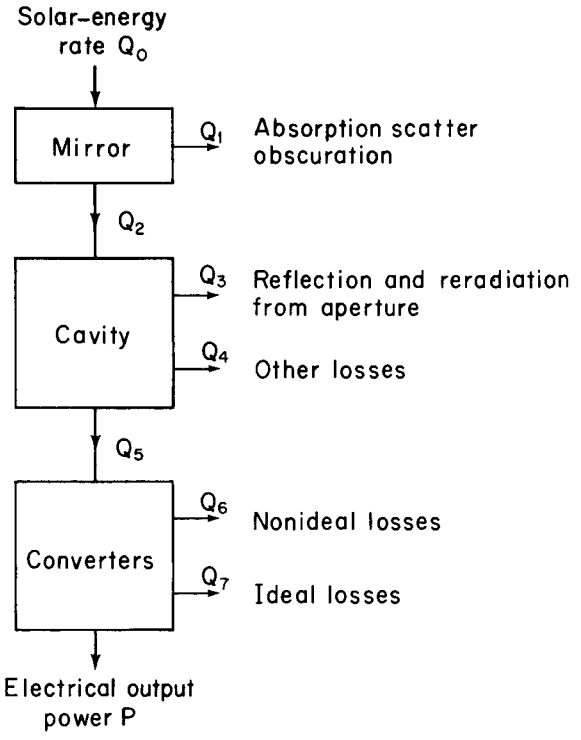


Fig. 17.4 Solar thermionic generator heat-transfer diagram.[2]

17.7 for both constant-output-voltage and constant-output-current conditions of operation.

Figure 17.7 shows that, for a given input-heat flux, the emitter temperature can be controlled by changes of the converter load conditions. Moreover, the figure indicates that, for a given change in input-heat flux, the change in emitter temperature is larger when the converter is operated under conditions of constant output current than under conditions of constant output voltage. Large excursions in emitter temperature are undesirable. For this reason it is preferable to control the generator to maintain a fixed output voltage. A value of 0.8 volt per converter gives the most uniform response over a broad range of input-heat fluxes.

Temperature control of the cesium reservoir can be relatively simple. For the converter shown in Fig. 17.6, a graph of the output current at maximum power vs. the cesium-reservoir temperature for this power is shown in Fig. 17.8. The figure shows that the relation between these two quantities is almost

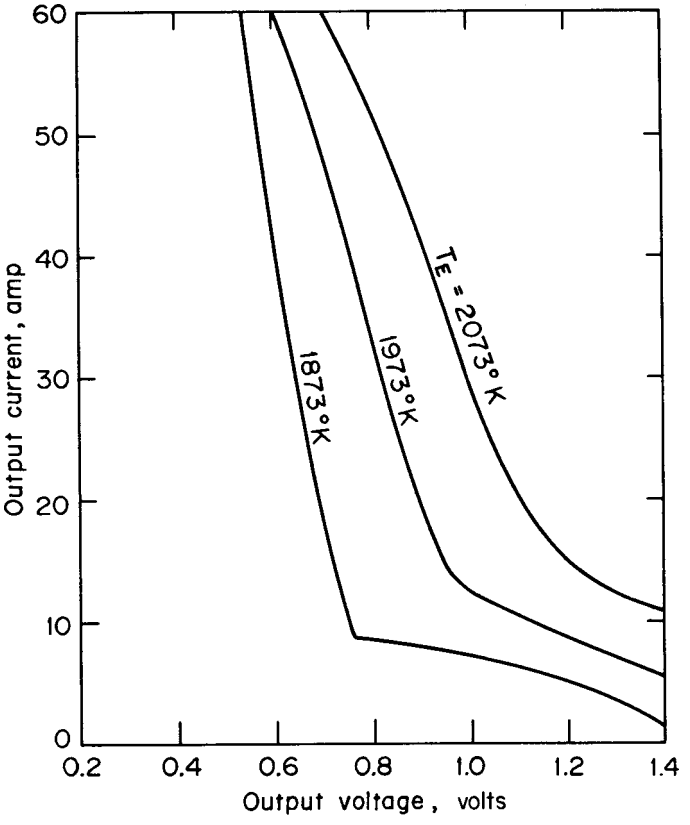


Fig. 17.5 Converter $I-V$ characteristics.[6]

independent of other conditions of operation; this is typical of the performance of many converters. It follows that the characteristic shown by the dashed line in the figure can be used to adjust the reservoir temperature in response to a measurement of the output current.

The reservoir temperature excursion required by a change in output current from 20 to 40 amp is only 17°K (Fig. 17.8). A fraction of the required temperature excursion occurs spontaneously because of changes in converter heat-transfer processes accompanying an increase in output current. The remaining fraction of the excursion is achieved by means of an electrical heater. The power needed for such a heater is small (less than 1 watt), as illustrated by Fig. 17.9. The data in Fig. 17.9 were obtained from a converter (Fig. 17.6) in which no effort was made to reduce the power required for electrical control.

Fig. 17.6 Solar thermionic generator prototype.

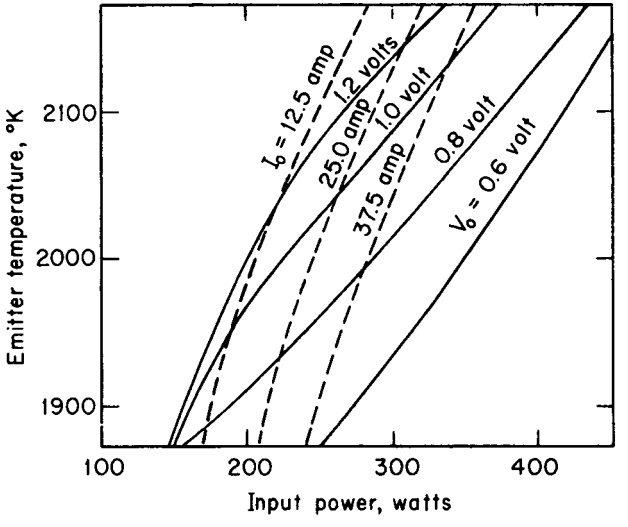


Fig. 17.7 Thermal response of converters.[6]

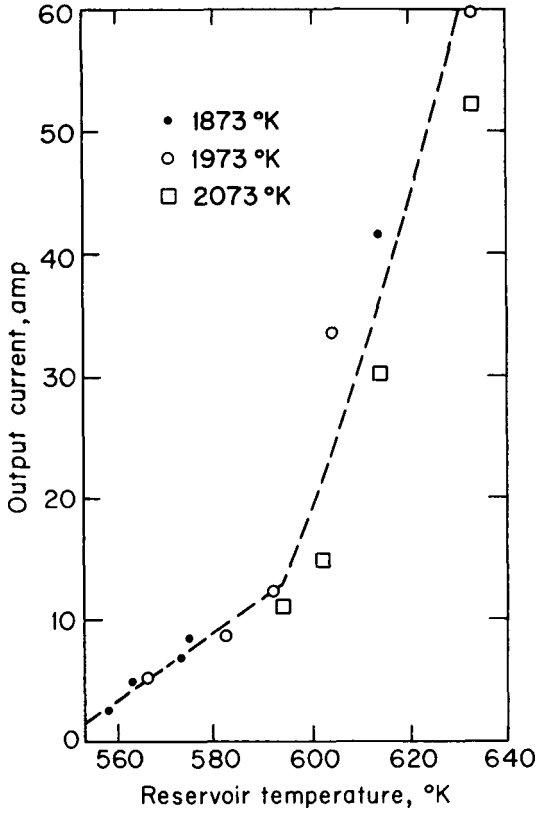


Fig. 17.8 Output current vs. optimum cesium-reservoir temperature.[6]

17.2.3 **The Solar-Energy Thermionic Generator.** The most intensive effort to evaluate solar thermionic generators as auxiliary power sources for space[7] is represented by the Solar Energy Thermionic (SET) program. Compared with all similar programs, the SET effort has the longest history and has explored most fully the potential of solar thermionics.[4,6,8-16] The initial SET generator design incorporated a 5-ft-diameter parabolic mirror made of electroformed nickel and a cavity made up of four or five converters. A plan view of the converter arrangement is shown in Fig. 17.10. The mirror diameter was dictated by the space available at the top of the Agena stage rocket, then being used for space exploration, and by the fabrication capability provided by the then available glass masters, which had been used for electroforming searchlight mirrors during World War II. Later the program incorporated designs with a 9.5-ft-diameter mirror and a cavity consisting of 16 converters.

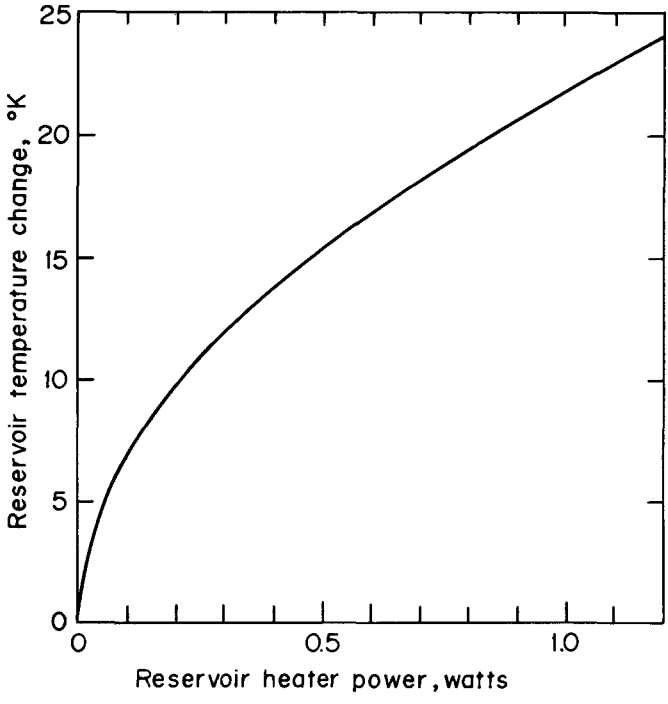


Fig. 17.9 Cesium-reservoir thermal characteristic.[6]

Advanced converter designs developed for the SET program, with or without heat pipes (Fig. 17.11) for heat rejection, can be found in the literature.

17.2.4 Solar Probes. Studies of solar thermionic generators have consistently shown that operating in close proximity to the sun enhances performance. Where such proximity is required, solar thermionic generators are particularly attractive because of their ability to withstand the expected severe thermal environment. Also, near the sun the requirements on shape and orientation errors are met with relative ease, and the mirror weight per unit of power output becomes very small. As a result, solar thermionic generators are suitable for use in Solar Probe and Mercury flyby missions.

Modular weight profiles for Solar Probe and Mercury flyby missions are listed in Table 17.1 for modules such as the one shown in Fig. 17.2, which consist of converters with heat pipes for heat rejection (Fig. 17.11). The characteristics of two four-module, hybrid, photovoltaic–thermionic generators designed to provide a minimum of 200 watts of power are listed in Table 17.2.

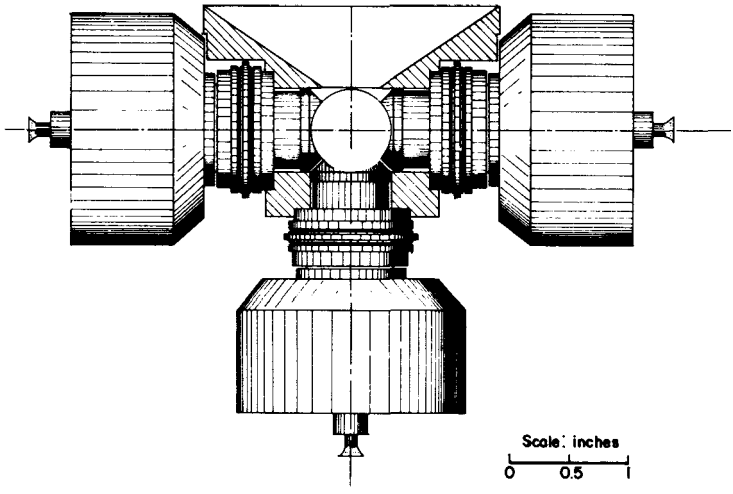


Fig. 17.10 Plan view of the Solar Energy Thermionic generator.

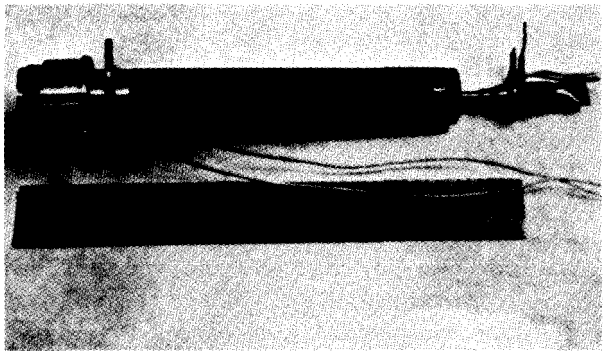


Fig. 17.11 Solar Energy Thermionic heat-pipe converter (weight, 110 g).

Table 17.1 Module Weight Profile[17]

Component	Solar Probe	Mercury flyby
Mirror and support	0.44 lb	2.34 lb
Attachment brackets	0.13	0.35
Support arms	0.10	0.69
Converter structure	0.43	1.19
Four converters	0.43	1.19
Output leads	0.32	1.20
Converter shielding	0.29	0.58
Flux control	0.83	1.59
Cesium-control electronics	0.20	0.20
Total	3.17 lb	9.33 lb
Module minimum power* (net)	56 watts	114 watts
Specific power	17.7 watts/lb	12.2 watts/lb

*Allows 2 watts for cesium-reservoir control.

Both generators include a 250-watt photovoltaic subsystem. In the 0.1-AU Solar Probe mission, the subsystem is controlled by continuous tilting until the panel temperature requires that the panels be folded away from the sunlight, whereas, in the Mercury flyby mission, the panels are maintained at full exposure until folded away.

17.3 Radioisotopic Thermionic Generators

The heat source of radioisotopic thermionic generators is a material heated by the decay of a radioactive element. All the input energy required for the life of the generator is stored in it during fabrication. For this reason radioisotopic thermionic generators are suitable for use in remote areas such as the Arctic, the ocean, and outer space. Moreover, they are practical for applications that require electrical power at a rate up to several kilowatts.

In the United States the development of radioisotopic thermionic generators has been based on modular designs.[18] A module consists of a fuel capsule, a converter, heat-transfer paths for the thermal coupling of the fuel to the converter and of the converter to the sink, a hermetically sealed housing to provide a vacuum environment for the high-temperature components of the module, and thermal shields to minimize heat losses. The design of a module is influenced by the inherent characteristics of each of its components. These characteristics and some module designs are discussed in Secs. 17.3.1 to 17.3.3.

17.3.1 Radioisotopes as Heat Sources or Fuels. Radioisotopes decay at rates fixed by the nature of their nuclei. As heat sources they are highly reliable and have accurately predictable characteristics.

Table 17.2 Design Characteristics of Photovoltaic–Thermionic Generators[17]

Characteristic	Solar Probe	Mercury flyby
Mirror diameter, in.	13	33
Mirror reflectivity	0.75	0.85
Blockage factor	0.63	0.85
Required orientation accuracy, deg	±0.30	±0.15
Converters per module	4	4
Cavity temperature, °K	1960	1980
Output current, amp	18.75	37.5
Converter voltage, volts	0.80	0.80
Module voltage, volts	3.10	3.10
Net module power, watts	56	114
Cesium-reservoir power, watts/module	2	2
Operating solar range, AU	0.30 to 0.10	0.70 to 0.35
Module weight, lb	2.97	9.13
Number of modules	1	4
Power-system voltage, volts	6.20	6.20
Total power, watts (thermionic)	224	456
Power-system weight, lb	12.7	37.3
Number of parallel connections	2	2
Specific power, watts/lb	17.7	12.2
Solar-panel power, watts	250	250
Solar-panel control	Tilting	Fixed
Solar-panel fold-away distance, AU	0.3	0.7
Solar-panel area, sq ft	29.8	29.8
Solar-panel weight (1 lb/sq ft)	30.0	30.0
Combined system weight, lb	42.7	67.3

Table 17.3 lists the principal characteristics of radioisotopes that have a sufficient rate of energy release and a sufficiently long half-life to be potentially suitable fuels in thermionic generators. These radioisotopes are produced in nuclear reactors either as fission products or by neutron irradiation of parent materials. For example, ^{90}Sr, ^{106}Ru, ^{144}Ce, ^{137}Cs, and ^{147}Pm are fission products, and ^{60}Co, ^{170}Tm, ^{210}Po, ^{238}Pu, ^{242}Cm, and ^{244}Cm are produced by neutron irradiation of parent materials.

The availability of radioisotopes tends to depend on the type of nuclear power plants in service and the separation techniques that have been developed. The power density of radioisotopes is determined by the form in which they can be used at the required converter temperatures and by the presence of materials that cannot be separated economically from the radioisotope.

Table 17.3 Characteristics of Radioisotopic Heat Sources[19]

	Class of emitter										
	Gamma	Beta						Alpha			
Characteristic	^{60}Co	^{90}Sr	^{106}Ru	^{137}Cs	^{144}Ce	^{147}Pm	^{170}Tm	^{210}Po	^{238}Pu	^{242}Cm	^{244}Cm
Half-life, years	5.3	27.7	1.0	30	0.78	2.6	0.35	0.38	86	0.45	18
Principal decay, Mev											
Alpha								5.3	5.5	6.1	5.8
Beta	0.31	2.26	3.35	1.17	2.98	0.23	0.97				
Gamma	1.33	1.73		0.67	2.18	0.12	0.08	0.8	0.04	0.04	0.04
Fuel form	Metal	$SrTiO_3$	Metal	CsCl	CeO_2	Pm_2O_3	Tm_2O_3	Metal	PuO_2	Cm_2O_3	Cm_2O_3
Density (fuel form), g/cm³	8.7	3.7	12.2	3.6	6.6	6.6	8.5	9.3	10	9	9
Isotopic purity, %	10	50	3.3	35	18	95	10	95	80	90	95
Power density (fuel form), watts/cm³	15.2	0.94	13.4	0.42	25.3	1.8	9.1	1210	3.9	882	20.4
Curies/watt	65	148	102	207	126	2788	500	32	30	28	29
Shielding required [lead shielding for 1-kw source (10 mr/hr at 1 m)], in.	9.5	6	9	4.6	10.2	1	2.5	1	0.1	0.4	2
Melting point, °C	1480	1900	2450	646	2680	2270	2300		2280	1950	1950
Biological hazard (MPC in air), μc/cm³	3×10^{-9}	10^{-10}	2×10^{-9}	5×10^{-9}	2×10^{-9}	2×10^{-8}	10^{-8}	7×10^{-11}	7×10^{-13}	4×10^{-11}	3×10^{-12}
Estimated future cost, $/watt	16	21	5	24	2	220	10	20	540	17	64

Radioisotopes have different modes of decay. For example, as seen in Table 17.3, ^{60}Co is principally a gamma-ray emitter, ^{170}Tm is principally a beta-particle emitter, and ^{238}Pu is principally an alpha-particle emitter. In addition, neutrons are produced as a result of spontaneous fission of ^{238}Pu, ^{242}Cm, and ^{244}Cm or of fission of low-atomic-weight elements by alpha–neutron reactions in the immediate vicinity of the radioisotope. The emitted rays and particles carry the energy released by the decay and have different modes of interaction with matter. Accordingly each mode of decay suggests a particular way to harness the energy. For example, gamma rays can be captured relatively far from the fuel capsule, whereas alpha particles impart their energy over a very short distance from their formation.

Because they are much more penetrating than alpha and beta particles, gamma rays and neutrons play a dominant role in determining the health-shielding requirements of a radioisotopic heat source. In addition, these requirements depend on the bremsstrahlung due to emission of beta particles. Basic design considerations for the shielding of radioisotopic heat sources are given in Ref. 20.

Alpha particles are energetic, ionized helium atoms that eventually take the form of helium gas. Since the alpha particles emitted by the various alpha emitters all have about the same energy, the helium gas produced is always about 0.15 cm^3/kw-hr(t) at standard temperature and pressure.

17.3.2 Handling of Radioisotopes. Because they continuously generate radiation and heat, radioisotopes require special handling procedures that place severe constraints on the design of generators. The radioisotopes that require substantial health shielding must be handled by remote control in "hot" facilities. Consequently the generators in which they are installed must be designed so that all final stages of assembly can be performed by remote handling. In addition, since heat is generated continuously, special heat sinks must be used any time the radioisotope is not in suitable thermal contact with the sink provided by the generator.

The most difficult problem to be faced in designing radioisotopic generators is that of containing the radioisotope in the event of an accident. The best example of such an accident is reentry, burnup, and dispersal of a radioisotopic space-power system in the earth's atmosphere. The maximum permissible concentration (MPC) of radiation is between 10^{-14} and 10^{-18} curie/cm^3, depending on the nature of the radiation. Table 17.3 shows that burnup of a 1-kw heat source usually releases about 100,000 curies of radioactivity, which results in radiation densities much larger than MPC. To achieve densities lower than MPC, we would have to disperse the radioisotope uniformly over an extremely large volume, about 2×10^3 to 2×10^7 cu miles.

Because of the danger of radioactive contamination of the surface and the atmosphere of the earth, radioisotopic generators must be designed to contain the radioisotope under all possible accident conditions. Methods of designing containers that will provide reentry without burnup are given in Ref. 20. Note, however, that reentry without burnup presents a problem. Under certain conditions of impact on soft soil, the generator can be buried in a medium of low thermal conductivity. When this happens, it is very difficult to prevent rapid degradation of the containers by overheating and the consequent release of the radioisotope.[21]

When alpha emitters are used as fuels, the disposal of the generated helium must be considered in the design for total fuel containment. If helium is allowed to accumulate in a void provided in the fuel capsule, the pressure P as a function of time t is given by the relation

$$P = 2.38\, T_c f (1 - e^{-t/\tau})\ \text{atm}$$

where T_c = capsule temperature in degrees Kelvin
f = fuel-to-void-volume fraction
τ = half-life of the radioisotope times 1.44

In some designs it may be possible to vent the helium gas so that the fuel capsules need not be designed to sustain the high pressure that would otherwise result.[21,22] Venting should not be in the region of the thermal shields, however, since they could lose their effectiveness because of heat conduction through helium.

17.3.3 Module Designs. The geometry of the converter of a module can be either planar or cylindrical; the choice depends on the heat-transfer paths used. For example, because of the high heat fluxes required at the electrode surfaces, a planar-geometry converter is preferred when conduction heat-transfer paths are used from the fuel capsule to the emitter and from the collector to the sink. A cylindrical-geometry converter is preferred when heat pipes are used to couple the fuel capsule to the emitter. The fuel capsule cannot be directly within the cylindrical emitter, however, because the power per unit length of the capsule is generally much smaller than that required by the emitter.

For a given module power rating, the power densities of the radioisotopes listed in Table 17.3 result in fuel capsules having surface areas larger than the surface area of the emitter of the required converter. If not properly shielded, the portion of the fuel-capsule surface area which is not covered by the emitter (or by the heat-transfer path from the capsule to the converter) can lead to substantial radiant heat loss. For this reason the uncovered area should be thermally shielded. Shielding with compact, high-temperature, multilayer,

foil-type insulation can reduce the losses to 10 to 15% of the fuel thermal power.

Of course, using relatively high power density radioisotopes results in small fuel capsules and small heat-loss areas. Small capsules are also desirable from the standpoints of fuel temperature and compatibility of materials. Since thermal-conduction paths are shorter and the resulting temperature differences are small, the fuel temperature is close to the required emitter temperature. This, in turn, increases the life of the module and the compatibility between the fuel and capsule materials.

The components of the module are structurally supported by the housing. When the fuel capsule is small and lightweight, the converter is attached directly to the housing and the fuel capsule is mounted on the converter. On the other hand, when the fuel capsule is large and heavy, it is attached directly to the housing and the converter is mounted on the fuel capsule. In either support arrangement the module design must provide for unequal thermal expansion of the components within the housing. Finally, the surface area of the housing can be used for heat rejection.

A great variety of module designs is possible, but only three of these designs are discussed here.

(a) Conduction-Coupled Planar Converter. The conduction-coupled planar converter is shown schematically in Fig. 17.12. The collector of the converter is directly connected to and supported by the housing. Waste heat is rejected from

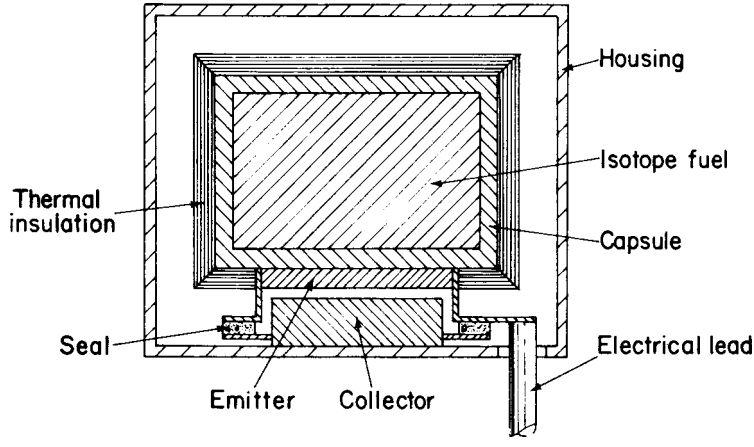


Fig. 17.12 Conduction-coupled planar converter.

the exterior of the housing, and the housing acts as the electrical lead to the collector. The fuel capsule is thermally attached to and supported by the emitter structure but is electrically insulated from the housing. Thermal insulation surrounds the free surface of the fuel capsule. If an alpha emitter is used, helium is formed in the fuel, and the capsule must be designed either to accommodate helium in the fuel region or to vent it into some low-temperature storage volume. As already stated, however, helium cannot be vented into the region of the thermal insulation. The emitter electrical lead penetrates the housing by means of a ceramic-to-metal seal or a sufficiently resistant metal seal. The radioisotope fuel-form power density must be larger than 20 watts/cm^3 for efficient operation of the converter.

Final assembly of the module is carried out by remotely controlled manipulators in a hot cell. The module without the insulation, the fuel, and the tops of the housing and capsule are located in the hot-cell fueling fixture. The radioisotope fuel form is inserted into the capsule, and the top of the capsule is welded to form a leak-tight container. Then the thermal insulation is installed, and the housing top is welded. The fueling must be carried out in a vacuum or an inert-gas environment to prevent oxidation of the refractory-metal components of the module. Finally, the housing must be evacuated through the pump-out conduit attached to the bottom of the module, and the conduit must be pinched off.

(b) Gamma-Coupled Planar Converter.[23] The gamma-coupled planar converter is shown schematically in Fig. 17.13. The fuel, a gamma-ray emitter, such as ^{60}Co, is outside the hermetically sealed housing. The housing contains the converter; an efficient gamma-ray absorber, such as tungsten; and thermal insulation around the absorber to prevent heat loss.

The distinguishing feature of the design is the thermal decoupling of the gamma-ray energy source from the region of absorption of the gamma rays. The decoupling is possible because of the long range of energetic gamma rays. Its advantage is that the absorber can operate at the high temperatures required by the emitter of the converter whereas the fuel can operate at much lower temperatures.

Low fuel temperatures allow a very important degree of freedom in the selection of the fuel form. For example, the fuel can be used in the metal form, and many of the difficulties associated with high-temperature operation are eliminated. Maximum fuel power densities can be achieved because of the high density of the undiluted metal. In addition, the problems associated with helium containment, which are inherent in the use of alpha emitters, are absent.

Finally, the fueling of the module is greatly simplified compared with modules using alpha or beta emitters. Since the fuel is located in a well outside the housing (Fig. 17.13), the module can be fabricated without concern about

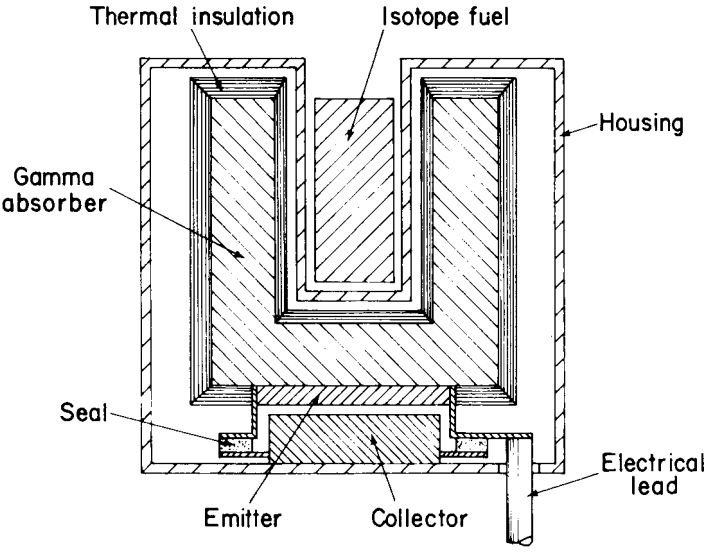


Fig. 17.13 Gamma-coupled planar converter.

constraints imposed by a radioactive material, and the fuel can be inserted after all welding and evacuation steps required in fabricating the vacuum-tight housing are completed. If desired, the fuel can be easily removed without affecting the integrity of the housing.

(c) Heat-Pipe-Coupled Cylindrical Converter. The heat-pipe-coupled cylindrical converter is shown schematically in Fig. 17.14. The heat pipe transfers heat from the core of a long radioisotope fuel capsule to the cylindrical converters.[24] This is an effective way of thermal coupling because the heat flux is high (e.g., 500 watts/cm^2 of heat-pipe cross section) and the temperature difference along the pipe is about 10 to 20°K.

Heat pipes can be operated as heat-flux concentrators, and thus they permit the use of low-power-density radioisotopes. In the fuel capsule the heat flux per unit area into the heat pipe can be relatively low as determined by the power density and configuration of the fuel. In the converter, however, heat removal from the heat pipe can be restricted to a small area and thus can lead to a high heat flux per unit area. For this reason heat-pipe-coupled modules can be built with fuels having power densities as low as 1 watt/cm^3.

The housing, structural-support, and fueling requirements of heat-pipe-coupled designs are similar to those for conduction-coupled designs. All the high-temperature refractory-metal components (fuel capsule, heat pipe, converter,

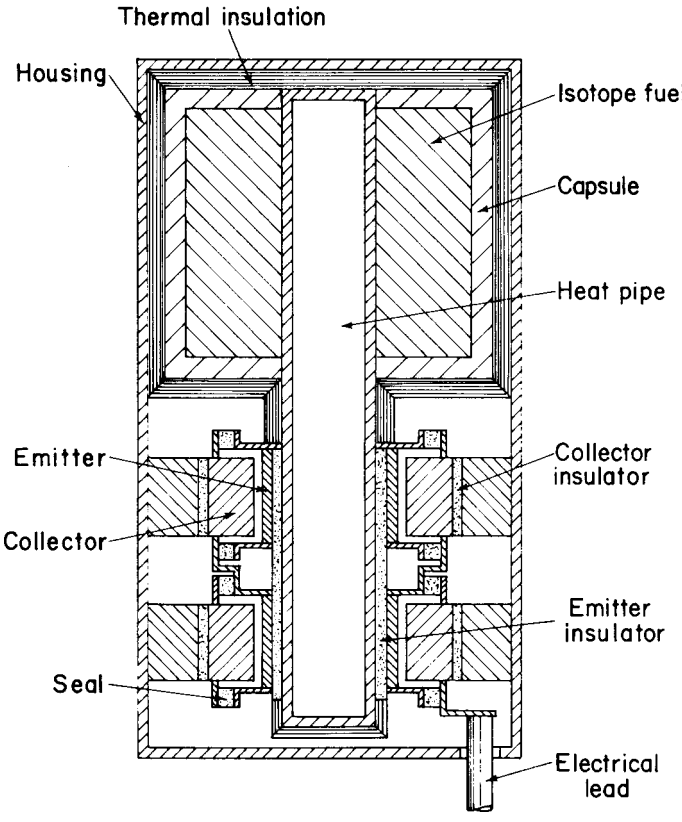


Fig. 17.14 Heat-pipe-coupled cylindrical converter.

etc.) must be contained in a vacuum-tight housing. Thermal insulation should be used around the fuel capsule and heat pipe to minimize heat loss. Fueling must be performed in a hot cell.

Finally, an interesting radioisotopic thermionic generator is the "*atomic battery*," which can produce small amounts of power (about 1 mw) for very long periods of time. Such a generator, described in Ref. 25, is useful for a number of applications, e.g., in implanted cardiac pacemakers.

17.4 Nuclear Reactor Thermionic Generators

In nuclear reactor thermionic generators, or simply *thermionic reactors*, the heat source is a nuclear reactor. The coupling of converters to nuclear reactors is the most attractive application of thermionic conversion.[26-33]

For space missions a thermionic reactor has the following desirable features: (1) no moving parts and a highly modular structure, which offer potentially high reliability levels; (2) inherently high rejection temperatures, which permit a major reduction in radiator area and in overall size of the power system; and (3) high conversion efficiency without complex turbomachinery, which provides for low system specific weight [lb/kw(e)].

Many of these same features are desirable for certain terrestrial and undersea power systems. For example, thermionic reactors can be considered as power sources for small deep-submergence vehicles where minimum volume is dictated by the pressure-hull design. Quiet operation, relative to turbines, is also a factor in this type of mission.

In the area of nuclear central-station power production, thermionic conversion may eventually be employed as a topping cycle for more-conventional steam- or gas-turbine plants. This application is not feasible with the low-temperature thermal reactors now in wide usage. Most converter materials have high capture cross sections for thermal neutrons and, hence, cannot be placed in thermal reactor cores without seriously affecting criticality. On the other hand, future developments in central-station power production are expected to emphasize the fast breeder reactor, the core of which is far less sensitive to the addition of high-density materials. Consequently, converters might be designed integral with the fuel elements in at least some regions of a fast breeder reactor core. Since fast breeders are in an early stage of development, it is not yet possible to forecast the overall economic benefit of the use of thermionic conversion in the system. Much of the technology being developed for thermionic reactors in space may be applicable to future central-station power reactors.

Some of the thermionic reactor designs that have been studied are reviewed in subsequent sections. These designs, although specifically created for space applications, provide a basis for understanding thermionic reactors in general. Two major categories can be identified:

1. *Out-of-core designs.* In out-of-core designs the converters are mounted external to the reactor core and are coupled to the core by means of conductive or convective heat-transfer paths.

2. *In-core designs.* In in-core designs the converters are integral parts of the nuclear fuel elements and are dispersed throughout the core.

Within each category several different designs have been proposed, each having special advantages and limitations. For example, some designs are simple in configuration and relatively easy to develop but are limited to low power levels. On the other hand, more complicated designs capable of high power ratings and attractive power-to-weight ratios require that difficult technical problems be solved. These points, as well as the various engineering compromises that must

be made, will become apparent from the discussions of out-of-core designs in Sec. 17.5 and in-core designs in Sec. 17.6.

17.5 Out-of-Core Thermionic Reactors

The out-of-core category includes all types of designs in which the converters are outside the reactor core. The major advantage of these designs is that certain interactions between the converters and the nuclear fuel are avoided; thus development of both converters and core is simplified. On the other hand, out-of-core thermionic reactors tend to be bulkier and heavier than in-core designs for electrical power ratings greater than about 100 kw(e).

Each design in the out-of-core category is given a different name depending on the method used to transfer energy from the core to the converters. *Core-surface* or *conduction designs, forced-convection designs,* and *heat-pipe designs* are discussed in Secs. 17.5.1 to 17.5.3.

17.5.1 Core-Surface or Conduction Designs. Core-surface or conduction designs were among the earliest configurations studied.[30,31,34] Fission heat is conducted directly from the core surface through a high-temperature electrical-insulator layer to the emitters of the converters. The converters are connected electrically in series-parallel arrays to achieve a balance between circuit reliability and output voltage. The collectors of the converters are extended to form radiator fins for waste-heat rejection. The cesium reservoirs are attached to the envelopes of the converters so that they can assume the correct operating temperatures as a result of their inherent heat balances.

In another design the core is divided into electrically isolated segments to eliminate the need for the high-temperature insulator layer. In still another design the emitters are physically separated from the core, and heat is transferred by thermal radiation. A temperature difference of 100 to 200°K between core surface and emitters is necessary to achieve the required heat flux. Even if all materials-compatibility and fabrication problems could be resolved, at reasonable specific weights the central-core temperatures limit the power ratings of core-surface designs to about 20 kw(e).

17.5.2 Forced-Convection Designs. In forced-convection designs the converters are completely separated from the core, receiving heat from a circulating fluid that, in turn, is heated either directly or indirectly by the nuclear reactor. Heat is rejected from the converters either by direct radiation or by a convective-fluid loop to a separate radiator. A wide variety of configurations has been suggested for each combination of heat-transfer paths. Only one of the configurations, shown schematically in Fig. 17.15, is discussed here.

A circulating fluid, such as liquid lithium, is heated in the reactor and supplied to a header. Attached to the header is a large number of tubes, each of which is

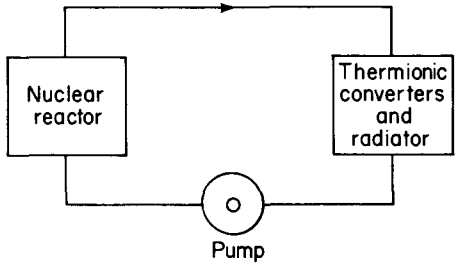


Fig. 17.15 Forced-convection thermionic reactor.

part of a converter module. Each tube is surrounded by an electrically *insulating sleeve* that is bonded to tubular emitters of individual converters. Each emitter is, in turn, surrounded by a concentric tubular collector. Heat flows from the hot fluid to the central tube by convection and from the central tube to the emitter by conduction across the insulating sleeve. If heat is rejected from the collectors by radiation to space, then two plate-type fins are bonded to each collector to provide the heat-transfer area required to radiate the waste heat at a desirable collector temperature.

A major difficulty with forced-convection designs is containing the circulating fluid at the temperatures required for efficient operation. For example, although lithium is basically compatible with such materials as niobium, molybdenum, tantalum, tungsten, and some of their alloys over the range of practical temperatures, severe corrosion and mass transport can occur when even trace quantities of impurities are present either in the materials of the container or in lithium. Even when special high-purity inert-gas or vacuum facilities are used, the introduction of impurities is very difficult to avoid in fabricating large power systems where thousands of welds and brazes must be made.

Another serious problem is the insulating sleeve. This sleeve must be both a good electrical insulator and a good thermal conductor. Significant temperature drops will occur across the sleeve unless it is very thin and is well bonded to both the circulating-fluid tube and the emitters. Typical differences in thermal-expansion characteristics between metals and insulators make it very difficult to maintain good bonds. Moreover, the integrity of the bonds is further affected by thermal cycling, which is inevitable in any practical system.

Problems are also encountered in pumping the high-temperature circulating fluid. Electromagnetic pumps can be used, but the magnets must be cooled. Developing mechanical pumps for use at the required temperatures is still a formidable task. Because of the difficulties with high-temperature convective loops, most recent studies of out-of-core designs have been based on the use of heat pipes. Such designs are discussed in Sec. 17.5.3.

17.5.3 Heat-Pipe Designs. In heat-pipe designs heat is transferred from the reactor core to the emitters of the converters by means of lithium heat pipes. In addition, current studies have indicated that the need for high-temperature electrical-insulator material can be eliminated by using the electrical resistance of the heat pipes to achieve the required electrical isolation between converters.

The elements of a possible heat-pipe design are as follows: A set of tungsten heat pipes with lithium as the working fluid carries heat from the core to positions outside the axial reflector. At these positions the tungsten heat pipes are coupled to a set of tantalum heat pipes, again with lithium as the working fluid. The coupling is in the form of a cross-flow heat exchanger. The tantalum heat pipes carry heat to the converters. Heat is rejected from the collectors by means of a vapor-chamber radiator. The results of a study of a 150-kw(e) unit are given in Ref. 35.

17.6 In-Core Thermionic Reactors

The in-core category includes all designs in which the converters are integrated into the fuel elements of the reactor. These designs are attractive from a volume and weight standpoint. The proximity of the nuclear fuel to the converters gives rise to severe chemical, mechanical, and nuclear compatibility problems, however. Despite these difficult problems, in-core thermionic reactors offer sufficient advantages in power-to-weight ratio to make them prime candidates for missions where performance is critical, for example, in electric propulsion for high-power planetary-exploration spacecraft.

Several in-core designs have been proposed which are based on cores with either a fast or an epithermal neutron spectrum. Each design is designated by a different name depending on the neutron spectrum and the geometric arrangement used to integrate the converters into the reactor core. Fast reactor core *flashlight, pancake,* and *externally fueled* designs are discussed briefly in Secs. 17.6.1 to 17.6.3. Epithermal reactor core designs are not discussed.

17.6.1 Flashlight Designs. Flashlight designs owe their designation to the appearance of the *thermionic fuel element* (TFE). Each TFE consists of a string of converters, each of which is heated by the fissioning of a nuclear fuel pellet enclosed within the emitter. Each emitter is surrounded by a tubular collector and is electrically connected to the collector of one adjacent converter. The collectors are surrounded by a ceramic tube that provides electrical isolation between collectors and between converters and coolant. The ceramic tube is housed in a metallic sheath that provides hermetic containment for the string of converters and structural integrity for the fuel element. Many TFE's are assembled to form a critical reactor. The reactor is cooled by a coolant that flows axially along the fuel elements.

The results of a study of a 4-Mw(e) flashlight design are given in Ref. 36.

A number of problems result from integrating converters in the reactor environment. For example, because of criticality considerations small thicknesses of structural materials are necessary, and this, in turn, leads to an emitter stability problem. Small-diameter thin-walled emitter structures are susceptible to dimensional distortion by mechanical interaction between the fuel and the emitter wall. The high operating temperatures of the emitter can also cause reactions between the nuclear fuel and the emitter materials. Diffusion of the fuel (uranium) through the emitter wall can alter thermionic performance. The primary fuels employed in flashlight designs are bulk UO_2 and UC–ZrC (90%–10%). The oxide fuel does not cause degradation of thermionic performance but produces dimensional distortion of the emitter. The carbide fuel (fabricated with ~10% interconnected porosity) has shown good dimensional stability up to desired burnup levels, but it interacts with tungsten and causes contamination of electrode surfaces. Preparing emitters by vapor deposition of tungsten from various halides seems to reduce this contamination substantially.

Again high-conductance heat-transfer paths are required for efficient heat rejection from the collectors. This, in turn, favors the fabrication of bonded structures to minimize temperature drops. Bonded collector-to-insulator interfaces are difficult to fabricate. On the other hand, collectors slip-fitted within the ceramic tube are easy to assemble but are heat-flux limited owing to yielding of the collector material. Significant quantities of gaseous fission products are produced in the nuclear fuel. For long operating periods this gas must be removed to prevent high internal pressures that cause emitter distortion. The fission gas can be removed through channels completely separate from the converters, or it can be bled into the converters and then removed at the end of the TFE.

Finally, the reactor environment subjects the TFE materials to high neutron fluences over extended operating periods. Insulator materials are prone to mechanical degradation by exposure to fast neutrons.[37,38] A fluence greater than 10^{21} neutrons/cm^2 by neutrons with energies greater than 0.1 Mev can cause substantial cracking in alumina. Much more work is needed to develop insulator materials suitable for long term use in a reactor environment.

Insulator cracking is undesirable because it can lead to electrical breakdown and arcing in cesium. This, in turn, can cause failure of the outer metal sheath and flooding of the interelectrode space by the coolant.

Variations of the basic flashlight design which avoid the problem with the insulator have been proposed. One of these variations uses two insulator layers separated by a metallic sheath. Thus the outermost layer, which must withstand the full output voltage, is not exposed to the easily ionizable cesium.

Although difficult, the problems cited are not of a fundamental nature. Much progress has been made in understanding them and reducing their severity. Flashlight-type converters (both single cell and multiple cell) have been tested in nuclear reactors and have achieved practical lifetimes.[39,40]

17.6.2 Pancake Designs. Pancake designs owe their designation to the appearance of the core, which consists of self-contained layers or pancakes of converters.[41] In each layer, individually fueled converters are brazed into back-to-back cylindrical receptacles. Each converter has its own cesium reservoir. Coolant flows in a cross-flow pattern across each layer from one side to the other. Each layer and its coolant loop can be electrically isolated.

Series-parallel electrical connections between converters in a layer are made in the space between layers and the liquid-metal coolant region. A highly redundant electrical matrix of series-parallel connections can be achieved since the terminals of all individual converters are accessible. Moreover, an overall output voltage greater than that of flashlight designs can be accomplished since the layers can be connected electrically in series without increasing the voltage between each converter and its coolant loop.

The advantages of the pancake design are offset by the complexity of its structure. Removing and replacing individual converters after the reactor has operated for some time are practically impossible. Even removing and replacing an entire pancake would require essentially complete disassembly of the reactor. Ability to accomplish these operations is important in the initial stages of reactor development and testing. In addition, although its ability to survive certain types of failures is undoubtedly improved, a pancake design requires so many joints, welds, and brazes that the probability of failures is increased.

17.6.3 Externally Fueled Designs. Externally fueled designs owe their designation to the appearance of the fuel element in comparison with a TFE. The fuel element consists of a coolant tube at the center of the element, the outer surface of which serves as the collector. Beyond the collector is the interelectrode space, the emitter (which is also the inner container of the fuel), the fuel, and the outer container of the fuel.

In this fuel-element geometry, the cross-sectional area of metal located on the emitter side is sufficient to permit long converters (10 to 12 in. long) to be used without excessive ohmic losses. For power ratings up to 300 kw(e), the converters can extend over the full length of the core. The insulating seals can then be located outside the core in a region of reduced fast-neutron flux.

Several full-length converters (four to seven) are grouped together and are electrically in parallel. The groups are connected electrically in series on both sides of the core by massive bus bars, which can form the axial reflectors. In this

arrangement the coolant loop of each group must be kept electrically isolated from the others.

Externally fueled designs have several potential advantages: (1) For a given limiting fuel temperature, fuel-volume fraction is higher than that of flashlight or pancake designs. For systems with power ratings smaller than 100 kw(e), this results in a smaller core and shielding; at power levels greater than 300 kw(e), however, this advantage is not significant. (2) Venting of fission products can be readily accomplished without their passing through the interelectrode space. (3) The need for an in-core insulator is eliminated; electrical insulation is provided by spacing between individual fuel elements. (4) If cesium is lost and fuel and emitter temperatures rise because of loss of electron cooling, the magnitude of the rise is reduced because each fuel element can transfer heat by radiation to neighboring fuel elements. (5) Fuel elements can be tested by means of electrical heating before they are assembled into a reactor. (6) Since several coolant loops can be employed, some coolant-loop leakage failures are tolerable.

Externally fueled designs also have a number of disadvantages: (1) Long converters can be more susceptible to emitter–collector shorting from small distortions. (2) Above about 300 kw(e) full-core-length converters are difficult, if not impossible, to fabricate. (3) The surface of the core is at emitter temperature and requires high-temperature thermal insulation. (4) Reactor access for fuel element removal and replacement is much poorer than in flashlight designs. (5) Testing experience is very limited; very few electrically heated tests and no in-core tests have been performed.

A study of a 120 kw(e) externally fueled design for space applications is discussed in Ref. 42.

17.6.4 Control of In-Core Thermionic Reactors. The control of fast-spectrum in-core thermionic reactors requires special attention. It differs from that of nonthermionic reactors because (1) the overall prompt temperature coefficient of reactivity is positive and (2) the inherent characteristics of converters result in direct coupling between electrical-load changes and reactor temperatures. The net effect of these differences is that an active control system must be used for an in-core thermionic reactor.

Investigations of various control techniques[43] indicate that, although active control is indeed necessary to cover the range of operating conditions from 10 to 100% full power, the time constants of the in-core thermionic reactors are sufficiently long that the control requirements can be easily met by a straightforward control-system design.

17.7 Chemical Thermionic Generators

In chemical thermionic generators heat is produced by combustion of hydrocarbon fuels.

17.7.1 General Configuration. The basic components of a combustion-heated thermionic generator are shown schematically in Fig. 17.16. Fuel and air burn in either a *combustion chamber* or the *combustion side of the emitter.* The products of combustion impart their energy to the emitter structure via the *oxidation barrier* and then are vented. Because of the required high emitter temperature and high input-heat-flux density, preheating of the air of combustion is necessary when hydrocarbon fuels and air are used. Preheating may not be needed in other reactions, such as the burning of hydrogen in oxygen, where high-flame temperatures can be achieved without oxidizer preheating. Typically, the emitter structure is operated at 1700°K with an input-heat-flux density of the order of 50 to 75 watts/cm^2 and a heat-to-electricity conversion efficiency of the order of 10%.

Because of the relatively small value of the conversion efficiency, recuperation should be used to recover a fraction of the waste heat from the combustion products and, thus, increase generator efficiency.

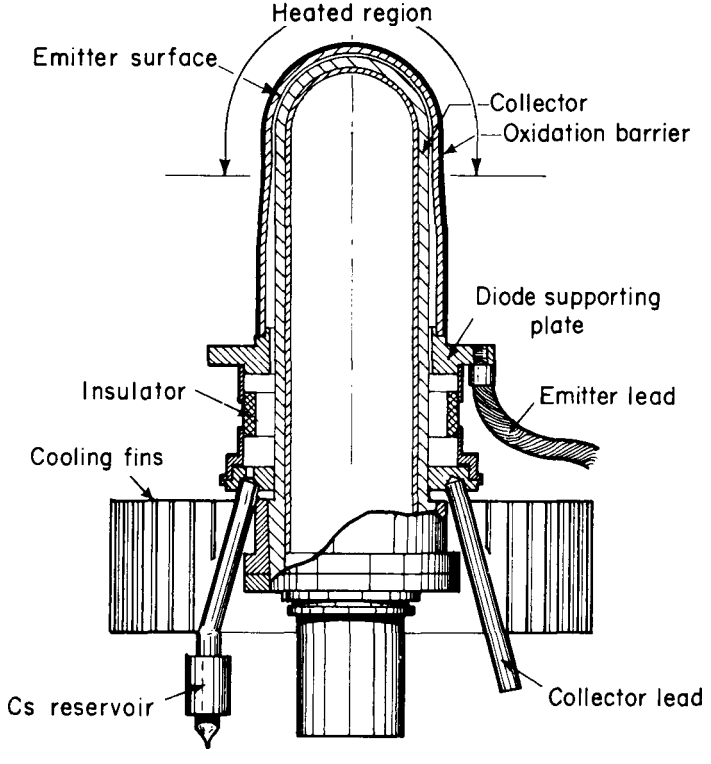


Fig. 17.16 Combustion-heated thermionic generator.

17.7.2 Development of Emitter Structure. The key technological problem of combustion thermionic generators is the development of a satisfactory emitter structure.

The high emitter temperature, the high input-heat flux necessary for effective emitter operation, and the drastically different environments that must be maintained on each of the emitter's sides impose severe requirements on materials. Specifically, at temperatures up to about 1800°K and for the desired lifetime of the generator, the emitter structure must be (1) resistant to thermal shocks and thermal gradients so that it can withstand the sudden application of high heat fluxes, (2) resistant to oxidation so that it is not affected by the oxygen in the air of combustion, (3) resistant to corrosion and erosion so that it is not affected by the combustion components, and (4) impermeable to all gases so that the required level of vacuum and the purity of the cesium vapor in the interelectrode space can be maintained.

Readily available materials, such as ceramics, refractory metals, and coatings, can satisfy some of these requirements. Materials that satisfy all the requirements simultaneously are not available, however, and must be engineered.

Several approaches investigated for the development of a satisfactory emitter structure are summarized in the following paragraphs.

1. *Metallized Ceramics.* A high-purity fully dense ceramic body is coated with a refractory metal on the thermionic-emission or interelectrode side of the emitter structure. Ceramics were selected because they are resistant to oxidation and impermeable to gases. They have limited resistance to thermal shock and thermal gradients, however.

2. *Coatings.* A refractory metal is coated with a suspension (slurry) of metal powders. After heat treatment the binder of the suspension evaporates and the metal particles interact with the base metal to form a protective coating. Combinations that appeared promising* are aluminum–tin on tantalum and mullite on molybdenum. The coating materials proved to be ineffective, however, because they decompose at temperatures of 1500 to 1700°K and some are permeable to gases. Coatings also proved to be edge sensitive at the interface between coated and uncoated metal and extremely vulnerable to erosion by high-velocity gases.

3. *Composite Structures.* A ceramic such as high-purity alumina (98 or 99.7% Al_2O_3) is bonded to a refractory metal. Composite structures proved to be corrosion resistant but incapable of withstanding thermal shock.

4. *Liquid-Metal-Bonded Layers.* Two separate parts, one having the desired thermionic properties and the other being oxidation resistant, are bonded

*Since materials and processes are proprietary, further information should be sought from Sylvania and Raytheon, who originally developed the coatings.

thermally by a liquid-metal layer. Such emitter structures can satisfy all the requirements cited earlier but present serious fabrication problems.

5. *Graded Structures.* Successive layers of materials are bonded together metallurgically. The layers change progressively from one with the desired thermionic properties to one that is oxidation resistant. An example of such a structure is formed by pyrolytic deposition of silicon carbide on tungsten coated with a thin layer of carbon. Tungsten is acceptable on the interelectrode side; silicon carbide is acceptable on the combustion side; and the interposed layer of carbon prevents the formation of the detrimental tungsten disilicide at the interface between tungsten and silicon carbide. Graded structures have met the requirements for emitter structures of combustion-heated generators for limited periods of time.

The best results were obtained with a composite structure where silicon carbide and tungsten were deposited on a shell of isotropic graphite.[44-46] Controlled tests with this structure have shown no deterioration of bond strength up to 2000 hr at 1700°K and no failures from thermal cycling to 1400°K in 40 sec with an axial gradient of 1300°K/in. In addition to its oxidation resistance, silicon carbide has a high conductivity, and therefore the temperature drop across the emitter structure is relatively low. The structure was also tested in converters that operated without degradation for up to 300 hr.

17.8 Topping Thermionic Generators

In *topping thermionic generators* heat is produced by combustion of hydrocarbon fuels or coal. Unlike the chemical units (Sec. 17.7), however, the heat rejected from the collectors of topping generators is used to raise steam, which, in turn, is used in a turbogenerator to produce additional electricity.

Topping thermionic generators can be combined with oil-fired or coal-fired central power stations and thus increase the efficiency of conversion of fuel to electricity from about 40%, the efficiency of modern power stations, to about 50%. They can also be combined on the emitter side with the exhaust of coal-fired magnetohydrodynamic generators and on the collector side with steam turbogenerators to achieve overall efficiencies of conversion from fuel to electricity greater than 60%.

Several studies have been made of coal-fired 1000-Mw(e) systems consisting of either a topping thermionic unit and a steam turbogenerator or a combination of MHD, thermionic, and steam units. The results are encouraging. The experimental work is currently concentrating on:

1. Developing low-temperature emitter-electrode surfaces capable of emitting about 10 amp/cm^2 at 1400 to 1600°K.

2. Developing low-temperature collector-electrode surfaces having work functions of 1 to 1.3 ev at about 700 to 800°K.

3. Developing high-temperature emitter-electrode coatings capable of withstanding combustion environment products.
4. Reducing plasma losses in the interelectrode space of the converters.

Progress is being made in each of these four efforts.

References

1. A. G. Conrad, System Engineering Considerations in Spacecraft Design, Technical Report No. 32-960, Jet Propulsion Laboratory, June 1966.
2. P. J. Brosens, Solar Thermionic Generators for Space Power, *J. Eng. Power,* 87(3): 281-289 (1965).
3. P. J. Brosens and S. S. Kitrilakis, Cesium Thermionic Converters and Generators for Solar Space Power Systems, ARS Preprint No. 2571-62, American Rocket Society, 1962.
4. W. R. Menetrey and A. H. Smith, Solar Energy Thermionic Conversion System, ARS Preprint No. 2499-62, American Rocket Society, 1962.
5. General Electric Company, Missile and Space Division, $9\frac{1}{2}$ Foot Diameter Master and Mirror, Contract 950239, March 1964.
6. Thermo Electron Corp., Design and Fabrication of Advanced Thermionic Converters, Contract No. 951263, November 1968.
7. A. H. Smith, Thermionics: Auxiliary Power Application for Space, in *Proceedings of the Design Engineering Conference of the American Society of Mechanical Engineers, New York, 1963.*
8. D. L. Purdy, in *Space Power Systems,* Proceedings of American Rocket Society Space Power Systems Conference, p. 153, Academic Press, Inc., New York, 1961.
9. Ref. 8, p. 791.
10. W. J. Leovic and C. Z. Kamien, Solar Thermionic Power Systems Development, in *Space Power Systems Engineering,* G. C. Szego and J. E. Taylor (Eds.), Progress in Astronautics and Aeronautics, Vol. 16, Academic Press, Inc., New York, 1966.
11. W. R. Menetrey, Summary of Final Report, May 22–Nov. 22, 1961. Solar Energy Thermionic System, Report 1850, Electro-Optical Systems, Inc., January 1962.
12. S. G. Merra and J. H. Weinstein, Recent Progress in the Development of Solar Thermionic Converters, in Report of the Thermionic Conversion Specialists Conference, Gatlinburg, Tenn., Oct. 7–9, 1963, pp. 295-304, General Electric Research Laboratory, 1963.
13. P. G. Pantazelos, S. G. Merra, P. J. Brosens, and B. Gunther, Final Report. Task 1, Solar Energy Thermionic (SET) Program, Report TE4034-73-67, Thermo Electron Corp., March 1968.
14. P. Rouklove, Thermionic Converter and Generator Tests, in *IEEE Thermionic Conversion Specialists Conference, Palo Alto, Calif., Oct. 30–Nov. 1, 1967,* pp. 75-85, Institute of Electrical and Electronics Engineers, 1967.
15. P. J. Brosens, Final Report. Heat Pipe Thermionic Converter Development, Report TE4067-61-68, Thermo Electron Corp., December 1967.
16. P. Rouklove, Status Report on Solar Thermionic Power Systems, in *Space Power Systems Engineering,* G. C. Szego and J. E. Taylor (Eds.), Progress in Astronautics and Aeronautics, Vol. 16, Academic Press, Inc., New York, 1966.
17. P. J. Brosens, Advanced Converter Development, in *IEEE Thermionic Conversion Specialists Conference, Palo Alto, Calif., Oct. 30–Nov. 1, 1967,* pp. 68-74, Institute of Electrical and Electronics Engineers, 1967.

18. J. B. Dunlay and R. C. Howard, SNAP-13 Generator Development Program, in *Thermionic Electrical Power Generation,* Proceedings of the Second International Conference, Stresa, Italy, May 27–31, 1968, pp. 297-304, Euratom Center for Information and Documentation, 1968.
19. C. A. Rohrmann, Radioisotope Heat Sources, USAEC Report HW-76323, Table IV (rev. 3/1/67), General Electric Company, Feb. 1, 1963.
20. H. M. Dieckamp, Nuclear Space Power Systems, Report AI-Memo-7585, Atomics International, September 1967.
21. E. W. Williams and R. C. Howard, Development of a 100 Watt(e) Isotope Thermionic Electrical Power Module, in *Thermionic Electrical Power Generation,* Proceedings of the Second International Conference, Stresa, Italy, May 27–31, 1968, pp. 281-296, Euratom Center for Information and Documentation, 1968.
22. A. de Troyer et al., Design and Characteristics of an Actinium Fueled Thermionic Generator, in *Thermionic Electrical Power Generation,* Proceedings of the Second International Conference, Stresa, Italy, May 27–31, 1968, pp. 305-336, Euratom Center for Information and Documentation, 1968.
23. J. B. Dunlay and A. E. Wallis, Description of a Co^{60}-Fueled Thermionic Generator Experiment, in *IEEE Thermionic Conversion Specialists Conference, Palo Alto, Calif., Oct. 30–Nov. 1, 1967,* pp. 97-101, Institute of Electrical and Electronics Engineers, 1967.
24. P. Rouklove, Radioisotope Thermionic Generator, in *Proceedings of the 9th Annual Thermionic Conversion Specialists Conference, Miami Beach, Fla., Oct. 26–27, 1970,* pp. 382-387, Institute of Electrical and Electronics Engineers, 1970.
25. N. S. Rasor, K. A. Gasper, and J. G. DeSteese, A Miniature Atomic Battery Based on Thermionic Conversion, in *Thermionic Electrical Power Generation,* Proceedings of the Second International Conference, Stresa, Italy, May 27–31, 1968, pp. 337-344, Euratom Center for Information and Documentation, 1968.
26. R. L. Hirsch and N. S. Rasor, Preliminary Design Study of a Reactor–Thermionic Space Power Plant, Report NAA-SR-5056, Atomics International, March 1960.
27. R. L. Hirsch and J. W. Holland, Problems Associated with the Development of a Thermionic Conversion Reactor, ARS Preprint No. 1338-60, American Rocket Society, 1960.
28. C. K. Smith, J. W. Holland, and R. L. Hirsch, Design Study of a Thermionic Reactor Power Plant for Space Applications, USAEC Report NAA-SR-6077, Atomics International, May 1, 1961. (Classified)
29. R. C. Howard, Thermionic Reactor Systems, *Trans. Amer. Nucl. Soc.,* 3: 2 (1960).
30. N. Rasor and C. Weeks, Thermionic Converters for Compact Nuclear Power Plants, USAEC Report NAA-SR-7144, Atomics International, Aug. 1, 1962.
31. B. G. Voorhees, Progress Toward Nuclear Thermionic Space Power, ARS Paper No. 2577-62, American Rocket Society, 1962.
32. C. K. Smith and R. Y. Parkinson, A 1.5 Mwe Thermionic Space Power System, in *American Rocket Society Space Power Systems Conference, Sept. 25–28, 1962—Classified Sessions,* USAEC Report TID-7661, U. S. Atomic Energy Commission, American Rocket Society, and National Aeronautics and Space Administration, 1962. (Classified)
33. N. Rasor and R. C. Howard, Nuclear Thermionic Space Power Systems, in *Nuclear Thermionic Space Power Systems,* proceedings of the 18th Annual IAS Propulsion Meeting, Cleveland, Ohio, Mar. 7–8, 1963, Institute of Aerospace Sciences, 1963.
34. E. P. Gyftopoulos and G. N. Hatsopoulos, Thermionic Nuclear Reactors, *Elec. Eng.,* 82: 108-116 (1963).

35. J. J. Ward, R. Breitwieser, and R. M. Williams, Conceptual Design of a 150 Kwe Out-of-Core Nuclear Thermionic Converter System, in *Proceedings of the 9th Annual Thermionic Conversion Specialists Conference, Miami Beach, Fla., Oct. 26–27, 1970*, pp. 179-184, Institute of Electrical and Electronics Engineers, 1970.
36. C. D. Sawyer, P. R. Hill, and D. R. Wilkins, Multimegawatt Thermionic Reactor Systems for Space Applications, in *Thermionic Electrical Power Generation*, Proceedings of the Second International Conference, Stresa, Italy, May 27–31, 1968, pp. 171-184, Euratom Center for Information and Documentation, 1968.
37. W. H. Reichelt et al., Radiation Induced Damage to Ceramics in the EBR-II Reactor, in *Proceedings of the 9th Annual Thermionic Conversion Specialists Conference, Miami Beach, Fla., Oct. 26–27, 1970*, pp. 39-44, Institute of Electrical and Electronics Engineers, 1970.
38. E. W. Salmi et al., Effect of Neutron Spectra on the Swelling of Alumina Under Irradiation, in *Proceedings of the 9th Annual Thermionic Conversion Specialists Conference, Miami Beach, Fla., Oct. 26–27, 1970*, pp. 45-48, Institute of Electrical and Electronics Engineers, 1970.
39. G. F. Tape, The United States' Thermionic Program, in *Thermionic Electrical Power Generation*, Proceedings of the Second International Conference, Stresa, Italy, May 27–31, 1968, pp. 1407-1416, Euratom Center for Information and Documentation, 1968.
40. R. C. Howard and J. B. Dunlay, Converter SD-4 Design and Summary of Test Results, in *Thermionic Electrical Power Generation*, Proceedings of the Second International Conference, Stresa, Italy, May 27–31, 1968, pp. 365-384, Euratom Center for Information and Documentation, 1968.
41. J. P. Davis, Thermionic Reactor Systems—A Review, *Mech. Eng.*, 92(6): 42-46 (1970). (The concept has been studied extensively by Gulf General Atomic, but at the present time reports of GGA are not available.)
42. W. Z. Prickett, A 120 Kwe Thermionic Reactor Ion Propulsion Spacecraft for the Comet Halley Rendezvous Mission, in *Proceedings of the 9th Annual Thermionic Conversion Specialists Conference, Miami Beach, Fla., Oct. 26–27, 1970*, Institute of Electrical and Electronics Engineers, 1970.
43. H. G. Gronroos and J. P. Davis, Stability and Control Considerations for Thermionic Reactors, in *Thermionic Electrical Power Generation*, Proceedings of the Second International Conference, Stresa, Italy, May 27–31, 1968, pp. 273-279, Euratom Center for Information and Documentation, 1968.
44. L. J. Lazaridis and P. G. Pantazelos, Tests on Flame-Heated Thermionic Diode, in *Proceedings of the 20th Annual Power Sources Conference, Atlantic City, N. J., May 24–26, 1966.*
45. L. J. Lazaridis, Life Test on Fossil-Fuel Thermionic Diode, in *Proceedings of the 21st Annual Power Sources Conference, Atlantic City, N. J., May 16–18, 1967.*
46. L. J. Lazaridis, P. G. Pantazelos, and I. Shai, Design of a Gas-Fired Thermionic Power Supply for Domestic Furnaces, Paper No. 66-WA/ENER-3, American Society of Mechanical Engineers, New York, November 1966.

Index

Accelerating collector sheath, 425, 429, 436
Accelerating emitter sheath, 425, 429, 435
Accelerating-field steady state, 267
Adams, 351
Additive property, 106
Adiabatic process, 105
Adsorbate, 310
Adsorbate surface, 312
  density of full monolayer, 310
Alleau, T., 335
Allowed state, 108
Amplitude, probability, 8
Anomalous Schottky effect, 306, 308
Antibonding orbital, 60
Antisymmetric determinantal function, 71
Antisymmetric energy eigenfunction, 170
Antisymmetric state, 49, 50
Apparent emission constant, 269
Apparent work function, 269
Approximation
  Boltzmann, 173, 174
  Born–Oppenheimer, 57, 75, 80, 98, 271, 275
  constant potential, 173, 175
  continuous-eigenvalue spectrum, 173, 176
  gas, 168
  one electron, 82, 97, 275
  one particle, 168, 169, 335
  point particle, 173, 176
  semiperfect gas, 179
  Sommerfeld gas, 265
Atomic battery, 635
Atoms
  configuration, 55
  orbital, 44, 54
    linear combinations, 63, 67, 73
  orbital type, 46
  radius, 74, 76
Auer, P. L., 371
Auger emission, 480
Available work, 111
Average collision frequency, 405
Average free path, 405
Average reflection coefficient, 263, 271, 487
Average value, 10
Avogadro's number, 145
Axes, in crystal lattice, 573

Bacal, M., 335
Baksht, F. G., 454
Bardeen, J., 276, 277, 279, 280, 282
Bare work function, 313, 320, 483, 582, 583, 586
Bernoulli, D., 119
Binding energy, 70
Bird, R. B., 414
Body-centered-cubic crystal, 294, 573
Bohm, D., 400
Bohr, N., 2, 4
Bohr magneton, 43
Bohr radius, 42, 44, 59, 74
Boltzmann, L., 102, 119, 233
Boltzmann approximation, 173, 174
Boltzmann constant, 126, 145
Boltzmann equation, 233, 236, 405
  steady state, 239, 241, 384, 440
Boltzmann line, 521, 523, 525
Bond
  covalent, 73
  partially covalent, 74
  partially ionic, 74
Bond number, fractional, 293, 294
Born–Oppenheimer approximation, 57, 75, 80, 98, 271, 275
Bornhorst, W., 454
Bose–Einstein factor, 172, 235
Bose–Einstein gas, 171, 173
Bosons, 49, 170
Bound state, 40
Boundary-value process, 203, 204, 208
Bravais lattice, 77
  body centered cubic, 78
  cubic, 77
  direction, 78
  face centered cubic, 78
  hexagonal close packed, 80
Brillouin zone
  central, 88, 89, 91, 93, 95
  second, 88
  third, 88

Bulk flow, 213, 214, 227
Bullis, R. H., 454

Canonical distribution, 121, 127, 137, 146
  grand, 129, 152, 272
Canonical equations of Hamilton, 6, 7
Canonical partition function, 127, 147 194
Capillarity, 583
Capturing tendency, 133
Caratheodory, C., 108
Carnot, S., 142
Carnot efficiency, 142
Casimir, H. B. G., 209
Cellular method, 94
Celsius temperature, 135
Central Brillouin zone, 88, 89, 91, 93, 95
Central force, 40, 396
Cesiated work function, 490, 517
Chapman, S., 414
Characteristic of electrode pair, 517
Characteristic equation, 122
Characteristic function, 128, 140
Charge cloud, 44, 53, 60
Charge density, 71
  exchange, 68
  overlap, 65, 68
Charge distribution
  self consistent field, 86
Charge transfer, 73
Chemical desorption, 478
Chemical etching, 586
Chemical potential, 159, 164, 175, 228, 258, 320, 324
Chemical thermionic generator, 642
Chemical valence, 55
Chemical vapor deposition, 580
Child, C. D., 345
Child–Langmuir equation, 350
Classification of orbitals, 54
Clausius, R., 102, 108, 119
Closed system, 122, 130
Coefficient
  electron transport, 419
  ion transport, 421
  Peltier, 220
  of probability, 120
  reflection, 26
    average, 263, 271, 487
  Seebeck, 220
  Thomson, 220
  transport, 227, 414, 415, 416, 417, 418
  volume expansion, 178, 285
Collector sheath, retarding, 426, 430, 436
Collector work function, 460, 525, 533
  effective, 539
  optimum, 554
Collision
  direct, 237, 238
  elastic, 396
  electron–neutral ionization, 402
  restituting, 237, 238
Collision frequency, 246, 395, 417
  average, 405
  effective, 418, 419, 431
  effective electron–ion, 418
  effective electron–neutral, 418
  effective ion–ion, 419
  effective ion–neutral, 419
Collision integral, 237, 241
Commutator, 21
Complementarity, 5
Complex surface, 250
Components, 113, 148
Composite surface, 250, 310
Composition
  of surface, 249
  of system, 148
Conductivity, 209
  electric, 209
  thermal, 209
Configuration interactions, 72
Conjugate fluxes and forces, 207, 217, 224
Conjugates, 207
Conservation
  energy, 21, 107, 217, 224, 391
  mass, 107
  momentum, 391
  particles, 217, 224, 432
  total probability, 15

Constant potential approximation, 173, 175
Constraints, 108, 113
Contact potential, 489
Continuous-eigenvalue spectrum approximation, 173, 176
Continuous one-dimensional rate process, 206
Coordinate space, 8
Correlation, 517
  effects, 100, 256, 265
  energy, 275, 276, 284
  energy function, 280
Correspondence principle, 5
Corresponding stable equilibrium state, 131
Coulomb integral, 68, 70
Coulomb logarithm
  for electrons, 400
  for ions, 401
Covalent binding, 73
Covalent bond, 73
Covalent radius, 74
Coverage, 192, 310
Cowling, T. G., 414
Criterion for stable equilibrium, 113
Critical current, 359, 484, 523
Critical point of operation, 359
Critical voltage, 359, 484
Cross section
  differential, 393
  differential elastic scattering, 238, 241
  differential scattering, 394
  elastic collision, 410
  electron–neutral elastic scattering, 401
  hard-sphere collision, 457
  ionization, 402
  momentum transfer, 394
  neutral–neutral, 401
  total microscopic, 394, 396
  total microscopic elastic scattering, 394
Crystal lattice, 76, 573
  axes in, 573
  body centered cubic, 573
  cubic, 294, 573
  face centered cubic, 573
  hexagonal close packed, 573
Crystals
  directions in, 78, 574, 581
  orientations in, 574, 575
    preferred, 581
  perfect, 572
  planes in, 574, 575, 581, 582
Curie symmetry principle, 226, 247
Current, critical, 359, 484, 523
Current density, probability, 16
Curry, R. H., 454
Curtiss, C. F., 414
Cutler, P. H., 277

Davisson, C., 3
deBroglie, L., 3
Debye, P., 2, 193, 194
Debye frequency, 196
Debye length, 389, 390, 400
Debye temperature, 196, 197, 301, 337
Debye theory, 193, 271, 283, 337
Degenerate state, 18, 35, 41
Degree of coverage, 310
Degrees
  Celsius, 135
  of freedom, 5, 37, 103, 120
    internal, 175
    translational, 109
  Kelvin, 134
DeGroot, R. S., 209
Deignition point, 518
Democritus, 119
Density
  of full monolayer, 310
  number, 386
  operator, 123, 232
  probability, 8, 16, 120
  saturation current, 487
  of states, 36, 84, 96
  surface, 583
Density matrix, 123, 124
  eigenvalues of, 125
  for stable equilibrium state, 127
Derby, S. L., 454, 461
Desplat, J. L., 335
Determinantal function, 71, 72
  antisymmetric, 71
Differential cross section, 393
Differential elastic-scattering cross section, 238, 241, 394

Differential solid angle, 394
Dilute gas, 169
Dilute surface, 191
Dirac, P. A. M., 121
Direct collision, 237, 238
Directions in crystals, 78, 574, 581
Dislocations, 576
Distribution function, 360, 374, 384, 385, 392, 393, 440
  one electron, 266
  one particle, 170, 172, 241, 244
  probability, 8
Double layer, 275, 276, 277, 290
  energy, 284
Double sheath, 367
Double-valued sheath, 438, 461, 539, 545
Dugan, J. V., 454

Effective collision frequency, 418, 419, 431
Effective free path, 421, 422, 427
Effective work function, 488
Efficiency, 141
  Carnot, 142
  reversible, 142
Eigenfunction, 19
  antisymmetric energy, 170
  ground state, 70
  momentum, 91, 92, 93
  one electron, 82
  orthogonal, 18
  single-particle energy, 23
Eigenstate, 19
Eigenvalue, 19
  one electron, 82
  problem, single particle, 23
Einstein, A., 2, 3, 193
Einstein frequency, 194
Einstein's theory of lattice vibrations, 193
Elastic collision, 390
  cross sections, 410
  hard sphere, 396
  integrals, 406
Elastic scattering
  ion–neutral, 401
  neutral–neutral, 401
Electric resistivity, 219, 599
Electrochemical etching, 586
Electrochemical potential, 154, 258
  temperature coefficient of, 336
Electrode output voltage, 454, 517
Electrode system, 260
Electron
  current density, 414
  density, 347
  distribution function, 351
  emission-flux density, 407
  energy-flux density, 427
  flux density, 427
  interelectrode motive, 371
  kinetic energy flux, 414
  mobility, 420
  motive, 229, 247, 254
  number density, 349, 366
  random current density, 190, 254, 257, 262, 324, 326
  random flux density, 260
  saturation current, 486, 519
  saturation-current density, 264, 269, 302, 358, 375
  shell, 54
  subshell, 54
  temperature, 457
  thermal conductivity, 247, 414
  transport coefficients, 419
  valence, 80, 275, 572
  work function, 326
Electron–neutral elastic scattering, 401
  cross section, 401
Electron–neutral ionization collisions, 402
Electron-rich condition, 523, 527
Electronegative surface, 310, 334
Electronegativity, 73, 154, 292, 293
Electronic structure, 58, 72, 83
Electropolishing, 589, 608
Electropositive surface, 310, 311
Emission constant, apparent, 269
Emitter
  saturation-current density, 517
  work function, 450, 460, 530, 554
Emitter sheath, retarding, 425, 429, 435
Energy, 106
  binding, 70

conservation of, 21, 107, 217, 224, 391
correlation, 275, 276, 284
of desorption, 341
eigenfunction, 18, 35, 82
antisymmetric, 170
of electronic structure, 71
single particle, 23
eigenstate, 18, 37, 146
eigenvalue, 18, 34, 35, 137, 146
eigenvalue equation, 17, 22, 57, 80
exchange, 100, 275, 276, 284
Fermi, 84, 96, 265, 278, 281
flux density, 207, 210, 266, 408, 423
gap, 92, 94, 95
Hartree potential, 275
internal, 140, 158
ionization, 45, 56, 187
kinetic, 6
one electron, 71
partial, 159, 165
representation, 126
specific, 167, 182, 213
surface, 583
of transport, isothermal, 213, 215, 217, 227
Energy band, 90, 92, 94, 95
first, 90
lowest, 90
second, 90
theory of, 82, 84
Engelmaier, W., 335
Enthalpy, 140, 165
of desorption, 323, 326, 339
partial, 165
specific, 182, 212, 227
Entropy, 111, 126
change, 112
generation, 203
rate of, 203, 206, 223
partial, 165
principle of increase of, 112
specific, 162, 180, 182
of transport, isothermal, 220
Epicurus, 119
Equation
Boltzmann, 233, 236, 405
characteristic, 122
of chemical equilibrium, 151, 187, 285
Child–Langmuir, 350
energy eigenvalue, 17, 22, 57, 80
Fourier, 215, 218
fundamental, 140, 172, 173
Gibbs, 150, 156, 157, 202
Gibbs–Duhem, 164, 184, 193
Hartree–Fock, 56, 71, 292
hydrodynamic, 242
of motion, 126, 198
Ohm's, 219
Poisson's, 188, 347, 349, 366, 375, 380, 386
Richardson, 264, 407, 450
Richardson–Dushmann, 264
Richardson–Dushmann–Laue, 264
Saha, 187
Saha–Langmuir, 326, 333, 407, 450
Schroedinger, 14, 16, 125, 126
secular, 90, 95
self-consistent-field plus exchange, 56
of stable equilibrium states, 130, 139
of state, 129
Vlasov, 352, 371
Equilibrium state, 17, 107, 199
criterion for stable, 113
stable, 105, 108
Escaping tendency, 134, 153, 168
Evaporation rate, 602, 603
Exchange
charge, 70
charge density, 68
effects, 99, 256, 265
energy, 100, 275, 276, 284
integral, 68, 70
symmetry rules, 49
Expectation value, 10, 14, 20, 103, 104, 235
Extensive property, 106, 154

Face-centered-cubic crystal, 294
Faceting, 583, 587
Fermi–Dirac factor, 172, 235, 265
Fermi–Dirac gas, 171, 173
Fermi energy, 84, 96, 265, 278, 281

Fermi factor, 265
Fermi surface, 84, 96, 277
Fermions, 49, 170, 177
Field emission, 269
First law of thermodynamics, 106
Flux, 200, 203
Flux density, 200, 207
Force, 207
  central, 40, 396
  generalized
    body, 114, 149, 156
    particle, 6
  short range, 237, 383
  spreading, 320
    surface, 190
Fourier equation, 215, 218
Fourier heat flux density, 215, 217
Fractional bond number, 293, 294
Free-electron gas, 82, 83
Free-electron model, Sommerfeld, 82, 83
Free particle, 23
Free path, 395, 417
  average, 405
  effective, 421, 422, 427
  mean, 405
Fundamental equation, 140, 172, 173

Gadzuk, J., 277
Gas
  dilute, 169
  Fermi–Dirac, 171, 173
  free electron, 82, 83
  kinetic theory of semiperfect, 233
  perfect, 145, 174, 181
  semiperfect, 145, 173, 179, 234, 235
  Sommerfeld, 173, 177
Gas approximations, 168
  semiperfect, 179
  Sommerfeld, 265
Gassendi, P., 119
Generalized body force, 114, 149, 156
Generalized coordinates, 6, 8
Generalized Gibbs equation, 150, 202
Generalized momenta, 8
Generalized particle force, 6
Generalized velocity, 6
Germer, L. H., 3
Gibbs, J. W., 119
Gibbs equation, 156, 157, 202
Gibbs free energy, 140, 162, 197, 572
  partial, 162, 283
  specific, 167
  surface, 583
Gibbs–Dalton law, 185
Gibbs–Dalton mixture, 185, 187
Gibbs–Duhem equation, 164, 184, 193
Gibbs–Thomson relation, 583
Gordy, W. E., 293
Grain boundary, 576, 577, 592
Grain growth, 578, 592
Grand canonical distribution, 129, 152, 272
Grand partition function, 129, 170
Grand system, 128
Ground state, 45, 54, 67, 108, 147
  eigenfunction, 70
Guard ring, 483, 490, 510
Gyftopoulos, E. P., 283, 293, 294, 314, 315, 454

Haas, G., 494
Half-Maxwellian distribution, 266, 353
Hamiltonian function, 7, 15, 22, 32
Hamiltonian matrix, 122, 126
Hamiltonian operator, 14, 15, 22, 32, 37, 43, 57, 68, 122
Hamilton's canonical equations, 6, 7
Hansen, L. K., 454
Hard-sphere collision cross section, 457
Hard-sphere elastic collision, 396
Harmonic oscillator, 146
Harris, L. P., 454, 506
Hartree, D. R., 53
Hartree–Fock equation, 56, 71
Hartree–Fock method, 56, 99, 100
Hartree–Fock one-electron equation, 292
Hartree method, 54
Hartree potential energy, 275
Hatsopoulos, G. N., 283, 333
Heat, 116, 136
  of desorption, 341
  engine, 142
  and flow of entropy, 117
  Joulean, 219, 595, 596
  specific, 180, 181

of transport, 214
treatment, 586, 587
Heisenberg, W., 4, 121
Heisenberg uncertainty principle, 5
Helmholtz, H. L. F., 119
Helmholtz free energy, 128, 139, 140, 145
Hermiticity, 13
Hermitian operator, 13
linear, 13, 122
Herring, C., 298, 308, 309
Heterogeneous system, 154, 155
Heteronuclear molecule, 73
Hexagonal-close-packed crystal, 294, 573
Hinnov, E., 454
Hirschberg, J. G., 454
Hirshfelder, J. O., 414
Homogeneous system, 154, 155
Homonuclear molecule, 66
Homopolar binding, 72
Hooke, R., 119
Houston, J. M., 401, 454, 494, 506
Hydrodynamic equations, 242

Ideal characteristic, 526
Identical state, 104
Ignited mode, 457, 518, 527, 528
Ignition, 527
current density, 565, 567
point, 518
temperature, 469
Image force, 256
Impact parameter, 238, 392, 396
Impurities, 576, 605
Independent properties, 105
Indeterminacy principle, 4
Inelastic collision, 390
integral, 406
Intensive property, 154
Interelectrode motive, 351
Interelectrode spacing, 517, 561, 591
Interelectrode system, 258
Interfacial motive, 275
Internal degrees of freedom, 175
Internal energy, 140, 158
Internal structure of particles, 175
Internal voltage drop, 406
Interstitial impurities, 576

Ion
current density, 414, 457
emission-flux density, 407
mobility, 401, 421
motive, 284
number density, 366
production rate, 458
random-flux density, 325, 326
saturation-current density, 375
transport coefficient, 421
work function, 313
Ionic binding, 74
Ionic particle, 316, 317
Ionic radius, 74, 317, 327
Ionization
cross section, 402
energy, 45, 56, 187
of orbital, 56
rate, 224
Ionized state, 401
Ion–neutral elastic scattering, 401
Ion-rich condition, 519, 526
Ion-richness ratio, 375, 376, 378, 457
Irreversible process, 109, 138
Isolated system, 107
Isothermal energy of transport, 213, 215, 217, 227
Isothermal entropy of transport, 220
Isothermal flow, 211

Joule, J. P., 119
Joulean heat, 219, 595, 596
Juretschke, H. J., 277

K-shell, 54
K-space, 84, 87, 88, 90, 92, 96
Keck, J., 439, 440, 443, 469
Kelvin temperature, 134
Kinetic energy, 6
flux density, 239, 266
Kinetic theory of semiperfect gas, 233
Knudsen effect, 214
Koopmans' theorem, 71, 99
Koskinen, M. F., 278

L-shell, 54
Lagrangian function, 6
Landau, L. D., 119
Langmuir, I., 314, 315, 321, 322, 333, 334, 345, 346, 351, 363, 367, 482

Langmuir's space-charge theory, 357
Laplacian operator, 57
Lapp, M., 506
Lattice, reciprocal, 86, 87
  See also Bravais lattice; Crystal lattice
Law of thermodynamics
  first, 106
  second, 108
Leibnitz, G. W., 119
Levine, J. D., 314, 315, 454
Lieb, D., 382, 438, 439, 454
Line defect, 578
Linear combinations of atomic orbitals (LCAO), 63, 67, 73
Linear Hermitian operator, 13, 122
Linear process, 204
Linearity, 13
Local phenomena, 221, 247
Local phenomenological equations, 228, 230
Long-range force, 236, 383
Lorentz plasma, 412, 417
Loucks, T. L., 277
Low-energy electron diffraction (LEED), 481
Luke, K. P., 319

M-shell, 54
McCandless, R. J., 454, 461, 559
McIntyre, R. G., 371
Magneton, Bohr, 43
Mass
  conservation of, 107
  reduced, 38, 397, 399
Maugis, D., 330
Maxwell, J. C., 119
Maxwellian distribution, 146, 235, 266, 403
Mean free path, 405
Mechanical damage, 586, 587
Mechanical polishing, 586
Metallic valence, 293
Method of configuration interactions, 72
Miller indices, 574
Mixed state, 125
Mixed surface, 310, 335
Mobility, 247, 414
  of electrons, 420
  of ions, 401, 421
Molecular orbital, 60, 71
  antibonding, 60
  bonding, 60
  classification, 60
  method, 72
Molecular statistics, 119
Momentum
  conservation of, 391
  eigenfunction, 91, 92, 93
  representation, 234
  space, 9
  transfer cross section, 394
Monocrystalline surface, 252
Morse curve, 54
Motive, 254
  electron, 229, 247, 254
  interelectrode, 351
  interfacial, 275
  ion, 284
μ-space, 235, 236, 239
Mullins, W. W., 585
Mutual stable equilibrium, 109, 132, 150
  condition for, 133, 150, 201

Neutral atom
  arrival-rate density, 330
  random flux density, 323, 326
Neutral–neutral cross section, 401
Neutral–neutral elastic scattering, 401
Nichols, M. H., 298, 308, 309
Nighan, W. L., 454
Nonadiabatic process, 106
Nonequilibrium state, 107, 131, 198
Nonlinear process, 204
Nonpure state, 125
Nonwork, 105, 114
Nottingham, W. H., 483, 484, 486
Nuclear reactor thermionic generator, 635
Number density, 386

Occupation number, 170
Ohm's equation, 219
One-dimensional rate process, continuous, 206
One-electron approximation, 82, 97, 275

One-electron distribution function, 266
One-electron eigenfunction, 82
One-electron eigenvalue, 82
  equation, 53, 54, 90
  problem, 80
One-electron energy, 71
One-electron potential energy, 82
One-particle approximation, 168, 169, 335
One-particle distribution function, 170, 172, 241, 244
One-particle energy eigenvalue problem, 169
One-particle partition function, 174
Onsager, I. L., 209
Onsager reciprocal relations, 209, 210, 217
Open-circuit voltage, 523
Open system, 128, 130, 147, 148
Operator, 12, 13
  single particle, 51
Optimization of performance, 554, 591
Orbit, 42
Orbital
  antibonding, 60
  of atom, 44, 46, 54
  classification of, 54
  molecular, 60, 71
  spin, 44, 53, 54, 71, 170, 275
Ordinary closed system, 122, 130
Ordinary open system, 128, 130, 147, 148
Orientations in crystals, 574, 575
  preferred, 581
Orthogonal eigenfunction, 18
Oscillator, harmonic, 146
Outgassing, 611
Output current
  characteristic, 451, 517, 518
  density, 357, 517
Overlap
  charge, 65, 70
  charge density, 65, 68
  charge fraction, 65
  integral, 68
Overlapping bands, 95
Oxide cathode, 250
P-language, 10
Parity, 15, 30
Partial energy, 159, 165
Partial enthalpy, 165
Partial entropy, 165
Partial Gibbs free energy, 162, 283
Partial pressure, 185, 188
Partial property, 159
Partial volume, 165
Partially covalent bond, 74
Partially ionic bond, 74
Particle-flux density, 207, 210
Partition function, 139, 144, 163, 319
  canonical, 127, 147, 194
  grand, 129, 170
  one particle, 174
Patchy surface, 303
Path, 105
Pauli exclusion principle, 43, 52, 54, 71, 83
Pauling, L., 294
Peltier coefficient, 220
Peltier effect, 220, 495
Perfect crystal, 572
Perfect gas, 145, 174, 181
Phase, 155
Phase space, 7, 120
Phenomenological equations, 208, 209, 210, 217, 416
  local, 228, 230
  spatial, 226, 248
Photoelectric effect, 3
Pigford, T. H., 382, 436, 439
Pines, D., 400
Planck, M., 2, 106
Plane wave, 33, 86, 88
Planes, in crystals, 574, 575, 581, 582
Plasma, 387, 447, 448
  boundary conditions, 424, 428, 434
  Lorentz, 412, 417
  motive difference, 455, 457
  region, 423
Point-particle approximation, 173, 176
Points just inside, 255, 276
Points just outside, 257, 261, 336
  patchy surface, 303, 305, 308
Poisson's equation, 188, 347, 349, 366, 375, 380, 386
Polycrystalline surface, 302
Potential energy, one electron, 82

Power relation, three halves, 350
Preferred orientation, 581
Pressure, 114, 140, 156
  partial, 185, 188
Prigogine, I., 209
Primitive vector, 77
Principle
  conservation of energy, 21
  correspondence, 5
  increase of entropy, 112
  indeterminacy, 4, 121
  state, 108, 113
  superposition, 13, 17
  uncertainty, 5
Pritchard, H. O., 293
Probability
  amplitude, 8
  coefficient of, 120
  conservation of total, 15
  current density, 16
  density, 8, 16, 120
  distribution, 120
  distribution function, 8
  total, 8, 15
Process, 105
  adiabatic, 105
  boundary value, 203, 204, 208
  continuous one-dimensional rate, 206
  irreversible, 109, 138
  linear, 204
  nonadiabatic, 106
  nonlinear, 204
  rate, 198
  reversible, 109
  steady rate, 198, 200
Property, 104
  additive, 106
  extensive, 106, 154
  intensive, 154
  partial, 159
  specific, 161
Pure nonuniform surface, 250, 301
Pure state, 125, 235
Pure uniform surface, 250, 252
Pyrolytic deposition, 580, 606, 607

Q-language, 10, 15
Quantum number, 35, 40, 42
Quasistable state, 130, 131

Radioisotopic thermionic generator, 627
Radius
  Bohr, 42, 44, 59, 74
  covalent, 74
  ionic, 74, 317, 327
Random flux density, 189
Rasor, N. S., 314
Rasor plot, 327
Rate of entropy generation, 203, 206, 223
Rate of ionization, 224
Rate process, 198
  continuous one dimensional, 206
Reciprocal lattice, 86, 87
Reciprocal space, 87
Reduced mass, 38, 397, 399
Reduced wave vector, 88, 277
Reflection coefficient, 26
  average, 263, 271, 487
Reitz, J. R., 454
Relaxation time, 403, 404, 405
Reservoir, 109
  standard, 111
Rest mass, 104
Restituting collision, 237, 238
Retarding collector sheath, 426, 430, 436
Retarding emitter sheath, 425, 429, 435
Reversible efficiency, 142
Reversible process, 109
Richardson–Dushmann equation, 264
Richardson–Dushmann–Laue equation, 264
Richardson equation, 264, 407, 450
Richardson plot, 270, 487
Rufeh, F., 333
Rydberg unit, 59

Saha equation, 187
Saha–Langmuir equation, 326, 333, 407, 450
Saturation-current density, 487
Schottky effect, 254, 267, 269, 460
  anomalous, 306, 308
Schottky plot, 269, 270, 306, 486, 487, 520
Schroedinger, E., 121
Schroedinger equation, 14, 16, 125, 126
Second Brillouin zone, 88

Second law of thermodynamics, 108
Secular equation, 90, 95
Seebeck coefficient, 220
Seebeck effect, 220, 495
Seebeck emf, 495, 496, 497, 499, 509
Seitz, F., 280, 282
Self-consistent-field charge distribution, 86
Self-consistent-field method, 54
Self-consistent-field plus exchange equation, 56
Semiperfect gas, 145, 173, 179, 234, 235
Semiperfect gas approximation, 179
Semiperfect surface, 191
Set of components, 113
Sheath, 387
  accelerating collector, 425, 429, 436
  accelerating emitter, 425, 429, 435
  double, 367
  double valued, 438
  height, 455
  retarding collector, 426, 430, 436
  retarding emitter, 425, 429, 435
Sheldon, J. W., 401, 454
Shell of electrons, 54
Short-range force, 237, 383
Simple system, 140
Single-particle eigenvalue problem, 23
Single-particle energy eigenfunction, 23
Single-particle operator, 51
Skinner, H. A., 293
Slater, J. C., 100, 274, 277, 280, 293
Slater determinant, 52, 56, 72, 275, 292
Smith, G. F., 277, 283
Smoluchowski, R., 289, 290, 291
Solar thermionic generator, 615, 616
Solid, surface of, 249
Sommerfeld free-electron model, 82, 83
Sommerfeld gas, 173, 177
Sommerfeld gas approximation, 265
Space-charge theory, 345
  Langmuir's, 357
Spatial phenomena, 221, 247
Spatial phenomenological equations, 226, 248
Specific energy, 167, 182, 213
Specific enthalpy, 182, 212, 227
Specific entropy, 162, 180, 182
Specific Gibbs free energy, 167
Specific heat, 180
  under constant pressure, 181
  under constant volume, 181
Specific property, 161
Specific volume, 162
Spin, 15, 42, 122
Spin-orbital, 44, 53, 54, 71, 170, 275
Spreading force, 320
  surface, 190
Stable equilibrium state, 105, 108
Standard reservoir, 111
State, of a system, 5, 7, 8, 104, 122
  allowed, 108
  antisymmetric, 49, 50
  bound, 40
  corresponding stable equilibrium, 131
  degenerate, 18, 35, 41
  equilibrium, 17, 107, 199
  ground, 45, 54, 67, 108, 147
  identical, 104
  mixed, 125
  nonequilibrium, 107, 131, 198
  nonpure, 125
  pure, 125, 235
  quasistable, 130, 131
  stable equilibrium, 105, 108
  stationary, 17
  steady, 107, 199
  types of, 105, 107
State principle, 108, 113
Steady-rate process, 198, 200
Steady state, 107, 199
  accelerating field, 267
  Boltzmann equation, 239, 241, 384. 440
  zero field, 263
Steiner, D., 293, 294, 454
Stickney, R. E., 335
Stone, P. M., 454
Substrate, 310
Surface, 190, 249
  adsorbate, 312
  complex, 250
  composite, 250, 310
  composition of, 249
  coverage, 192
  density, 583
  dilute, 191
  electronegative, 310, 334

electropositive, 310, 311
energy, 583
Fermi, 84, 96, 277
Gibbs free energy, 583
just inside, 255
just outside, 261, 262
mixed, 310, 335
monocrystalline, 252
patchy, 303
polycrystalline, 302
pure nonuniform, 250, 301
pure uniform, 250, 252
semiperfect, 191
of solid, 249
spreading force, 190
structure of, 249
system, 190, 314
valence, 291, 293
System, 104
closed, 122, 130
heterogeneous, 154, 155
homogeneous, 154, 155
interelectrode, 258
ordinary closed, 122, 130
ordinary open, 128, 130, 147, 148
simple, 140
surface, 190, 314

Taylor, J. B., 314, 315, 333, 334
Temperature, 113, 134
Celsius, 135
Debye, 196, 197, 301, 337
equality, 135, 150
Kelvin, 134
Temperature coefficient
electrochemical potential, 336
work function, 269, 295, 323
Theory of energy bands, 82, 84
Thermal conductivity, 215, 218, 408, 573, 599
Thermal diffusion ratio, 247, 414
Thermal expansion, 573, 599
Thermionic generator
chemical, 642
nuclear reactor, 635
solar, 615, 616
topping, 645
Thermodynamic potential, 149, 201, 202, 415
Thermodynamic temperature scales, 134
Third Brillouin zone, 88
Thomson coefficient, 220
Thomson effect, 220, 495
Three-halves power relation, 350
Topping thermionic generator, 645
Total microscopic cross section, 394, 396
Total microscopic elastic-scattering cross section, 394
Total potential, 114, 158, 162, 165, 167, 168, 316, 415
equality, 151
Total probability, 8, 15
Trace, 124
Transport coefficients, 227, 414, 415, 416, 417, 418
Types of states, 105, 107

Uncertainty principle, 5
Unit cell, 76, 572
Wigner–Seitz, 77, 78, 80, 87, 89
Useful work, 143

Vacancies, 576, 577
Valence
chemical, 55
metallic, 293
surface, 291, 293
Valence electron, 80, 275, 572
orbitals, 67
Vapor deposition, chemical, 580
Vector, reduced wave, 88, 277
Velocity, generalized, 6
Vlasov equation, 352, 371
Voltage
critical, 359, 484
drop, internal, 406
open-circuit, 523
Volume
expansion coefficient, 178, 285
specific, 162
von Neumann, J., 119

Wang, C., 382, 438, 439
Warner, C., 454
Wave function, 8, 14, 121, 125

Wave vector, 83
  lattice, 87
  reduced, 88, 277
Webster, H. F., 494, 506
Wigner, E. P., 274, 279, 280, 282
Wigner–Seitz unit cell, 77, 78, 80, 87, 89
Wilkins, D. R., 453, 454, 455, 459, 461, 559
Work, 105, 136
  available, 111
  useful, 143
Work function, 262, 270, 277, 281, 284, 294, 295, 488, 530
  apparent, 269
  bare, 313, 320, 483, 582, 583, 586
  cesiated collector, 460, 525, 533, 554
  effective, 488
  electron, 326
  emitter, 450, 460, 530, 554
  ion, 313
  temperature coefficient of, 269, 295, 323
Work hardening, 579

Zero-field steady state, 263
Zero rest mass, 104