MASTERS SERIES IN PHYSICS AND ASTRONOMY

MODERN VACUUM PHYSICS

AUSTIN CHAMBERS

CHAPMAN & HALL/CRC

Vacuum Karan

وب سایت وکیوم کاران

مرجع فناوری خلاء

به زبان فارسی

مقالات علمی و فنی فناوری های تولید، اندازه گیری و کاربردهای خلاء

تاسیس۱۳۸۹

www.vacuumkaran.com

MASTERS SERIES IN PHYSICS AND ASTRONOMY

MODERN VACUUM PHYSICS

MASTERS SERIES IN PHYSICS AND ASTRONOMY

Edited by David S. Betts

Department of Physics and Astronomy,
University of Sussex, Brighton, UK

Core Electrodynamics
Sandra C. Chapman

Atomic and Molecular Clusters
Roy L. Johnston

Quantum Theory of Solids
Eoin P. O'Reilly

Basic Superfluids
Anthony M. Guénault

Modern Vacuum Physics
Austin Chambers

MASTERS SERIES IN PHYSICS AND ASTRONOMY

MODERN VACUUM PHYSICS

AUSTIN CHAMBERS

Department of Physics
University of York, UK

CHAPMAN & HALL/CRC

A CRC Press Company
Boca Raton London New York Washington, D.C.

Library of Congress Cataloging-in-Publication Data

Chambers, A.
Modern vacuum physics / Austin Chambers.
p. cm. -- (Taylor & Francis masters series in physics and astronomy ; 4)
Includes bibliographical references and index.
ISBN 0-8493-2438-6 (alk. paper)
1. Vacuum. I. Title. II. Master's series in physics and astronomy ; 4.

QC166.C444 2004
533′.5—dc22 2004049375

International Standard Book Number 0-8493-2438-6
Library of Congress Card Number 2004049375
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

Dedication

To

Oliver Heavens,

Founding Head at the University of York in 1964 of the

Department of Physics,

a place of scholarship, fellowship, and humanity.

Series Preface

The Masters Series of textbooks is aimed squarely at students taking specialized options in topics either within the primary areas of physics and astronomy, or closely related to them in areas such as, for example, physical chemistry and environmental science. Appropriate applied subjects are also included. The student interest group will typically be studying in the final year of their first degree or in the first year of postgraduate work. Some of the books may also be useful to professional researchers finding their way into new research areas, and all are written with a clear brief to assume that the reader has already acquired a working knowledge of basic core physics.

The series is designed for use worldwide in the knowledge that wherever physics is taught at degree level, there are core courses designed for all students in the early years followed by specialized options for those consciously aiming at a more advanced understanding of some topics in preparation for a scientific career. In the U.K., there is an extra year for the latter category, leading to an M.Phys. or M.Sci. degree preceding entry to postgraduate M.Sc. or Ph.D. degrees, while in the U.S., specialization is often found mainly in master's or doctorate programs. Elsewhere, the precise modulations vary, but the development from core to specialization is normally part of the overall plan.

Authors for the series have usually been able to draw on their own lecture materials and experience of teaching in preparing draft chapters. It is naturally a feature of specialist courses that they are likely to be given by lecturers whose research interests relate to them, so readers can feel that they are gaining from both teaching and research experience.

Each book is self-contained beyond an assumed background to be found in appropriate sections of available core textbooks on physics and useful mathematics. There are, of course, many possible preferences, but examples might well include Richard P. Feynman's classic, three-volume *Lectures on Physics* (R.P. Feynman, R.B. Leighton, and M. Sands, published by Addison-Wesley, volume 1 [1963], volume 2 [1964], volume 3 [1965]), and Mary L. Boas' *Mathematical Methods in the Physical Sciences* (2nd ed., John Wiley & Sons, 1983). The primary aim of books in the series will be to enhance the students' knowledge base so they can approach the corresponding research literature in their chosen field with confidence. It is not intended to offer major treatises at the forefront of knowledge, accessible only to a few world experts. Rather, the books are student-oriented, didactic in character, and written to build the confidence of their readers. Lengths are mostly in the region of 200 pages, with generous numbers of figures where these are helpful.

Different topics may have different styles of questions and answers, but authors are encouraged to include questions at the ends of most chapters, with answers at the end of the book. I am inclined to the view that simple numerical answers, though essential, are often too brief to be fully satisfactory, particularly at the level of this series. At the other extreme, model answers of the kind that examination boards normally require of lecturers would make it too easy for lazy readers to think they had understood without actually trying. So the 'house style' includes, in addition to numerical answers, advice about difficult steps in calculations, lines of detailed argument in cases where the author feels that the 'typical' reader is likely to get stuck, and any algebraic manipulation which may get in the way of proceeding from good understanding to the required answer. Broadly, what is given is enough help to make it possible for the typical reader to experience the feel-good factor of actually finishing questions, but not so much that all that is needed is passive reading of a model answer.

David S. Betts
University of Sussex

Preface

The extent to which vacuum contributes to progress in scientific research and in the technology of manufacturing has increased markedly in recent decades. Its applications in the large machines of particle physics and in the fabrication of microelectronic devices are just two of many. From its origins in the mid-17th century, the long history of the subject includes its role in the discovery of the electron and, in one of the earliest examples of industrial mass-production, the manufacture of electric light bulbs. Vacuum technology, which provides the equipment and expertise to create the vacuum environment appropriate for a particular purpose, has evolved in response to needs that arise in a wide variety of research and development laboratories and in industrial production. This evolution has been bound up with related progress in vacuum science, the understanding of the physical principles that underlie the subject.

Vacuum technology is interdisciplinary, with a broad foundation in physics that extends to embrace aspects of mechanical, electrical, and chemical engineering, chemistry, and materials science. Its subject matter is arguably "applied physics," and its practical realization requires engineering of the highest quality. As it continues to grow in scope and sophistication, the number of people it involves increases and, correspondingly, so does the importance of teaching its fundamental principles and methods. This book is an attempt to present the concepts, principles, and methods of the subject's broad foundation for those with backgrounds in physics, engineering, or materials science, and for students of these subjects. Much of the underlying physics is of the long and well-established "classical" sort, applied in a modern context. It has to be said, however, that due to the increasing pressure that has been evident in recent years to change physics curricula in both schools and universities in the U.K., there has been a tendency to give these classical matters less attention than hitherto. For this reason I have felt it proper to deal with some of them at greater length than might otherwise have been the case.

The first half of the book is intended to give the reader a thorough grounding of basic knowledge about gases and vapors under rarefied conditions, how they behave when they come into contact with surfaces, and how they flow. These matters are necessary for understanding the means of producing and measuring vacua — pumps and gauges — which is the subject of the book's second half, together with examples of the analysis of representative systems and descriptions of some of the exciting current developments in science and manufacturing in which vacuum plays an important part. In describing pumps and gauges, emphasis is again on explaining how they

work in terms of first principles and basic physics. In conclusion, there is a chapter about strictly practical matters.

The book is the outcome of experience gained over many years in teaching the subject to undergraduate and graduate physics students in the University of York and, with colleagues, to graduate students from other universities and technical staff in various industrial and research establishments in the U.K. Particularly in the early chapters, many of the points made are illustrated numerically within the text. At the conclusion of these chapters, a set of exercises and problems is given whose completion is important for the reinforcement of learning and for acquiring the skill to analyze and predict the performance of components and systems. I hope that the book will equip readers to approach vacuum systems confidently, prepare them for the study of more specialized texts, some of which are mentioned in the Bibliography, and persuade them of the scope and value of vacuum technology in general scientific and technological progress.

Acknowledgments

I wish to record thanks to many individuals who, in various ways, have made the writing of this book possible. At the University of York, successive heads of the Department of Physics — Oliver Heavens, Michael Woolfson, James Matthew, Geoff Pert, and Rex Godby — have encouraged and supported teaching and research in vacuum physics. I particularly appreciate James Matthew's collaboration in teaching and also the experience of working alongside colleagues in surface and thin film research — Martin Prutton, Tom Gallon, and, more recently, Sarah Thompson. Dave Coulthard's practical knowledge and expertise have frequently helped me out of experimental difficulties and, in doing so, deepened my understanding of the subject. Ian Wright, and also Mark Wilson and Peter Main, have given valuable assistance in creating many of the figures for the book, and Yvonne Cook, Academic Administrator, has rescued me from word-processing crises on many occasions. I owe particular thanks to former project students, especially Neil Condon, and research students, Andrew Chew and David York, the first two now at BOC Edwards, for their enthusiasm for the subject and collaboration.

To John Colligon of Manchester Metropolitan University and Ron Reid of CLRC Daresbury Laboratory, who some years ago were instrumental in setting up short-course teaching in vacuum technology that was associated with the Vacuum Group of the Institute of Physics and the British Vacuum Council, I have a profound debt. My involvement in this activity was a considerable stimulus that led to valued collaborations with Bryce Halliday and Keith Fitch (now deceased).

It is also a pleasure to record thanks to various individuals in manufacturing companies and other organisations. In a period of teaching involvement with BOC Edwards, I had the good fortune to collaborate with Gordon Livesey, Alan Troup, Malcolm Baker, Dick Amos, Steve Hoath, and others. They will recognize their influence in parts of the text. Among a number of people in the vacuum community who have assisted teaching activities I am particularly grateful to David Fellows of Arun Microelectronics Ltd., Rob White of Pfeiffer Vacuum, Chris Stone and Neil Henderson of MKS Instruments, and Vahid Krupic of Leybold. In similar ways, John Greenwood at NPL and Robert Pearce at JET have generously given their time.

This book therefore reflects the influence of many experts, though I have not sought their assistance during its preparation and hope that I have done justice to their specialisms. I also wish to acknowledge the debt that a textbook writer has to established works in the subject. Many are mentioned in the bibliography, but notably they include, in a number of editions, the

books of O'Hanlon, Hablanian, Harris, and Roth. All the above being said, the responsibility for any flaws is, of course, wholly mine.

At Taylor and Francis, in the long period of the book's gestation, Tony Moore, Luke Hacker, Grant Soanes, Frances Horrocks, and Janie Wardle have been patient and helpful, and more recently at CRC Press, Sunil Nair, Jasmin Naim, Jamie Sigal, and Gail Renard. And I have been grateful throughout to the series editor, David Betts of the University of Sussex, for his encouraging comments and sound advice, which have been of critical importance.

But, above all, the book would not have been possible without the wise counsel and constant support and affection of my wife Brenda.

Contents

1 Introduction
1.1 What Is a Vacuum? ... 1
1.2 History and Present Importance of Vacuum Science and Technique ... 2
1.3 Units ... 5
1.4 The Ranges of Vacuum ... 6
1.5 Plan of This Book ... 8

2 The Physics of Gases and Vapors
2.1 Specifying the Quantity of a Gas, the Mole, Ideal, and Real Gases ... 11
2.2 Gas Mixtures and Partial Pressures ... 16
2.3 Vapors and Vapor Pressures ... 17
Exercises ... 21
Problems ... 23

3 The Molecular Description of Gases
3.1 Introduction ... 25
3.2 Kinetic Theory of Gases ... 26
3.3 The Maxwell–Boltzmann Distribution of Molecular Speeds ... 27
3.4 Molecular Impingement Rate, J ... 30
3.5 Pressure and Molecular Number Density, n ... 32
3.6 Molecular Collisions and Mean Free Path, λ ... 34
3.7 Summary ... 38
3.8 Evaporation and Condensation ... 39
3.9 Knudsen Number: Continuum and Molecular States of Gas ... 41
3.10 Internal Friction and Viscosity in Gases ... 43
3.11 Heat Conduction in Gases ... 45
3.12 Thermal Transpiration ... 46
Exercises ... 47
Problems ... 48

4 Surface Processes and Outgassing
4.1 Introduction ... 49
4.2 The Scattering of Molecules by a Surface ... 50
4.3 Diffuse Scattering from a Surface — the Knudsen Cosine Law ... 53
4.4 Adsorption and Desorption ... 55
4.5 Outgassing ... 62

4.6 Molecular Drag 70
4.7 Sputtering 72
Exercises 75
Problems 75

5 Gas Flow and Pumping
5.1 Introduction: Flow Regimes 77
5.2 Measures of Flow: Throughput and Pumping Speed 78
5.3 Conductance 80
5.4 Continuum Flow 82
5.5 Dynamical Analysis of Continuum Flow 84
5.5.1 Bernoulli's Equation 85
5.5.2 Effects of Compressibility 86
5.5.3 Flow through Short Ducts and Apertures: Choked Flow 88
5.5.4 Continuum Flow through Pipes 92
5.6 Molecular Flow 95
5.6.1 Molecular Flow through a Long Pipe 95
5.6.2 Molecular Flow through an Aperture 97
5.6.3 The Concept of Transmission Probability 98
5.6.4 Molecular Flow through Short Pipes 100
5.6.5 Transmission Probability by Monte Carlo Methods 102
5.6.6 Entrance and Exit Effects 102
5.6.7 Molecular Beaming 104
5.6.8 Combining the Conductances of Elements in Series 105
5.7 The Pumping Process, Pump-Down Time, and Ultimate Pressure 108
5.8 Differential Pumping 113
Exercises 114
Problems 116

6 Creating a Vacuum — Pumps
6.1 Introduction 119
6.2 Positive Displacement Pumps 122
6.2.1 The Rotary Vane Pump 122
6.2.2 The Roots Pump 126
6.2.3 Other Dry Pumps 130
6.3 Momentum Transfer Pumps 138
6.3.1 The Molecular Drag Pump 138
6.3.2 The Turbomolecular Pump 143
6.3.3 The Turbo-Drag Pump 153
6.3.4 The Vapor Jet (Diffusion) Pump 153
6.4 Capture Pumps 160
6.4.1 The Sorption Pump 160
6.4.2 The Cryopump 162
6.4.3 Water Vapor Pumps 171

6.4.4 The Titanium Sublimation Pump 171
6.4.5 Nonevaporable Getter (NEG) Pumps 173
6.4.6 The Sputter-Ion Pump 176

7 Measuring a Vacuum
7.1 Introduction 183
7.2 Gauges Measuring Pressure Directly 186
7.2.1 U-Tube Manometers 186
7.2.2 The McLeod Gauge 187
7.2.3 Gauges Depending on Deformation of a Sensor 189
7.2.4 The Capacitance Diaphragm Gauge 190
7.3 Gauges Depending on Thermal Conduction 195
7.3.1 The Pirani Gauge 198
7.3.2 The Thermocouple Gauge 200
7.4 The Spinning Rotor Gauge (SRG) 201
7.5 Ionization Gauges 205
7.5.1 The Hot Cathode Gauge 205
7.5.2 Cold Cathode Gauges 215
7.6 The Calibration of Vacuum Gauges 218
7.7 Partial Pressure Gauges 221
7.7.1 General 222
7.7.2 The Quadrupole Instrument 224
7.7.3 Magnetic Sector and Time-of-Flight Analyzers 230
7.7.4 Analysis and Display 232

8 Illustrative Examples and Representative Laboratory Systems
8.1 Introduction 237
8.2 Pumping Water Vapor 237
8.3 Pumping through a Long Pipe 240
8.4 Differential Pumping in an Electron-Optical Application 244
8.5 High Flow Rates at Low and Medium Vacuum with Roots and Rotary Pumps in Combination 246
8.6 System for Vacuum Coating at High Vacuum, 10^{-6} mbar 247
8.7 High and Ultrahigh Vacuum Pumping with Turbomolecular Pumps 253
8.8 Ion-Pumped Ultrahigh-Vacuum System 257
Questions 258

9 Applications in Industrial Products and Scientific Instruments
9.1 Introduction 261
9.2 Vacuum-Insulated Panels 261
9.3 Vacuum Deposition of Thin Film Coatings 264
9.4 Microelectronic Fabrication 269
9.5 Flat-Panel Displays 272
9.6 Mirrors of the Very Large Telescope (VLT) 278

9.7 Synchrotron Radiation Sources 281
9.8 Controlled Thermonuclear Fusion 285
9.9 Gravitational Wave Detection 289
9.10 Particle Physics: the LEP and LHC 292

10 Vacuum Components, Construction, and Leak Detection
10.1 Introduction 299
10.2 Materials 300
10.3 Demountable Connections 302
10.4 Valves 305
10.5 Bellows and Motion Feedthroughs 306
10.6 In-House Design 310
10.7 Cleaning 312
10.8 Leaks and Their Detection 312

Answers to Exercises and Problems 321

Bibliography 325

References 327

List of Symbols 333

Index 337

1

Introduction

1.1 What Is a Vacuum?

From the practical point of view, a vacuum is defined as a space in which the pressure of gas is low compared with what is normal — that is to say, atmospheric pressure. This definition correctly allows a rather broad interpretation because the vacuums encountered span a wide range of pressure values. Thus, a pressure of the order of one tenth of atmospheric is quite adequate to exploit the force due to the difference between this and atmospheric pressure, as in vacuum molding applications. By contrast, for the acceleration of charged particles to relativistic energies in modern high-energy physics experiments, much more rarefied conditions are necessary, with gas densities that are approximately 15 orders of magnitude less than that prevailing at atmospheric pressure, and therefore at very low pressure indeed. Between these limits there is a whole spectrum of other applications at various degrees of vacuum, examples of which will be given in due course.

The measure of a vacuum, therefore, is its associated pressure p, the SI unit of which is the Pascal (Pa), equal to a force of 1 N/m^2. It should be stressed that, with the single exception of applications of the type mentioned above that involve force, the quality that makes a vacuum suitable for a particular purpose is not primarily its pressure — which is caused by molecular impacts on bounding surfaces — but the state of rarefaction of the gas itself, as measured by the number of molecules contained in an evacuated volume, and specified by the number of molecules per unit volume n. Thus, it is n that determines how far a beam of electrons or neutral atoms can travel through low-pressure gas without being seriously attenuated, or how rapidly a clean surface is contaminated by molecular bombardment. These are matters of concern in the manufacture of television tubes, the coating of items such as compact discs with a thin reflecting layer of aluminum, and in surface science investigations. However, the measure p of vacuum and the determining quantity n are simply related in the fundamental equation $p = nkT$ (to be discussed), in which k is

Boltzmann's constant and the temperature T can, in practice, be taken to be constant at 295 K typically, so that p is directly proportional to n and provides an appropriate measure of the vacuum.

A widely known fact from introductory courses about gases is that under standard conditions of temperature and pressure (T = 0°C = 273 K and atmospheric pressure p = 101.3 kPa), a centimeter cube of gas contains 2.5×10^{19} molecules, and therefore 2.5×10^{25} per m^3. These huge numbers will be put into context in Chapter 3, but it is clear that, even for a vacuum 15 orders of magnitude lower than atmospheric pressure, there will still be $n = 2.5 \times 10^{10}$ molecules per m^3, so that the vacuum space is far from empty. The lowest laboratory-made vacuum achieved thus far, under very specialized conditions, corresponds to n being of the order 10^7 molecules per m^3. The reasons for these limiting values relate to the nature of the surfaces that confine the vacuum, which do not have a passive role and which will be discussed later. At altitudes of about 800 km where observational earth satellites typically orbit in a virtually friction-free atmosphere, n is approximately 10^{15} m^{-3}. In interplanetary space it is about 10^7 m^{-3}, due mainly to the solar wind. In the space between galaxies there is thought to be on average less than one atom of hydrogen per cubic meter.

1.2 History and Present Importance of Vacuum Science and Technique

The concepts of rarefied gas described above may seem quite reasonable to us today. Even a perfect vacuum can be contemplated as an abstract ideal. But in ancient times, to most philosophers trying to understand the nature of the world, matter, and the heavens, a space empty of matter was inconceivable. To Aristotle (384 to 322 B.C.) vacuum was a logical impossibility, and even up to medieval times the dominant thinkers of the day followed him in this belief. Thought and argument were sufficient grounds on which to deny the existence of vacuum or the possibility of its existence. Medieval scholars expressed this as *horror vacui* — nature renounces the possibility of vacuum. Fortunately, in the late 16th and early 17th centuries, with the stimulus of Galileo (1564 to 1642) and others who were concerned with matters such as the weight of the air and the workings of water siphons and pumps, methods of inquiry started to change. Natural philosophy began to acquire a new dimension in which ideas were tested by carefully designed experiments, thus establishing the framework of interplay between theory and experiment by which science has progressed ever since.

The story of how vacuum science and technology started in the 17th century and developed through the 18th and 19th centuries is related in an article by Madey (1983) and by Andrade (1960). An early event of

profound significance was Torricelli's experiment of 1643, shown in Figure 1.1. A strong glass tube about 1 m long and closed at one end was filled with mercury and temporarily closed with a finger at the open end while the tube was inverted and the open end placed below the level of mercury in a bowl. As a result, some mercury flowed out of it into the bowl to leave a column of mercury about 760 mm above the level in the bowl and a space above it containing apparently nothing — a vacuum. As well as demonstrating the creation of a vacuum (even though, as we now know, with a very small amount of mercury vapor and perhaps the vapors of impurities, especially water), this also provided direct evidence of the pressure of the atmosphere and was a measure of it, in effect a barometer. The vertical height of the column remained unchanged if the tube was tilted or raised slightly to lengthen the section above the level of the mercury in the bowl. Tubes of greater length could be used, but the result was the same, and water introduced from below would fill the empty space. A vacuum created in this way is described as a Torricellian vacuum. The experiment was a profound stimulus to others and was repeated and developed by Pascal who famously organized an experiment in which such a device was carried up to the top of the Puy-de-Dome volcano in Auvergne, France, at an altitude of about 1 km, during the course of which a progressive shortening of the mercury column was observed, amounting at the top to about 75 mm. This was a critical experiment that confirmed the existence of a pressure due to the air and its origin as the weight of the air in the atmosphere above.

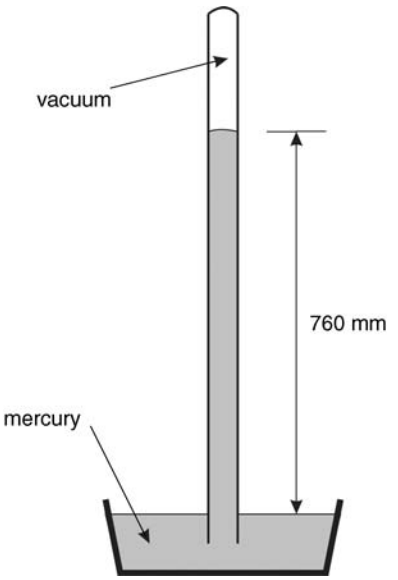

FIGURE 1.1
Torricelli's experiment.

The first successful air pumps, much later to be called vacuum pumps, were also made at about this time by von Guericke in Germany, who adapted the design of water pumps for the purpose. By improving the wet leather seal of the piston with a cylinder wall and incorporating valves appropriately, he was able to remove air from sealed vessels. The action of these pumps in moving air was similar to that of a bicycle pump in which the seal has been reversed so that work is done by pulling the piston out rather than pushing it in. In the course of developing his pumps and carrying out various investigations, Guericke devised some notable demonstrations, including that of the Magdeburg hemispheres which, when assembled as a sphere and evacuated, could only be separated by the opposite pull of two teams of horses, to great percussive effect. Larger hemispheres could not be separated until the vacuum was released.

The historical development of vacuum pumps up to the present day has been described by Hablanian in his book of 1997 and an article from 1984 which give many references that enable the subject to be pursued. The range of modern pumps may be broadly divided by their working principle into the following three types: those which involve displacing gas by repetitive mechanical actions, those which induce directed molecular motion in the gas by its interaction with high-speed fluid streams or solid surfaces, and those that capture molecules at surfaces by cryogenic condensation and other surface processes, sometimes with ionization prior to capture. Redhead (1984) reviewed the measurement of vacuum pressures from early days up to modern times and more recently (1999), the history of vacuum devices. In the 18th and 19th centuries, advances in vacuum science and technique enabled the investigation of many phenomena such as thermal conduction in gases, low-pressure electrical discharges (the study of whose optical spectra laid the foundations for the understanding of atomic structure), and the discovery in the late 19th century of x-rays and the electron. At the end of this century, the manufacture of electric light bulbs became the first large-scale industrial application of vacuum technique.

In the early 20th century the thermionic valve (vacuum tube) in various forms enabled currents to be manipulated and amplified in the developing subject of electronics to provide the basis of radio communication and many other devices. Alongside this were many applications of vacuum in scientific research, including the early charged-particle accelerators and electron microscopes. The Second World War stimulated many technologies, one of which was the development of the first electronic computers using vacuum tubes. In the years that followed and ever since, progress in semiconductor electronics and subsequently in microelectronics and optoelectronics has been possible only in association with developments in vacuum technology and ultraclean manufacturing methods. In this same period the scope of the technology widened with general technological progress so that it is now a necessary part of many industrial manufacturing processes; television tubes, high-purity alloys, pharmaceuticals, and optical coatings are just a few examples. In scientific research, progress in fields such as fundamental particle

physics, nuclear fusion, surface science, x-ray lasers, and nanotechnology all depend, in various ways, on the creation and control of vacuum conditions.

Evidently, a vacuum is necessary for things to happen in a range of diverse applications that nowadays embrace activities in research, development, and manufacturing. It is an enabling technology which has become part of the advance of technology as a whole in the infrastructure of work, communication, travel, daily living, and leisure. Computers, which permeate all these, do so with hardware that owes its existence to vacuum technology. As has been the case since the introduction of the electric light bulb, scientific research and development work in the industrial laboratories of manufacturers of special products and vacuum equipment have played an important and arguably dominant part in the progress of vacuum science and technology.

1.3 Units

As noted in Section 1.1, the SI unit of pressure is the Pascal (Pa). In this subject, as one might expect, there is much concern with gas flow, which may be specified by a volumetric flow rate, for which the appropriate SI unit is cubic meters per second, m^3s^{-1}. Other quantities naturally arise to describe, for example, how flow through pipes is determined by their dimensions and the pressure difference across them. All are, of course, expressible in SI units. However, the subject developed historically in a way that did not naturally lead to these units of pressure and volume. Thus, Torricelli's experiment, in which the pressure of the atmosphere supported a column of mercury approximately 760 mm high, led to the unit of pressure in a vacuum context being the pressure that would support a column of mercury 1 mm high, and in due course named the *torr* in his honor. A *standard atmosphere* is defined to be exactly 101,325.0 Pa and from this 1 torr is defined as 101325/760 = 133 Pa (to three significant figures). This is done because evaluating the torr as a pressure $h\rho g$, where $h = 10^{-3}$ mm, ρ the density of mercury, and g the acceleration due to gravity, leaves some small but significant arbitrariness in the choice of values for both ρ and g. The former varies slightly with temperature, and the latter from place to place, though taking the 0°C value of ρ and standard gravity g gives the pressure of 1 mm of mercury differing from the defined torr by only one part in a million. The torr, with origins in the history of the subject, is today still a widely used unit in vacuum practice, especially in North America. To recapitulate, its relationship to the SI unit is:

$$1 \text{ torr} = 133 \text{ Pa } (= 133 \text{ Nm}^{-2})$$

Alongside this,

$$760 \text{ torr} = 101{,}325 \text{ Pa} = \text{a standard atmosphere}$$

In characterizing the pressure of the atmosphere in meteorological practice, the unit used, familiar from weather maps and broadcasts, is the millibar. A bar is defined to be 100,000 Pa exactly. Thus, 1 mbar = 100 Pa, and the standard atmosphere is a pressure of 1013.25 mbar. The mbar is a widely used unit of pressure in vacuum practice in Europe and, to an accuracy of four significant figures, 1013 mbar = 760 torr, so that 1 mbar = (760/1013) = 0.75 torr. These widely used practical units, the mbar and the torr, are therefore of similar size, a fact that may have helped to perpetuate their use in preference to the Pa from which they differ by two orders of magnitude. It is therefore necessary to be aware of three units of pressure and the relationship between them:

$$1 \text{ mbar} = 100 \text{ Pa} = 0.75 \text{ torr}$$

Some latitude also exists in the use of units of volume, so that in addition to the m^3, the *liter* finds widespread use, especially for specifying the volumetric flow rates into high vacuum pumps, which are expressed almost universally in liters per second, ls^{-1}. Being the volume of a 10-cm cube (1000 cc), a liter was historically a more natural unit in which to express the volume of apparatus of typical laboratory size, and it remains an easily visualized quantity in this practical context. Since its use is widespread, it needs to be incorporated into our system of analysis, and it is appropriate to remember that 1 m^3 = 1000 l.

Vacuum practice, therefore, involves some well-established non-SI units. They may eventually be consigned to history with the universal adoption of the SI units but, in the meantime, practitioners need to be able to convert speedily between them when necessary. In this book, the mbar, liter, and secondary units derived therefrom will be used when strictly practical matters are being dealt with, as distinct from derivations and analyses having to do with basic physics. This is in keeping with the majority of other texts, information from suppliers describing equipment, and much of the published literature. Of course, in inserting practical values into equations derived from first principles, quantities have to be expressed in SI units. Many examples will be given in the text.

1.4 The Ranges of Vacuum

As mentioned in Section 1.1, the range of vacuum presently explored and utilized spans a range of about 15 orders of magnitude below atmospheric pressure. This corresponds therefore to pressures from 1000 mbar down to 10^{-12} mbar or less, or 10^5 to 10^{-10} Pa. This range is subdivided into smaller ranges as defined below. The boundaries are rather arbitrary, and there are

no sharp distinctions in crossing them, but the ranges they define do correspond roughly to regions in which different physical properties of the gas present are important. The divisions are:

Low (rough) vacuum	Atmospheric pressure to 1 mbar
Medium vacuum	1 to 10^{-3} mbar
High vacuum (HV)	10^{-3} to 10^{-8} mbar
Ultrahigh vacuum (UHV)	10^{-8} to 10^{-12} mbar
Extreme high vacuum (XHV)	Less than 10^{-12} mbar

Many applications in the low-vacuum region exploit the force created by the pressure difference with atmospheric pressure — for example, in vacuum molding and mechanical handling using suction pads. Also, oils may be degassed to rid them of dissolved air. Applications in the medium vacuum range include vacuum- and freeze-drying in the pharmaceutical industries in which vacuums have to be sufficiently low to permit relatively free evaporation from surfaces, and sputtering and chemical vapor deposition processes, which will be discussed later.

High vacuum has a very large number of applications that include the manufacture of television tubes; the deposition of thin film coatings by evaporation of materials from the bulk, as in lens blooming; and semiconductor technology in which there is an additional strict requirement for "clean" vacuum, free of critical impurities. Instruments such as electron microscopes and mass spectrometers for chemical and forensic analysis operate in this region. A significant property of the high vacuum is that it is sufficiently rarefied to allow free passage of particles — atoms, molecules, or electrons — from one place to another within it; therefore, mean free paths, to be discussed later, are of importance.

Ultrahigh vacuum is widely used in research applications where it is important that surfaces are subject to minimal contamination by capture of molecules of the residual gas. Surface science was one of the early driving forces that stimulated progress in this field. The *monolayer formation time*, to be discussed later, relates to this concern. Such vacuums are also a prerequisite in thermonuclear fusion experiments, prior to their back-filling with ultrapure gases for fusion studies, and vacua of 10^{-12} mbar are necessary in order that charged particle beams can be accelerated to very high energies and stored for reasonable times with acceptable rates of attenuation due to scattering, as in the large hadron collider, the LHC, at the CERN laboratories. As is pointed out by O'Hanlon (2003), manufacturing processes for some sophisticated semiconductor and optoelectronic devices now necessitate the ultraclean conditions of the UHV environment, a trend that will continue. An introduction to the techniques and problems in the XHV region below 10^{-12} mbar is given in an article by Redhead (1998).

1.5 Plan of This Book

Because our basic working commodity is gas, it is appropriate to begin with a study of the relevant properties of gases and also vapors. This is done in two chapters. Chapter 2 deals with the quantity of gas, moles, mixtures, and vapors from a macroscopic point of view. In Chapter 3, the molecular picture is presented, and kinetic theory exploited to get important results that relate pressure, impingement rate, and mean free path to the number density of molecules. These are of central importance to an understanding of how gases behave in vacuum systems. Condensation and evaporation, viscosity, thermal conduction, and thermal transpiration — phenomena that arise in various vacuum contexts — are also discussed. The Knudsen number, which distinguishes between the fluidic state of a gas and more rarefied conditions, is introduced. The distinctions are crucial and reflect regimes of quite different physical behavior.

Surfaces play a critical role in vacuum technology, and Chapter 4 is about relevant molecular processes at surfaces. These include the surface scattering of molecules and their capture by adsorption and release by desorption, which lead to the concepts of monolayer formation time and the mean time of adsorption. Adsorption and desorption are fundamental in the outgassing phenomenon. In many circumstances, it is outgassing that determines and limits the vacuum attainable. Also described are the interaction of molecules with high-speed surfaces — molecular drag — which is important in some pumps and measuring devices, and the sputtering of surface atoms by ion bombardment that finds application in ion pumps and, very extensively, in vacuum coating.

Equipped with this knowledge, one can proceed to consider gas flow and pumping, the subject of Chapter 5. The different regimes of flow that arise from a consideration of the Knudsen number are established: continuum flow at higher and molecular flow at lower pressures. Measures of flow are introduced, and continuum and molecular flows under various conditions analyzed. Some of the analysis for continuum flow is rather heavy going, and the reader for whom it is not a prime concern might simply note from Figure 5.7 the existence of choked flow and the useful related Equation 5.29. Some further aspects of molecular flow, including the concept of transmission probability and its application, are introduced. The pumping process and important matters such as pump-down time and ultimate pressure, together with a brief discussion of differential pumping, bring this chapter to a conclusion.

Chapter 6 and Chapter 7 are about the means for creating a vacuum and measuring it. They are both relatively lengthy, reflecting the variety of devices available given the large range of vacuum, which has already been noted. The natural division of pumps into three broad classes allows some rationalization of the subject. In a similar way, gauges to measure pressure

may do so directly or, equivalently, as discussed in Section 1.1, depend on the measurement of a physical property that is determined by the number density n of molecules. In describing all these devices, I have attempted to explain how they work in terms of basic physics, as well as dealing with their performance and limitations.

Illustrative examples and the analysis of representative laboratory systems are the subject of Chapter 8, which builds on what has gone before. The purpose of Chapter 9 is to describe the role of vacuum in a selection of industrial products and scientific instruments. These range from flat-panel displays in the domestic domain to the Large Hadron Collider as an example in the domain of "big science," as it has been called, and demonstrate the scope of vacuum technology. The final chapter deals with some strictly practical matters.

At the end of certain chapters, where appropriate, I have included some short exercises and problems, answers to which are supplied at the end of the book. The purpose of the exercises is to encourage the student to think quantitatively about the subject matter and gain a feel for orders of magnitude, while the problems are a little more challenging, requiring in some cases the different aspects of a situation to be recognized and analyzed, and a path formulated to a solution.

2

The Physics of Gases and Vapors

In this chapter, the properties of gases, vapors, and gas mixtures relevant to vacuum practice will be considered from a macroscopic point of view. In addition to setting up various quantitative measures, we shall be concerned with understanding the general behavior of gases and vapors, and the range of behavior encountered because of the variety of gases. The description of gases from a molecular standpoint is the subject of the next chapter.

2.1 Specifying the Quantity of a Gas, the Mole, Ideal, and Real Gases

Specifying the quantity of gas at a pressure p and temperature T in a volume V is straightforward because of the existence of the ideal gas equation:

$$pV = n_M R_0 T \tag{2.1}$$

which is quite accurately obeyed for most gases at pressures of order atmospheric. In this equation, n_M is the quantity of gas expressed in moles, to be discussed in the following text, R_0 is the universal gas constant = 8.314 JK^{-1} $mole^{-1}$, and T is the absolute temperature in degrees K = °C + 273.15, usually taken to be 273 with sufficient accuracy. In the SI system of units, p is measured in Pa (1 Pa = $1N/m^2$) and V in m^3.

The origins of the ideal gas equation lie in the work of Charles Boyle and Gay-Lussac. In 1662, Boyle demonstrated that at a fixed temperature the product of pressure and volume of a fixed mass of gas was constant. In his first experiments, using air, the pressure ranged from a low value of about 1 in. of mercury (30 mbar) up to a few atmospheres. Much later, in 1807, Gay-Lussac, extending the work of Boyle, established that at constant pressure the volume of a fixed mass of gas increases linearly with temperature, and in direct proportion to its temperature as measured on a scale that is now identified with the absolute scale referred to earlier. These were the experiments that suggested that there might be an absolute zero of temper-

ature — minus 273°C — at which the volume of a gas, had it not first condensed, would shrink to zero.

The mole as a concept originated in the work of the great chemists of the late l8th and early l9th centuries that led to an understanding in terms of an atomic theory of matter of the laws of chemical combination. In particular, Gay-Lussac's observations that in gaseous reactions the constituents combine in simple proportions by volume as well as by weight led Avogadro in 1811 to the hypothesis that *equal volumes of different gases under the same conditions of temperature and pressure contain the same number of molecules*. This proposal is now amply verified so that the statement is known as Avogadro's law. From the work of this time it emerged that a number of the common elemental gases such as hydrogen, oxygen, and nitrogen must exist in the form of diatomic molecules. For hydrogen, the lightest gas, it was determined that under STP conditions (1013 mbar, 273 K), a mass of 2 g occupied 22.4 l. The same volume of oxygen, for example, under these conditions, contains the same number of oxygen molecules but weighs 32 g. The weight of any gas that occupies 22.4 l under STP conditions became known as its *gram molecular weight*, but it is nowadays referred to as the *molar mass, M*, and is expressed in kg. The *mole* (mol) is a concept that embraces all substances. It is the amount of a substance that contains Avogadro's Number N_A of constituent particles, be they atoms, molecules, or whatever. One may speak of a mole of ions, electrons, fission fragments, etc. By definition, N_A is the number of ^{12}C atoms in 0.012 kg of the pure carbon isotope ^{12}C. It has the value 6.022×10^{23}, which has been determined by experiment in a number of ways (see, for example, Haken and Wolf [1996]). Associated with this scheme of measurement and deriving from it is the *atomic mass unit* (amu or u), defined as $1/12$ of the mass of the carbon ^{12}C isotope, and therefore equal to $1/12$ of $0.012/6.022 \times 10^{23} = 1.66 \times 10^{-27}$ kg. Nominally, the hydrogen atom has a mass of 1 amu.

The molar masses of some of the gases and other substances relevant to vacuum practice are shown in Table 2.1. For inert gases such as argon that are monatomic, molar mass also replaces the older term *gram atomic weight*. Also shown in the table are the corresponding *relative atomic* and *relative molecular masses*, symbol M_r. These express the mass m of an atom or molecule in amu. Strictly, the mass of the hydrogen atom on this scale (C = 12.000 by definition) is 1.008 and that of helium 4.002, but for many purposes the values stated are sufficiently accurate. Exact values can be found in Kaye and Laby (1996). The relationship between the molar mass M and the mass m of a molecule is simply $M = N_A \times m$. Air, being a mixture of nitrogen and oxygen in the approximate proportions 4:1, may be taken to have $M = 0.029$ kg and $M_r = 29$.

Returning to Equation 2.1, note that it also contains the familiar and useful statement that for a fixed mass of gas, pV/T is constant, so that $p_1V_1/T_1 = p_2V_2/T_2$. As an example, let us work out the mass of helium that would be needed to fill a container of volume 40 m^3 to a pressure of 1013 mbar at room temperature, taken as 22°C (295 K). From Equation 2.1

TABLE 2.1

Molar Masses of Various Substances

Substance	Formula	Molar Mass M kg	Relative Mass M_r
Atomic hydrogen	H	0.001	1
Hydrogen	H_2	0.002	2
Helium	He	0.004	4
Carbon	C	0.012 exactly	12.00
Methane	CH_4	0.014	14
Atomic oxygen	O	0.016	16
Water	H_2O	0.018	18
Neon	Ne	0.020	20
Nitrogen	N_2	0.028	28
Carbon monoxide	CO	0.028	28
Air	mixture	0.029	29
Oxygen	O_2	0.032	2
Argon	Ar	0.040	40
Carbon dioxide	CO_2	0.044	44

$$n_M = \frac{1.013 \times 10^5 \times 40}{8.314 \times 295} = 1652$$

and the mass of helium is therefore $W = n_M \times M = 1652 \times 0.004 = 6.61$ kg. Because 1 mol of any ideal gas occupies 22.4 l at STP, such computations can be approached in an alternative way. Thus, 1 mol occupies $22.4 \times 295/273 = 24.20$ l at 1013 mbar and 22°C, and therefore the number of moles of gas in a volume 40 m^3 is $40/0.0242 = 1652$, as before. In a small room of volume 40 m^3, the mass of air would be $1652 \times 0.029 = 48$ kg. In the form of a solid mass, this would be quite heavy to lift!

We deal in vacuum practice with gases at low pressure, and so it will be safe to assume that gas in such a state is accurately described by Equation 2.1. Moreover, since the gas flow in systems of interest normally takes place at constant temperature, usually room temperature, another very straightforward way of specifying quantity of gas emerges. At fixed T, the product $(p \times V)$ varies only with n_M and in proportion to it. Therefore, at fixed temperature, usually taken to be 295 K, the product pV serves as a measure of quantity, albeit in unconventional units of pressure × volume. In the SI system, the unit will be the Pascal m^3. A more practical unit, and that most frequently used, is the millibar liter, mbar l. In terms of this measure, the 1652 mol of helium (or air) referred to above would be $1.0135 \times 10^5 \times 40 = 4.05 \times 10^6$ Pa m^3 or $1013 \times 40 \times 10^3 = 4.05 \times 10^7$ mbar l. Clearly, subject to the temperature being constant and specified, the pV measure of amount is a quickly assembled and useful quantity.

The behavior of real gases is described accurately by Equation 2.1 only in the limit of low pressure. The data which most clearly demonstrate deviations from the ideal behavior are plots of the product pV for 1 mol at given

temperature vs. the pressure p, as shown, for example, in Zemansky and Dittman (1997). At pressures of a few tens of atmospheres and at room temperature for gases such as hydrogen and oxygen, which have boiling points below 100 K, pV differs from its low-pressure value R_0T by amounts typically ~1%. For gases such as carbon dioxide, the deviations are more marked. Data on numerous gases are extensively tabulated because of the industrial importance of characterizing gas at high pressures and over a large range of temperature. Their behavior is described comprehensively in a power series expansion called the *virial equation of state* that has the form, for 1 mol:

$$pV = R_0T(1 + Bp + Cp^2 + \ldots) \tag{2.2}$$

The virial coefficients B, C, etc. depend on the particular gas and the temperature. Zemansky and Dittman give data for nitrogen. In the accurate calibration of vacuum gauges using isothermal expansion to generate known low pressures (see Chapter 7), Equation 2.2 is used up to the second virial coefficient B in order to determine as accurately as possible the amount of gas prior to the expansion. When the virial equation is alternatively formulated as a power series in terms of volume rather than pressure, the second virial coefficient is important in relating the cohesive and compressive properties of gases to theoretical models (see, for example, Reif [1965] and Present [1958]).

The deviations from ideal behavior that are empirically described in the virial expansions are caused principally by two effects: the existence, at small separations, of weak forces of attraction between molecules and the fact that molecules, although very small, have a finite size. These effects become important in states of relatively high compression when molecules, in spite of their considerable kinetic energies, are forced into close proximity. Although virial expansions may accurately describe the behavior of real gases, they neither have a simple analytical form nor do they directly convey the underlying physics. It is valuable to have an equation with an analytical form that does have a physical basis. Such an equation is that of van der Waals (1846), which for n_M moles is:

$$\left(p + \frac{n_M^2 a}{V^2}\right)\left(V - n_M b\right) = n_M R_0 T \tag{2.3a}$$

and for 1 mol

$$\left(p + \frac{a}{V^2}\right)\left(V - b\right) = R_0 T \tag{2.3b}$$

As we shall discuss in more detail in the next chapter on the molecular description of gases, the origin of the pressure that a gas exerts is the impact and rebound of its molecules at the container wall. In the equation above, the measured pressure p is augmented by the a/V^2 term because the cohesive intermolecular forces between the gas molecules, by slightly slowing them down as they approach the container surface, reduce the pressure compared with gas in the interior. The effect depends on the square of the density of the gas (and hence the inverse square of V) because density determines the average separation and therefore magnitude of the intermolecular force as well as also the rate of impact. The value of a depends on the gas. The physical cause of these weak forces is the attraction between the transient electric dipole states induced in the molecules by random thermal fluctuations. These were first discussed by van der Waals and bear his name.

The quantity b in the volume term arises because, when the gas is highly compressed, the space in which its molecules are free to move is reduced significantly by the presence of the molecules themselves. The magnitude of b, called the covolume, is determined by collision processes as shown in Jeans (1982) and is four times the volume of the molecules themselves if they are regarded as hard spheres. At high pressure and small volume, these additional terms a and b are significant and cause differences from the ideal behavior. At low pressures the added terms are negligible and the equations reduce to $pV = n_M R_0 T$. Values of a and b deduced from experiment are shown in Table 2.2. Notice that a values range widely, being small for helium and large for water vapor, as one might anticipate. The values of b are roughly comparable, reflecting that the molecules of simple substances are all of a similar size. For nitrogen, b is 3.91×10^{-5} m^3 per mol, which leads to an estimate of effective diameter of a nitrogen molecule of the correct order.

TABLE 2.2

Van der Waals Constants

Substance	a/l^2 atm mol^{-2}	100b/lmol^{-1}
He	0.034	2.37
H_2	0.244	2.66
O_2	1.36	3.18
N_2	1.39	3.91
CH_4	2.25	4.28
CO_2	3.59	4.27
H_2O	5.46	3.05

2.2 Gas Mixtures and Partial Pressures

In most vacuum systems, the gases encountered are mixtures, and specifying their constituents — sometimes minority constituents which might have an adverse effect in a particular process if they exceed a specified small amount — is increasingly a matter of importance. Equally, in some technological processes, it is necessary to supply gaseous ingredients into vacuum vessels in known proportions. Therefore, a means of quantifying mixtures is necessary, and specifying the *partial pressures* fulfils this need.

Following Dalton (1814), a partial pressure is defined as the *pressure which a gas (or vapor) would exert if it occupied the volume of the mixture on its own*. It arises in the following way. Consider a mixture of gases A, B, and C in a volume V at temperature T. Suppose there are n_1 moles of A, n_2 of B, and n_3 of C, and that each obeys the ideal gas law. The total pressure will be proportional to the total number of moles of gas present, and therefore one can write

$$p = \frac{\left(n_1 + n_2 + n_3\right)R_0T}{V}$$

which can be rearranged as

$$p = \frac{n_1R_0T}{V} + \frac{n_2R_0T}{V} + \frac{n_3R_0T}{V}$$

so that if, following Dalton's definition, we define a partial pressure p_i by

$$p_i = \frac{n_iR_0T}{V} \tag{2.4}$$

which is pressure that would be exerted if gas i occupied the volume V on its own, then

$$p = p_1 + p_2 + p_3 \tag{2.5}$$

and the sum of these partial pressures is the total pressure exerted by the mixture, which is Dalton's Law. It has same validity as the ideal gas law on which it is based, and so for gases at normal pressures and temperatures and vapors that are reasonably dilute, it is quite accurate. If one defines a *mole fraction* x by the relationship

$$x_i = \frac{n_i}{\Sigma(n_i)} \tag{2.6}$$

then $p_i = x_i p$

To make a mixture consisting of n_1 moles of gas A and n_2 of gas B at pressure p in a volume V at a temperature T, one can imagine taking a container of volume V_A of gas A at pressure p and temperature T such that $V_A = n_1 R_0 T/p$ and, similarly, a container of volume $V_B = n_2 R_0 T/p$ of gas B, also at pressure p. Connecting the two volumes so that A and B mix would give, as required, a mixture of volume $V = (V_A + V_B)$ of total pressure p with partial pressures $p_1 = n_1 R_0 T/V$ and $p_2 = n_2 R_0 T/V$. It follows that

$$V_A = \frac{n_1 R_0 T}{p} = \frac{V p_1}{R_0 T} \times \frac{R_0 T}{p} = \frac{p_1}{p} V$$

Evidently, in ideal gas mixtures, the mole fractions are equal to the volume fractions. Dry air has the composition by volume: N_2 78%, O_2 21%, and Ar 1%. At a pressure of 1000 mbar, therefore, the partial pressures of N_2, O_2, and Ar will be 780, 210, and 10 mbar, respectively. Air normally contains water vapor as an important minority constituent. The saturated vapor pressure of water (see Section 2.3) at 22°C is 26 mbar, so that for air with 50% relative humidity under these conditions — again with a total pressure of 1000 mbar — the partial pressures would be 13 mbar of water vapor and 770, 207, and 10 mbar of N_2, O_2, and Ar, respectively. Note that the atmosphere typically contains amounts of about 1% of water vapor. At these dilutions, it will be a reasonable approximation at room temperature to regard it as an ideal gas.

2.3 Vapors and Vapor Pressures

Liquid ether poured into an open dish has two very obvious properties: It has a strong characteristic smell, and it quickly evaporates. Ether molecules readily leave the liquid and enter the surrounding space as vapor. The tendency to vaporize is a property of all materials, but we fix our attention on ether because the various aspects of vapor behavior are easily demonstrated.

If resources (and regulations) allow, it is instructive to set up the simple apparatus of Figure 2.1(a), though it is slightly hazardous, and is best considered as "a thought experiment." It consists of two 6-mm-bore glass tubes, L and R, about 1 m long, the lower ends of which are connected by a strong rubber hose, H, also about 1 m long. The upper end of tube L is open to the atmosphere whereas that of tube R contains a small tap T and above it a

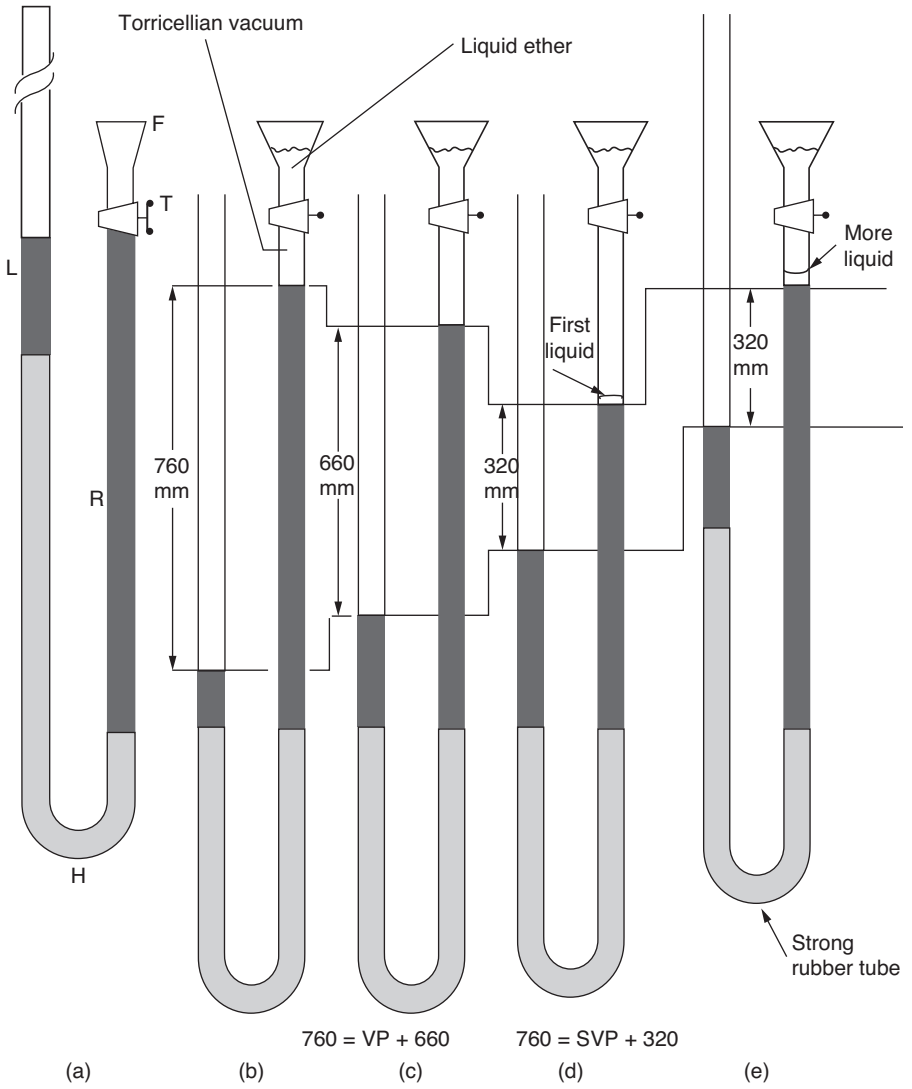

FIGURE 2.1
Experiment to demonstrate the vapor pressure of ether.

funnel F. With the top of L at about the same height as F, and with T open, it is possible to pour mercury through F until it forms a *U* column about 1.3 m in length, filling H and being visible at the same height in both L and R. A Torricellian vacuum, as described in Chapter 1, can then be created in R by raising L until the mercury just reaches the tap T, as in Figure 2.1(a), closing it, and then lowering L to achieve the condition shown in Figure 2.1(b). On a day when atmospheric pressure is 1013 mbar = 760 torr, it

supports a column of mercury 760 mm high with a Torricellian vacuum in the upper part of tube R below tap T.

Some liquid ether is now poured into F. If tap T is carefully opened so that just a few drops of ether enter the vacuum, two things happen. The drops disappear and the level of mercury in R falls. The ether is vaporizing and exerting a pressure. For example, the mercury level in R may drop (and that in L rise) so that the difference in mercury levels is reduced from 760 to 660 mm, as in Figure 2.1(c). Ether is thus exerting a pressure of 100 mm of mercury (100 torr). The addition of ether in such small amounts may be continued with its complete vaporization and further depression of the mercury until a point is reached, as in Figure 2.1(d), where not all the added liquid vaporizes, and a small pool of it collects on top of the mercury meniscus. The ether vapor is said to have become saturated and is exerting its *saturated vapor pressure*, denoted SVP. Further addition of liquid would simply add to the liquid pool and not to the vapor. At the first appearance of a tiny amount of liquid, the difference in mercury levels at 22°C is 320 mm, so that the pressure of saturated vapor is 440 mm. Thus, the existence of vapor pressure, and its maximum saturated value at a given temperature, are demonstrated. If, as in Figure 2.1(e), limb L is raised to reduce the volume of the vapor in R, some of it condenses and the vapor pressure exerted is unaltered.

Thinking now of a fixed mass of any vapor contained in a cylinder closed by a light, frictionless, nonleaking piston of area A, as shown in Figure 2.2(a), we can imagine the state of the cylinder contents at various degrees of compression. Initially, the volume is large, the vapor is unsaturated, $p < p_e$ the SVP, and a force $F = pA$ would be necessary to contain the vapor at this volume and temperature. By slowly increasing the force to compress the vapor and maintaining the temperature constant by removing the heat produced due to the work of compression, a state as in Figure 2.2(b) is eventually reached where the pressure has increased to its saturation value p_e at this temperature, with application of a force $F_0 = p_e A$. Further reduction of the volume occurs at constant pressure p_e with condensation of saturated vapor to liquid, as in Figure 2.2(c). Eventually, with further compression, the contents of the cylinder would be entirely liquid, with the force still F_0. Further reduction of the volume would require considerably increased force because liquids are almost incompressible. It is useful to plot such a sequence of actions on a p-V diagram, as shown in Figure 2.3(a). The line ABCD is a typical isotherm at temperature T and section AB represents compression from an unsaturated state at A to a saturated state at B. At this point, condensation starts and volume is reduced at constant pressure until point C is reached. The isotherm then becomes almost vertical, reflecting the incompressibility of the liquid.

All vapors of pure substances demonstrate this sort of behavior, but it depends in detail on the temperature and the family of isotherms of a typical substance as shown in Figure 2.3(b). Only at temperatures below a critical temperature T_c is there the two-phase behavior just described, in which liquid

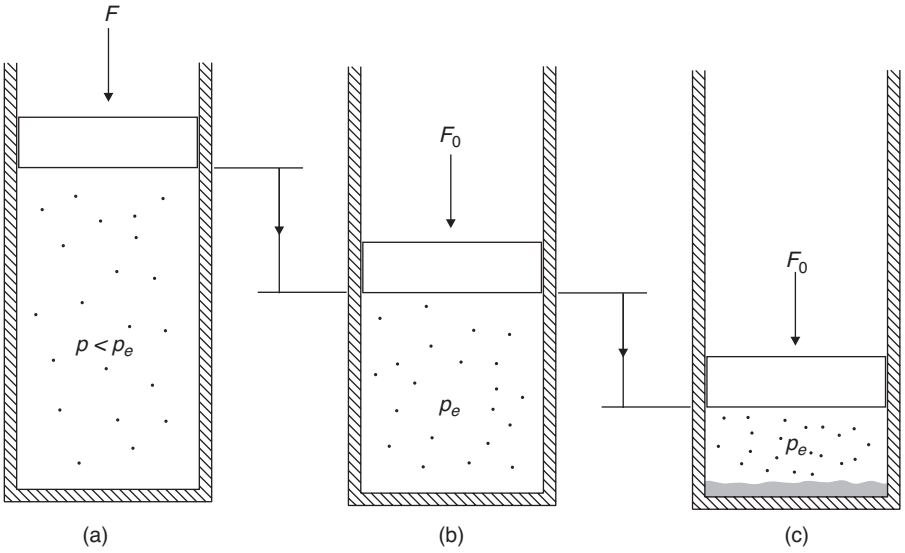

FIGURE 2.2
Equilibrium states of a vapor. (a) Unsaturated vapor, (b) saturated vapor, (c) saturated vapor + liquid.

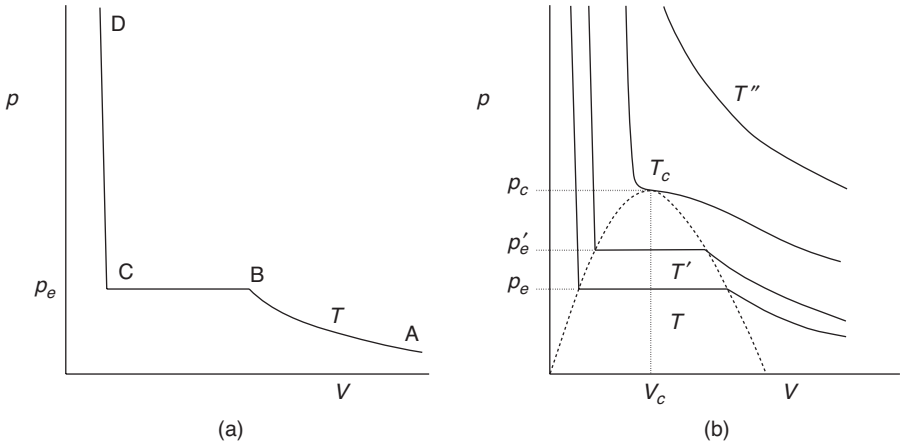

FIGURE 2.3
(a) Isotherm at temperature T; (b) family of isotherms.

and vapor are in equilibrium with each other and separated by a meniscus. With increase in temperature from T to T', a higher pressure p_e' is needed for the onset of condensation, and the two-phase region correspondingly occurs over a smaller range of volume. At the critical temperature T_c, the critical isotherm has a point of inflection at specific values p_c and V_c that specify the critical point. The vapor and liquid have the same density and

TABLE 2.3

Critical Point Data

Substance	Temperature K	Pressure/MPa	Density/kg m^{-3}
Hydrogen	33	1.29	31.1
Nitrogen	126	3.39	313
Oxygen	155	5.04	436
Carbon dioxide	304	7.37	467
Water	647	22.0	322

are indistinguishable; there is no two-phase region and no separating meniscus. At temperatures greater than T_c, it is not possible to produce two coexisting phases, and the state of the substance is referred to as a gas. When T is sufficiently greater than T_c, as at T″, Boyle's law is obeyed, and the p-V plot becomes a rectangular hyperbola. Below T_c it is a vapor that can be liquefied by compression. The critical point data of some substances are shown in Table 2.3. It is a unique condition of particular importance in the theory of phase equilibrium and allows to some extent the normalization of the behavior of substances to a universal model. Walton (1983) may be consulted to pursue this further.

Evidently room temperature is considerably greater than the critical temperatures of nitrogen and oxygen but much less than that of water, as we might expect on the basis of the foregoing.

The diagram of Figure 2.3(b) is a projection from the representation of the state of a substance on a three-dimensional p-V-T surface. It is instructive to study these representations. Roth (1990) gives a good example and they are extensively described in Sears (1965) and Sears and Salinger (1986). Some relevant p vs. T data are shown in Figure 2.4. Extensive vapor pressure data compilations are to be found in many texts, notably O'Hanlon (2003) and Wutz, Adam, and Walcher (1989).

Exercises

2.1 Convert a pressure of 400 mbar to (a) torr, (b) Pa.

2.2 A high-pressure gas storage cylinder contains 8 g of helium in a volume of 2 l at a temperature 22°C. What pressure (in Pa) will the gas exert? Also give the answer in bar.

2.3 If the cylinder in the above question had contained 8 g of argon, what would the pressure have been?

2.4 How many mbar liters of air occupy a volume of 1 m^3 at a pressure of 10^4 Pa?

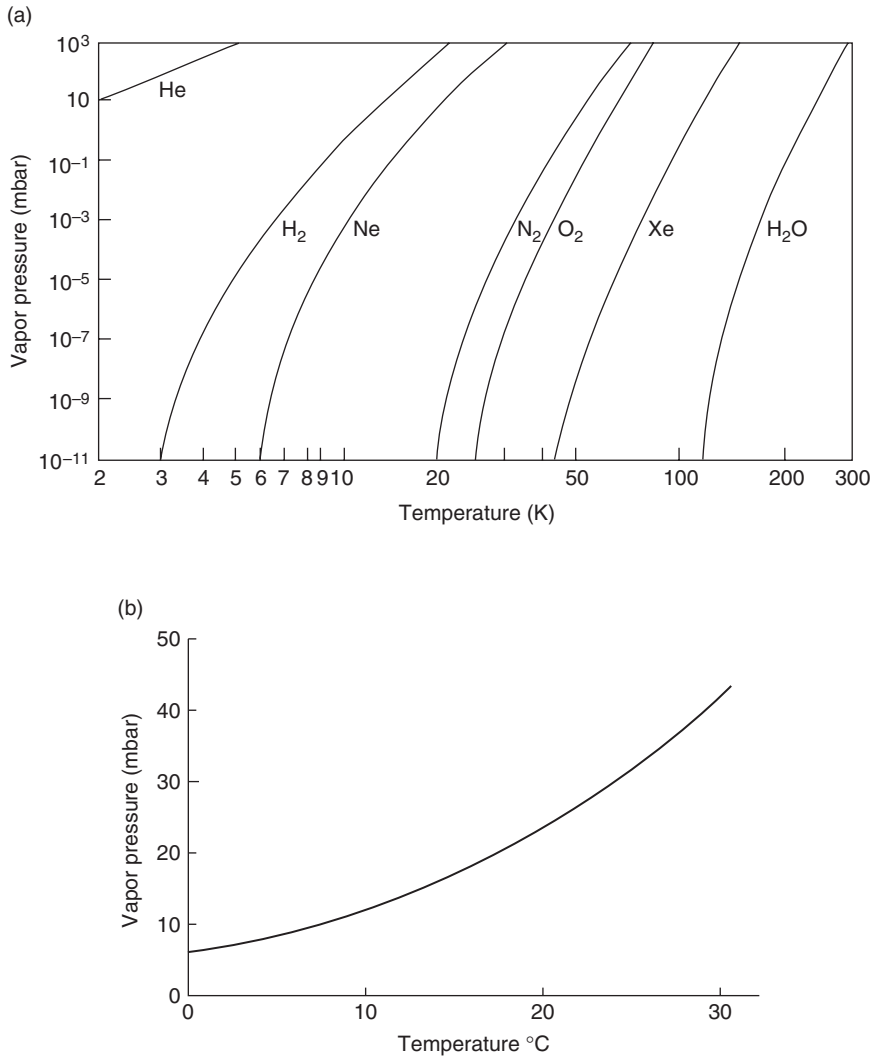

FIGURE 2.4
Vapor pressure data as a function of temperature for various gases and for water vapor.

2.5 A mixture consisting of 40 g of argon and 48 g of methane exerts a pressure of 40 mbar. What are the partial pressures?

2.6 Calculate the partial pressures and hence the total pressure in a gaseous mixture consisting of 32 g oxygen, 14 g nitrogen, 14 g carbon monoxide, and 40 g argon in a volume 2.23 m^3 at a temperature 295 K.

2.7 A volume V_0 of atmospheric air is isolated in a cylinder and piston at a temperature 20°C and a total pressure of 1000 mbar. The partial

pressure of water vapor is 10 mbar. Determine the pressures of water vapor and the total pressure when the volume is decreased to (a) $V_0/2$, (b) $V_0/4$, (c) $V_0/8$, with the heat of compression being removed so that the temperature remains constant. The vapor pressure of water at 20°C is 23.4 mbar.

Problems

2.8 Two vessels, one of volume 3 l containing helium at a pressure 1200 mbar, the other of volume 10 l containing oxygen at a pressure 800 mbar, are connected by a valve, initially closed, and are at room temperature 295 K. Calculate the number of moles of each and the total and partial pressures after the valve has been opened and equilibrium restored. Give pressures in mbar.

2.9 Liquid ether at room temperature is slowly introduced through a fine-leak valve into a glass container of volume 1 l, initially evacuated to high vacuum, where it vaporizes. The container is maintained at a temperature 295 K. How many cm^3 of liquid will be introduced before the first liquid droplets are detectable? What will be the pressure when (a) 1 cm^3, (b) 5 cm^3 have been introduced? Data for ether: SVP at 295 K, 560 mbar (5.6×10^4 Pa); density of liquid ether, 714 kg m^{-3}; molar mass 0.074 kg.

2.10 Determine the volume occupied by 1 g of saturated water vapor at 20°C, assuming that it behaves as an ideal gas. The vapor pressure of water at this temperature is 23.4 mbar. In an otherwise empty container of volume 1 m^3, what pressure will 1 g of water exert at this temperature? Will the volume occupied by 1 g of saturated water vapor at 10°C be greater or less than its volume at 20°C?

3

The Molecular Description of Gases

In this chapter the molecular description of gases will be considered in some detail, not only because it underlies the macroscopic properties described in the previous chapter but also because of its importance in understanding, for example, the flow of gas at low pressures and the details of molecule–surface interactions.

3.1 Introduction

The molecular-kinetic description of matter recognizes that the constituent atoms or molecules of a substance are in ceaseless motion, with a kinetic energy that increases with increase of temperature. The cohesive intermolecular forces that hold molecules close together in the condensed state are opposed by this thermal agitation. In the solid state of a substance, its molecules vibrate about average positions of a fixed array, which is ordered if the material is crystalline, disordered if it is not. At higher temperatures, in the liquid state rigidity is lost and molecules move more freely but are still held together in close proximity by cohesive forces. Thus, volume changes when a solid melts are very small. However, at the boiling temperature of a liquid, the intermolecular bonds are broken and the energy of the gas becomes kinetic energy of free translation of its molecules, now occupying a much greater space than they did as a liquid. For example, liquid nitrogen at 77 K has a density of 81 kgm^{-3} or 0.81 gcm^{-3}, so that a mole has a volume of $28/0.81 = 34.6\ cm^3$. If we divide this volume into N_A equal cubes, and imagine nitrogen molecules to be spheres located inside them, their diameter would be the length of the cube edge, i.e., $(34.6 \times 10^{21}/6 \times 10^{23})^{1/3} = (0.0576)^{1/3} = 0.38$ nm. As a gas at room temperature the same number of molecules occupy 24.2 l, as noted earlier, and there is, therefore, a several 100-fold increase of volume between liquid and gaseous states. By similarly assigning a cubical volume to each molecule for this gaseous state, the cube edge has length $(24.2 \times 10^{24}/6 \times 10^{23})^{1/3} = (40.2)^{1/3} = 3.4$ nm. Thus the average

distance of a molecule from its neighbors in the gaseous state at atmospheric pressure is about 10 times the molecular diameter.

Freed from the bonding of the condensed state, the energy of a simple monatomic gas is purely kinetic energy of translation, and its molecules travel with speeds of typically several hundreds of meters per second. Diatomic molecules have similar speeds but rotational energies in addition. The density of molecules is such that frequent collisions are inevitable, and so they travel in straight "free paths" between the successive collisions that cause their directions and speeds to be changed. Because of the enormous number of molecules involved, we must expect that there will be a range of molecular speeds and of the free paths traveled between collisions.

3.2 Kinetic Theory of Gases

The kinetic theory of an ideal gas, which is the source of the general description given above, is presented in the following text. It was developed primarily by Clausius, Maxwell, and Boltzmann, and good descriptions may be found, for example, in Zemansky and Dittman (1997). A number of its results are of central importance in vacuum physics. The distribution of molecular speeds will be presented without derivation, but others will be derived because they afford important insights into molecular processes. In addition to the form of the distribution of molecular speeds, the results of prime importance are:

$$J = \frac{n\bar{v}}{4} \tag{3.1}$$

which relates the molecular impact rate per unit area per second, J, to the number of molecules per unit volume, n (usually called the *number density*), and the average or mean molecular velocity $\bar{v}$. Secondly,

$$p = nkT \tag{3.2}$$

which relates pressure p to n, T, and Boltzmann's constant k ($= R_0/N_A$). Third,

$$\lambda = \frac{1}{\sqrt{2}\pi d^2 n} \tag{3.3}$$

which relates mean free path λ to n and the molecular diameter d. Note that the molecular number density n appears in all three formulae. At room temperature and atmospheric pressure, $n = p/kT = 1.013 \times 10^5 / 1.38 \times 10^{-23} \times 295 = 2.5 \times 10^{25}$ molecules per cubic meter. Equation (3.2) may be proved

most directly by working out the magnitude and rate of molecular impacts on a boundary surface, considering averages, which leads to the result that $p = mn\overline{v^2}/3$, where m is the mass of a molecule and $\overline{v^2}$ is the mean square velocity, and then making comparison with the ideal gas Equation (2.1). However, for the reasons explained, it will be derived, together with the result $J = n\bar{v}/4$, by considering molecular trajectories within a gas and then incorporating the results of the Maxwell–Boltzmann distribution of molecular speeds.

3.3 The Maxwell–Boltzmann Distribution of Molecular Speeds

Consider an ideal gas in which N molecules are contained in a volume V. The number density n is N/V. The assumptions of the kinetic theory are:

1. That the number of molecules is sufficiently large that in small volume dV of the gas the number $dN = ndV$ of molecules is large enough to be typical of the gas as a whole so that within it all directions of motion are equally probable, and the whole range of velocities encountered in the gas as a whole is represented
2. That molecules behave as smooth, hard elastic spheres and that when they collide the directions and the velocities of the colliding molecules change in accordance with the laws of classical mechanics as applied to elastic collisions
3. That, other than the forces involved in the hard-sphere impacts, there are no other intermolecular or external forces, so that after collision the molecules travel in straight lines at the constant velocities that are the outcome of the collision, until the next collisions
4. That collisions with boundaries are also elastic, and boundaries are considered to be smooth so that the rebound is specular

The analysis of this model yields the Maxwell–Boltzmann distribution of molecular speeds that is expressed as follows: A quantity $f(v)dv$ is defined as the probability that a molecule will have a velocity between v and $v + dv$. Of the total of N molecules, therefore, and counting all directions, a fraction dN/N will have velocities between v and $v + dv$ so that $dN = N\, f(v)dv$. We divide N and dN by V in order to have numbers per unit volume, giving $N/V = n$ and $dN/V = dn_v$, say, as the number of molecules per unit volume with velocities between v and $v + dv$. Thus, $dn_v = n \times f(v)dv$, and writing $n \times f(v) = n(v)$, we have $dn_v = n(v)dv$, so that $n(v) = dn_v/dv$ is the number of molecules per unit volume per unit velocity range.

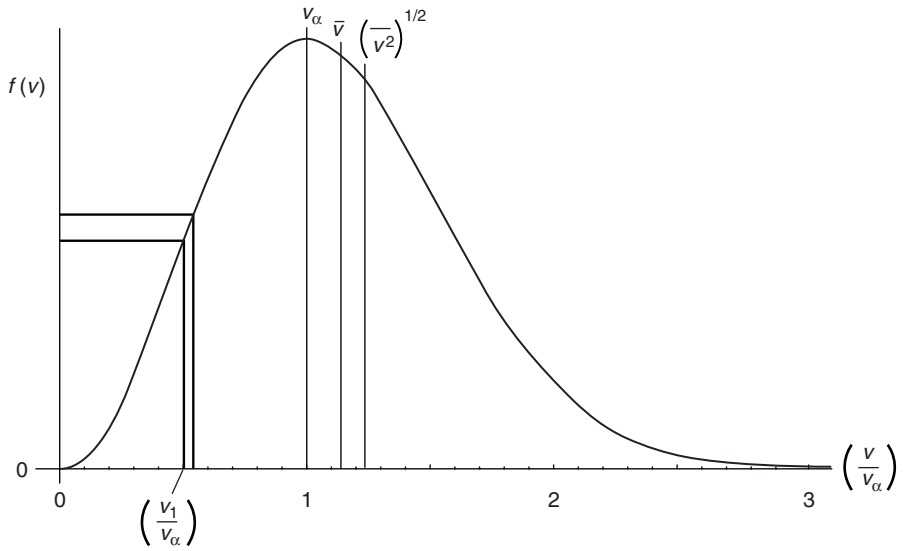

FIGURE 3.1
The Maxwell–Boltzmann distribution of molecular speeds.

As derived by Maxwell and Boltzmann, the probability distribution function of molecular speeds is

$$f(v) = \left(\frac{4}{\sqrt{\pi}}\right)\left(\frac{m}{2kT}\right)^{\frac{3}{2}} v^2 \exp\left(-mv^2/2kT\right) \tag{3.4}$$

It is plotted in Figure 3.1. The probability curve rises from zero to a maximum at the most probable speed, v_α, say, and then falls away as the exponential term in v^2 becomes dominant. Molecular speeds on the horizontal axis are normalized to v_α so that the maximum occurs at $(v/v_\alpha) = 1$. The curve is used as follows. Suppose one needs to know how many molecules per unit volume have speeds in the small interval between values, say, v_1, and a slightly larger value $v_1 + dv$. These would correspond to the two values (v_1/v_α) and $(v_1 + dv)/v_\alpha$ drawn as the two vertical lines in the figure. The value of f_1 at v_1 multiplied by the quantity dv/v_α is the probability of a molecule having a speed in this small range. Multiplying this by n then gives dn_v, the number of molecules per unit volume with speeds in this range.

By straightforward application of calculus it may be shown that the most probable speed, v_α is

$$v_\alpha = \sqrt{\frac{2kT}{m}} \tag{3.5}$$

and defining the mean speed $\bar{v}$ by

$$\bar{v} = \frac{\int vn(v)dv}{n} \tag{3.6}$$

yields

$$\bar{v} = \sqrt{\frac{8kT}{\pi m}} \tag{3.7}$$

Because $R_0 = N_A \times k$ and $M = N_A \times m$, multiplying inside the square root by N_A in both numerator and denominator gives an alternative form

$$\bar{v} = \sqrt{\frac{8R_0T}{\pi M}} \tag{3.8}$$

A quantity we shall need in due course, the mean square velocity $\overline{v^2}$, is

$$\overline{v^2} = \frac{\int v^2 n(v)dv}{n} = \frac{3kT}{m} \tag{3.9}$$

It is also readily shown that the mean speed $\bar{v} = 1.128\, v_\alpha$ and that the root mean square speed is $(\overline{v^2})^{1/2} = 1.225 v_\alpha$. They are indicated in Figure 3.1.

Consider now the ordinate of the curve of Figure 3.1 to be multiplied by n so that it becomes $n \times f(v) = n(v)$. It now represents the whole distribution, and the number of molecules per unit volume with velocities between v and $v + dv$ is $n(v)dv$. It is clear that there is a broad range of speeds centered on the most probable value. A substantial majority of the molecules have speeds within about ±50% of the most probable speed. One can understand this in a general way; it is quite difficult to move either very slowly or very fast in a densely packed, slow-moving crowd without being jostled towards the mean speed. Analysis shows (Dayton, 1998) that the number of molecules having speeds greater than twice the most probable value is less than 5% of the total, and with speeds greater than $4v_\alpha$, about 2 in ten million. It is important to note that although the number of molecules in any velocity range stays constant, their identity is continually changing because of collisions; the rates at which molecules are scattered into and out of a particular range by collisions are equal.

For nitrogen at room temperature 295 K, inserting $k = 1.38 \times 10^{-23}$ JK^{-1}, and $m = 28 \times 1.66 \times 10^{-27}$ kg from Table 2.1 in Equation (3.5) and Equation (3.7) gives most probable and mean speeds of 418 and 472 ms^{-1}, respectively. The correctness of the distribution and its predictions has been amply verified

by very elegant and direct experiments that are described, for example, in Pendlebury (1985).

3.4 Molecular Impingement Rate, *J*

To proceed with the discussion of molecular trajectories, we need to introduce the concept of solid angle. Imagine a sphere of radius r and an area A_r, of any shape, inscribed on its surface. The solid angle Ω subtended at its center O by the area A_r is defined as the dimensionless ratio $\Omega = A_r/r^2$ and has the unit steradian. As an example, one can imagine on a globe of the world the solid angle subtended at its center by, say, the African continent. Because the surface area of a sphere is $4\pi r^2$, the solid angle surrounding a point is 4π, whereas that surrounding a point on a plane is 2π. In evaluating *small* solid angles, such as that subtended at the eye by a coin of area dA = 1 cm^2 at a distance of 1 m, the area dA_r, which is strictly that of a spherical cap defined by the rim of the coin on a sphere of radius 1 m, is negligibly different from the area of the coin itself, and therefore one can write $d\Omega = dA/r^2$ [$=1/(10^2)^2 = 10^{-4}$ steradian]. If, as in some applications, the small area dA is such that the normal drawn at its center makes an angle θ with the line of reference (e.g., if the coin were tilted with respect to the line of sight), then $d\Omega = dA\,cos\theta/r^2$.

Now consider the collection of velocity vectors of varying magnitude and direction that represent the velocities of the n molecules per unit volume at an instant, and imagine them translated to have a common origin in order to get a measure of the way they are distributed in solid angle. All directions are equally probable, and therefore, in a small solid angle $d\Omega$, thought of as a slender cone centered on *any* chosen direction, the number of molecules per unit volume whose velocity directions are within it will be $n(d\Omega/4\pi)$. The smaller number per unit volume with speeds between v and $v + dv$ will be $dn_v(d\Omega/4\pi)$ where $dn_v = n(v)dv$ as previously defined.

Now let us imagine the collisions that occur with an element of plane surface of area dS as indicated in Figure 3.2(a). This might be part of a real surface bounding the gas *or an imaginary small area within it through which molecules pass*. It is bombarded continuously by molecules that come from all directions in the solid angle 2π above it, and in each direction there is a range of speeds given by the Maxwell–Boltzmann distribution.

Construct a cylinder based on area dS of length vdt, in a direction making an angle θ with the surface normal. The volume of this cylinder is $dS \times vdt\,cos\theta$. It contains many molecules but only those traveling within it in a direction θ with velocities between v and $v + dv$ will strike dS in time dt. This number is

$$(dS \times vdt\,cos\theta) \times dn_v(d\Omega/4\pi) \tag{3.10}$$

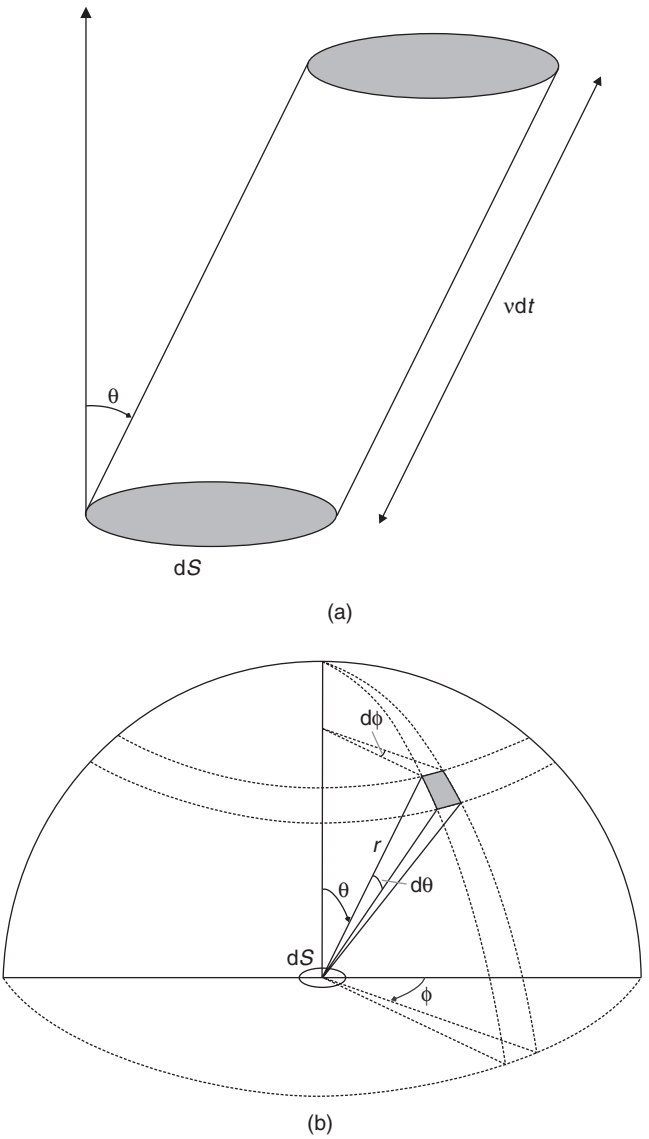

FIGURE 3.2
The geometry of molecular impingement.

But there are many directions of approach to dS characterized by angle θ, as Figure 3.2(b) shows. The element $d\Omega$ of solid angle in that diagram, our slender cone, becomes that subtended at the center of dS by the shaded area $r\,d\theta \times r\,sin\theta\,d\phi$ and is therefore $sin\theta\,d\theta\,d\phi$. The *whole* solid angle between θ and $\theta + d\theta$, which resembles a funnel shape bounded by two cones with apex at the center of dS, involves integration at constant θ of $sin\theta\,d\theta\,d\phi$ around the azimuthal angle ϕ, and is therefore just $2\pi\,sin\theta\,d\theta$. Thus, the total

number of molecules with velocities between v and $v + dv$ striking dS in time dt at an angle θ may be obtained by replacing the element $d\Omega$ in expression (3.10) by the larger element $2\pi \sin\theta\, d\theta$ to obtain, after some rearrangement,

$$(1/2)\, v\, dn_v \sin\theta \cos\theta\, d\theta\, dS\, dt$$

Defining $J_{v,\theta}$ as the *rate* of impingement *per unit area* of molecules with velocity v from angle θ and so dividing by dt and dS we get

$$J_{v,\theta} = (1/2)\, v\, dn_v \sin\theta \cos\theta\, d\theta \tag{3.11}$$

This is straightforwardly integrated with respect to θ between the limits 0 and $\pi/2$ to cover the whole 2π solid angle above the surface, and therefore to get J_v, the total number of molecules with velocities between v and $v + dv$ striking unit area per second, as

$$J_v = (1/4) v\, dn_v \tag{3.12}$$

Substituting $dn_v = n(v)dv$ and integrating now over the whole range of velocities gives J, the total number of molecules striking the surface from all directions and with all speeds, as

$$J = (1/4) \int v\, n(v) dv$$

But from Equation (3.6) the integral is equal to $n\bar{v}$. Finally, therefore,

$$J = \frac{n\bar{v}}{4} \tag{3.13}$$

Thus, the result quoted earlier as (3.1) is proved. It is extremely important and can be used to calculate numerous other quantities such as condensation and evaporation rates, monolayer adsorption times, and flow properties of gases, as we shall see later.

3.5 Pressure and Molecular Number Density, *n*

Let us now calculate the pressure exerted on the surface by the impact of molecules. A molecule of mass m traveling with velocity v at angle θ to the surface normal has components of momentum $mv \cos\theta$ perpendicular to the surface and $mv \sin\theta$ parallel to it. In the assumed elastic collision between the hard-sphere molecule and the smooth surface, the parallel component is unchanged but the perpendicular component is reversed as the molecule

bounces back off the surface into the gas. The associated change of momentum is therefore $2mv\ cos\theta$ perpendicular to the surface. In reversing the momentum of the molecule, the surface, by Newton's third law, experiences an equal and opposite impulse. The arrival rate of molecules with velocities between v and $v + dv$ at unit area of the surface from all angles θ in the solid angle $2\pi\ sin\theta\ d\theta$ is $J_{v\theta}$ given by Equation (3.11). By Newton's second law, force is equal to the rate of change of momentum, and so the force on unit area due to all the molecules with velocity v that strike the surface at angle θ is just $2mv\ cos\theta$ multiplied by their arrival rate $J_{v,\theta}$ given by Equation (3.11); that is

$$2mv\ cos\theta \times (1/2)\ v\ dn_v\ sin\theta\ cos\theta\ d\theta = mv^2\ dn_v\ sin\theta\ cos^2\theta\ d\theta$$

This is the force per unit area and therefore the pressure exerted by molecules with velocities between v and $v + dv$ at angles between θ and $\theta + d\theta$. Hence, by straightforward integration, the pressure p_v exerted by molecules of velocity v to $v + dv$ coming from all angles θ between 0 and $\pi/2$ is

$$p_v = mv^2 dn_v \int \sin\theta \cos^2\theta\ d\theta = \frac{mv^2 dn_v}{3}$$

The total pressure p is obtained by substituting $dn_v = n(v)dv$ and integrating the expression for p_v over the whole range of possible velocities. Thus

$$p = \frac{m}{3}\int v^2 n(v) dv$$

But the mean square velocity is defined by

$$\overline{v^2} = \frac{\int v^2 n(v) dv}{n}$$

and therefore

$$p = \frac{1}{3} mn\overline{v^2} \tag{3.14}$$

The quantity mn is just the gas density ρ, and $\overline{v^2}$ relates to the average kinetic energy per molecule. From Equation (3.9) we have an expression for $\overline{v^2}$ in terms of T and Boltzmann's constant k. Substituting leads to

$$p = nkT \tag{3.15}$$

Evidently, from Equation (3.14) and Equation (3.15)

$$\frac{1}{3}mn\overline{v^2} = nkT$$

so that, canceling n's and rearranging

$$\frac{1}{2}m\overline{v^2} = \frac{3}{2}kT \tag{3.16}$$

Thus the average kinetic energy per molecule is (3/2)kT and energy (1/2)kT may be associated with each of the three translational degrees of freedom of the motion of a molecule. The energy of motion of molecules is proportional to the absolute temperature. The principal result of this analysis is Equation (3.15). One might think of it as Boyle's Law expressed in molecular rather than macroscopic terms. Indeed, it can be deduced algebraically from Equation 2.1 by manipulation. It shows that the pressure exerted by an ideal gas is directly proportional to its absolute temperature and the number density of molecules. This is very reasonable because we have seen that n directly determines the rate of molecular impacts, and that T determines the energy.

3.6 Molecular Collisions and Mean Free Path, λ

The final result to be derived is Equation (3.3) for the mean free path λ the average distance traveled by a molecule between successive collisions. Let us focus attention on just one of our hard-sphere molecules of diameter d. Construct an imaginary disc of *radius d* and therefore area $\sigma = \pi d^2$ with the molecule at its center, and suppose that its plane is perpendicular to the molecule's direction of travel. In moving a distance l, the molecule will sweep out a cylinder of volume σl and the number of other molecules whose centers lie within it will be the number of collisions sustained by the moving molecule. Each collision will, of course, change the direction of motion and the velocity, and so the trajectory will be a crooked one made up of linear sections. In one second the total length of the cylinder will be $\bar{v}$, and so the number of collisions will be $n\sigma\bar{v}$. The mean free path is thus the distance traveled in one second divided by the number of collisions. Thus $\lambda = 1/n\sigma$. This argument ignores the motion of the other molecules. When their Maxwellian distribution is allowed for a factor $\sqrt{2}$ appears in the denominator so that

$$\lambda = \frac{1}{\sqrt{2}n\sigma} = \frac{1}{\sqrt{2}n\pi d^2} \tag{3.17}$$

Notice that λ depends on n and σ, the collisional cross section, but not on the temperature, and not, therefore, on the velocity of molecules. It is determined just by the spatial distribution of targets and their size. Because $n = p/kT$, it is evident that λ is inversely proportional to pressure, $\lambda \propto 1/p$, and that the product λp is constant for a given gas. We can estimate the value of λ for nitrogen at room temperature and pressure, taking n as previously worked out, and the diameter of a nitrogen molecule to be 0.376 nm. The result is $\lambda = 64$ nm, which is about 200 molecular diameters and significantly greater than the average molecular separation under these conditions (which we worked out to be 3.4 nm). As earlier noted, the molecules of the more common gases are roughly comparable in size — the effective diameters of helium, nitrogen, and methane molecules, for example, are 0.218, 0.376, and 0.414 nm, respectively. Thus, from Equation (3.17), the mean free path in helium will be about three times larger than that for nitrogen at any pressure. Nevertheless, for common gases at a given pressure, the mean free paths will be of similar magnitude. The results of this simple model are usually adequate in vacuum practice.

It is worthwhile to explore further the relationship of mean free path λ to the probability of a molecule having a collision. Consider what happens to a molecule that, having just had a collision, sets off along its next free path in some new direction, with coordinate x. Let $s(x)$ be the probability that the molecule reaches x without collision, and (αdx) the probability that it will have a collision in the next dx of its path, where α is some constant to be determined. The probability that it reaches $x + dx$ without collision may be expressed in two ways: as $s(x + dx)$ or as $s(x)[1 - \alpha dx]$ where the latter is the product of the probabilities that it reached x without collision and that it does *not* have a collision in the next dx. Thus

$$s(x + dx) = s(x)[1 - \alpha dx]$$

But the left-hand side may also be formally expressed as

$$s(x) + \left(\frac{ds}{dx}\right)dx$$

from which it follows that

$$\frac{ds}{s} = -\alpha\, dx$$

Integrating this from 0 to x and remembering that $s(0) = 1$ leads to

$$s(x) = \exp(-\alpha x) \tag{3.18}$$

The constant α remains to be determined. The difference ds between the probabilities $s(x)$ and $s(x + dx)$ is the probability of a collision between x and $x + dx$ and therefore of a free path of length x. Its *magnitude* is

$$ds = \alpha \exp(-\alpha x)dx$$

Therefore, the average path length, the mean free path λ, will be

$$\lambda = \int_0^\infty x\,\alpha \exp(-\alpha x) \Big/ \int_0^\infty \alpha \exp(-\alpha x)dx$$

The integral in the denominator is unity and that in the numerator $1/\alpha$ so $\lambda = 1/\alpha$. Thus, Equation (3.18) becomes

$$s(x) = \exp(-x/\lambda) \tag{3.19}$$

and the probability of a collision in a length dx of path is $\alpha dx = (dx/\lambda)$. This is in accord with intuition — on average the probability of a collision after traveling a distance λ must by definition be unity.

Equation (3.19) is called the "survival equation." It gives the probability that a molecule will have a free path equal to or greater than x. When applied to a group N_0 of molecules, it gives the fraction which will have free paths of x or longer:

$$N = N_0 \exp(-x/\lambda) \tag{3.20}$$

By differentiating, the number of free paths dN between x and $x + dx$ is

$$dN = \left(\frac{dN}{dx}\right) dx = \left(\frac{-N_0 \exp\left(-x/\lambda\right)}{\lambda}\right) dx = -\left(\frac{N}{\lambda}\right) dx \tag{3.21}$$

The familiar exponential curve of Equation (3.20) is reproduced with x/λ as ordinate in Figure 3.3. Remembering that exp (–1) = 0.37, we may deduce that 63% of free paths are shorter than λ. Since $e^3 = 20$ to a very good approximation, 95% are shorter than 3λ; only 1 in 22,000 is greater than 10λ.

The hard-sphere model of collisions is, of course, a fairly gross approximation, especially for molecules such as nitrogen N_2 that do not have any spherical symmetry. One aspect of the approximation is the neglect of the forces of attraction between molecules in close proximity and the effect of the speed of the molecules. A fast incoming molecule with so much momentum that it passes close by another molecule without appreciable deflection would, on the same track at a lower speed, have been deflected more as the forces involved would act for a longer time. Thus, one might expect that with an increase of temperature, which determines average molecular speeds, the cross section for scattering will be decreased slightly to give a

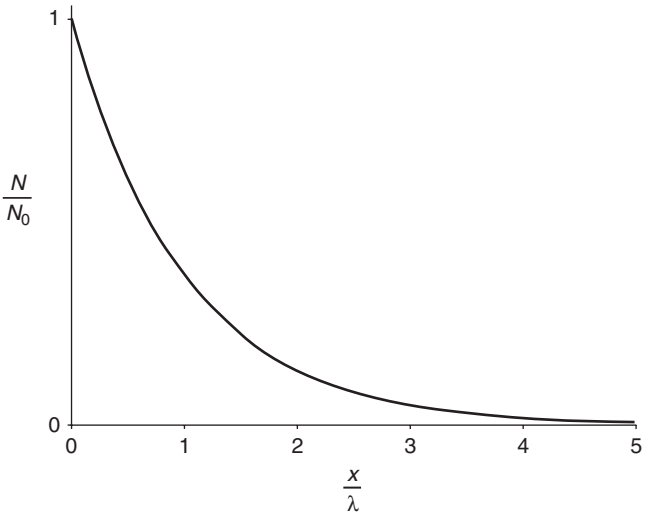

FIGURE 3.3
Fraction of molecules having free paths longer than x.

mean free path that increases with temperature. The reader will find a temperature correction due to Sutherland discussed in Kennard (1938) and detailed discussion of free paths in Jeans (1982) and Present (1948). Because most vacuum practice is carried out at room temperature, and great precision is not asked of a quantity whose simple derivation is relatively crude, the formulas given are adequate.

The basic formula $\lambda = 1/\sqrt{2}\, n\sigma$ in which $\sigma = \pi d^2$ was derived for a gas of like molecules so that d, the sum of the radii of the two interacting particles, was therefore just the molecular diameter. For a molecule of radius r_1 colliding with another of radius r_2, the dimension d would be replaced by $(r_1 + r_2)$. Mean free paths in mixtures are complicated to analyze, but Jeans gives some results for particular cases. For example, in a mixture of small, relatively light (and therefore fast-moving) helium molecules with a gas of larger, much heavier molecules of diameter d_2 and number density n_2, the mean free path for helium is approximated by

$$\lambda_1 = \frac{4}{\pi n_2 d_2^2}$$

For a fast-moving point particle such as an electron moving through a gas of molecules of diameter d, the target area σ will simply be the cross sectional area $\pi d^2/4$, and because of the electron speed, the factor $\sqrt{2}$ that allows for their motion can be ignored, so that $\lambda = 4/n\pi d^2$. For fast-moving ions, σ would be $\pi(d/2 + r)$ where r is the ionic radius and the factor $\sqrt{2}$ would again be ignorable to give $\lambda = 1/n\pi(d/2 + r)^2$.

3.7 Summary

The three results for J, p, and λ expressed in terms of n are very basic to an understanding and analysis of phenomena in various degrees of vacuum. Table 3.1 gives numerical values of J, n, and λ at various pressures.

As gas becomes more rarefied, quantified by the reduction in its molecular number density n, impact rates J fall, and mean free paths naturally increase. At atmospheric pressure, the mean free path, 64 nm, is submicroscopic — 64 nm = 64×10^{-7} cm — and therefore, there are of the order of 100,000 such distances in just 1 cm. Likewise, the number of collisions which a molecule has in one second, on average $\bar{v}/\lambda$, will be $472/64 \times 10^{-9}$, that is, $\sim 10^{10}$ per second. The number density of 2.5×10^{25} m^{-3} implies that the number of molecules in a 1-cm cube is $\sim 10^{19}$, and these have as an average speed of order several hundred meters per second. Thus, the molecular picture of a gas that kinetic theory gives us is of an enormous number of molecules, themselves very tiny and traveling at high speeds, having very many collisions per second between which they travel only small distances, with each collision altering the direction and speed of their motion. It is a picture of complete molecular chaos in which individual motions cannot be followed but for which there are very good average values. Figure 3.4(a) represents, at an instant, molecular trajectories in a gas under these conditions. The arrows suggest the direction and magnitude of individual velocities. Figure 3.4(b) follows the motion of a typical molecule along an element of its zigzag path. Figure 3.4(c) will be discussed in Section 3.9.

The importance of J has been previously emphasized. It can be put into a form that is frequently more useful by substituting for n from $p = nkT$ and $\bar{v}$ from Equation (3.7) to give

$$J = \frac{p}{\sqrt{2\pi mkT}} \tag{3.22}$$

TABLE 3.1

n, λ, and J at various p for N_2 at 295 K

p mbar	n (m^{-3})	λ	J (cm^{-2} s^{-1})
10^3 = 1 atm	2.5×10^{25}	64 nm	2.9×10^{23}
1	2.5×10^{22}	64 μm = 0.064 mm	2.9×10^{20}
10^{-3}	2.5×10^{19}	64 mm	2.9×10^{17}
10^{-6}, HV	2.5×10^{16}	64 m	2.9×10^{14}
10^{-10}, UHV	2.5×10^{12}	640 km	2.9×10^{10}

This may also be written as

$$J = \frac{pN_A}{\sqrt{2\pi MR_0 T}} \tag{3.23}$$

3.8 Evaporation and Condensation

As mentioned in the Introduction, for a substance in a condensed state the thermal energy of its atoms or molecules manifests itself as localized vibratory motion restrained by cohesive binding forces. At the surface of a solid or a liquid, however, the environment is different from the bulk, and molecules are less restrained. Some molecules at the surface will, by chance and favorable collisions with neighboring molecules, acquire sufficient energy to break their bonds and escape from the surface to enter the gas phase, collectively forming a continuous evaporation flux. The rate at which this happens depends on the temperature and the strength of binding at the surface, which depends on the material. The rapid evaporation of ether was noted in Chapter 2. Water evaporates more slowly. Solids also release their molecules in the same way, though at a rate that is many orders of magnitude smaller at room temperature. For evaporation rates from solids to be appreciable, their temperature usually has to be quite high. For many materials, this is exploited in evaporation sources for thin film deposition.

A number of useful results may be deduced by considering the equilibrium of a liquid (or solid) with its vapor in a closed volume maintained at constant temperature, as shown schematically in Figure 3.5. Imagine that, as first set up, there is just liquid in the bowl and high vacuum in the enclosure with no vapor in the space above it. As molecules escape from the liquid surface,

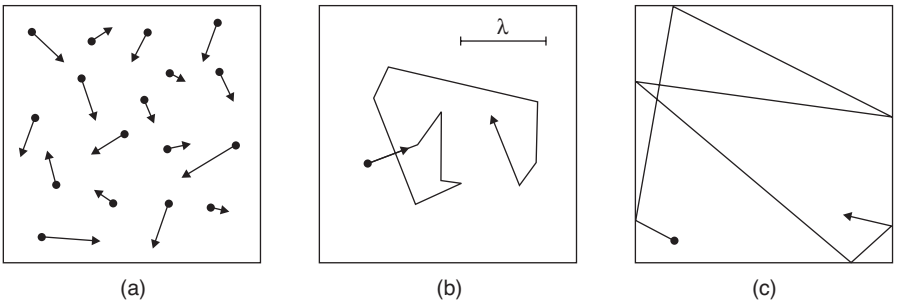

FIGURE 3.4
Molecular trajectories in a gas. (a) Molecular chaos in a gas, (b) trajectory of a typical molecule, (c) trajectory in a rarefied gas, λ greater than container size.

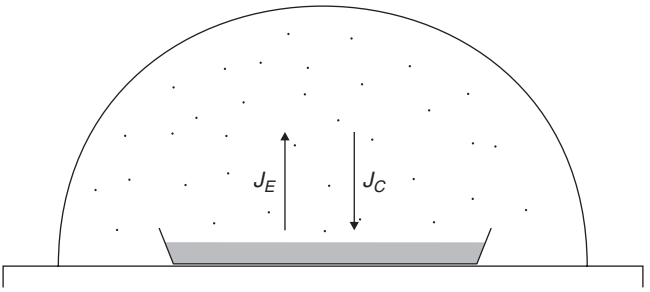

FIGURE 3.5
Evaporation and condensation fluxes from a liquid surface into a closed volume.

they establish a vapor in the enclosure. Some of them will return to the liquid surface again and condense; at their prevailing density n, as a vapor, at any instant, there will be an associated impingement rate back onto the liquid surface. Thus, there is a constant evaporation flux, J_E, from the liquid, and a condensation flux, J_C, which builds up from nothing as the vapor phase is established. When the condensation flux has increased to a value J_C equal in magnitude to J_E, a state of equilibrium exists. As many evaporate as condense, and the enclosure contains saturated vapor with a constant molecular number density n exerting its saturation vapor pressure p_e, often simply referred to as its vapor pressure. A pressure gauge attached to the enclosure would have recorded the change from high vacuum to p_e. Although on a macroscopic scale nothing seems to be happening once this equilibrium is reached, at a molecular level two vigorous processes are in exact balance.

By Equation (3.23) with $J = J_C$, the condensation flux will be

$$J_C = \frac{p_e N_A}{\sqrt{2\pi M R_0 T}} \tag{3.24}$$

where p_e is the saturation vapor pressure at temperature T. Since $J_C = J_E$, the evaporation flux J_E is also given by this expression. Therefore,

$$J_E = \frac{p_e N_A}{\sqrt{2\pi M R_0 T}} \tag{3.25}$$

This essentially thermodynamic argument using measured vapor pressure data is the only route by which J_E can be predicted. It is not possible to determine vapor pressures by first-principles calculations.

The vapor pressure of a substance increases very rapidly with temperature because the release of molecules from a solid or liquid surface is a thermally activated process whose probability rises dramatically with increase of temperature, so much so that graphical presentation of data requires logarithmic

pressure scales. Typical data for various materials and the important vapor pressure curve for water were given at the end of Chapter 2.

In nonequilibrium evaporation conditions, for example, where the vapor released from a surface is pumped away, Equation 3.25 gives the maximum possible evaporation flux from the surface at temperature T. It finds useful application in predicting the mass loss per unit area per second from the surface of a substance of molar mass M at temperature T. The associated value of p_e is determined from vapor pressure data. Assuming that the evaporating molecules are directly pumped away or captured elsewhere so that there is no returning flux, the mass flow rate away from the surface is just

$$m \times J_E = \frac{p_e N_A m}{\sqrt{2\pi M R_0 T}} = p_e \sqrt{\frac{M}{2\pi R_0 T}} = \frac{p_e}{7.2}\sqrt{\frac{M}{T}}\ \text{kg m}^{-2}\text{s}^{-1} \qquad (3.26)$$

where p_e is in Pa and M in kg.

Another example of nonequilibrium evaporation is the *vacuum drying* process, in which water is removed at a constant temperature from wet pharmaceutical pastes by pumping. It is necessary to maintain the temperature of the surfaces involved at the designated value by supplying heat energy to provide the latent heat of vaporization. Otherwise, cooling occurs due to pumping, the vapor pressure is reduced, and the efficiency of removal falls. In other applications, the cooling may be desirable. In low-temperature physics, for example, pumping hard over thermally well-isolated liquid helium is used to achieve temperatures below 4 K.

In the nonequilibrium process of condensation of a gas or vapor on to a colder surface, such as arises in vapor traps and condensation pumping (to be discussed), Equation (3.24) also gives the maximum possible condensation rate per unit area, assuming that the vapor pressure of the condensed species at the temperature of the condensing surface is negligibly small, which is frequently the case.

3.9 Knudsen Number: Continuum and Molecular States of Gas

At atmospheric pressure, we think of gases as fluids. Air offers tangible resistance to motion through it at speed, and communication by sound waves is possible because speech and music from loudspeakers induce wave motion through the air in the form of pressure/density waves. The criterion that gas should behave in this familiar way is that our picture of Figure 3.4(a) should be valid, the essence of which is that molecule–molecule collisions be the characteristic process that determines gas behavior. In other words, the mean free path λ in the gas should be very much less than the characteristic

dimension of its container. For a cubical box, that would be the length of a side, or for a pipe, its diameter. But from Table 3.1 it is evident that at low pressures, mean free path will become comparable with or greater than a container's characteristic dimension. Molecule–surface collisions will now dominate gas behavior and molecule–molecule collisions become quite rare. Our physical picture would become that of Figure 3.4(c). Characterized by $\lambda \geq D$, this is called the *molecular* state of gas. Table 3.1 indicates that this state would be reached in a pipe of 50 mm diameter at a pressure of 10^{-3} mbar. In cube of side 0.5 m, it would be reached at 10^{-4} mbar (we interpolate to deduce that $\lambda = 0.64$ m at 10^{-4} mbar).

It is useful to have a criterion for distinguishing between the fluid (or what is better called the *continuum* state) and the molecular state. This is provided by the Knudsen Number Kn defined as $\mathrm{Kn} = \lambda/D$, where λ is the mean free path and D is a characteristic dimension. There is, of course, in the nature of things, no sharp change between the continuum and molecular states, but rather a gradual transition between them. These states, or *regimes* as they are sometimes called, are conventionally taken to be

$$\mathrm{Kn} < 0.01\text{: CONTINUUM}$$

$$\mathrm{Kn} > 1\text{: MOLECULAR}$$

and in between, $0.01 < \mathrm{Kn} < 1$: TRANSITIONAL

Kn is a useful dimensionless number because it automatically incorporates information about the gas condition in relation to the container size. For example, at a pressure such that λ is 5 cm, the gas in a 1-cm cube would be in a molecular state, but in a large vacuum tank of dimension 3 m at the same pressure, conditions would be more nearly continuum. For the purpose of making quick estimates of Kn, it is useful to memorize a value of λ at some particular pressure. Since $\lambda \times p$ is constant, then knowing that $\lambda = 64$ m at 10^{-6} mbar means that the other values, for example, 64 mm at 10^{-3} mbar, 6.4 m at 10^{-5} mbar, etc., are easily deduced.

It is possible to compare the number of molecule–molecule collisions with the number of molecule–wall collisions in containers of simple geometry. Suppose a cubic container of side L contains gas at a pressure and associated number density such that its mean free path in unrestricted, continuum conditions is λ. The number of collisions per second with the wall is just the area $6L^2$ multiplied by the impact rate of Equation (3.1) i.e., $J = n\bar{v}/4$, which is $3nL^2\,\bar{v}\,/2$. Within the volume L^3 of gas at this pressure, there are nL^3 molecules, and each makes $\bar{v}/\lambda$ collisions per second. Thus the number of molecule–molecule collisions per second within the gas will be $n\,\bar{v}\,L^3/2\lambda$, and the ratio of the number of surface collisions to the number of volume collisions is $3\lambda/L$ or 3 Kn.

3.10 Internal Friction and Viscosity in Gases

When there is relative motion within the bulk of a gas, such as is produced by the action of opening a door, the disturbance to the gas at a molecular level is a very small perturbation superposed on what we know to be a vigorous, chaotic, and random molecular motion, and it is soon dissipated by viscous effects. The mechanism by which these changes are caused and dissipated is the propagation of change from molecule to molecule.

Consider, as in Figure 3.6, a stationary horizontal plate in the plane $z = 0$ and a large plate of area A parallel and close to it at $z = t$, moving in the horizontal x direction with velocity u_0. The gas between them is in a pure shearing mode, and an applied force F is needed to maintain the motion against the gas friction. Continuum conditions are assumed ($\lambda << t$), and edge effects are ignorable as the plates are supposed to be large and close. Gas in immediate contact with both plate surfaces is taken to be stationary with respect to them, and so across the gap, a gradient in the gas velocity of magnitude (u_0/t) exists, which is assumed to be uniform. Experimentally, it is found that F depends on A, u_0, and t according to

$$F = \eta A\left(\frac{u_0}{t}\right) \tag{3.27}$$

The quantity u_0/t is called the velocity gradient, and the equation defines η, the viscosity coefficient.

Kinetic theory is applied as follows: The gas is imagined to be divided into layers of thickness λ so that a molecule moving in a downward direction, having just had a collision in one layer, travels on average a distance λ before having its next collision in the layer below. At an intermediate plane such as z_i between two such layers, the imposed horizontal drift velocity caused by the motion of the upper plate is u. It is supposed that molecules arriving here from above had their last collision at $z_i + \lambda$ and have drift velocity $u + \delta u = [u + \lambda(du/dz)]$, and those arriving from below come from $z_i - \lambda$ with drift velocity $[u - \lambda(du/dz)]$. The velocity gradient $(du/dz) = u_0/t$. From above

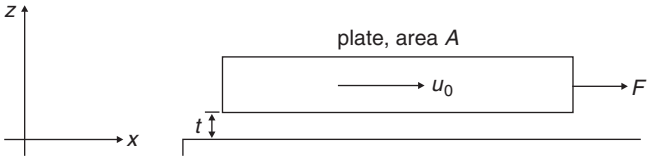

FIGURE 3.6
Viscous effects due to relative motion.

the plane z_i, the flux J per unit area per second of downward moving molecules with horizontal drift momentum $m(u + \delta u)$ has collisions that reduce their drift velocity on average to u. In doing so, this flux exerts a force in the $+x$ direction on the layer below it, and by Newton's third law, itself feels a force in the $-x$ direction. Momentum is therefore being transported across the plane. There is a similar upward flow of drift momentum, on average $m(u - \delta u)$, from the plane below due to the equal flux of upwards traveling molecules so the total force on the plane is zero. The net rate of transport of the drift momentum is therefore

$$JAm\left[u + \lambda\left(\frac{du}{dz}\right)\right] - JAm\left[u - \lambda\left(\frac{du}{dz}\right)\right] = 2JAm\lambda\left(\frac{du}{dz}\right)$$

By Newton's second law, this is the force transmitted between layers. It is transmitted to the surfaces of the bounding plates where it is balanced by the external force F. But $J = n\bar{v}/4$ from Equation (3.13), and so by substitution

$$F = \left[\frac{nm\bar{v}\lambda}{2}\right]A\left(\frac{du}{dz}\right) \tag{3.28}$$

By comparison of Equation (3.27) and Equation (3.28), with $(du/dz) = u_0/t$,

$$\eta = \frac{nm\bar{v}\lambda}{2} \tag{3.29}$$

The dependence on the variables is correct, but the method of averaging is rather crude. Better averaging of the collision processes within a layer leads to a factor of (1/3) rather than (1/2), and the result

$$\eta = \frac{nm\bar{v}\lambda}{3} \tag{3.30}$$

This latter value is usually quoted for use, although sophisticated calculations of Chapman and Cowling (1991) for the hard-sphere model come closer to Equation (3.29)!

The fact that λ is inversely proportional to n means that n cancels in Equation (3.30) so that the viscosity coefficient is predicted not to depend on pressure, which seems quite surprising. One might expect viscous effects to depend strongly on gas density. But this counter-intuitive prediction was verified experimentally for gas in continuum conditions. This was quite important historically in the development of the kinetic theory. The reason for the independence of pressure is that, if the pressure is doubled, say, so that there are twice the number of molecules carrying the drift momentum,

their free paths are halved, with the result that the rate of transport of momentum is unchanged. The dependence of η on $\bar{v}$ implies that η increases with temperature. This is also verified experimentally, though the temperature dependence is slightly greater than the predicted $\sqrt{T}$.

At pressures lower than that discussed above, such that gas is *not* in a continuum condition, the viscosity coefficient is no longer constant and starts to decrease. The details of the interaction of molecules with the bounding surface begin to play a critical role. In the transitional regime (Kn = λ/t in the range 0.01 up to about 1), the processes are rather complex, and the concept of "slip" is introduced into the analysis. Delchar (1993) and Walton (1983) may be consulted for a detailed treatment. When gaseous conditions are molecular so that the separation of surfaces in relative motion is less than the mean free path, the description of phenomena becomes more straightforward. Viscosity as described above ceases to have any meaning and the tangential forces that the molecules of a gas create on a surface by their movement relative to it are described as "molecular drag" forces. They are discussed in Chapter 4.

3.11 Heat Conduction in Gases

In just the same way as viscosity in gases at normal pressures can be analyzed as the transport of imposed drift momentum through the gas by molecular collisions, so can thermal conduction be viewed as the transport of energy by a similar mechanism. Again, this is strictly a nonequilibrium process, but the deviations from equilibrium are typically very small. We have seen that the mean energy of the molecules in a gas is proportional to the gas temperature T. When a temperature gradient is maintained in a gas the more energetic molecules at a region of higher temperature pass on energy by collisions. The property being transported is the excess energy of motion associated with the temperature difference δT over distances of the order of the mean free path. An analysis similar to that above for viscosity yields, for the thermal conductivity K

$$K = \frac{1}{3} nm\bar{v}\lambda c_v \tag{3.31}$$

where c_v is the specific heat per kg at constant volume. This result again correctly predicts that in continuum conditions K is independent of pressure, though, in a similar way to viscosity, this ceases to be true at low pressures, when its decrease with decreasing pressure can be exploited in pressure gauges, as will be discussed later.

3.12 Thermal Transpiration

This phenomenon does not occur in the continuum state. It is the establishment of a pressure difference between two connected regions at different temperatures and is fully developed in molecular conditions. Consider the arrangement shown in Figure 3.7 in which two vessels maintained at different temperatures T_1 and T_2 contain gas at low pressures p_1 and p_2 and are connected together through a very short pipe, so short that it can be considered as a simple aperture of area A.

Suppose that the pressures on both sides of the aperture are such that mean free paths are greater than the aperture diameter. In its vicinity, therefore, gas is in a molecular state. The eventual steady-state condition achieved is determined by the fact that the number of molecules per second passing through the aperture in opposite directions, that is $J_{12}A$ and $J_{21}A$, must be equal. Thus, from either Equation (3.22) or Equation (3.23) for J, we have

$$\frac{p_1}{\sqrt{T_1}} = \frac{p_2}{\sqrt{T_2}}$$

and so

$$\frac{p_1}{p_2} = \sqrt{\frac{T_1}{T_2}} \tag{3.32}$$

The phenomenon described by this equation is called thermal transpiration. By convention, T_2 is taken as the higher temperature. It is different from the situation that would exist if the gas were in a continuum state, which would be characterized by p_1 and p_2 being equal. Careful experimentation has verified Equation (3.32) in equipment where the idealized situation of

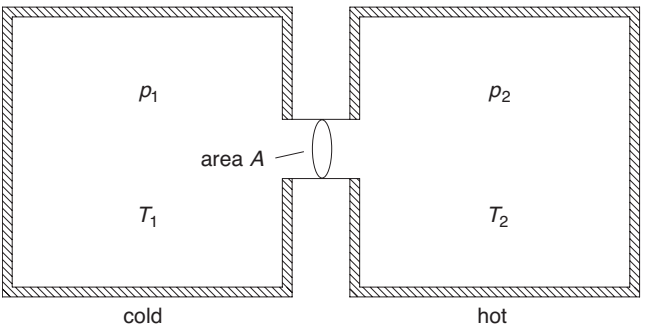

FIGURE 3.7
Thermal transpiration.

Figure 3.7 is closely approximated. It has relevance for vacuum systems that have regions which are not at room temperature, for example, in cryogenic applications. Because vacuum gauges are normally operated at room temperature, the pressure in parts of the system at a different temperature has to be inferred from the reading of the gauge, allowing for transpiration effects. For example, if region 1 in Figure 3.7 is at liquid nitrogen temperature 77 K and region 2 at room temperature 295 K where a pressure gauge measures p_2, then $p_1 = \sqrt{77/295} \times p_2 = 0.51\ p_2$. The significance of the effect for pressure measurements will be discussed later.

To demonstrate an application rather than a consequence of the effect, it is instructive to consider the following "thought experiment." Consider the aperture in Figure 3.7 to be a variable aperture or "iris," which could be opened up from a minimum size, as discussed, to a much larger value so that conditions became continuum. Starting from this state with the temperature difference set, the pressures would be equal, say, p_{10} and p_{20}. If the aperture were now closed down again to the molecular condition, the molecular fluxes through the aperture would become unequal. Remembering that T_2 is greater than T_1 we see from Equation (3.32) that the molecular flux from cold to hot, which is proportional to $p_{10}/\sqrt{T_1}$, is greater than the flow from hot to cold, which is proportional to $p_{20}/\sqrt{T_2}$ because $p_{10} = p_{20}$ initially. Thus, because of the temperature difference, molecules are driven from cold to hot, and eventual steady state is only achieved when p_1 has decreased and p_2 increased to the values given by Equation (3.32). This movement of molecules from cold to hot that precedes the achievement of the steady state is known as thermomolecular pumping, a subject discussed by Hobson and Salzman (2000). The classic text of Loeb (1961) discusses thermomolecular matters in detail.

Exercises

3.1 Use Equation 3.8 to verify that the average speed of nitrogen molecules at room temperature 295 K is 472 ms^{-1}. Then, use the relative mass ratios to calculate the average velocity at this temperature for molecules of (a) hydrogen, (b) water vapor, (c) argon.

3.2 What is the average speed of nitrogen molecules at (a) 300°C, (b) 77 K?

3.3 If the mean free path for nitrogen molecules is 64 mm at 10^{-3} mbar, what will it be at (a) 0.1 mbar, (b) 0.3 mbar?

3.4 How many collisions will a nitrogen molecule have in one second at a pressure of 0.1 mbar?

3.5 At a pressure 10^{-3} mbar, what fraction of nitrogen molecules will travel for 192 mm or more without having a collision?

3.6 The mean free path in helium gas at 10^{-3} mbar is 173 mm. Is this bigger or smaller than that for nitrogen under the same conditions? Is it very different? What will be the mean free path in helium at 10^{-4} mbar?

3.7 If a well-built concert hall of typical dimension 100 m contained just nitrogen gas at a pressure 10^{-7} mbar, what state would the gas be in?

3.8 How many moles of nitrogen impinge in one second on an area of one square cm of surface at pressure of 1000 mbar and at room temperature? (Table 3.1)

Problems

3.9 Estimate how long it would take for a fingerprint to evaporate from the internal wall of a high vacuum vessel at room temperature. Assume that it is 2 μm thick and that the fatty acid involved has relative molecular mass 202.5, density 1000 kgm^{-3}, and a vapor pressure of 10^{-8} mbar.

3.10 Calculate the maximum rate of evaporation in kgs^{-1} per m^2 of surface for water at 20°C into a perfect vacuum. The vapor pressure of water at 20°C is 23.4 mbar. Assume (rather artificially) that the energy for evaporation can be supplied at a rate sufficient to maintain the surface temperature at 20°C. On the basis of this calculation, by how much would the water level in an open expanse of water fall in one minute under these conditions?

3.11 A spherical vacuum vessel of diameter D contains gas at a molecular number density n and mean free path λ. Derive expressions for the number of intermolecular collisions within the gas per second and the number of collisions per second with the container wall and show that their ratio is $3\lambda/D$. For such a vessel of diameter 10 m containing nitrogen gas at 10^{-5} mbar, would the gaseous conditions be continuum, transitional, or molecular?

4

Surface Processes and Outgassing

4.1 Introduction

Although an obvious and important role of surfaces in vacuum practice is that of the inner wall of the vessel in providing the boundary of the vacuum within, there are a number of processes occurring between gases and surfaces, both inside the vacuum and at its boundary wall, that make the role of surfaces more important than their mere mechanical presence. There are few circumstances in which surfaces play a strictly neutral role. For example, as we shall see shortly, the rebound of gas molecules from surfaces is not as simple as it is modeled to be in the kinetic theory presented in Chapter 3, though the results of that theory (be assured) are correct.

In addition to scattering molecules in non-simple ways, surfaces may, depending on a number of factors, trap molecules in bound states, a process called *adsorption*. The reverse process, the release of molecules from trapped states, is called *desorption*. Adsorption finds applications in pumping, while, less benignly, desorption is the principal feature of the process of outgassing, a property of fundamental importance and characteristic of all surfaces in vacuum. It describes the prolonged desorption of gas molecules from a surface, some of which reach it from subsurface regions in which they have been earlier stored by exposure to the atmosphere and other mechanisms. Thus, after the air in a vacuum vessel has been pumped away, pumping is still needed to remove the continual outgassing load from surfaces exposed to the vacuum. These matters, together with sputtering, which is the removal of surface atoms due to energetic particle bombardment, and the subject of molecular drag, which arises from the interaction of gas molecules with fast-moving surfaces, are to be discussed in this chapter.

4.2 The Scattering of Molecules by a Surface

To an incoming molecule, the surface of a solid presents a much more substantial target than it encounters in the gas phase when it collides with a single molecule of the same kind. It is essentially approaching a wall of atoms and, as it gets close, will come under the influence of many of them. There are three possible outcomes to its collision with the surface.

1. An *elastic* rebound without energy loss, in which the angle of incidence made with the normal to the surface at the point of impact equals that of reflection, with the directions of incidence, reflection, and the normal all being in the same plane. In this "specular" reflection, the molecule's linear momentum parallel to the surface is unchanged. Perpendicular to it, the momentum is unchanged in magnitude but reversed in direction, as shown in Figure 4.1(a).
2. Although the rebound of the molecule to the gas phase is more or less immediate, some energy is either lost to the surface or gained from it (for example, by exchange of energy with vibrations of the surface atoms), in which case the collision is described as *inelastic.* Such an outcome is represented in Figure 4.1(b).
3. The molecule may lose sufficient kinetic energy that it becomes bonded to the surface in an adsorbed state. Its stay in this trapped state will not be permanent. A molecule that has lost sufficient energy to bring it into equilibrium with the host surface at the prevailing temperature will nevertheless have kinetic energy of vibrational motion under the restraints imposed by its bonding to the surface and will eventually, by statistical chance, acquire enough energy from favorable collisions with its neighbors to break free. Depending on the identities of the molecule and the surface atoms, its stay may be brief or long. When the molecule does leave the surface, its direction of departure will be random and unrelated to its direction of arrival, because its initial momentum parallel to the surface, like the perpendicular component, will have been dissipated in the collisions that led to its capture. Thus, whereas elastic and inelastic rebounds off the surface will occur with either complete or partial preservation of the parallel component of momentum in the plane of incidence, the direction of release from a trapped state can be anywhere into the three-dimensional solid angle of 2π above the surface, as suggested in Figure 4.1(c).

For a gas at the same temperature as its container and in equilibrium with it, whose molecules will have a velocity distribution appropriate to that temperature, these different processes must separately be occurring at an unchanging rate. Thus, the rate of inelastic energy loss from gas to surface

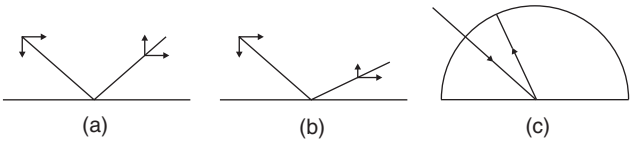

FIGURE 4.1
Rebound of a molecule from a surface. (a) Specular, (b) inelastic, (c) via a trapped state.

must be balanced by an equal and opposite gain from surface to gas, and similarly the rates of trapping and release of molecules must be in balance.

The facts presented above are at variance with the assumption used in the kinetic theory model of Chapter 3 that all collisions are elastic and specular, but that does not invalidate the results derived, for the following reason: Consider an imaginary plane in the gas phase, parallel to the surface and at a distance of, say, 10 molecular diameters. This distance is smaller than the mean free path of molecules in the gas, even at atmospheric pressure. Molecules that leave the surface pass through the plane after either elastic or inelastic collision or delayed release from a trapped state, as described above. Now, the theory describes a gas and a surface in equilibrium at temperature T, and so by definition requires that at this plane the number of molecules and their distribution in energy and angle going towards the surface and coming away from it are the same. The *principle of detailed balance* asserts that, because the number of molecular events involved is so large in the time of observation (at atmospheric pressure there are in one second approximately 10^{21} impacts per cm^2, equivalent to one million on each atom of the surface) that the conditions assumed in the model are, in effect, complied with. For a molecule incident on a small area of surface in a particular direction with a particular speed that departs in a non-specular direction, there is within a short time another molecule and a similar non-specular event whose result is to provide a "partner" specular reflection, so to speak, to the first. At the plane of reference, therefore, even though identities differ, it is as though all reflections were specular, in accord with the model, and the results such as $p = nkT$ are sound.

The nature of the molecule–surface interactions that determine the various possibilities for scattering or capture, described above, has of course been of concern since the earliest days of the subject. A thorough discussion of them is given by Redhead (1968) in his classic text. The experiments of Stern and his collaborators in the 1930s (see, for example, Haken and Wolf, 1996) pioneered the use of molecular beam scattering techniques as an investigative tool. Essentially, a stream of molecules of thermal energy is directed onto a surface in vacuum and the characteristics of the scattered molecules measured by various means. In recent decades, the developments of this technique for applications in surface science have become highly sophisticated and informative. Experiments carried out in ultrahigh vacuum in which molecular beams, sometimes modulated, are directed onto clean, well-prepared surfaces and the scattered fluxes detected reveal detail about elastic

and inelastic scattering phenomena, and in certain cases diffraction. Nowadays, incident beams can be produced with energies up to about 10 eV and relatively small energy spread, in contrast to thermal beams. An introduction to these matters may be found in the text of Hudson (1992).

In the context of molecules losing energy to a surface, one may define an energy accommodation coefficient α_E by the equation

$$\alpha_E = \frac{E_1 - E_2}{E_1 - E_S} \tag{4.1}$$

in which E_1 and E_2 are the energy of the incident molecule before and after scattering, respectively, and E_S the energy associated with equilibrium at the surface. Thus, for elastic reflection $E_1 = E_2$ and $\alpha_E = 0$, while if $E_2 = E_S$, the other extreme, $\alpha_E = 1$ and accommodation is complete. In nonequilibrium situations — for example, if a gas is hotter than a surface on which it impinges — the energies E_1 and E_2 can represent the mean kinetic energies of translation associated with the fluxes of arriving and scattered molecules. Because the mean kinetic energy of translation is $3kT/2$, Equation 4.1 may then also be expressed as

$$\alpha_E = \frac{T_1 - T_2}{T_1 - T_S} \tag{4.2}$$

For diatomic and polyatomic molecules like N_2, CO, and H_2O, for example, there are vibrational and rotational kinetic energies in addition to that of translation, although these may not be excited at 300 K. Potentially, however, the complexity of the scattering process is greater than for monatomic species such as argon, and accommodation coefficients appropriate to different aspects of the scattering process need to be defined. Beam scattering experiments that give state-specific information are discussed by Somorjai (1994).

While modern experiments on well-characterized surfaces yield data that indicate the complexity of many phenomena that take place on them, it is perhaps fortunate that the surfaces of the inner wall of the vacuum enclosure (frequently constructed of stainless steel) and the surfaces of components within are usually quite rough and irregular on a microscopic scale because of the ways in which they have been manufactured and prepared for use. Usually, they have microcracks and other flaws, and are covered with stable layers of oxide, on top of which, moreover, adsorbed gas will generally reside. One result of this roughness is that scattering effects tend to be randomized in direction. The evidence is (see, for example, Steckelmacher [1986]) that the scattering of molecules from such surfaces is diffuse and follows the cosine law of Knudsen, described below. Another related result is that, before departing randomly, molecules in colliding with such surfaces transfer all their incoming momentum. Evidence for this comes directly from

devices exploiting what is called "tangential momentum accommodation" in molecular drag, discussed in Section 4.6, and in connection with the spinning rotor vacuum gauge, discussed in Chapter 7.

4.3 Diffuse Scattering from a Surface — the Knudsen Cosine Law

In order to formulate this law, it is helpful to consider first the process of effusion, which is the escape of molecules from an enclosure into a surrounding high vacuum when the conditions inside it and close to the exit aperture are such that gas is in a molecular state, as discussed in Chapter 3.

In Figure 4.2(a) an enclosure contains gas or vapor at pressure p and associated number density n. Through a small hole of area A and diameter D in a thin part of the wall, whose thickness is negligible compared with D, molecules escape into the surrounding high vacuum, sufficiently rarefied that there is negligible flux returning. If p is sufficiently low that the mean free path λ of the molecules inside the enclosure is at least 10 times larger than the hole diameter D, the gas in its vicinity is in a molecular state. There will therefore be a negligible amount of molecule–molecule scattering in this region, and molecules in the bulk of the gas within a hemispherical volume of radius $\sim\lambda$ centered on the hole that are heading towards it will pass straight through to the high vacuum on the other side. This is called *effusion* and contrasts with the situation at higher, non-molecular pressures in which gas would stream through it as a fluid. If a mechanism exists to replace the molecules lost from the enclosure and maintain the number of molecules inside constant, there is a steady effusive

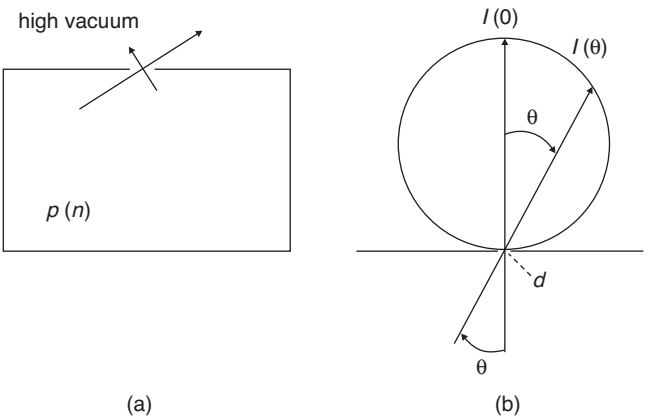

FIGURE 4.2
Experiment to illustrate the cosine law. (a) Effusion, (b) representative sphere.

flow out of it. A vapor in equilibrium with a parent liquid or molten phase within the enclosure can be thus maintained by supplying the necessary latent heat of evaporation, a property that is exploited in the design of Knudsen effusion cells.

Under the conditions described above, because the gas or vapor is essentially in equilibrium, all directions of individual molecular motion are equally probable. The flux of molecules escaping through the hole in a direction perpendicular to the surface, which constitutes the flux $I(0)$ that leaves the enclosure in the normal direction, is the flux that travels towards the opening in the same direction inside the enclosure. It is proportional to the area A of the hole. In a similar way, the flux emerging at an angle θ to the normal is determined by the apparent size $A\cos\theta$ of the aperture for the oncoming gas flux traveling in this direction inside the enclosure. The flux in any direction within the enclosure is the same, but the "escape window" appears progressively smaller as the direction becomes more oblique. Thus, the flux $I(\theta)$ emerging through the hole at angle θ may be expressed as

$$I(\theta) = I(0)\cos\theta \tag{4.3}$$

$I(\theta)$ is the number of molecules per second from an aperture of area A traveling into unit solid angle at angle θ to the normal. Equation 4.3 is the statement of Knudsen's cosine law of effusion. From the geometrical properties of a circle, the effusion profile has the circular (strictly spherical) form shown in Figure 4.2(b). The flux into the solid angle $2\pi\sin\theta d\theta$ at angle θ is $I(\theta) \times 2\pi\sin\theta d\theta$, and the total emerging flux I_T is determined by integrating over the total solid angle (2π) available. It emerges that $I_T = \pi \times I(0)$. But, seen from the inside of the container, this is the flux impinging on the hole and is $J \times A$, with J, the impingement rate on unit area of molecules at pressure p and temperature T, given by Equation 3.22. Hence $I(0)$ can be expressed in terms of the temperature and pressure of the gas or vapor contained.

Consider now the scattering of molecules from a small area A of a surface, when gas and surface are in equilibrium. In the argument used above, a small hole of area A permitted the escape by effusion of molecules from a region at pressure p into high vacuum, and these were imagined to be replaced in order to keep the pressure constant. But now suppose that the hole is closed. The flux that previously passed through is now turned back. As previously discussed, the distribution of back-scattered molecules in angle and velocity at some reference plane a small distance from the surface must, overall, be the same as that which is incident, and therefore the contributions to this flux due to impingement on area A and back-scattering from it must be the same. The form of this back-scattered distribution must therefore be identical with that incident on it, namely, the cosine distribution of Equation 4.3. Hence, in the direction θ, $I(\theta) = I(0)\cos\theta$ is the cosine law of molecular scattering. It is a characteristic of scattering from the technically rough surfaces encountered in vacuum practice and of desorption from a

fully equilibrated phase. In optics, where it is referred to as Lambert's law, a related phenomenon is the equal brightness, irrespective of the angle of view, of illuminated matte surfaces.

Evaporation from surfaces has been verified to follow the cosine law in many cases, and relevant information will be found in the article by R. Glang (1970). A most elegant experimental verification on the evaporation of mercury was described by Knudsen (1934) in his classic monograph. The reader should be aware that the subject has complexities that are revealed in some molecular beam experiments. The status of the cosine law itself has been the subject of long-standing discussions, and the refinements of experimental techniques in surface science that probe molecule–surface interactions reveal increasing detail of basic scattering processes. The article by Comsa (1994) is a good introduction to these matters.

4.4 Adsorption and Desorption

The character of the adsorbed state in which a molecule may be trapped on a surface will depend, of course, on the identity of the partners. For example, one would expect that the interaction between an inert gas such as argon and a glass surface would differ from that of a chemically active gas, such as oxygen, and a metal. The bonds established between a molecule and a surface are not intrinsically different in nature from those existing within molecules and in bulk phases. Thus, in the weakest form of adsorption, van der Waals forces, which arise from the attraction between fluctuating dipole moments and are responsible for the condensation of gas to liquid, are the source of the attractive force between a molecule and a surface. This type of bonding leaves the structure of the molecule essentially unchanged and is purely physical in origin; the process is described as *physical adsorption*, sometimes shortened to *physisorption*. In contrast, when bonding involves electron transfer or sharing between the molecule and atoms of the surface, it is referred to as *chemisorption* and can be thought of more as a chemical reaction between the two. Bonding in chemisorption is generally much stronger.

The interaction between a gas molecule and a surface can be usefully considered in terms of an energy diagram in which the potential energy E of the molecule is plotted as a function of its perpendicular distance z from the surface. A simple example for physical adsorption is shown in Figure 4.3(a). For separations greater than that shown as A, the potential energy of the interaction is zero because there is no force between the two. At separations less than this, the weak but relatively long range van der Waals force starts to operate, and E becomes negative. The value of the associated force, the derivative $-dE/dz$ at any point, is negative in this region, signifying an attractive force. Much closer to the surface, in region D, the energy increases

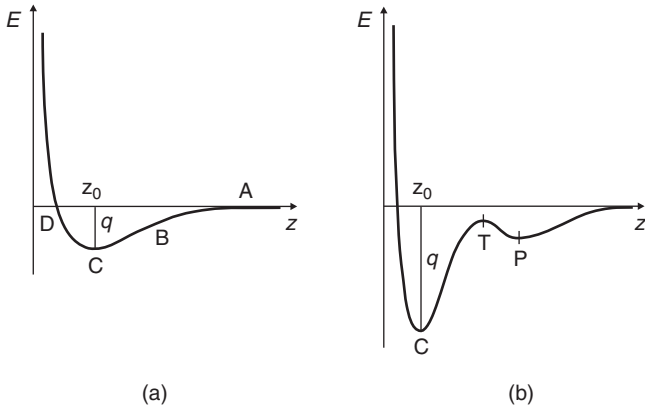

FIGURE 4.3
Energy–distance diagrams. (a) Physical adsorption, (b) chemisorption.

sharply as strong repulsive forces arise from the close proximity of the electron clouds of the interacting particles, and, correspondingly, $-dE/dz$ is positive. Point C, at distance z_0, where $-dE/dz = 0$, is the equilibrium distance of zero force, where attractive and repulsive forces are in balance, and it locates the energy minimum, an energy "well" of, say, depth q. It represents the energy given up to the surface, the "heat of adsorption," when the molecule adsorbs. Equally, it is the energy that has to be supplied if the molecule is to break its bond and desorb.

Heats of adsorption in physical adsorption, referred to one mole, are in the range of 6–40 kJ per mole. For nitrogen on various surfaces, it is about 15 kJ mol^{-1}. This is comparable with, but naturally greater than, the molar heat of vaporization, 6 kJ mol^{-1} approximately, which also depends on van der Waals forces. The specification as kJ per mole may also be expressed as eV per molecule; 97 kJ per mole is equivalent to 1 eV per molecule. A typical potential well depth for physisorption is therefore about 0.2 eV. This compares with the thermal energy kT at room temperature of 0.025 eV, and its significance will become apparent when desorption is considered shortly. Depending on conditions of temperature and pressure, the number of adsorbed molecules per unit area may build up at equilibrium to be less than a monomolecular layer or more, leading to multilayer adsorption. In the nature of the bond there will be van der Waals forces between the adsorbed molecules themselves as well as those made with the underlying surface. If more than one layer is established, bonding in outer layers will weaken to approach that which is characteristic of the liquid state.

The forces involved in chemisorption are significantly greater than for physisorption with heats of adsorption ranging widely, from about 40 up to 1000 kJ per mole. In many cases the alteration to a molecule's structure is quite drastic and occurs with dissociation of the molecule into its constituent atoms. In others, the molecular identity is retained. For example, when adsorbing at room temperature on nickel, hydrogen dissociates, whereas

carbon monoxide on nickel does not. The reader may consult Somorjai (1994) for a comprehensive compilation of the experimental findings. When dissociation occurs, the representative energy–distance diagrams for chemisorption become more intricate as shown, for example, in Hudson (1992) or McCash (2001).

The energy diagram of Figure 4.3(b) represents a relatively simple chemisorption without dissociation. The chemisorbed state C is closer to the surface and much deeper than a physisorbed state P that precedes it, which the adsorbing particle may occupy briefly before either desorbing or dropping into the chemisorbed state. The two are separated by an activation barrier, peaking at a point T below the zero level of potential energy. When this peak comes above it, the adsorption is said to be activated; incoming molecules need to have sufficient kinetic energy to surmount the barrier. In chemisorption, the heat of adsorption q has now a rather larger value than in Figure 4.3(a) and desorbing from this state will be correspondingly more difficult. In contrast with physisorption, the surface coverage in chemisorption is frequently complete when one layer has been formed, which is not surprising because the process exploits particular chemical attributes of the host surface.

To get a general understanding of the dynamical behavior of molecules in an adsorbed state and the orders of magnitude involved, it is instructive to consider a simple model following the treatment of de Boer (1953). Consider the survival of a molecule in the adsorbed state at temperature T, assuming that the adsorbed layer of such molecules, the adlayer, is sufficiently dilute that there are no complicating interactions with other adsorbed molecules. Once attached to the surface, the molecule is incorporated with the surface atoms and shares their vibratory motion, the frequency f of which can be estimated by equating a quantum of vibrational energy hf with kT at room temperature. Thus, $f = kT/h = 0.025/(6.63 \times 10^{-34} \div 1.6 \times 10^{-19}) = 6 \times 10^{12}$ s^{-1}, which can be taken as 10^{13} s^{-1} to order of magnitude. It is convenient to introduce $\tau_0 = 1/f$. Trapped in a potential well of depth q, the molecule's vibrations at this frequency can be thought of as the number of attempts it makes per second to overcome the potential barrier and break free of the surface. The probability that fluctuations in the energy sharing will result in an energy q or more is given by the Boltzmann factor as $\exp\{-q/kT\}$. The probability per second that a molecule will desorb — say, σ — is, therefore, as originally proposed by Frenkel (1924)

$$\sigma = f \exp\{-q/kT\} = (1/\tau_0) \exp\{-q/kT\} \tag{4.4}$$

For a molecule that becomes adsorbed at $t = 0$, the average time it stays adsorbed, τ_a, may be evaluated as follows. Defining $p(t)$ as the probability that it is still adsorbed after elapsed time t, it follows that $p(t + dt)$, the probability of survival to $t + dt$, must be equal to $p(t) \times [1 - \sigma dt]$, the product of it being adsorbed at t and of not desorbing in the next dt. But because the increment in p can also be expressed as $(dp/dt)dt$, there results $(dp/dt) = -\sigma p$. This integrates to give $p(t) = \exp\{-\sigma t\}$ and leads to an average time of stay

$\tau_a = 1/\sigma$. The analysis, which is about survival before desorption, is essentially the same as that which leads to Equation 3.20 for the mean free path, which is about survival before collision. Thus

$$\tau_a = 1/\sigma = 10^{-13} \exp\{+q/kT\} \tag{4.5}$$

Evidently, this equation makes the sensible predictions that the stronger the binding to a surface and the lower the temperature, the greater is the average time of stay. Variously, τ_a is called the mean stay time, the sojourn time, or the adsorption time. Because of the exponential factor, the variation of τ_a with binding energy is quite dramatic, as Table 4.1 shows for species adsorbed at room temperature. The ratio of binding energy to the thermal energy is critical. For molar rather than molecular specifications of q, kT is, of course, replaced by R_0T.

Notice the enormous range of times, from small fractions of a nanosecond in physisorption for the first two entries, through micro- and millisecond times in weak chemisorption to seconds, days, and years, with progressively stronger chemisorption. But, as well as that of q, the effect of temperature is profound. Consider physisorption at 15 kJ mol^{-1} at room temperature ~300 K. Lowering the surface temperature to that of liquid nitrogen 77 K will give a roughly fourfold increase in q/T and raise the stay time from 5×10^{-11} s to that of the fourth entry in the table, 5.5×10^{-3} s, i.e., increase it by a factor ~10^8. Such is the behavior of thermally activated processes. The strong effect of temperature is also evident in Table 4.2, which shows the stay time at 77, 295, and 600 K for a q value of 80 kJ mol^{-1} (~0.8 eV per molecule) which is approximately the binding energy of water on many surfaces.

While the analysis presented above is oversimplified in a number of aspects, it provides a guide that is essentially correct to the magnitudes of the phenomena. In more sophisticated treatments, absolute reaction rate theory is used, together with the formal language of thermodynamics. Thus, the adsorption process is more properly viewed in terms of changes in the

TABLE 4.1

Dependence of Stay Time on Binding Energy at Room Temperature

q in kJ mol^{-1}	τ_a
6	1.2×10^{-12} s
15	5×10^{-11} s
40	1.2×10^{-6} s
60	5.5×10^{-3} s
80	15 s
90	15 min
95	2 h
100	8×10^4 s ~ 1 d
120	2×10^8 s ~ 10 years
150	7×10^{13} s ~ 20,000 years

TABLE 4.2

Effect of Temperature on Stay Time for $q = 80$ kJmol^{-1}

T K	τ_a
77	10^{41} s !
295	15 s
600	1 μs

free energy $G = H - TS$ between gaseous and adsorbed states, where H and S are the enthalpy and entropy, respectively. In the process of adsorption, entropy decreases, $\Delta S < 0$, because of the reduced disorder of the adsorbed state. For a process to proceed spontaneously, it requires that the free energy change $\Delta G = \Delta H - T\Delta S$ be negative, and so ΔH must be negative, signifying that heat is released on adsorption. A good practical example of this occurs in the action of the cryosorption pump (see Chapter 6) when the release of heat on adsorption is accompanied by the copious boiling off of the chilling agent, liquid nitrogen. In the analysis presented above, the energy q is just ΔH_a for adsorption and numerically equal to ΔH_{des} for the case discussed. The principal deficiency of the treatment is the preexponential factor, frequently symbolized as τ_0, which has the value 10^{-13} s in Equation 4.4. In practice, τ_0 can be extracted from experimental data and in some cases has values greater or less than 10^{-13} s by orders of magnitude. The basic reason for this is the entropy change ΔS between gaseous and adsorbed states, which appears as $\exp\{\Delta S/k\}$ with $\exp\{-\Delta H_{des}/kT\}$ in the rate expression. Although we have no reason to discuss it further, adsorbed molecules may be mobile across the surface to a greater or lesser degree, driven by the same thermal fluctuations that eventually cause desorption and depending on the variation of binding energy between adjacent adsorption sites, which is specific to the gas and surface involved. The barrier to lateral movement across a surface due to these variations of binding energy will, of course, be smaller than the barrier to desorption. Entropy changes between the gaseous state and the adsorbed state may, therefore, be expected to depend on the mobility in the latter, and will vary from case to case. These matters are discussed by de Boer (1969) and Redhead (1968).

Turning to matters of the population of adsorbed molecules, it is appropriate to first work out how many molecules there are in a complete monomolecular layer, or *monolayer* as it is called, with the symbol ML. The unit of surface area in this context is usually taken to be 1 cm^2. Taking a typical molecular diameter to be about 3 Å = 0.3 nm, there will $10^8/3 = 3.3 \times 10^7$ molecules in a line 1 cm long. If we imagine such molecules to be close-packed in a square array, as they might well be if adsorbed on a surface whose atoms formed a square lattice, then the number per cm^2 will be $(3.3 \times 10^7)^2 = 10^{15}$, approximately. Although the exact number of molecules in a complete layer — say, N_0 — will vary a little from case to case, for the

purposes of general discussion, $N_0 = 10^{15}$ is usually taken to be the number of molecules in a monolayer.

Consider now the rate of impact of molecules on one cm^2 of surface due to the adjacent gas at pressure p and temperature T. This is the quantity J of the kinetic theory presented in Equation 3.22 of Chapter 3, and Table 3.1 gives values at various pressures. At $p = 10^{-6}$ mbar, $J = 2.9 \times 10^{14}$. This value is particularly significant because, if all the arriving molecules stick to an adsorbing surface at this pressure (a "worst-case" assumption), then a complete monolayer will be formed in about 3 seconds. The so-called *monolayer formation time* is, therefore, of the order of seconds at a pressure of 10^{-6} mbar (high vacuum). This is a useful benchmark in a number of contexts. In surface science, for example, it indicates that experiments on clean surfaces need to be carried out in ultrahigh vacuum ~10^{-10} mbar, in order that the times for study of freshly prepared surfaces can be extended by factors of 10^4 or more.

To return to the matter of the number of molecules adsorbed, several quantities need to be defined. One is the number adsorbed per unit area, n_a, and arising from it the *fractional coverage*, $\theta = n_a/N_0$. On area 1 cm^2, N_0 would be close to 10^{15}, as described in preceding text. In addition, the *sticking coefficient* $s = s(\theta)$ is defined as the probability that, on striking the surface already having coverage θ, a molecule becomes adsorbed. In terms of Figure 4.3(b), it is the probability that a molecule will reach the deeper state C rather than the precursor state P. We took $s = 1$ in our worst-case assumption for the monolayer formation time but, in general, it is less. Whatever its initial value, one would expect it to change with coverage, becoming less as the number of available sites is reduced. Measurements show that for chemisorption at room temperature, initial sticking coefficients $s(0)$ on fresh surfaces are usually between 0.1 and 1 and that they start to decrease at θ ~0.5, falling to very low values as θ approaches complete coverage.

Provided we restrict ourselves to dilute adsorbed layers, $\theta << 1$, consistently with our earlier assumption for the desorption process, it is possible to create a simple model for the equilibrium state as follows. The rate of adsorption at coverage θ will be

$$\frac{dn_a}{dt} = s(\theta) \times J \tag{4.6}$$

The rate of desorption, according to the definition of σ in Equation 4.4 and the result in Equation 4.5, will be

$$\left(\frac{dn_a}{dt}\right)_{des} = -n_a\sigma = -\frac{n_a}{\tau_a} \tag{4.7}$$

At equilibrium the rates balance and so

$$n_{a,eq} = s_{eq} J\tau_a \tag{4.8}$$

Thus, for example, for nitrogen impinging on a surface at room temperature at 10^{-5} mbar, $J = 2.9 \times 10^{15}$ cm^{-2} s^{-1}. If s_{eq} and τ_a were known to be, say, 0.2 and 5×10^{-3} s, respectively, the equilibrium population would be $n_{a,eq} = 2.9 \times 10^{12}$ per cm^2, or $\theta = 0.003$, a very dilute layer.

Equation 4.8 can be developed further. Expressing $n_{a,eq}$ as $\theta \times N_0$ and substituting for J and τ_a from Equation 3.22 and Equation 4.4, we get

$$\theta N_0 = \frac{p}{\sqrt{2\pi mkT}} s_{eq} \tau_0 \exp\{q/kT\} \tag{4.9}$$

This relationship between θ, p, and T is effectively an equation of state for the adsorbed phase, determining the amount adsorbed at equilibrium in terms of the prevailing temperature and pressure. When regarded, for example, as relating θ and p at constant temperature, it is described as an adsorption isotherm. Subject to the assumptions made about diluteness of the layer, at constant T, Equation 4.9 asserts that coverage is proportional to pressure, an example of Henry's law. Considering physisorbed species, whose adsorption energies are less than 40 kJ mol^{-1}, we can deduce (most directly from Equation 4.8 using the data of Table 3.1 and Table 4.1) that their adsorption in high vacuum at room temperature is negligible. The much greater adsorption at low temperatures is exploited in the cryosorption pump discussed in Chapter 6. The strong temperature dependence of the desorption process can frequently be used to rid contaminated surfaces of trapped gas by heating them in vacuum to a high temperature.

The model just described does not work at high coverages when complications arise due to the interactions between adsorbed molecules so that, for example, heats of desorption depend on coverage. Furthermore, taken out of context, it would predict unlimited adsorption with increase of pressure. In many cases, chemisorption is complete at one monolayer, as previously mentioned, and this is modeled by the Langmuir isotherm, the formulation of which incorporates a constraint that assures completion at a monolayer. Physisorption occurs up to several layers, but only up to a value, as the binding progressively weakens, that the intermolecular binding in the condensed state is not disrupted by fluctuations ~kT at the prevailing surface temperature. The empirically based and widely used BET isotherm models multilayer physical adsorption by assuming that binding in layers beyond the first occurs with energy equal to the heat of condensation.

The simple ideas underlying the model given above are also restricted to what is called first-order desorption which means that, as in Equation 4.7, the desorption rate is directly proportional to n_a. However, for molecules that dissociate on adsorbing, the recombination of constituent atoms, often from a mobile population, will depend on the square of their concentration, and desorption is a second-order process, depending on n_a^2 and leading to different equilibrium equations. These more complicated adsorption

phenomena are dealt with in many texts (see, for example, Hudson [1992] or the article by the same author in Lafferty [1998].)

4.5 Outgassing

Surfaces facing a vacuum release gas. If they did not, practical vacuum technology would be much simpler. The inner wall of a vacuum container would serve simply as a clean and inert barrier to gaseous molecules, and by the selection of a suitable pump all the air and vapor initially contained could be rapidly removed and very high vacuum attained quickly. In practice this is not so. Gas is continuously released, albeit at relatively small rates, but they prove to be a limitation on the attainable vacuum. Even with careful preparatory cleaning of surfaces, the steady release of gas, principally water vapor, limits the degree of vacuum achievable in reasonable times to values ~10^{-6} mbar or a little less, the conventional high vacuum. If vacua substantially lower than this are to be achieved, special procedures to reduce outgassing are necessary. Thus, boundary surfaces play an active role in the attainment of vacuum.

It is molecules desorbing from bound states at vacuum-facing surfaces that constitute the outgassing flux. Although their point of departure into the vacuum space is at the surface, they have various origins. Thus, a surface that has been exposed to normal air at atmospheric pressure for some time will be in equilibrium with it and covered by an adsorbed phase, dominantly of water molecules from atmospheric water vapor, which adsorb more strongly than other atmospheric species. If the pressure of the adjacent gas and vapor to which it has been exposed is now lowered by the action of pumping, this equilibrium is upset and molecules desorb into the vacuum space. Exposure to atmosphere will also have caused diffusion of gases into the near-surface region of the bulk material, assisted perhaps by the presence of rough, microcracked surface structure. Metals frequently have an oxidized surface layer, the passivation layer, typically a few nanometers thick, that absorbs gases by their inward diffusion following initial adsorption. From this near-surface region it will tend to diffuse back to the surface and desorb under low-pressure conditions.

The material of the boundary wall of the vacuum, typically stainless steel and several millimeters or more in thickness, will also contain gas that was trapped inside it at the time of its manufacture. Under certain conditions its slow diffusion to the vacuum interface and desorption can be significant. Dissolved hydrogen, occurring as interstitial H atoms rather than H_2, is present in this way and is particularly important. Also, and at what one may suppose to be a very small rate, gas adsorbed on the outer wall of the container may permeate the wall to arrive by similar diffusive processes at the inner wall, prior to desorption. Permeation through the relatively small

areas associated with elastomer seals that join metallic components can be a significant contribution to outgassing even when, as is usually the case, permeation through bulk walls is not. Thus, the desorption of gas, its diffusion out of near-surface regions and the bulk, and its permeation through bulk solid material are matters of concern.

Although the phenomena of adsorption, diffusion, and permeation in their own right are each quite well understood, having been extensively investigated for many years in a variety of industrial and scientific contexts, their combination in the outgassing process is quite complex. The search for means of reducing outgassing, via a fuller understanding and control of it, remains a matter of active investigation. We will return to the matter of modeling of detailed physical processes after describing the principal experimental findings. Measurements show that under the action of pumping, the outgassing of surfaces diminishes with time approximately according to the relationship

$$q_G = \frac{q_1}{t^\alpha} \tag{4.10}$$

In this equation, q_G is the rate of release of gas per cm^2, referred to as the specific gassing rate, and q_1 its value after 1 h of pumping. The quantity t is the dimensionless ratio of the time in hours to 1 h, and α is an exponent that is frequently about 1 for gassing from metallic surfaces, and nearer 0.5 for elastomers. This dependence is followed for tens of hours and then evolves exponentially into a sensibly constant value. The quantity of gas q_G is expressed in terms of pressure-volume units that are fully discussed in the next chapter. A frequently used unit of q_G is the mbar × liter/second/cm^2. Typical values of q_1 for well-precleaned metal are about 10^{-7} mbar l s^{-1}/cm^2. Empirical data available on a variety of surfaces, subjected to various treatments and a selection, are given, for example, in O'Hanlon's text (2003), which is a good introduction to this subject, as are the classic papers of Elsey (1975, (a) and (b)). Berman (1996) discusses the role of water vapor in vacuum systems. Figure 4.4 shows the general pattern of behavior with outgassing rates and time plotted on logarithmic scales, and plots of slope approximately –1.

Based on the fact that the molar volume of any gas contains $N_A = 6 \times 10^{23}$ molecules and occupies 22.4 l at STP, it may be shown that 1 mbar liter at room temperature contains 2.46×10^{19} molecules. Thus an outgassing rate of 10^{-7} mbar l s^{-1}/cm^2 implies the release of about 10^{12} molecules in one second, a substantial quantity, and as we have seen in Section 4.4, this corresponds to about one thousandth of a monomolecular layer per second of released gas, principally water vapor. The outgassing rates referred to in Equation 4.10 are "free" or "intrinsic" rates in the sense that the measurements made to determine them have been corrected to allow for not all desorbing gas being recorded; some eludes measurement by readsorbing on system surfaces. Such concerns also arise when using the data in the design

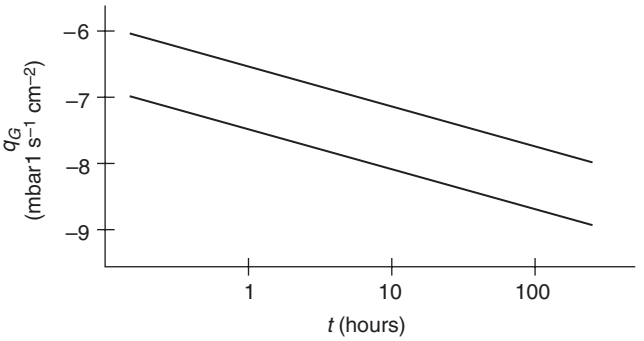

FIGURE 4.4
Typical outgassing rates as a function of time.

of systems, when allowance should be made for not all desorbing molecules reaching the pump directly.

As noted in Section 4.4, rates of desorption depend sensitively on the strength of the bond that attaches a molecule to the surface, and to the temperature. For molecules already on the surface that do not have to reach it from the interior, it is possible to make some order of magnitude estimates of the outgassing process. Because the binding energy of water molecules in the outermost layers of the multilayered coverage that exists after atmospheric exposure will be of the order of the condensation energy, namely, 41 kJ mol^{-1}, one can deduce from Table 4.1 that the stay time will be very short, about 10^{-6} s, so that their release is effectively immediate once the equilibrium is disturbed. However, after these weakly bound layers have desorbed, there remain the water molecules that were underneath them, attached to surfaces in more strongly bound chemisorbed states. The energies of their binding to typical surfaces are in the range of 85–90 kJ mol^{-1} and, correspondingly, stay times are of the order of minutes to hours. Thus, keeping in mind that only a fraction of the desorbing molecules are pumped away, it is evident that water vapor in particular, and other species with binding energies of this order, will remain in the vacuum environment for extended periods.

To analyze the desorption of a dilute monolayer of gas that is either physically adsorbed or chemisorbed in an undissociated form, so that first-order kinetics apply, Equation 4.7 may be used, together with Equation 4.4. This gives the desorption rate as

$$\left(\frac{dn_d}{dt}\right) = -\left(\frac{dn_a}{dt}\right)_{des} = n_a\sigma = 10^{13} n_a \exp\left(-E/kT\right) \tag{4.11}$$

showing, as one might expect, that the desorption rate is proportional to the number of adsorbed molecules remaining. Integration of Equation 4.7 in the form

$$\left(\frac{dn_a}{dt}\right)_{des} = -\frac{n_a}{\tau_a} \tag{4.12}$$

shows that the population n_a decays exponentially from the initial state with a time constant equal to the stay time τ_a. In a notable contribution to this subject, Hobson (1961) modeled the fall of pressure at room temperature in an idealized ultrahigh vacuum system that consisted of a volume of 1 l and vacuum-facing area of 100 cm^2, with the only gas source being an initially complete monomolecular layer of specified binding energy adsorbed at the wall, assumed ideal and impenetrable. Gas was pumped away at a rate of 1 l per second (see Chapter 5), and the sticking coefficient s for molecules which desorbed to another part of the wall rather than into the pump was taken to be 0.5. The results for pressure as a function of time are reproduced in Figure 4.5 with the heat of adsorption as parameter.

Evidently, gas that is adsorbed relatively weakly, with q less than 60 kJ mol^{-1}, is quickly pumped away consistent with the discussion above. Strongly absorbed gas, with adsorption energy 100 kJ mol^{-1} or greater, desorbs at such a slow rate that its contribution to the pressure, though prolonged, is negligibly small. But gas adsorbed with q between these values, and particularly at about 80 kJ mol^{-1}, has a sizeable contribution and is long lasting. This is approximately the energy with which the water molecule binds to many surfaces.

Because the desorption of gases is a thermally activated process whose rate increases dramatically with temperature, as reflected in Equation 4.11,

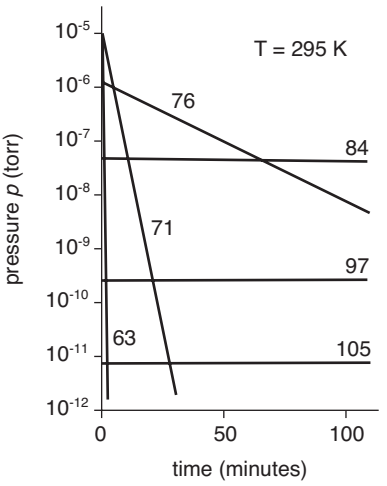

FIGURE 4.5
Pressure vs. time in a modeled system. Energies in kJ mol^{-1}. (Adapted from Hobson, J.P., *Trans. 8th Natl. Vac. Symp.*, Pergamon Press, New York, 1961. With permission.)

it is possible by raising surface temperatures to greatly accelerate the desorption process and, indeed, other thermally activated processes such as diffusion that occurs within the bulk of the vacuum wall. This process is called "baking" and consists of heating vacuum systems, walls, and contents, to temperatures of 150–250°C for times of order 10 h, with continual pumping. Surface adsorbed gas is removed at a much greater rate than at room temperature, and gas absorbed in the bulk diffuses at a much greater rate to the interior surface where it desorbs and is pumped away. Both the surface and its hinterland in the vacuum wall are therefore depleted of gas molecules, and on cooling back to room temperature the outgassing rate is dramatically reduced by factors ~10^3 or more, consistent with the attainment of ultrahigh vacuum.

Diffusive processes that contribute to outgassing are governed by Fick's two laws of diffusion, which for diffusion in one dimension are:

$$j = -D\left(\frac{dc}{dx}\right) \quad \text{and} \quad \frac{dc}{dt} = D\left(\frac{d^2c}{dx^2}\right) \tag{4.13}$$

The first describes diffusion in the steady state in direction x and relates the flow of molecules as a number j per unit area perpendicular to x per second to how their concentration c, expressed as a number per unit volume, varies with x. The rate of flow is proportional to the concentration gradient $\mathrm{d}c/\mathrm{d}x$, with flow in the direction of the decreasing concentration; the random migratory process favors the direction of flow in which more empty sites are available. The constant of proportionality defines the diffusion coefficient D. The second law describes conditions of change before a steady state is reached. The coefficient D has units $\mathrm{m}^2\ \mathrm{s}^{-1}$ and, because it relates to a thermally activated process, has a temperature-dependent form so that $D = D_0 \exp\{-E_D/R_0T\}$ where E_D is a molar activation energy. Because diffusion coefficients are very small, steady-state conditions are seldom achieved, and analyses involve solving the second equation with appropriate boundary conditions as discussed, for example, in Hudson (1998). Thus, the diffusion of adsorbed atmospheric water vapor into and out of the near surface region can be modeled, and the slow diffusion of gas initially uniformly distributed through the bulk of the vacuum wall. Dayton (1962) showed that the observed $1/t$ time dependence of outgassing could be accounted for as an aggregate effect of gassing from a distribution of pore sizes with specific binding energies in the near-surface layer. In a series of papers, Li and Dylla (1993, 1994, 1995) have measured and modeled the outgassing of water vapor from stainless steel. This work gives much detailed information about the processes at work and confirms the important role of moisture content of the venting gas when surfaces are reexposed to atmosphere in determining subsequent outgassing behavior in vacuum. Water vapor dominates the

outgassing, other gases such as hydrogen, methane, carbon monoxide, and carbon dioxide contributing only about 10% of the total. The amounts of water outgassed from surfaces exposed to atmospheres of typical moisture content are of the order of 100 monolayers equivalent, and the exponent α is approximately 1. When venting to very dry nitrogen, much less water is absorbed, and the subsequent pumpdown is comparatively rapid. The observed α values approach 0.5 so that the outgassing diminishes more slowly, but the magnitudes q_1 are less by more than an order of magnitude.

The analysis that deals with diffusion out of the bulk from initially uniform concentrations is applicable to the out-diffusion of hydrogen dissolved in the chamber wall and to diffusion from slabs of material inside the vacuum. There is increasing evidence (Fremerey, 1999) that hydrogen is the only gas that diffuses out of stainless steel, and that other surface desorbed gases in typical ultrahigh vacua, such as CO, CH_4, and CO_2, are artifacts caused by the action of hot filaments. How concentrations $c(x, t)$ vary due to diffusion from a semi-infinite solid with vacuum on one side and a finite solid with vacuum on both sides are shown in Figure 4.6(a) and (b). They are analogous, respectively, to the temperature distributions inside a solid when one face of an extended hot solid initially at uniform temperature throughout is suddenly clamped at a lower temperature and when a finite solid has both faces so clamped. Fourier's laws of heat conduction have the same form as Fick's laws for diffusion; heat flows in response to temperature differences just as particles migrate in response to concentration differences.

When, as is frequently the case, the desorption of molecules that diffuse out of the solid is determined by their rate of arrival at the surface rather than by the rate of the desorption process, i.e., when diffusion is the rate-limiting step, analysis shows that the outgassing rate, determined by the value of (dc/dx) at the surface, and the initially uniform concentration C_0 is given in the early stages by

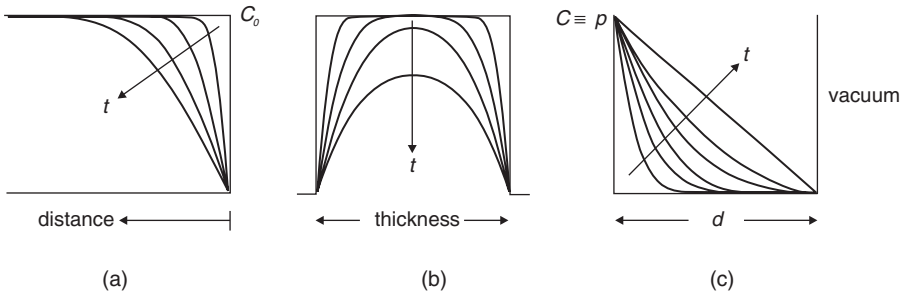

FIGURE 4.6
Evolution of absorbed gas profiles with time.

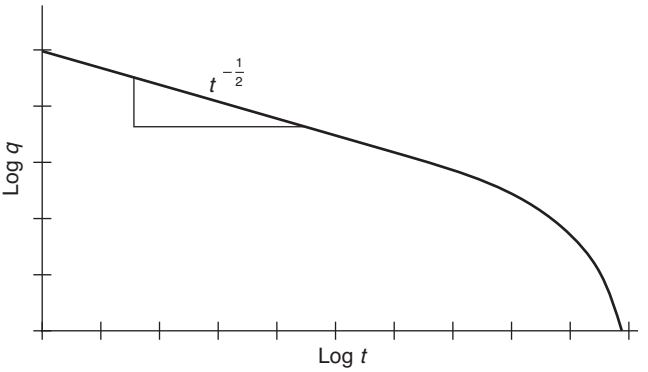

FIGURE 4.7
Outgassing rate as a function of time.

$$q = C_0\left(\frac{D}{t}\right)^{1/2} \tag{4.14}$$

This $1/t^{1/2}$ dependence changes at later times into a faster exponential fall $\exp(-\alpha Dt)$ as the quantity of dissolved gas becomes substantially depleted. The form of this variation, which is discussed in detail by Holkeboer et al. (1993) is shown in Figure 4.7 using logarithmic scales. Initially, the slope is –1/2, changing eventually to a steeper linear portion. It describes the hydrogen outgassing of well-baked stainless steel at long times and after all near-surface adsorbed gases have been pumped away.

Permeation is the diffusive flow of gas through a vacuum wall of thickness t driven by the pressure difference P across it. For monatomic gases and those that do not dissociate on diffusing into the solid, the outgassing flux is described by the equation (O'Hanlon, 2003)

$$q_{perm} = \frac{KP}{t} \tag{4.15}$$

where K is the permeation constant with unit $m^2\ s^{-1}$, q_{perm} expressed in Pa $m\ s^{-1}$, P in Pa, and t in m. Weston (1985) gives a good introduction to these matters. Figure 4.6(c) shows steady-state permeation and the time-dependent changes in concentration that precede it.

The broad relevance of the above to vacuum practice is as follows. It should be stressed that in typical systems the condition of interior surfaces is usually not well known because of the variety and complexity of the processes that occur in the course of operating the system. When vacuum systems are vented back up to atmospheric pressure, direct adsorption from the gas phase leads to the adsorption and absorption into the near-surface of atmospheric gases, principally water vapor in relatively large amounts. It is prudent to try to minimize water vapor adsorption by venting to dry nitrogen.

In pumping down from atmospheric pressure, most of the free gas in the pumped volume is soon removed, and in typical systems, sub-millibar pressures of 10^{-2} mbar or better are achieved in times of order of minutes. Desorbing gas starts to contribute to the gas load below about 10^{-1} mbar, and as the pressure continues to fall into the region below 10^{-4} mbar the gas load (assuming no leaks) becomes increasingly due to outgassing. Of the molecules that desorb, a number will find the entrance to the pump and be removed immediately. But others, a majority in vacuum chambers of typical proportions, will traverse the chamber to another part of the vacuum wall and further surface interactions before being pumped away. It is this traffic of molecules in transit to and fro across the chamber, fed by desorption and diminished by pumping, that constitutes the number density n of molecules in the vacuum and determines the pressure achieved.

Gas that is loosely bound on internal surfaces is pumped away quickly. Gas that is tightly bound desorbs at a very slow rate and does not contribute a significant load. But water vapor, which has an appreciable probability of desorption, has been stored in large amounts during exposure to the atmosphere. The result is that there is protracted gassing of water vapor from structural materials such as stainless steel and glass. The outgassing rate does decrease, but only slowly, and desorbing water vapor accounts for the dominant gas load for times of order tens of hours, unless bakeout procedures are used. In circumstances in which pump performance did not limit the vacuum achieved, outgassing would be observed to diminish as time progressed, with the contribution to it of the small amount of diffusion of hydrogen from the interior becoming more important, and eventually, after very long times, becoming the dominant contribution. Beyond that, if all the gas stored initially in the wall were removed by diffusion, the ultimate source of gas would be permeation.

In practicable experimental times, therefore, which are of order hours in typical laboratory vacuum systems, the main source of gas from unbaked surfaces is surface-desorbed gas, principally water vapor. Typical values of q_G for well-precleaned surfaces of stainless steel, glass, and ceramic, after a few hours pumping, are of the order of 10^{-7} mbar l s^{-1} per cm^2. In high vacuum practice, permeation of atmospheric gases through elastomer gaskets may, if there are many of them, contribute significantly to the ultimate gas load. Permeation through vessel walls is not significant.

If UHV conditions p ~10^{-10} mbar and less are to be achieved, outgassing has to be reduced by factors of 10^3 or more below the values that are achieved after a few hours' pumping, and which are characteristic and limiting in the high-vacuum regime. This is done, once the pressure has reached the HV regime, by baking. Surface adsorbed gas is removed at a much greater rate than at room temperature, and gas absorbed in the bulk diffuses at a much greater rate to the interior surface where it desorbs and is pumped away. Both the surface and its hinterland in the vacuum wall are, therefore, depleted of gas molecules, and on cooling back to room temperature, the outgassing rates are dramatically reduced. Water vapor is almost completely

removed and the composition of the outgassing from stainless steel is dominantly hydrogen. In UHV systems where pressures of 10^{-11} mbar or less are achieved, permeation of atmospheric hydrogen and helium through the vacuum wall at room temperature becomes significant, even though the concentration of these gases in the atmosphere is minute, of the order a few parts per million (ppm). Hydrogen permeates stainless steel and helium permeates glass. The permeation of other gases is insignificant.

4.6 Molecular Drag

As was discussed in Section 4.2, when gas molecules impinge on a stationary surface, they do not usually rebound elastically but, rather, if the surface is rough on a microscopic scale, they depart in a random direction as suggested in Figure 4.1(c). On average, therefore, their direction of departure is perpendicular to the surface. Molecules arriving in a given direction bring momentum with a component parallel to the surface mv^{pll}, but when they leave, because they depart on average in the perpendicular direction, they take away no parallel momentum. Therefore, in effect, the momentum mv^{pll} of an incoming molecule is destroyed, and during the brief time in which it is brought to rest, a force is exerted on the surface in the direction of the incoming motion. Because molecules arrive from all directions, there is no net force in a direction parallel to the surface, in contrast with the pressure force due to reversal of perpendicular components.

Consider now the situation in which the surface on which molecules impinge is not stationary, and the gas is rarefied so that conditions are molecular as described in Section 3.9. When surfaces move at speeds of order molecular speeds, appreciable molecular drag effects are observed, whose origin may be understood as follows. Consider the hypothetical device shown in Figure 4.8, which we imagine as part of a "thought experiment." A continuous belt moves in vacuum at constant speed U comparable with molecular speeds, guided on power-driven rollers that maintain the motion. The mechanical arrangement resembles that of a belt sander. Consider the molecules in the upper region between the top surface of the belt and a stationary plate. We suppose that the separation of the surfaces is less than the mean free path so that molecules leaving the belt travel to the plate without being scattered by other molecules, i.e., Kn > 1 and local conditions are molecular.

Consider the arrival of a gas molecule at the moving belt, event (1) in Figure 4.8. All arrival directions are equally probable and so the average direction of arrival is perpendicular to the belt as indicated. In becoming attached to the belt, the average molecule experiences a "sideways" force to the right necessary to accelerate it to the belt velocity U. An equal but

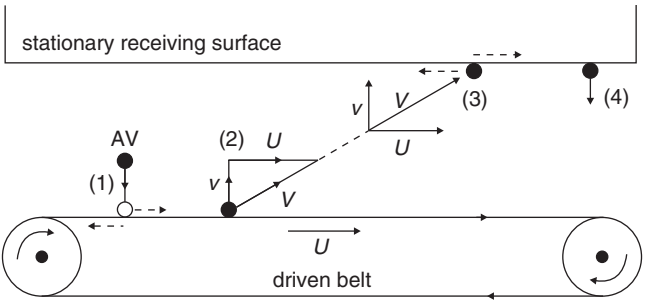

FIGURE 4.8
Device to illustrate molecular drag. (From Chambers, A., Fitch, R.K., and Halliday, B.S., *Basic Vacuum Technology*, 2nd ed., Institute of Physics Publishing, Bristol, 1998. With permission.)

opposite slowing force is exerted on the belt by the molecule so that the motor has to do the work necessary to maintain the belt speed.

At a later time, event (2), this average molecule will depart from the belt surface, again on average in a perpendicular direction with velocity v with respect to the belt, but in addition with a sideways velocity U imposed by the belt. The total velocity will be V, compounded of v and U, as shown. The belt thus drags molecules in the direction of its motion, imposing an orderly component to the right on the otherwise haphazard molecular motion.

The average molecule next arrives (3) at the upper stationary surface, bringing an imposed parallel momentum mU to it and, in coming to rest, gives up this momentum, creating a force that tends to move (drag) the plate to the right. Because, on average, the direction of departure from the stationary plate will be perpendicular to it, the total change of parallel momentum is just mU.

What happens to the average molecule represents what happens to all molecules impinging on the fast-moving surface. There is a cycle of arrival onto the belt, right-biased departure from it, followed by momentum loss to the upper stationary surface, then downward perpendicular departure from the stationary plate, continually repeated. The overall effect, therefore, is that of dragging molecules that are incident on the belt towards the right. This effect is exploited in the "molecular drag" stages of certain pumps and is discussed in Chapter 6.

Thus, three distinct aspects of molecular drag phenomena may be identified:

1. The imposed bodily movement of gas discussed above
2. The tendency of stationary object, in this example the upper plate, to be dragged in the direction of the biased motion
3. The slowing effect on fast-moving surfaces due to the arrival of random molecular flux onto them, exemplified in this case by the retarding force on the belt in event (1).

This last effect is exploited in one type of pressure gauge, the spinning rotor gauge, discussed in Chapter 7.

Molecular drag analyses are relatively straightforward for gas in a molecular state, as will be seen in the examples of Chapter 6 and Chapter 7. As might be expected, their magnitude is directly proportional to the number density n, and therefore to pressure. However, at higher pressures, in transition to the continuum state where the forces due to relative motion are described as viscous and become independent of pressure, matters become more complicated. Their basis is well discussed in Walton (1983).

4.7 Sputtering

In earlier sections of this chapter we have considered the scattering or trapping that may occur when molecules with typical thermal energies are incident on a surface. Under these conditions, there is no possibility that the surface atoms themselves might be so disturbed by the event as to be ejected into the gas phase. But for energetic incident particles, usually ions, with energies from a few tens of eV and upwards to keV and more, there is such ejection of surface atoms or molecules, and the process is called *sputtering*. It is used in ion pumps and in surface cleaning, as will be described later. It is also the basis of important analytical techniques in surface science, but dominantly, and in highly developed form, devices that exploit sputtering lie at the heart of the huge worldwide activities of vacuum coating and semiconductor fabrication.

The impact of an energetic ion on a surface and its stopping in the subsurface destroys its momentum and deposits much energy rapidly into a small region of connected atoms. We will restrict our attention to the energy range up to a few keV, which is of most interest for vacuum technology. In the primary collision with a surface atom, the ion loses a fraction of its incoming energy that is rapidly further distributed in a developing cascade of subsequent collisions. It is quickly neutralized by acquiring an electron and becomes buried below the surface. The escape of a target atom to the vacuum requires, of course, the acquisition of outwardly directed momentum by a suitable sequence of collisions among the disrupted target atoms, as well as sufficient energy, ~eV, to break the bond. If the initial collision of an incoming ion in a direction perpendicular to the surface with a surface atom is regarded as an isolated head-on two-body collision (isolated so as to conserve energy and momentum), then simple mechanics shows that if the ion is lighter than the struck atom, it rebounds back to the vacuum with reduced energy while the struck atom moves towards the interior with the remainder. However, if the ion is heavier than the struck atom, both move towards the interior to give greater energy deposition into the solid. Thus, as one might intuitively expect, the sputtering yield, defined as the number

of sputtered atoms per incident ion, is greater, the heavier the incident ion, and this is borne out in practice. Yields with argon ions, for example, are significantly greater than those for helium.

The sputtered species are dominantly neutral atoms, but in general may also include some ions, ionized clusters, and secondary electrons. For normal incidence, the spatial distribution of sputtered material roughly follows the cosine law, with energies of order tens of eV and a large energy spread. Although the effects of the primary collision cause disturbances for some distance into the solid, the sputtered particles come chiefly from the surface layer of atoms. The surface is thereby eroded and also roughened. As sputtering continues, the now embedded and neutralized incoming particles may themselves, after a while, be sputtered away. For incidence at oblique angles up to about 60°, sputtering is peaked in the forward direction and yields are considerably greater, by a factor of up to 10. This may be understood in terms of the beam-induced cascade of collisional excitation staying closer to the surface and the consequent greater availability of surface atoms for acquiring outward momentum.

Sputtering is one of the most complex of surface processes. There is no complete theory of it. Shah (1995) gives a general introduction to the subject, and Carter and Colligon (1968) and Wehner and Anderson (1970) discuss the basic nature of the phenomenon in an informative way. Theoretical aspects are discussed in Kaminsky (1965). The diverse possibilities for the multiple pre- and post-collision trajectories are best modeled by Monte Carlo simulations.

In Figure 4.9 the salient experimental features of the phenomenon are illustrated for argon ions, which find frequent application and may be generated in plasma discharge or as ion beams. The sputtering yield from copper, for example, is zero below a threshold energy of about 25 eV, but then increases linearly with energy reaching a values of unity at 200 eV and 5 at 1.2 keV approximately. At 5 keV it becomes roughly constant, with a value of about 6 in the range of 10 to 60 keV, before starting to fall as processes become more remote from the surface. Yields vary widely depending on the material and the identity and energy of the bombarding ion. Stuart's (1983) text contains various useful tables and graphs and the following useful benchmark. Because a yield of unity means that one atom is sputtered for each incident ion, and because one amp of current corresponds to $1/1.6 \times 10^{-19} = 6.25 \times 10^{18}$ charges, then to sputter away one monolayer of surface atoms (taken to be 10^{15} per cm^2 from earlier discussion) requires a current density of 0.16 mA/cm^2 at the energy for unity yield.

Sputtering of titanium by ions of the residual gas is the means for continuously maintaining a nascent chemically active "gettering" surface in ion pumps (Chapter 6). Glow discharge cleaning, in which a discharge in argon at ~10^{-3} mbar is set up between an internal anode at a potential ~1 kV and the gas-covered vacuum chamber walls as the grounded cathode, enables adsorbed gas to be removed by light sputtering, and is widely used. In the context of preparing clean surfaces for surface science investigations in

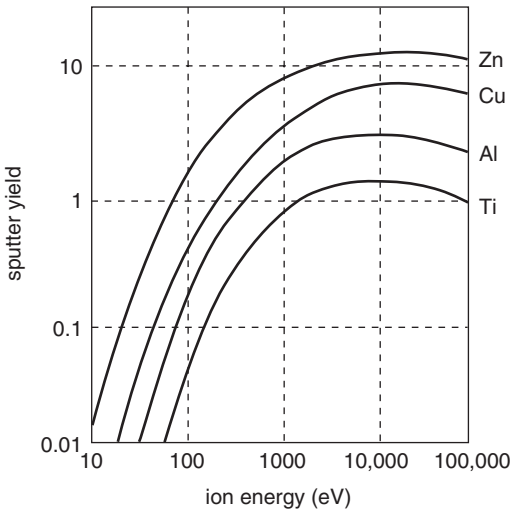

FIGURE 4.9
Sputtering yields for Ar^+ ions as a function of energy for various materials. (From Hoffman, D.M., Singh, B., and Thomas, J.A. III, *Handbook of Vacuum Science and Technology,* Academic Press, San Diego, 1998. With permission.)

ultrahigh vacuum procedures may be developed for specific materials for inert gas ion bombardment to sputter away contaminants and surface atoms, alternating with annealing periods at temperatures close to that of melting to heal the damage. In surface analysis, sputtering at a controlled rate is exploited for depth profiling, and the analysis of the sputtered yield under appropriate ion bombardment is the basis of the very important technique of secondary ion mass spectrometry, used to determine, albeit destructively, the composition of surfaces. The text of Woodruff and Delchar (1986) describes techniques for surface analysis.

In the semiconductor fabrication and vacuum coating industries referred to earlier, a central process is the deposition of materials as thin films onto suitable substrates. This may be accomplished by various means, which are described in Chapter 9. They include chemical vapor deposition that exploits chemical reactions in flowing gases at pressures ~1 mbar, and thermal evaporation in which materials evaporated from source crucibles or as molecular beams from Knudsen cells traverse the vacuum to condense on substrates. But the most widely used techniques exploit the sputtering process, in a variety of source configurations, which include the glow discharge, magnetron, and ion-beam sputtering. A good introduction to this subject will be found in the book of Maissel and Francombe (1991).

Exercises

4.1 If a monolayer of gas forms on a surface in 3 seconds at 10^{-6} mbar, how long would it take to form at 5×10^{-8} mbar?

4.2 A cubic vessel of side 10 cm contains a perfect vacuum and its internal walls are covered by a complete monolayer of gas molecules that suddenly desorb and stay in the vacuum space. How many desorb? What pressure will they create in the volume, assumed to be at 295 K?

4.3 For nitrogen gas at 295 K, the expression for the impingement rate of molecules on a surface, from Table 3.1, is $J = 2.9 \times 10^{20}\ p$ per cm^2 per second, where the pressure p is in mbar. Reexpress this so that the units of J are per m^2 per second, with pressure in Pa.

4.4 Verify that the specification of energies as "eV per molecule" is equivalent to 96.5 kJ per mole and that the value of kT at room temperature is 0.025 = 1/40 eV.

4.5 For molecules that adsorb on a surface with energy 1.0 eV, estimate the average stay time at (a) 295 K, (b) 285 K, (c) 305 K.

4.6 Molecules whose average stay time on a surface is 1 ms impinge on it with sticking coefficient 0.1 at a rate 10^{17} cm^{-2} s^{-1}. What is the equilibrium population of adsorbed molecules per cm^2? To what fraction of a monolayer does this correspond?

4.7 Molecules of helium and carbon dioxide effuse through an aperture into a very high vacuum from an enclosure in which they originally exert equal partial pressures. What is the ratio of the helium and carbon dioxide effusion rates initially?

Problems

4.8 The molar adsorption energy of CO_2 on carbon is approximately 34 kJ/mol. Calculate the mean stay time of CO_2 on carbon at (a) 295 K, (b) 90 K, assuming a preexponential factor 10^{-13} s.

4.9 Estimate the total area of surface in the "sponge-like" structure of a 1-cm cube of porous molecular sieve with a 5-nm pore diameter by modeling it as a set of spherical voids packed in a cubic arrangement (i.e., in a simple cubic structure). Note the length of a roll of carpet 1 m wide of the same area. Assuming that this material presents 10^{14}

adsorption sites per cm^2, estimate the maximum number of molecules that might be captured by it. To how many cm^3 of gas at room temperature and atmospheric pressure would this correspond? (Relevance: the cryogenic sorption pump, Section 6.4.1.)

4.10 With reference to Equation 4.3 and the associated text, prove by the use of solid angles and integration that the total flux I_T per second from unit area is, as stated, $\pi \times I(0)$. Hence, deduce that, for effusion through an aperture in a very thin plate from an enclosure at temperature T containing a vapor of molecules of mass m at pressure p, the flux from unit area per second into unit solid angle at angle θ is $I(\theta) = (A/\pi)(p/\sqrt{2\pi mkT}) \cos\theta$.

4.11 Use a result of the analysis carried out in Question 4.10 to show that for molecules that are scattered/desorbed/evaporated from a small area of surface according to the Knudsen cosine law, the fraction within a cone of semi-angle θ made with the surface normal is $^1/_2[1 - \cos 2\theta]$. Hence, deduce that, within cones of semiangle 30, 45, and 80° the fractions are 25, 50, and 97%, respectively.

4.12 A gas initially at a pressure p_0 in an enclosure of volume V effuses into high vacuum through an aperture of area A in a very thin part of its enclosing wall. If the mean thermal velocity of its molecules is $\bar{v}$, prove that the pressure in the enclosure will fall according to the equation $p = p_0 \exp\{-(\bar{v}A/4V)t\}$. Evaluate the time-constant of this process for hydrogen molecules ($\bar{v} = 1766$ ms^{-1} at room temperature) effusing from a volume of 1 l through a square aperture of side 0.1 mm. Also estimate the pressure below which the mean free path in hydrogen will be greater than 1 mm, thus making effusion rather than pressure-driven fluid streaming the mechanism of flow.

4.13 Show that the energy per unit area per second brought to (and taken away from) a surface at temperature T by the molecules that impinge on it from a gas with which it is in equilibrium is $(^1/_2)\, n\bar{v}kT$ and hence that the average energy per incident molecule is $2\,kT$.

4.14 The permeation constant K of helium through Viton at room temperature is about 10^{-11} m^2s^{-1}. Use Equation 4.15 to estimate the flow of helium into high vacuum through a large circular Viton sealing ring with circumference 1256 mm, assuming that it is squashed to a section 4 mm thick and depth 6 mm in the direction of flow. The partial pressure of helium in atmospheric air is 5×10^{-3} mbar. (Units of flow are discussed in Section 5.2.)

5

Gas Flow and Pumping

5.1 Introduction: Flow Regimes

How does the flow of gas in vacuum systems compare with familiar examples of fluid flow in the world around us? At pressures sufficiently low that the molecular mean free path is comparable with or exceeds the size of the equipment, that is for Kn values >1, it is totally different, as we shall discover in detail in due course. But for vacua not so rarefied, in which mean free paths are such that gases still demonstrate fluid behavior (Kn < 0.01), the concepts and measures developed to describe fluid flow at atmospheric pressure remain appropriate.

To fix ideas we can imagine gas flow through a pipe of diameter D. This will be its characteristic dimension to be used in evaluating the Knudsen number as $\mathrm{Kn} = \lambda/D$ and will determine what is called the flow *regime*. As earlier introduced in Section 3.9 to describe the condition of static gas, flow regimes are defined by

$$\mathrm{Kn} < 0.01 \quad \text{continuum flow regime}$$

$$\mathrm{Kn} > 1 \quad \text{molecular flow regime}$$

whereas for $0.01 < \mathrm{Kn} < 1$ the flow regime is described as transitional. These defining values are not as sharp as is implied, but their general correctness is founded in experimental results, particularly those involving viscous effects. Flow has distinct characteristics, to be discussed, in each regime. Flow in the transitional regime is difficult to analyse.

When considering flow through pipes of diameter D, it is useful to be able to determine the flow regime directly in terms of the prevailing pressure. Because for air $\lambda = 64$ mm at $p = 10^{-3}$ mbar, so that $\lambda p = 0.064$ mm mbar, the criteria for continuum flow and molecular flow become

$$pD > 6.4 \text{ mbar mm} \ldots\ldots\ldots\ldots \text{ continuum flow}$$

$$pD < 0.064 \text{ mbar mm} \ldots\ldots\ldots\ldots \text{ molecular flow} \qquad (5.1)$$

Vacuum systems of typical ~0.5 m dimension frequently operate at pressures of 10^{-5} mbar and less so that conditions are molecular; even at 10^{-4} mbar, λ has increased to be 0.64 m. High vacuum conditions are, of course, reached by initial pumping from atmospheric pressure through the continuum regime. And even in the ultimate steady-state, gas that flows out of the main high vacuum vessel in the course of pumping is subsequently compressed so as to reach a continuum state prior to being expelled at atmospheric pressure. Furthermore, there is an increasing number of applications, particularly in the semiconductor industry, in which gas flow takes place in low vacua ~1 mbar or more with mean free paths less than a millimeter. For these various reasons, therefore, it is appropriate in this chapter to consider continuum as well as molecular flow.

5.2 Measures of Flow: Throughput and Pumping Speed

Before considering details of flow in the different regimes, it is necessary to set up measures of flow. These will be applicable in all regimes. Gases, of course, are compressible, in contrast to liquids that, to a very good approximation, are not. Thus, gas will expand in flowing from a higher to a lower pressure through a pipe. The magnitude of the effect depends on conditions and may be quite small. Figure 5.1 depicts a pipe that connects two large volumes and through which gas flows at a steady rate and at constant temperature. In the volume at the left, the pressure is taken to be higher with a value p_U, where the subscript U signifies upstream and flow is from left to right. In the other volume, the downstream pressure is p_D. In the pipe at a cross-sectional plane 1 near the entrance, the pressure is p_1; at plane 2 further downstream, it is p_2.

The mass of gas flowing per second through plane 1 at pressure p_1 would have an associated volume V_1 at that pressure. Downstream at plane 2 and the lower pressure p_2, the associated volume V_2 would be larger. Under

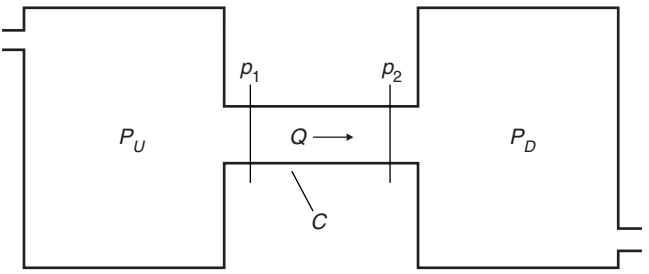

FIGURE 5.1
Flow of gas through a pipe.

conditions of steady isothermal flow, and assuming that the gas behaves ideally, $p_1V_1 = p_2V_2$. Denoting the volumes per second as *volumetric flow rates* $\dot{V}$ at the associated pressures, this becomes $p_1 \dot{V}_1 = p_2 \dot{V}_2$. As was discussed in Section 2.1, under static isothermal conditions the product $p \times V$ gives a simple and direct measure of the quantity of a gas. Here, similarly, the product $p \times \dot{V}$ of pressure at any cross-section multiplied by the volumetric flow rate there, which is called the throughput Q, gives a straightforward measure of the rate at which gas flows. Thus, defining throughput,

$$Q = p \times \dot{V} \tag{5.2}$$

For steady flow, Q is continuous, i.e., it has the same value at every position along the pipe, reflecting the conservation of mass. In particular, $Q_{in} = Q_{out}$ — as much gas leaves the pipe downstream as enters it upstream.

The unit of Q depends on the base units used. In the SI system it is the Pascal meter^3 per second (Pa m^3 s^{-1}). The more practical unit, widely accepted in Europe, and used henceforth in the text, is the millibar liter per second (mbar l s^{-1}). Throughput is an easily assembled and manipulated measure of flow and is extensively used. In the rare cases of flow that is not isothermal, it is frequently useful in a modified way, as will be noted.

When mass flow rates need to be specified directly in units of kg per second, conversions are easily made. Let $\dot{W}$ be the mass flow rate of gas in kg s^{-1}. We saw in Section 2.1 that the mass W of a gas may be expressed as $W = n_M \times M$, the product of the number of moles and the molar mass. Now $n_M = pV/R_0T$, and the flow rate of in moles may be denoted as $\dot{n}_M$ moles per second. At a particular plane of measurement where the pressure is p, this will become $\dot{n}_M = p\,\dot{V}/R_0T = Q/R_0T$ from Equation 5.2. Thus

$$Q = \dot{n}_M \times R_0T \tag{5.3}$$

and because $\dot{W} = \dot{n}_M \times M$, we have

$$\dot{W} = \frac{M}{R_0T} \times Q \tag{5.4}$$

It is sometimes useful to be able to relate throughput Q to (dN/dt), the number of molecules flowing per second, also called the particle flow rate. Dividing both numerator and denominator in the right-hand side of Equation 5.4 by Avogadro's number N_A, we get $\dot{W} = (m/kT) \times Q$ where m is the mass of a molecule and k is Boltzmann's constant. But also $\dot{W} = m \times (dN/dt)$ and therefore

$$Q = kT\left(\frac{dN}{dt}\right) \tag{5.5}$$

Two other measures of flow, included for completeness, are based on the definition of the *standard liter* and the *standard cc* and are the standard liter per minute (*slm*) and the standard cc per minute (*sccm*). The standard liter is the quantity of ideal gas that occupies a volume of 1 l under STP conditions, that is, at a pressure of 1013 mbar and temperature 273 K. The mass of a standard liter of any gas is thus its molar weight divided by 22.4. A flow rate expressed in terms of slm or sccm is thus a mass flow rate and independent of the actual conditions of flow. For example a flow rate of 1 slm at 273 K and 101.3 mbar would be a volumetric flow rate of 10 l per minute.

The volumetric flow rate $\dot{V}$ is frequently given the symbol S and called the *pumping speed.* This is particularly so when it refers to the intake port of a pump or the entrance to a pipe that has a pump connected to its other end. In pumping practice, typical units used are liters per second, liters per minute, and m^3 per hour; m^3 per second is rare!

Remembering that S is a volumetric flow rate, the defining Equation 5.2 now becomes

$$Q = S \times p \tag{5.6}$$

This is the usual form of the first of two basic defining equations that describe gas flow in vacuum practice. It expresses the quantity of gas flowing as the product of the pressure and the volumetric flow rate at that pressure. Allied with the condition for continuity, it is an important tool for analysis.

5.3 Conductance

The other fundamental equation of flow relates throughput Q to the *difference* between the upstream and downstream pressures p_U and p_D in the two volumes that the pipe connects and serves to define the quantity *conductance*. Thus, referring again to Figure 5.1,

$$Q = C(p_U - p_D) \tag{5.7}$$

Evidently, C has the same dimensions as S, i.e., volume per second. A typical unit is liter per second.

It may be helpful to note that although in dc electrical circuits the connection between current (analogue: Q) and potential differences (analogue: $p_U - p_D$) is expressed in Ohm's law by a resistance, in vacuum practice the link is made by its inverse, the conductance. Thus, conductance is a measure of ease of flow in response to a pressure difference, and the greater the conductance for a given pressure difference, the greater the throughput.

Accordingly, and pursuing circuit analogies, one may expect that there are simple rules of combination for conducting elements in series and in parallel.

It is easily shown, and intuitively reasonable, that for conductances C_1, C_2, etc., in parallel, the effective conductance of the combination is given by

$$C = C_1 + C_2 + \text{etc} \ldots \tag{5.8}$$

while for elements in series

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \ldots \tag{5.9}$$

Proving Formula 5.8 and Formula 5.9 is a good exercise in the use of Equation 5.6 and Equation 5.7. Some caution must, however, be exercised in using these formulas in practice because the entry conditions into pipes in the various flow regimes are not as simple as the flow of charge in the electrical case. Thus, in the molecular regime, for example, for Formula 5.9 to be accurate in describing flow through two pipes in series requires that they be separated by an intermediate buffer volume sufficiently large to ensure the existence of near-equilibrium and random entry conditions into the second pipe. These matters will be discussed in Section 5.6.8. It is worth noting that when Formula 5.9 is used for just two components, the "product over sum" rule enables quick — especially, back-of-the-envelope — calculations. That is:

$$C = \frac{C_1 \times C_2}{C_1 + C_2} \tag{5.10}$$

Evaluating conductance in various circumstances will be a principal concern in later sections of this chapter.

Equation 5.6 and Equation 5.7 are the fundamental ones with which all flow problems can, in principle, be solved. They correspond to the statements in circuit analysis that currents are continuous and to Ohm's law. We can use them at once to determine how pumping speed at a vessel is affected by the presence of a pipe that connects the vessel to the pump. This is a widely applicable result of general validity. Consider, as in Figure 5.2, a vessel within which the pressure is p connected via a pipe of conductance C to a pump of speed S^*. The pumping speed at the vessel is S. Let the pressure at the pipe exit and the entrance to the pump be p^*. The throughput from the vessel, through the pipe and into the pump, is Q.

Using Equation 5.7 and Equation 5.6 (twice, exploiting continuity) we have

$$Q = C\,(p - p^*) = S^* \times p^* = S \times p$$

This yields for S, after a little algebra,

$$S = \frac{S^* \times C}{S^* + C} \tag{5.11}$$

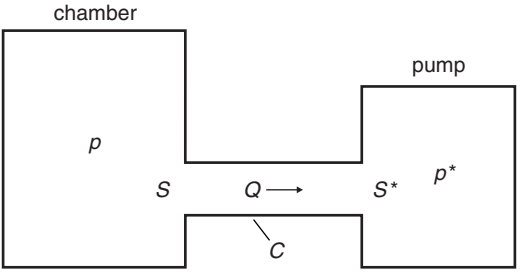

FIGURE 5.2
Effect of conductance on pumping speed.

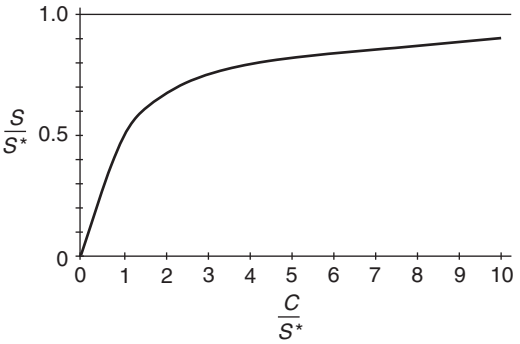

FIGURE 5.3
Variation of pumping speed at the vessel with connecting pipe conductance.

Necessarily, therefore, S is less than S^*. The significance of this formula is illustrated in Figure 5.3 in which S/S^*, the effective pumping speed at the vessel expressed as a fraction of the pump speed S^*, is plotted against C/S^*, the ratio of the connecting pipe conductance to S^*. Clearly, only when C is appreciably greater than S^* is the pumping speed at the vessel comparable with that of the pump. It is halved if the conductance and the pump speed are equal. We may visualize this result in terms of the volumes pumped at the vessel and at the pump. Gas expands as it moves downstream through the pipe to the pump, and the volumetric flow rate that is fixed at the pump end of the pipe by the pump's speed must, therefore, be smaller at the vessel.

5.4 Continuum Flow

In this regime, gas behaves as a fluid, and molecule–molecule collisions with mean free path much less than the equipment size determine gas behavior.

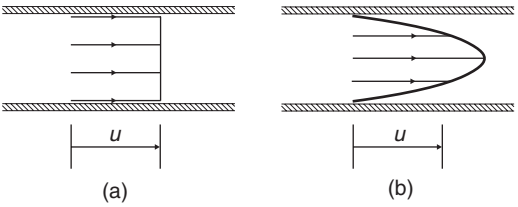

FIGURE 5.4
Velocity profiles for streamlined flow in a pipe. (a) Zero viscosity, (b) finite viscosity.

The characteristic dynamic property of a gas is its viscosity. Even though air is not particularly viscous — one is not normally aware of its viscosity when walking through it — the existence of this property, however small, has a profound effect on the pattern of gas movement in the presence of surfaces. This is that at any boundary surface the gas flow is arrested so as to be stationary at the boundary, causing velocity gradients to be established in the body of the gas, with associated frictional loss, as discussed in Section 3.10.

If a gas had zero viscosity (or were *inviscid*, to use the technical term), its steady flow through a pipe would be characterized by uniform parallel streamlines as shown in Figure 5.4(a), and the gas velocity u would be constant over any cross-section. In reality, however, viscosity causes the gas at the pipe wall to be stationary, so that the velocity profile is developed as shown in Figure 5.4(b), which has a maximum value at the center and some value u averaged over the section. This orderly flow, controlled by viscous forces and following stable, well-defined streamlines (such as may be observed in wind-tunnel studies of models of well-shaped cars or aircraft) is called *laminar* flow. Under these conditions the pressure difference driving the flow and the consequent speed of flow down the pressure gradient are sufficiently small that any velocity fluctuations transverse to the general direction of flow are damped out by viscous forces. The work done by the source of the pressure difference is used in overcoming viscous forces within the gas. However, at some greater and critical driving pressure difference, the stabilizing role of viscous forces is insufficient to do this, and the motion becomes *turbulent*. Viscous forces cease to play a significant role, and the motion of the gas in the general direction of flow is accompanied by disorderly and irregular large swirls and eddies. The concept of a streamline becomes inapplicable. The considerable irregular motion transverse to the general direction of flow consumes much of the work done.

Whether flow of gas with viscosity coefficient η and density ρ through a pipe of diameter D with velocity u is turbulent or not is determined by the value of the dimensionless Reynolds' number Re. Defined by

$$\mathrm{Re} = \frac{\rho u D}{\eta} \tag{5.12}$$

it is the ratio of the inertial forces to the viscous forces in a problem. Flow is laminar for Re values less than about 2000 and turbulent for those above. In the context of vacuum practice, this criterion can be reexpressed in terms of throughput Q and pipe diameter D. It is straightforwardly shown that for air at room temperature, flow is turbulent when, with Q in mbar l s^{-1} and D in mm

$$Q/D > 24.4 \text{ mbar mm} \tag{5.13}$$

As a numerical example, let us consider flow at a pressure of 10 mbar and a volumetric rate of 100 l s^{-1} in a pipe of diameter 50 mm. Using Equation 5.1, the pD product is 500, indicating continuum flow. The throughput is 1000 mbar l s^{-1}, and therefore $Q/D = 20$, so that flow is laminar but nearing the turbulent condition. Steady flows of this magnitude are rarely encountered in vacuum practice; conditions are usually laminar with Reynolds' number well below the critical value. The velocity of flow u is evaluated as the volumetric rate divided by the cross-sectional area of the pipe. In the example above, the flow rate is 0.1 m^3 s^{-1} and the area approximately 0.02 m^2, giving $u = 5$ m s^{-1}. To evaluate Reynolds' numbers directly for air, it may be noted that the viscosity of air (independent of pressure in the continuum regime) is 1.8×10^{-5} Ns m^{-2} at 295 K and its density (which is proportional to pressure) is 1.18 kg m^{-3} at 1000 mbar. Local or transient turbulence may be caused by abrupt changes in flow conditions due to an obstacle in the path of flow or the sudden opening of a valve, for example, but laminar conditions are soon restored as viscous forces overcome inertial forces. It may be noted in passing that flow velocities even as relatively high as ~10 m s^{-1} are still very much smaller than typical molecular velocities so that the bodily movement of gas is a minor perturbation on its intrinsic molecular motion.

5.5 Dynamical Analysis of Continuum Flow

Turbulent motion has defied detailed theoretical analysis since its discovery and early investigation by Osborne Reynolds, and it continues to do so. The analysis of laminar motion in general is also greatly complicated, particularly by the effects of viscosity. Analysis becomes less problematic if a fluid can be assumed to be inviscid, and the results of such analysis can be helpful in understanding some of the broad features of real flows in which viscosity does play a part, even though the viscous effects are ignored. We will pursue these matters to some extent because they are relevant to determining when the effects of compressibility may be ignored, to a particular type of flow called "choked" flow that has important applications, and to the interpretation of pressure measurements. Our starting point is to revisit Bernoulli's equation.

5.5.1 Bernoulli's Equation

Consider steady laminar flow depicted in Figure 5.5 by streamlines tangential to the flow at any particular point. Let PQ and RS be two such streamlines that, together with others, enclose a stream tube. We will suppose that this is horizontal although for gases, even if flow is not horizontal, the gravitational term that arises is often ignorable compared with others. Suppose that AB and CD define cross-sectional planes of area A and $A + dA$ at separation ds where s serves as a coordinate along the stream tube in the direction of flow. At section AB, let the velocity be u, the pressure p, and the fluid density ρ. At CD, the corresponding quantities are $u + du$, $p + dp$ and $\rho + d\rho$. The mass entering the section between AB and CD per second is $A\rho u$, and this same mass leaves at CD because the flow does not cross streamlines. This mass acquires increased momentum $A\rho u \times du$, and therefore applying Newton's second law in the direction of motion

$$-A\,dp = A\,\rho\,u\,du$$

whence

$$u\,du + dp/\rho = 0$$

Integrating,

$$\frac{u^2}{2} + \int \frac{dp}{\rho} = \text{constant} \tag{5.14}$$

This equation relates velocity, pressure, and density at any point along coordinate s in the stream tube and is the basic form of Bernoulli's equation. We can imagine the stream tube becoming sufficiently narrow that the equation relates quantities along the streamline coordinate s. The condition of continuity that applies to the whole flow, namely, that $A\rho u$ = constant, reduces to $\rho\,u$ = constant if flow occurs with unchanging cross-section.

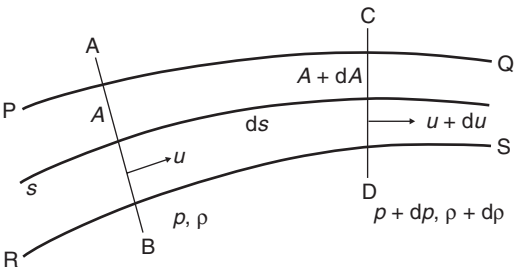

FIGURE 5.5
Flow through a stream tube.

For *incompressible* flow, as in liquids, ρ is constant and Equation 5.14 becomes

$$\frac{1}{2}\rho u^2 + p = \text{constant}$$

This is Bernoulli's equation in its most widely used form. When $u = 0$, as it would be, for example, upstream in a reservoir that acts as the source of the flow or at a downstream "stagnation" point at the entrance to a Pitot tube, the pressure has some value p_0 such that

$$p + \frac{1}{2}\rho u^2 = p_0 \tag{5.15}$$

The total pressure p_0 is therefore the sum of the static pressure p in the undisturbed flow and a kinetic pressure $\rho u^2/2$. A hypothetical nonintrusive directional pressure gauge would measure the total pressure when facing upstream and the static pressure, if directed perpendicularly to the streamline. It will be useful, as is customary, to refer to quantities at points on a streamline where the fluid is stationary as *stagnation* values.

5.5.2 Effects of Compressibility

Because gas is *compressible*, its density will vary from place to place with pressure change when it flows, as remarked in Section 5.2. For flow due to relatively small pressure differences, the effects of compressibility are small, and we shall deduce shortly when they may be ignored. For many cases of compressible flow, especially for that at relatively high speeds, it is often a good approximation to assume that conditions are adiabatic and reversible and hence isentropic. When gas flows in this way from a source at high pressure, a "particle" of gas accelerates along a streamline, expands, and becomes less dense, and its temperature falls. The physical requirement for reversibility is that the heat transfer along or transverse to the streamline shall be negligible. We will deduce the form of Bernoulli's equation for these conditions. If we assume perfect gas behavior, change obeys the equation $pV^\gamma = \text{constant}$, where γ is the ratio of principal specific heats C_p/C_v. In terms of gas density $\rho = M/V$ (molar mass divided by molar volume), this becomes $p = k\rho^\gamma$, where k is a constant. Substituting this in the fundamental Equation 5.14 and carrying out the integration gives Bernoulli's equation for isentropic flow:

$$\frac{u^2}{2} + \frac{\gamma}{\gamma - 1}\frac{p}{\rho} = \frac{\gamma}{\gamma - 1}\frac{p_0}{\rho_0} \tag{5.16}$$

in which, as before, the right-hand side refers to a point along the streamline where $u = 0$. We will take this to be upstream at a point in a reservoir from which flow starts. We may use this result first to demonstrate when it is permissible to regard gas flow as incompressible.

Equation 5.16 may be rearranged as

$$p_0 = \frac{\rho_0}{\rho} p\left(1 + \frac{\gamma - 1}{2\gamma}\frac{\rho}{p}u^2\right) \tag{5.17}$$

It follows from the equation $p = k\rho^\gamma$ that values of p and ρ at a general point along the streamline are related to stagnation values (subscript 0) by $(\rho_0/\rho) = (p_0/p)^{1/\gamma}$. Substituting to replace (ρ_0/ρ) in Equation 5.17 yields

$$p_0 = p\left(1 + \frac{\gamma - 1}{2\gamma}\frac{\rho}{p}u^2\right)^{\frac{\gamma}{\gamma-1}} \tag{5.18}$$

Now, as is shown in texts on wave motion and acoustics — for example, Feather (1961) — the speed of sound c in a perfect gas is

$$c = \sqrt{\frac{\gamma p}{\rho}} = \sqrt{\frac{\gamma R_0 T}{M}} \tag{5.19}$$

For the stagnation condition at temperature T_0, $c_0 = \sqrt{\gamma R T_0/M}$, and for places downstream at pressure p, density ρ, and temperature T, the local value of c is $\sqrt{\gamma R_0 T/M}$. Thus, in Equation (5.18)

$$p_0 = p\left(1 + \frac{\gamma - 1}{2}\frac{u^2}{c^2}\right)^{\frac{\gamma}{\gamma-1}} \tag{5.20}$$

In terms of the Mach number Ma defined as the ratio u/c of the gas velocity to the sonic velocity

$$p_0 = p\left(1 + \frac{\gamma - 1}{2}\mathrm{Ma}^2\right)^{\frac{\gamma}{\gamma-1}} \tag{5.21}$$

Expanding by the binomial theorem and using $c^2 = p/\rho$ again gives

$$p_0 = p + \frac{1}{2}\rho u^2\left(1 + \frac{\mathrm{Ma}^2}{4} + \frac{2-\gamma}{24}\mathrm{Ma}^4 + \ldots\right) \tag{5.22}$$

The velocity of sound in air at 295 K and atmospheric pressure is about 340 m s^{-1}. In the third term of the expansion, γ for air is 1.40, so that it becomes $\mathrm{Ma}^4/40$. For values of Ma up to 1/5 corresponding to air speeds of nearly 70 m s^{-1}, the term in Ma^2, therefore, causes Equation 5.22 to differ from Equation 5.15 for incompressible flow by only 1%. This is the justification for neglecting the effects of compressibility for Mach numbers less than ~ 0.2–0.3.

5.5.3 Flow through Short Ducts and Apertures: Choked Flow

Compressibility cannot be neglected in dealing with flows through apertures or short ducts that separate regions between which there are large pressure differences of the order of atmospheric pressure. In typical situations, gas flows from a relatively large volume at constant high pressures ~ atmospheric, or from the open atmosphere, through a constriction to a region where a substantially lower pressure is maintained by a pump. Returning to the Bernoulli Equation 5.16, we can use it to determine downstream velocity and other properties in the assumed isentropic flow. Thus it may be rearranged to give

$$u^2 = \frac{2\gamma}{\gamma - 1}\left[\frac{p_0}{\rho_0} - \frac{p}{\rho}\right] = \frac{2\gamma}{\gamma - 1}\frac{R_o T_0}{M}\left[1 - \frac{p/\rho}{p_0/\rho_0}\right]$$

where we have used the perfect gas law expressed in terms of density: $p_0/\rho_0 = R_0 T/M$. Using $(\rho_0/\rho) = (p_0/p)^{1/\gamma}$ again to replace the densities, the downstream velocity u as a function of downstream pressure p (strictly the pressure ratio) becomes

$$u = \left[\frac{2\gamma}{\gamma - 1}\frac{R_o T_0}{M}\left\{1 - \left(\frac{p}{p_0}\right)^{\frac{\gamma-1}{\gamma}}\right\}\right]^{\frac{1}{2}} \tag{5.23}$$

To determine the mass flow rate per unit area j, we multiply u by the density ρ to get

$$j = \rho u = \rho_0\left(\frac{p}{p_0}\right)^{\frac{1}{\gamma}} u = \rho_0\left[\frac{2\gamma}{\gamma - 1}\frac{R_o T_0}{M}\left(\frac{p}{p_0}\right)^{\frac{2}{\gamma}}\left\{1 - \left(\frac{p}{p_0}\right)^{\frac{\gamma-1}{\gamma}}\right\}\right]^{\frac{1}{2}} \tag{5.24}$$

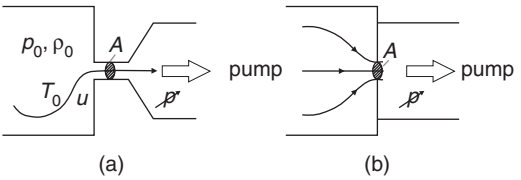

FIGURE 5.6
Choked flow through (a) short duct, (b) aperture.

To interpret these equations, let us consider that flow starts from a reservoir where the pressure, density, and temperature have constant values p_0, ρ_0, and T_0. Let us assume that that flow is down a short duct of constant area A and that it is uniform over any given cross-section as in Figure 5.6(a) and therefore one-dimensional in the sense that one coordinate is sufficient to describe it. At the end of the duct is a volume in which the pressure is determined by a pump. We will imagine that its speed is variable and can be adjusted so that the downstream pressure, and therefore the pressure at the duct outlet, can be set at any desired value.

Starting from the reservoir, the velocity of a given fluid "particle" increases as it accelerates down its streamline through the duct, and its pressure, density, and temperature fall in accordance with the governing adiabatic equations. For modest pressure differences, the flow velocity and rate increase as the downstream pressure set by the pump is lowered below p_0, as indicated in Figure 5.7. However, a maximum flow velocity derived by inserting $p = 0$ in Equation 5.23 is not achievable. If the pressure set at the outlet is sufficiently low that the pressure ratio is equal to or less than some critical value p^*/p_0, to be determined shortly, the velocity of the accelerating flow increases up to a maximum equal to the local speed of sound, beyond which it cannot be increased because downstream conditions of lower pressure to further accelerate the flow can no longer be communicated back upstream. The associated mass flow rate is also a maximum.

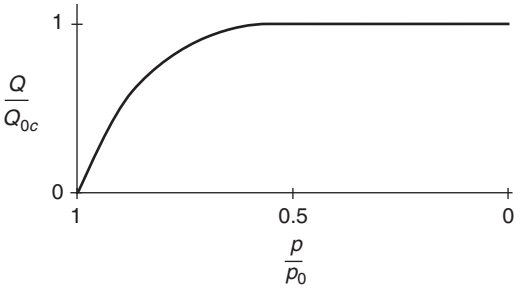

FIGURE 5.7
Flow as a function of downstream pressure.

Differentiation of Equation 5.24 shows that the maximum value of j occurs for a critical pressure ratio p^*/p_0 such that

$$\left(\frac{p^*}{p_0}\right) = \left(\frac{2}{\gamma + 1}\right)^{\frac{\gamma}{\gamma - 1}} \tag{5.25}$$

For air, taking $\gamma = 1.4$, this critical pressure ratio is $(2/2.4)^{(1.4/0.4)} = 0.53$. Substituting Equation 5.25 in Equation 5.24, the associated maximum flow rate per unit area is

$$j_{\max} = \rho_0 \sqrt{\frac{\gamma R_o T_0}{M}} \left(\frac{2}{\gamma + 1}\right)^{\frac{\gamma + 1}{2(\gamma - 1)}} \tag{5.26}$$

From the governing equation pV^γ = constant, it is easily shown that $(T/T_0) = (p/p_0)^{(\gamma-1)/\gamma}$. Hence at the critical condition, using Equation 5.25, $T^*/T_0 = 2/(\gamma + 1)$, which for air is $2/2.4 = 0.833$. Thus, reservoir air initially at a temperature 295 K would cool to 246 K, i.e., from 22°C to –23°C. From Equation 5.23 the value of u at the critical condition is

$$u = \sqrt{\frac{2\gamma}{\gamma + 1} \frac{R_o T_0}{M}}$$

Substituting for T_0 gives

$$u = \sqrt{\frac{\gamma R_o T^*}{M}} = c$$

proving that Ma = 1. Therefore, for the critical pressure ratio, the flow is a maximum, and when the critical pressure along the streamline has been reached, the flow velocity has increased to the velocity of sound. Flow is subsonic until it becomes sonic at the duct exit and pressure p^*. In these circumstances the flow is said to be *choked*. Increasing the pump speed to set a lower pressure in the downstream volume does not increase the flow rate, which stays constant at the maximum value. It can only be increased by increasing the reservoir pressure p_0.

These results apply equally to the aperture of area A shown in Figure 5.6(b), similarly located between a reservoir and a region of variable lower pressure. In this case the streamlines develop in the manner shown to make the flow sonic at the aperture. (Strictly, the inward momentum of the flow results in a narrowing and subsequent oscillation in width of the gas stream — the *vena contracta* phenomena — and a reduction in the flow rate accounted for by inserting a discharge coefficient of typically 0.85 in Equation

5.26. If the rim of the aperture is carefully shaped, however, the coefficient can be close to unity and, for our purposes, it may be ignored.)

Multiplying j by the area A of Figure 5.6(a) or Figure 5.6(b) gives the mass flow rate $\dot{W}$, and then Equation 5.4 can be used to find equations for throughputs Q. These are referred to the upstream condition at temperature T_0 because temperature falls downstream. For choked flow in the duct of constant area or the aperture, from Equation 5.24, using subscript C for the choked condition,

$$Q_{0C} = p_0 \times A\sqrt{\frac{\gamma R_0 T_0}{M}}\left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{2(\gamma-1)}} \tag{5.27}$$

When the downstream set pressure is greater than p^* so that the pressure ratio is greater than the critical value and flow is not choked, then from Equation 5.24,

$$Q_0 = p_0 A\left[\frac{2\gamma}{\gamma-1}\frac{R_o T_0}{M}\left(\frac{p}{p_0}\right)^{\frac{2}{\gamma}}\left\{1-\left(\frac{p}{p_0}\right)^{\frac{\gamma-1}{\gamma}}\right\}\right]^{\frac{1}{2}} \tag{5.28}$$

From Equation 5.6, pumping speeds $S = Q/p$ at the aperture may be defined as $S_A = Q_0/p_0$ for the unchoked condition and $S_{AC} = Q_{0C}/p_0$ for the choked condition. From Equation 5.28, S_A increases from zero up to a maximum at the choked value as p is reduced below p_0. Once the aperture has become choked, it has a constant pumping speed, S_{AC}, which follows directly from Equation 5.27. Inserting numerical values for air gives

$$S_{AC} = 20 \times A \text{ l s}^{-1},\ A \text{ in cm}^2,\ p/p_0 \leq 0.53 \tag{5.29}$$

It is a useful exercise to show that flow through an aperture or short duct of area A will be choked if the speed S of the pump that determines downstream conditions is such that

$$S \geq \gamma^{\frac{1}{2}}\left(\frac{\gamma}{\gamma+1}\right)^{\frac{1}{2}} A\sqrt{\frac{R_o T_0}{M}} \tag{5.30}$$

For air at 22°C, $S \geq 38 \times A$ l s^{-1} where A is in cm^2.

Choked flow is encountered in a number of applications of vacuum technology. For example, constant pumping speed at pressure p_0 upstream of a choke with a value determined by Equation 5.29 is guaranteed if $p < 0.53\ p_0$ is maintained downstream. Equally, as Equation 5.27 shows, a throughput

exactly proportional to upstream pressure is achieved if the choked condition is maintained over a range of upstream pressure values, and this is exploited in one type of flow-measuring device. When large pumps are needed for applications in which initial pumping from atmospheric pressure might create turbulence and airborne debris, choking can be used to limit the throughput until nonturbulent conditions are achieved. Choked flow also occurs when vacuum systems are vented back to atmospheric pressure. This is the subject of Problem 5.27.

5.5.4 Continuum Flow through Pipes

Flow through a pipe of any length may become choked if a pump of sufficiently great speed is connected downstream and if the pipe has limited conductance. Although flow in the initial stages of pumping may approach these conditions in some systems, flow in typical systems under normal operating conditions is viscous laminar and driven by relatively modest pressure differences. The formula of Poiseuille and Hagen, which describes, as initially formulated, the flow of liquids in long pipes, is adaptable to the gas flow but is unfortunately of limited value. This is because the usual requirement in vacuum technology is that pipes that connect work chambers to pumps shall offer as little impedance as possible to gas flow, and so they are made as short and wide-bored as possible, consistent with spatial constraints. Typical conditions do not, therefore, involve long pipes. Nevertheless, because it is the only pipe flow problem with a simple analytical solution and because its use illustrates methods of analysis, it is presented here. Occasionally, however, "long pipe" applications do arise. Prior to doing this, it is necessary to consider how flow in a pipe develops from the entrance.

When gas enters a pipe from a volume where it has negligible initial motion, driven by relatively small pressure difference so that flow is not turbulent, the gas starts to accelerate down the pipe and viscous forces come into play because flow at the pipe walls is arrested. This results in a nonuniform velocity profile in the pipe cross-section. Work is done to give the gas kinetic energy as it accelerates and against viscous forces that operate between the layers brought into relative motion. The profile changes progressively downstream, with gas in the central region acquiring the greatest velocity until, if the pipe is long enough, the velocity profile becomes constant. Such a flow is described as "fully developed" and the distance down the tube for this to be established is called the "entry length," denoted X. It is related to flow conditions (see Fay [1994]) by the approximate formula $X/D = \mathrm{Re}/16$, where Re is Reynolds' number. Thus, for example, for the flow of air at a pressure of 10 mbar and volumetric rate 10 l s^{-1} through a pipe of diameter 25 mm the flow velocity is 2 m s^{-1}, Re = 32, and the entry length is about 2 diameters. Through a 10-mm pipe a flow of 1 l s^{-1} has Re = 82.5 and the entry length is 5 diameters. In many circumstances the entry length

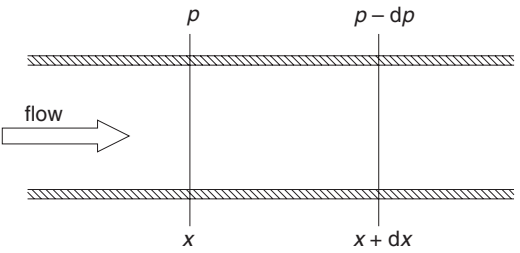

FIGURE 5.8
Poiseuille flow.

exceeds the pipe length and no simple analysis is possible. A regime is not approached for which the entry length becomes ignorably small with respect to the total length, thus making a long pipe formulation valid. The reader should consult the authoritative article by Livesey (1998) for a comprehensive discussion of these problems and how calculations may be formulated in practical situations.

Returning to simple Poiseuille flow, consider, as in Figure 5.8, steady flow in a long pipe of diameter D in a section of length dx between positions x and $x + dx$, in which the pressure falls from p to $p - dp$. As shown, for example, in Temperley (1961), the volumetric flow rate of fluid through this section is

$$\dot{V} = -\frac{\pi D^4}{128\eta}\frac{dp}{dx} \tag{5.31}$$

For an incompressible fluid, the volumetric rate and hence dp/dx are constant and integration over the length L of the pipe from $x = 0$ ($p = p_1$) to $x = L$ ($p = p_2$) gives

$$\dot{V} = \frac{\pi D^4}{128\eta L}\left(p_1 - p_2\right) \tag{5.32}$$

For a gas, multiplying Equation 5.31 by p gives the throughput Q at the section x, which is constant along the pipe, and also formally allows for the change in volume with pressure. Thus

$$Q = p \times \dot{V} = -\frac{\pi D^4}{128\eta} p\frac{dp}{dx} = -\frac{\pi D^4}{128\eta}\frac{d}{dx}\left(\frac{p^2}{2}\right)$$

Integrating between x = 0 and $x = L$ gives

$$Q = \frac{\pi D^4}{128\eta L}\left(\frac{p_1^2 - p_2^2}{2}\right) = \frac{\pi D^4}{128\eta L}\left(\frac{p_1 + p_2}{2}\right)\left(p_1 - p_2\right) \tag{5.33}$$

The middle term on the right-hand side is the mean pressure $\bar{p}$. In accordance with the earlier discussion about compressibility, it may be noted that this flow could have been treated as incompressible with volumetric rate given by Equation 5.32 and the pressure taken as the average value. From Equation 5.33 the conductance is

$$C = \frac{\pi D^4}{128\eta L}\bar{p} \tag{5.34}$$

Note the fourth-power dependence on pipe diameter D and the inverse dependence on L. Doubling the pipe diameter increases the conductance by a factor of 16! This is one origin of the guidance to make pipes as short and wide-bored as possible, though as we have seen, this formula may not be appropriate when one has done so! Evidently, conductance depends on mean pressure and is therefore a changing quantity as flow conditions alter, adding a complication to analysis. Inserting numerical values for air at 22°C gives

$$C = 133\left(D^4/L\right)\bar{p}\ \text{ls}^{-1} \text{ with } D \text{ and } L \text{ in cm and } \bar{p} \text{ in mbar} \tag{5.35}$$

At higher pressures the conductance may be so large as not to be a restricting factor, i.e., the speed at the pumped chamber may effectively be the speed of the pump itself (see Figure 5.3). Thus, in the context of primary pumps of typical size having speeds (as we shall see in Chapter 6) of the order of a few liters per second, for a pipe 1 m long and 1 cm in bore, the conductance for air is 650, 65, 6.5, and 0.65 l s^{-1} at mean pressures of 500, 50, 5, and 0.5 mbar, respectively. At high pressures the conductance is considerably larger than the pump speed, and at the lower pressures becomes comparable with it. For this rather large L/D ratio = 100, formula 5.35 would be quite accurate.

Returning to fundamental Equation 5.7, let us consider how it might be used in this context of viscous flow where conductances depend on pressure. In terms of upstream and downstream pressures p_U and p_D,

$$Q = C(p_U - p_D)$$

Various situations may arise. If p_U and p_D have known values in a particular problem, then it is a straightforward matter to evaluate C from Equation 5.35 and then the throughput Q. However, if either of p_U and p_D are not known, then neither is the conductance and matters are not straightforward. In spite of this, if, say, the throughput and downstream pressure p_D, at a pump perhaps, are known, it is possible by a clever manipulative device described by Livesey to determine the pressure ratio p_U/p_D. It is used in Section 8.3. As he points out, while conductance has considerable value as a concept, its prime purpose is to relate the throughput to the pressure conditions, and if this can be accomplished by other means, then that purpose is served. Knowing its value in a practical

problem may not be crucial because it is the means to an end. Livesey's article is also a source of formulas for flow in pipes with cross-sections other than circular and contains data for modifying them for gases other than air.

In problems that cannot be formulated in an analytically solvable form because of the dependence of conductance on pressure, the process may be modeled by setting intervals in which conductance has a constant and appropriate specified mean value. An example of this is given later.

Having considered the principal attributes of continuum flow and some of its complexities, we move on to the quite different matter of molecular flow. As we shall discover, this is less complicated in character, and analysis of the simpler situations involving apertures and pipes is more straightforward.

5.6 Molecular Flow

Molecular flow is characterized by Knudsen numbers Kn greater than unity, which physically means that the mean free path associated with the prevailing number density of the contained gas molecules is greater than the size of the container, with the consequence that molecule/wall collisions dominate gas behavior, as discussed in detail in Chapter 3. All semblance to fluid behavior is lost because there are no molecule–molecule collisions. These are the conditions in the work chambers of many vacuum systems because, as previously noted, the mean free path for nitrogen at 10^{-4} mbar is 0.64 m, so that for a chamber of typical size, conditions are molecular at pressures below this value.

The other feature of molecular behavior that determines the nature of flow in this regime is the randomizing nature of the molecule–surface collision discussed in Chapter 4 and as described by the Knudsen scattering law. In these circumstances, flow is not driven by the force created by pressure differences as it was for continuum flow, but is the outcome of statistical chance. In a pipe, for example, a molecule arriving at the pipe wall is as likely to rebound in a backward direction from which it came as in a forward direction along its original direction of travel. The flow that occurs under these conditions between two regions at different (low) pressures connected by a pipe is due to steady net diffusive drift of molecules and is discussed in detail later. We will derive the expression for the molecular flow conductance of a long pipe using a simple classical approach that ignores fine detail of the flow mechanism, before considering better and more modern methods of analyzing conductance problems.

5.6.1 Molecular Flow through a Long Pipe

In a pipe of length L and diameter D, consider a short section between coordinates x and $x + dx$, across which pressure changes from p to $p - dp$. There

will be an associated change n to $n - dn$ in the number density of molecules, but n may be taken as the density in the section for the purposes of the calculus. Molecular flow in the pipe may be considered to occur with a mean drift velocity $\bar{u}$, superposed on the thermal velocities, that is reduced to zero by the collisions that molecules have with the wall. By Newton's second law, equating the rate of change of momentum due to the loss of $\bar{u}$ in wall collisions to the force across the element, gives

$$\left(n\bar{v}/4\right)\left(\pi D\, dx\right)\left(m\bar{u}\right) = \left(\pi D^2/4\right)dp$$

in which the first two terms on the left-hand side give the number of wall impacts per second in the element dx. After cancellations, $n\bar{v}m\bar{u} = D(dp/dx)$. The number of molecules per second passing through the plane at x is

$$\frac{dN}{dt} = \left(\frac{\pi D^2}{4}\bar{u}\right)n = \frac{\pi D^2}{4}\frac{D}{m\bar{v}}\frac{dp}{dx}$$

Substituting $\bar{v} = \sqrt{8kT/\pi m}$ from Equation 3.7, multiplying by kT to convert to a throughput, and integrating over the pipe (and ignoring end effects) leads to

$$Q = \frac{\pi D^3}{16L}\sqrt{\frac{2\pi kT}{m}}\left(p_1 - p_2\right)$$

The conductance is therefore

$$C = \frac{\pi D^3}{16L}\sqrt{\frac{2\pi kT}{m}} = \frac{\pi D^3}{16L}\sqrt{\frac{2\pi R_0 T}{M}}$$

Because this treatment is oversimplified in its assumptions about the drift velocity, the factor π/16 should be replaced by 1/6 (see Loeb, 1961) to give

$$C_L = \frac{D^3}{6L}\sqrt{\frac{2\pi R_0 T}{M}} \tag{5.36}$$

This formula was first proposed by Knudsen and is correct for long pipes. The geometrical dependence on D^3/L is its most important feature and again points towards making pipes as short and fat as possible to maximize conductance. The factor $\sqrt{T/M}$ indicates the dependence on the particular gas and its temperature and is directly related to the molecular velocity. Note that, as one would expect, the conductance does not depend on pressure. Inserting numerical values for nitrogen at 295 K gives

$$C_L = 12.4\frac{D^3}{L}\ \text{ls}^{-1},\ \text{for } D, L \text{ in cm} \tag{5.37}$$

The modern approach to analyzing the conductance of pipes and other pipeline components involves the concept of *transmission probability*, but in order to introduce that, it is necessary first to consider flow through apertures.

5.6.2 Molecular Flow through an Aperture

Consider an aperture of area A in a very thin wall separating two regions maintained at different pressures p_1 and p_2, with $p_1 > p_2$ and the gas in both regions sufficiently rarefied that conditions are molecular, as shown in Figure 5.9. The molecular mean free path is greater than the diameter of the aperture, and there are no molecule–molecule collisions. From each side, molecules approach the opening from all directions within a 2π solid angle and with a range of speeds. The fluxes are represented by the arrows J.

Corresponding to p_1 and p_2 are number densities n_1 and n_2 and associated fluxes J_1 and J_2, where $J = n\bar{v}/4 = p/\sqrt{2\pi mkT}$, as discussed in Section 3.4 and Section 3.7 of Chapter 3. Molecules heading towards the aperture opening from both sides will pass through it, and so with $J_1 > J_2$ as indicated, there will be a net flow of molecules from left to right. The number per unit time will be

$$\frac{dN}{dt} = \left(J_1 - J_2\right)A \quad .$$

Using Equation 5.5 to convert to a throughput and substituting for J from Equation 3.22 gives

$$Q = kT \times \frac{\left(p_1 - p_2\right)}{\sqrt{2\pi mkT}}A = \sqrt{\frac{kT}{2\pi m}}A\left(p_1 - p_2\right) = \sqrt{\frac{R_0T}{2\pi M}}A\left(p_1 - p_2\right)$$

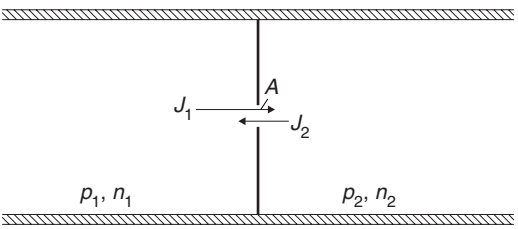

FIGURE 5.9
Molecular flow through an aperture.

Comparing this with the basic defining Equation 5.7 for conductance, we have, introducing the symbol C_A for the molecular flow conductance for an aperture,

$$C_A = A\sqrt{\frac{R_0 T}{2\pi M}} \tag{5.38}$$

This is an important result, exploitable not only in its own right but also because the entrances into pipes and pumps can be regarded as apertures. The presence of the factor $\sqrt{T/M}$ that will also occur in other formulas for molecular flow conductance is noteworthy because it enables conductance values for other gases to be quickly computed once values for a particular gas of reference, usually nitrogen, are known. For nitrogen at 295 K, substituting numerical values gives

$$C_A = 11.8\ A\ \text{l s}^{-1}, \text{ with } A \text{ in cm}^2 \tag{5.39}$$

For a circular aperture of diameter D cm, it takes the useful form

$$C_0 = 9.3\ D^2\ \text{l s}^{-1}, D \text{ in cm} \tag{5.40}$$

The flows in each direction through the aperture are quite independent and do not interfere with each other. They are determined solely by the number of densities on the side from which they originate. On each side, therefore, it is as though there are no molecules in the region beyond, and we can, therefore, express the left-to-right and right-to-left throughputs at the plane of the aperture as $Q_1 = C\ p_1$ and $Q_2 = C\ p_2$, consistently with $Q = Q_1 - Q_2 = C\ (p_1 - p_2)$.

The aperture can be thought to have a pumping speed. If the pressure p_2 on the right-hand side is very much less than p_1, the pumping speed is $Q_1/p_1 = C_A$ as given by Equation 5.38. Although their regimes are quite different, it is interesting to compare Equation 5.38 and Equation 5.39 with Equation 5.27 and Equation 5.29 for continuum choked flow.

5.6.3 The Concept of Transmission Probability

Consider, as in Figure 5.10, a pipe of length L, diameter D, and cross-sectional area A connecting two regions of low pressure p_1 and p_2, such that $\lambda >> L$, D, and conditions are molecular.

The total number of molecules per second crossing the plane EN to enter the pipe is $J_1 A$. They approach it from all directions within a solid angle 2π in the left-hand volume. Relatively few molecules, like molecule (1), will be traveling in such a direction as to pass right through the pipe without touching the sides, but most will not. The majority, like molecule (2), will

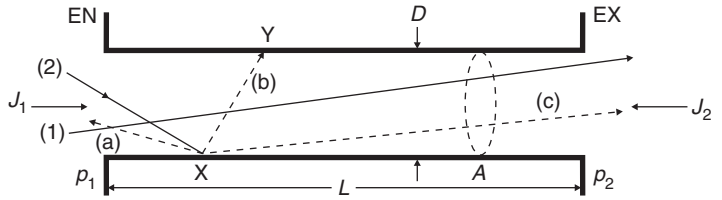

FIGURE 5.10
Molecular flow through a pipe. (From Chambers, A., Fitch, R.K., and Halliday, B.S., *Basic Vacuum Technology,* 2nd ed., Institute of Physics Publishing, Bristol, 1998. With permission.)

collide with the wall at a place such as X and return to the vacuum in a random direction as discussed earlier. There are now three possible outcomes (a), (b), or (c) as shown by the dashed trajectories. The molecule may (a) return to the left-hand volume, (b) go across the pipe to Y, and then another "three-outcome" event, or (c) leave the pipe through the exit plane EX into the right-hand volume. These three outcomes occur with different probabilities, each of which has two aspects: the probability of take-off in a particular direction as determined by the cosine law (Section 4.3) and whether that direction is within the solid angle subtended at X by the entrance area, the exit area, or the rest of the pipe. Furthermore, for a molecule which goes to Y at a different distance along the pipe, the balance of probabilities for where it next goes will have changed accordingly. A little reflection shows that this three-dimensional problem is quite complex to analyze. The ways in which it has been tackled are described in Section 5.6.5.

Considering the total number of molecules per second, J_1A, which cross plane EN and enter the pipe, and the diverse possibilities for their subsequent trajectories, it is clear that some molecules will eventually be transmitted through the exit plane EX. The remainder will return through the plane EN. The fraction that does pass through EX into the right-hand region may be defined as the *transmission probability* α of the pipe so that the transmitted flux is $\alpha(J_1A)$. We expect that α will be large for short pipes, and that for $L << D$ it will approach unity, corresponding to the flow through an aperture in a thin wall. Increase of L, because of the increased number of randomizing collisions with the pipe wall, must cause α to decrease.

For flow in the right-to-left direction, similar considerations must apply. The transmission probability of the pipe must be the same in both directions, but the flux J_2 corresponds to the lower pressure p_2. The right-to-left flow is, therefore, $\alpha(J_2A)$, and the net observable flow is the difference of the flows in each direction. Thus, multiplying this net flow rate $\alpha(J_1 - J_2)A$ by kT to get a throughput specification gives

$$Q = kT(J_1 - J_2)A\alpha$$

Substituting for J from Equation 3.22 gives

$$Q = \sqrt{\frac{kT}{2\pi m}} A\alpha\left(p_1 - p_2\right) = \sqrt{\frac{R_0 T}{2\pi M}} A\alpha\left(p_1 - p_2\right)$$

and therefore from Equation 5.38 we have

$$Q = \alpha C_A (p_1 - p_2)$$

Thus, the molecular flow conductance of a pipe C may be expressed in terms of the conductance of its entrance and a transmission probability, that is:

$$C = \alpha C_A \tag{5.41}$$

The concept of transmission probability, though introduced in terms of a circular pipe, is quite generally extendable to other pipeline components, such as elbows and components with, say, rectangular openings. For an aperture, clearly, $\alpha = 1$. For *long* circular pipes, from Equation 5.37 and Equation 5.40

$$\alpha = \frac{C_L}{C_A} = \frac{4D}{3L} \tag{5.42}$$

Modern methods described in the next section show, however, that even for $L/D = 20$, which would normally be regarded as a safe approximation to being long, this formula gives values that are 10% too high.

5.6.4 Molecular Flow through Short Pipes

A serious difficulty with Equation 5.36 is that because of the assumptions involved in its derivation it is likely to be in error for pipes that are not long, particularly so for pipes of typical size used in vacuum technology, where $L \sim D$. For $L = 0$, it wrongly predicts an infinite conductance. We know that in this case we must simply have an aperture conductance as given correctly by Equation 5.38. In the early days of the subject, Dushman in 1922 argued empirically that this short-pipe problem might be addressed by considering its conductance to be that of an entrance aperture with conductance C_A given by Equation 5.38 in series with a pipe conductance given by Equation 5.36 (see Dushman, 1962). Essentially, the argument is that molecules have to find the pipe entrance before they can proceed down it. Thus, using Equation 5.10, the conductance of a short pipe is

$$C = \frac{C_A C_L}{C_A + C_L} = \frac{C_L}{1 + C_L/C_A} = \frac{C_L}{1 + 4D/3L} \tag{5.43}$$

For nitrogen at 295 K, substituting numerical values gives

$$C = \frac{12.4\,D^3/L}{1 + 4D/3L}\ \mathrm{ls}^{-1},\ \text{for } D, L \text{ in cm} \tag{5.44}$$

Comparing this with Equation 5.37, the effect of introducing the entrance conductance may be interpreted as an effective increase of the pipe length. This approximate, almost venerable equation gives the correct C_A and C_L values in the limits $L = 0$ and $L >> D$. In the range $L \sim$ a few D of typical pipes, it may be up to 12% in error as will be discussed shortly. Nevertheless, for "back-of-the-envelope" calculations, it is often very useful and sufficiently accurate.

The transmission probability $\alpha_D = C/C_A$ associated with this approach is easily shown to be

$$a_D = \frac{1}{1 + 3L/4D} \tag{5.45}$$

The history of attempts to solve the finite pipe problem more accurately are reviewed by Steckelmacher (1986). Livesey (1998) gives an up-to-date survey of developments. Clausing (1932) first introduced the concept of a transmission probability, and using analysis based on integral equations, derived alpha values for round pipes of arbitrary L/D ratio that are considered to be a few percent accurate and are still widely used. These methods were further developed and refined by Cole (1976), whose data are considered to be the most accurate. A selection of α values is given in Table 5.1 together with values from Dushman's formula. Cole's data is the most accurate available. The last column gives the error (C estimated too high) in using Dushman's formula. Evidently the error is greatest for $L \sim 4D$.

More comprehensive tabulations are given in various texts — for example, O'Hanlon (2003). Transmission probabilities for other pipes of uniform rectangular and annular cross-sections are given by Livesey (1998).

TABLE 5.1

Transmission Probabilities of Circular Pipes

L/D	$\alpha_{Clausing}$	α_{Cole}	$\alpha_{Dushman}$	Error, %
0	1	1	1	0
0.2	0.8341		0.87	4
0.5	0.672	0.671984	0.727	8
1	0.5136	0.514231	0.571	11
2	0.3589	0.356572	0.4	12
5	0.1973	0.190941	0.21	10
10	0.1135	0.109304	0.118	8
20	0.0613	0.059422	0.0625	5
50	0.0258	0.25258	0.026	4

5.6.5 Transmission Probability by Monte Carlo Methods

By computer simulation of the trajectories of molecules through a component using Monte Carlo techniques, it has been possible to determine transmission probabilities for a variety of components and shapes, as well as validating the analytical results for the simpler shapes. These methods were first used by Davis (1960) and Levenson et al. (1963). The data of Davis for an elbow and Levenson for a rectangular duct are presented graphically in Figure 5.11, from which it can be seen that they are well parameterized for easy use.

Once a transmission probability has been determined from tabulated or graphical data, it is simply a matter of multiplying it by the aperture conductance ($11.8A$ l s^{-1} for nitrogen) to find the conductance of the component. For example, from Table 5.1 a circular pipe 10 cm in diameter and 50 cm long will have a transmission probability of 0.191, and therefore a conductance for nitrogen, using Equations 5.41 and 5.40 of $0.191 \times 9.3D^2 = 177$ l s^{-1}. The use of the graphical data may be illustrated by determining the conductance of a pipe 50 cm long with rectangular cross section 20 cm × 5 cm. From Figure 5.11 $\alpha = 0.2$ and so $C = 0.2\ (11.8 \times 5 \times 20) = 1180/5 = 236$ l s^{-1} for nitrogen. The accuracy obtained by careful use of the graphical data will be good enough for most purposes.

5.6.6 Entrance and Exit Effects

Figure 5.12(a) shows flow through a pipe that connects two large volumes in which the gas may be considered to be at rest and maintained at constant pressures p_U and p_D by appropriate supply and withdrawal devices. The planes of entry to and exit from the pipe are labeled E and X.

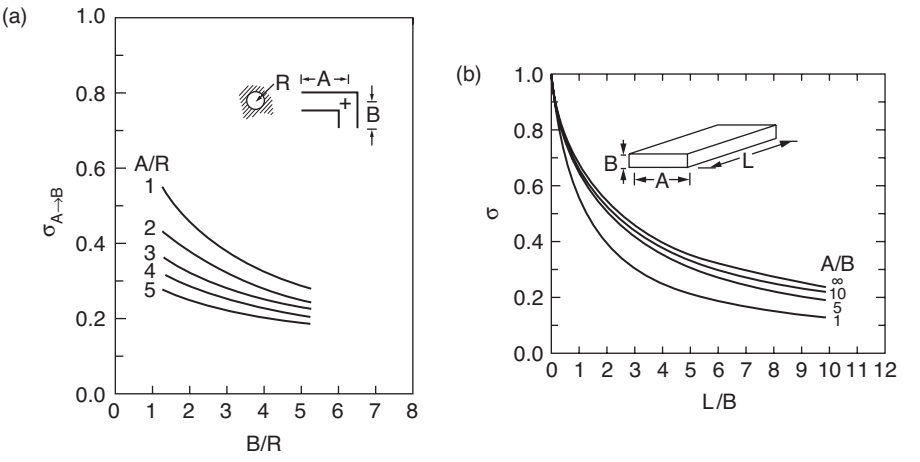

FIGURE 5.11
Transmission probabilities (a) for an elbow (from Davis, D.H., *J. Appl. Phys.*, 31, 1169-1176, 1960; with permission) and (b) for a rectangular duct (from Levenson, L.L. et al., *Le Vide*, 103, 42, 1963; with permission).

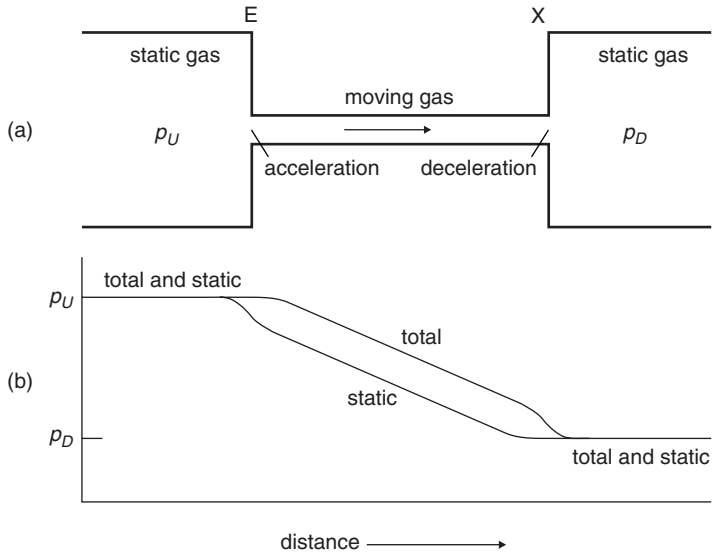

FIGURE 5.12
Variation of total and static pressure in gas flow between two reservoirs.

Let us consider how pressures would be measured and what the measurements taken in various places would mean, bearing in mind the analysis of Section 5.5.1 in which it is shown that in flowing gas the total pressure is the sum of static and kinetic contributions. Though derived for continuum flow, the concepts remain valid in any regime. They simply distinguish the effect of molecular impacts on a measuring device for zero and finite *bulk* motion of the gas. Suppose that we have a simple first-principles directionally sensitive pressure gauge. It might consist (hypothetically) of a suitably supported and housed circular membrane, one side of which faces the gas under test while the other is held at a suitable reference pressure with its distortion providing the measure. We imagine further that it is sufficiently small not to alter the pattern of flow, i.e., that it is nonintrusive. It is a direct mechanical device recording the effect of molecular impacts and can be imagined to be arbitrarily sensitive. When the gauge is located in static gas, the pressures measured do not depend on direction. The total pressure is the static pressure and there is no kinetic component. But when located in moving gas and facing upstream, the pressure measured is a total pressure which will have a kinetic component caused by the bodily movement of gas in the direction of flow superposed on the intrinsic molecular motion. When located in moving gas but pointing perpendicularly to the flow, just the static component of the total pressure is measured.

For continuum flow, as in Figure 5.12(b), measurements taken between the two volumes with a gauge pointing upstream would give the total pressure represented by the upper line, starting and finishing with the static pressures p_U and p_D, respectively. This pressure is the sum of static and kinetic

contributions, but the latter exist only in the pipe where the gas is in motion. The lower line is the static pressure only. Bodily gas movement will start just in front of the pipe and be fully developed at some point inside, as indicated by the separation of the total and static pressure lines in the figure. In the vicinity of the entrance plane, the static pressure falls as the gas accelerates and the fall in static pressure that is used to bring the gas into motion is referred to as the "entry loss." In the central part of the pipe, pressure falls due to viscous effects. In the vicinity of the exit plane, the kinetic component of the pressure falls to zero as the gas slows down and emerges into the large volume. There is a smooth variation in static pressure at the exit, but the drop in total pressure at this location may be regarded as an exit loss. Depending, therefore, on which pressures are measured, one can regard the static pressure change associated with acceleration near the entrance as an entry loss, or the change in total pressure at the exit associated with deceleration as an exit loss.

For molecular flow in this geometry, the pressure variations are basically similar but have a different interpretation in detail and are more correctly thought of as variations in molecular number density. There will be a reduction in number density at the pipe entrance because, compared with bulk gas in the volume, some molecules proceed down the pipe. As the net drift velocity of molecules becomes established, it would be possible to measure static and kinetic pressure contributions as before. The steady fall in pressure across the pipe is now attributable to molecular drag. In particular, a gauge located at the exit plane X and looking upstream will measure, say, a higher pressure p_2 than the pressure p_D in the downstream volume. This difference $p_2 - p_D$ may be thought of as an exit loss, associated with the motion of the gas molecules becoming once more totally randomized as the drift motion associated with the flow disappears. At the exit plane the gas about to emerge has to pass through an aperture equal in area to the pipe cross-sectional area, with throughput Q and pressure drop $p_2 - p_D$. Therefore, one can regard the exit loss as associated with an exit conductance equal to $C_A (= 9.3\ D^2)$.

In defining the conductance of the pipe, the pressures referred to are the static pressures in the large volumes that the pipe connects. Any measurement or theoretical derivation of the pipe conductance in terms of the defining Equation 5.7, therefore, includes this exit loss. When a pipe does not terminate in a large volume that enables randomization of the molecular motion, due allowance must be made. This arises when, for example, a pipe is connected to a pump or to another series component. Santeler (1986) discusses these matters.

5.6.7 Molecular Beaming

As molecules make their exit from a pipe into a downstream volume, the spatial distribution of their velocities will have been modified by their passage through the pipe, controlled by the randomizing surface-scattering process.

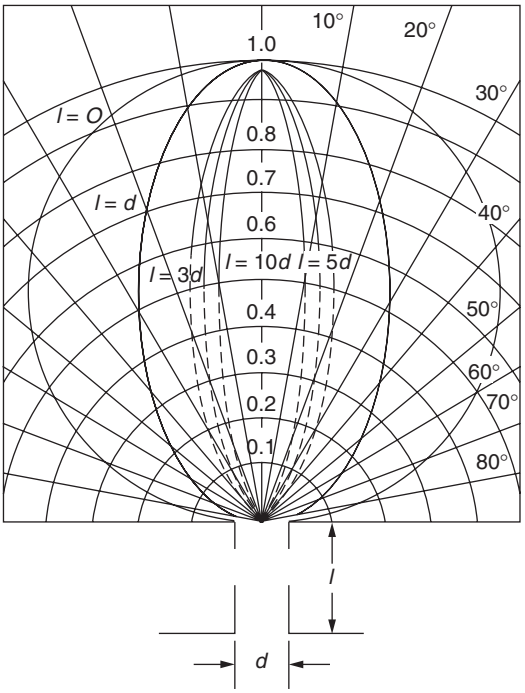

FIGURE 5.13
Angular distribution of molecules from apertures of finite thickness. (From Valyi, L., *Atom and Ion Sources*, Akademiai Kiado, Budapest, 1977. With permission.)

Through an ideal circular aperture, the distribution is, of course, the spherical one of the Knudsen cosine law as discussed in Chapter 4, but for apertures in walls of finite thickness, changing into short pipes with finite L/D ratio, there are beaming effects that cause the distribution to narrow. They were analyzed by Dayton (1957) and some data are presented graphically in Figure 5.13. For $L = D$, there is a substantial change compared with the cosine distribution of $L/D = 0$. For $L = 5D$ the beaming effect is pronounced.

5.6.8 Combining the Conductances of Elements in Series

The foregoing descriptions of exit loss and beaming are important when evaluating the molecular flow conductance of directly connected elements. The problem that arises if the simple "sum of reciprocals" formula of Equation 5.9 is used without care is illustrated by the following example. Suppose that we wish to evaluate the transmission probability α of a short pipe for which $L/D = 2$. From Cole's data in Table 5.1, it is 0.356, to three significant figures. If one regards the pipe as made up of two pipes with $L/D = 1$ connected together and uses the sum of reciprocals formula, then the two equal α values of 0.514 combine to give an α value $0.514/2 = 0.257$, smaller than

it should be, and a serious discrepancy. The reason is that, as noted earlier, the sum of reciprocals formula requires the presence of an intervening volume to randomize the conditions of exit from the first pipe and entry into the second, and such a volume is not present in the direct combination. Furthermore, because the α values of the separate components each incorporate an exit loss, then the combination of the two directly means that an extra exit loss at the join has been included. An exit loss should therefore be subtracted. The transmission probability α of an aperture is simply unity. For the combined pipes with $L/D = 1$, therefore

$$\frac{1}{\alpha} = \frac{1}{0.514} + \frac{1}{0.514} - \frac{1}{1} = \frac{2}{0.514} - 1 = 2.89 \text{ giving } \alpha = 0.346$$

This is much closer to the correct value 0.356. The remaining small difference may be attributed to beaming, which is not easy to compute, though beaming errors are likely to be small, especially when tubes are short. But the importance of removing the extra exit loss is evident.

Oatley (1957) presented a way of evaluating the conductance of series-connected elements in terms of their individually known transmission probabilities, which deals with matters differently. It applies to series-connected elements that have a common area at their join, as in Figure 5.14(a), as distinct from the discontinuous case of Figure 5.14(b), where there is an abrupt change in cross-sectional area.

In Figure 5.14(a), two elements F and G are connected at plane 2 with area A_2. The entrance of F and exit of G have areas A_1 and A_3, respectively. The separately evaluated transmission probabilities in each direction for F are α_{12} and α_{21}, and for G are α_{23} and α_{32}, as indicated. For F or G on their own, the conductance in either direction must be the same, and so

$$\alpha_{12}A_1 = \alpha_{21}A_2 \text{ and } \alpha_{23}A_2 = \alpha_{32}A_3 \tag{5.46}$$

The α's will have been evaluated for random entry conditions into the component and for exit into a randomized molecular distribution in the downstream volume.

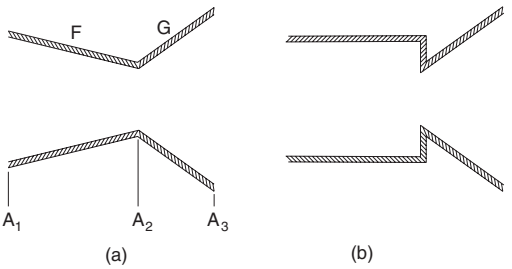

FIGURE 5.14
Series-connected elements. (a) Common junction area, (b) discontinuous area.

Consider N molecules entering F randomly through plane 1. A number $\alpha_{12}N$ will be transmitted to G and $\alpha_{23}\alpha_{12}N$ will be transmitted through G to emerge in the right-hand volume. The number $(1 - \alpha_{12})$N will return to the left-hand volume as though "reflected" at plane 2, and are of no more account. But at plane 3 $(1 - \alpha_{23})\alpha_{12}$N will be returned to plane 2, and of these, a fraction $(1 - \alpha_{21})$ will be returned to travel forward through G again. A fraction α_{23} will be transmitted through plane 3, so that $\alpha_{23}(1 - \alpha_{21})(1 - \alpha_{23})\alpha_{12}N$ will reach the right-hand volume. This partial reflection and transmission continues so that, of the original N molecules, the total number transmitted is

$$N\alpha_{12}\alpha_{23}[1 + (1 - \alpha_{21})(1 - \alpha_{23}) + (1- \alpha_{21})^2(1 - \alpha_{23})^2 + \ldots]$$

This geometric series sums to

$$\frac{N\alpha_{12}\alpha_{23}}{\alpha_{21} + \alpha_{23} - \alpha_{21}\alpha_{23}}$$

If an overall transmission coefficient α_{13} is defined, then

$$N\alpha_{13} = \frac{N\alpha_{12}\alpha_{23}}{\alpha_{21} + \alpha_{23} - \alpha_{21}\alpha_{23}}$$

from which, using Equations 5.46 to express all α's in the direction of flow, gives

$$\alpha_{13} = \frac{\alpha_{12}\alpha_{23}}{\left(A_1/A_2\right)\alpha_{12} + \alpha_{23} - \left(A_1/A_2\right)\alpha_{12}\alpha_{23}} \tag{5.47}$$

When $A_1 = A_2$ so that $\alpha_{12} = \alpha_{21} = \alpha_1$, say, and $\alpha_{23} = \alpha_2$, then overall with $\alpha_{13} = \alpha$, Equation 5.47 may be rearranged to give

$$\frac{1}{\alpha} = \frac{1}{\alpha_1} + \frac{1}{\alpha_2} - 1 \tag{5.48}$$

Thus, considering the "sum of reciprocals" form and the earlier discussion of this section, Oatley's procedure in effect takes care of the absence of an intermediate exit loss automatically.

For discontinuous junctions as depicted in Figure 5.14(b), the method may be further developed. Formulas for some of the more frequently met situations are given, for example, by O'Hanlon (2003) and Livesey (1998).

5.7 The Pumping Process, Pump-Down Time, and Ultimate Pressure

Having introduced the quantities throughput Q and pumping speed S earlier in this chapter, and the subject of vapor release and outgassing from surfaces in Chapter 4, we can now set up the basic equation of the pumping process. Figure 5.15 is a schematic representation of a pumping system.

A vessel of volume V is connected via pipe of conductance C to a pump of speed S^*. The pumping speed at the vessel will be S as given by Equation 5.11. Various sources may contribute to the gas load that has to be pumped. In addition to gas originally in the volume, outgassing from the interior surfaces will commence as soon as the pressure is reduced, as discussed in Chapter 4. Its magnitude may be represented by a throughput Q_G. There may be gas entry into the volume with throughput Q_L by unintended leaks or, in some applications, the intentional steady inflow of a specific gas. At some stage, gas may be produced internally as a result of an operating process for which the system has been designed, and when activated, it will contribute a throughput, say, Q_P, to the load. These contributions are represented schematically in the figure. Let Q_T be the total of all such contributions and any others, such as vaporization, that cause the entry of gas into the volume: $Q_T = Q_G + Q_L + Q_P$ + etc. In some cases, for example, for a system with no leaks in which there are no gas generating processes, Q_T will simply be due to outgassing.

The pumping equation assumes the isothermal conditions normally encountered and expresses the fact that the change in the quantity of gas in the volume V, which is associated with a change dp in the pressure p in a small time-interval dt, must be the difference in the quantities entering the volume and leaving it. Thus in pressure–volume units:

$$V dp = Q_T\, dt - Sp\, dt$$

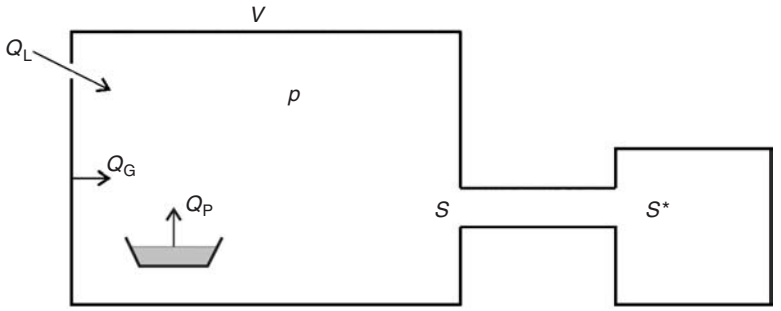

FIGURE 5.15
Schematic representation of a pumping system.

In the context of pumping to evacuate the vessel, the rate of exit of gas exceeds that of entry, and pressure will be falling so that dp and therefore dp/dt are negative. The equation may be written to express the positive rate of reduction of gas in the volume, which is $-V(dp/dt)$, as

$$-V\left(\frac{dp}{dt}\right) = Sp - Q_T \tag{5.49}$$

This differential equation is the fundamental pumping equation, expressing the fact that the rate of change of the amount of gas in the volume at any instant is the difference between the rate of its removal $S \times p$ by the pump and the influx rate Q_T. Although this is an exact equation true at all times t, integrating it to get realistic information about how pressure falls with time is often complicated for a number of reasons. In many applications, pumping speed S at the vessel depends on pressure. This may be due to either the pressure dependence of the speed S^* of the pump itself or, unless flow is in the molecular regime, of the conductance of the connection, or both. Pump characteristics are described in Chapter 6. Secondly, the gas influx rate due to outgassing varies significantly with time as indicated in Section 4.5, and will depend on the previous conditions of use of a system, slowly diminishing as pumping proceeds to a small and sensibly constant value, but only being dramatically reduced in normal experimental times if special procedures such as baking are adopted. There are, nevertheless, two results of prime importance that may be obtained from Equation 5.49. They relate to the lowest pressure achievable and the pump-down time when pumping speed can be considered constant.

It is evident from Equation 5.49 that when pressure eventually ceases to fall so that dp/dt becomes zero, the steady pressure achieved in the vessel, called the system's *ultimate pressure* or its base pressure, and denoted p_u is given by

$$p_u = \frac{Q_T}{S} \tag{5.50}$$

This confirms common-sense thinking that, for a given pumping speed, low pressures will be achieved for small gas loads. Equally, for a given gas load, the best vacuum is obtained for the largest pumping speed. This simple but important equation has applications in various contexts in all ranges of vacuum. Note that it could have been deduced directly from the defining equation $Q = S \times p$ applied to the steady ultimate pressure condition. When steady state is eventually attained, the gas load and the pump's gas handling capacity are in balance. For gas mixtures that are pumped under conditions of molecular flow, for which, depending on the choice of pump, the pumping speed may depend on the identity of the gas, Equation 5.50 would be reexpressed as a sum of ultimate partial pressures, each of which would be the ratio of the throughput for a particular gas divided by its pumping speed.

Returning to Equation 5.49, we may note that in the early stages of pumping, starting at atmospheric pressure, because the system will be free of large leaks and the contribution of outgassing negligibly small, the term $S \times p$ will be very much larger than the term Q_T, which can be ignored. Therefore, with rearrangement,

$$\frac{dp}{p} = -\left(\frac{S}{V}\right)dt \tag{5.51}$$

Furthermore, many of the primary pumps used in these early stages of pumping, and particularly the rotary pump (see Chapter 6), have pumping speeds that are sensibly constant over several decades of pressure, from 1000 down to 10^{-1} mbar or less. Therefore, with S constant, the above equation may be straightforwardly integrated to give

$$p = p_0 \exp\{-(S/V)t\} \tag{5.52}$$

where p is the pressure at time t and p_0 its value at $t = 0$ when pumping starts. Under these conditions, therefore, pressure falls exponentially with time, $p = p_0 \exp(-t/\tau)$, and with a time constant $\tau = V/S$.

Equation (5.52) may be restated as

$$t = (V/S) \ln(p_0/p) \tag{5.53}$$

so that the time taken for the pressure to fall from p_0 to p may be determined. For example, for typical values $V = 50$ l and $S = 1$ l s^{-1}, the time taken for the pressure to fall from $p_0 = 1000$ to 1 mbar is $t = (50/1) \ln 10^3 = 50 \times 2.3 \log_{10} 10^3 = 50 \times 2.3 \times 3 = 345$ s ~ 6 min. The associated time constant V/S is 50 s, and the time per decade $2.3 \times 50 = 115$ s. Within its range of validity, Equation 5.52 may also be used to determine the pumping speed necessary to pump down a volume to a given pressure in a specified time. Thus

$$S = (V/t) \ln(p_0/p) \tag{5.54}$$

If a volume of 1 m^3 = 10^3 l has to be pumped down from 1000 mbar to 10 mbar in 10 min = 600 s, then the necessary pumping speed is $S = (1000/600) \ln(10^2) = 1.66 \times 2.3 \times 2 = 7.5$ l s^{-1} = 450 l min^{-1} = 27 m^3 h^{-1}. The variety of units used reflects common practice. The speeds of primary pumps that work in the low vacuum range are frequently specified in either l min^{-1} or m^3 h^{-1}, whereas for pumps working in higher vacua where flow is molecular, l s^{-1} units are usually used. It is useful to remember that 1 l s^{-1} = 3.6 m^3 h^{-1}.

In Figure 5.16, curve A shows how pressure falls with time under the conditions discussed above, with pressure displayed on a logarithmic scale. The linear region corresponds to constant pumping speed and typically occupies approximately 3 decades down to about 1 mbar or a little less.

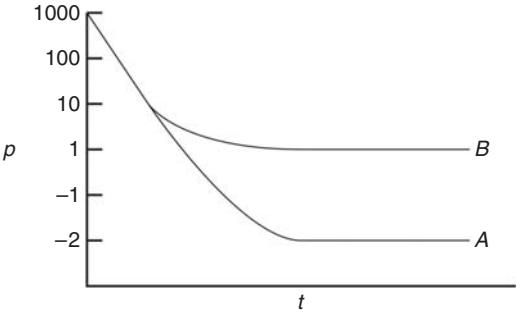

FIGURE 5.16
Fall of pressure with time, with and without a leak.

Thereafter the rate of fall starts to diminish as outgassing from the walls, particularly of water vapor, becomes a significant load comparable with that being pumped directly out of the volume. Also, at pressures of $10^{-1}/10^{-2}$ mbar, the speed of a typical rotary pump begins to fall below its previously constant value. The system pressure will fall to an ultimate pressure of the order of 10^{-3} mbar, slightly greater than that of the pump. In circumstances where a pump of this type handles a purposely introduced and relatively large constant gas load Q_L and the pressure stays within the range of constant pumping speed, we may write $Q_L/S = p_L$, the ultimate pressure reached. Equation 5.49 then becomes, with some rearrangement,

$$\frac{dp}{p - p_L} = -\left(\frac{S}{V}\right)dt$$

which integrates to give, with pumping from $p = p_0$ at time $t = 0$,

$$p = p_L + (p_0 - p_L)\exp\{-(S/V)t\} \tag{5.55}$$

Thus, as shown in curve B, pressure falls on the same time constant (V/S) but to the higher ultimate pressure p_L.

In systems designed to reach high vacuum, pressures of 10^{-4} mbar and less, a secondary high-vacuum pump is usually required to take over the pumping as the falling pressure approaches the lower limit of operation of the primary pump. This is more fully discussed in Chapter 6. Thus, at some stage of the pumping process, the identity of the pump in the schematic diagram of Figure 5.15 changes. The secondary pump, assuming that there are no leaks, will pump the outgassing load to reduce the system pressure to an ultimate $p_u = Q_G/S$. Achieving ultimate pressures of the order of 10^{-6} mbar or a little less in the high-vacuum region typically takes an hour or two. There is no generally valid formula for pump-down time in this region because the outgassing rate $Q_G(t)$ depends markedly on previous surface treatments and operating procedures. In later stages, as the ultimate pressure

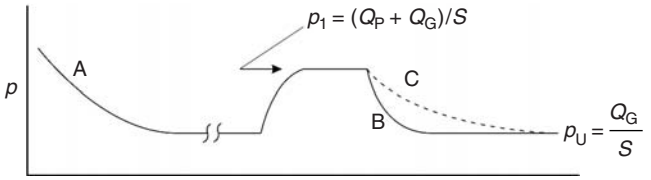

FIGURE 5.17
Response to a temporary increase in gas load.

is approached, the term $V(dp/dt)$ of Equation 5.49 is much smaller than the other two and is ignorable numerically. For example, for a system of volume 50 l pumped with a speed 100 l s^{-1} and achieving an ultimate pressure of 1.0×10^{-6} mbar, the final total gassing load $Q_T = S \times p_u = 10^{-4}$ mbar l s^{-1}. At an earlier time, it is reasonable to assume that a pressure fall of 1×10^{-6} mbar from, say, 4 to 3×10^{-6} mbar, would occur in a time of order 10 min, so that the quantity $V(dp/dt) = 50 \times (10^{-6}/600) = 8 \times 10^{-8}$ mbar l s^{-1}, less by a number of orders than the other two terms, of which at any stage it is the difference. Thus, at these late stages the system pressure is determined by $Q_G(t)$, and one may write $p(t) = Q_G(t)/S$ during the eventual slow approach to p_u. Modeling may be carried out for highly characterized special systems, and the resulting equations are rather complicated. Under the assumption that Q_G falls exponentially with time, Hablanian (1997) gives a formula that illustrates typical behavior reasonably well. The general form of the dependence of pressure on time is shown in Figure 5.17, curve A.

It may be noted that the *form* of Equation 5.55 may be adaptable to other contexts where the pumping speed S can be considered constant and if the only gas to be pumped is that present in the volume. Thus, consider a system of constant speed S (for all gases) at its ultimate pressure $p_u = Q_G/S$ in which a process with gas load Q_P, perhaps caused by local heating, is started. The pressure will rise to $p_1 = (Q_P + Q_G)/S$, and after Q_P is turned off, provided that the gas produced by the process has not adsorbed on the system surfaces so as to change their outgassing properties, it will fall back to p_u according to

$$p = p_u + (p_1 - p_u) \exp\{-(S/V)t\} \tag{5.56}$$

Curve B of the figure illustrates this situation, though it is likely in these circumstances that the recovery of the ultimate pressure will take longer because of the adsorption and subsequent protracted desorption of gases produced in the processing stage, as suggested by curve C.

It may be noted that the ratio of system volume to pumping speed, V/S, may frequently be a useful indicator of the time scale of change in a more general context. Even in circumstances where S is not constant, it may be possible to assign a reasonable average value so that times of change can be estimated to order of magnitude.

5.8 Differential Pumping

In a significant number of applications of vacuum technology, the need arises to operate adjacent parts of a vacuum system at distinctly different pressures so that the result of some process at relatively high pressure (say) can be communicated to another part maintained at a much lower pressure. The difference in pressure may be several orders of magnitude. Examples are the extraction of charged particles from a low-pressure gas discharge and their injection into a region of high vacuum and, conversely, the injection of an electron beam from an electron gun operating in high vacuum into a relatively high-pressure cell containing biological material, as in some applications of electron microscopy. Industrial applications are typified by the process for coating float gas with a thin transparent low emissivity layer for energy-saving purposes. This involves feeding glass sheets up to 2 m wide from atmospheric pressure through a thin horizontal slot into a coarse vacuum from where they are transported on rollers through similar slots to other pressure regions in which they are heated and cleaned in preparation for entry into the region where the coating process is performed. With fabrication completed, they are transported on through further stages to reappear at atmospheric pressure at the equipment's exit.

As a simple illustration of the type of analysis involved consider the arrangement of Figure 5.18. Two vacuum chambers A and B equipped with pumps of speed S_1 and S_2, respectively, are connected by a small aperture of conductance C that allows a direct line of sight connection between them so that some quantity, a beam perhaps, can pass between them without

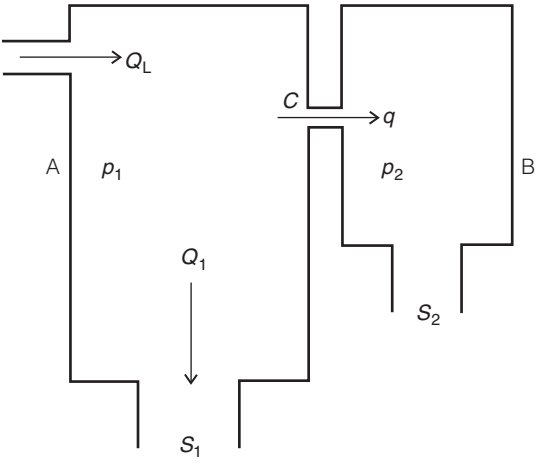

FIGURE 5.18
Differential pumping.

physical obstruction. Suppose that A and B are to be maintained at pressures p_1 and p_2, respectively, with p_2 much less than p_1, and that the process occurring in A requires the entry of gas with a throughput Q_L. From chamber A there will be throughputs Q_1 to its pump and q through the aperture into chamber B, as shown, so that $Q_L = Q_1 + q$, but with $q << Q_1$. In chamber B the pump has to maintain a pressure p_2 while pumping the load q. Hence $S_2 = q/p_2$. (Wall outgassing is ignored for the purpose of illustration, but could be straightforwardly included in the analysis.) In terms of the conductance C, the load $q = C(p_1 - p_2)$, but because $p_1 >> p_2$, this can be taken as just $C \times p_1$. In a practical design the throughput Q_L to sustain the process would be known, and also the pressure p_1 required. Hence $S_1 = Q_1/p_1$ is determined as Q_L/p_1 because Q_L and Q_1 differ by only q. The size of the aperture will be dictated by its function in context and hence its conductance C determined. Thus $q = Cp_1$ is known, and because p_2 is prescribed, the speed of the pump for chamber B is determined as $S_2 = Cp_1/p_2$.

The principles involved in this analysis are applicable in more complicated systems with several stages. The analysis is less problematic when flow is molecular so that conductance values do not depend on pressure.

Exercises

The relatively large number of exercises below reflects the breadth of subject matter in this chapter and its importance for subsequent application in making and measuring vacuum.

5.1 Below what pressure approximately will the flow of air in a pipe 25 mm in diameter cease to be continuum and become transitional? Below what pressure will it become molecular?

5.2 At a particular cross section in a pipe the volumetric flow rate of a gas is 30 l s^{-1} at a pressure 10^{-4} mbar; what is the throughput?

5.3 Convert (a) a throughput of 1 mbar liter per second to Pa m^3 s^{-1} and (b) a volumetric flow rate of 1 l s^{-1} to m^3/h.

5.4 The throughput of oxygen in a certain process is 766 Pa m^3 s^{-1}. Find the mass flow rate in kg s^{-1}.

5.5 In a controlled steady leak into a vacuum system, 1 cm^3 of atmospheric air at 1000 mbar is sucked into the vacuum in a period of 5 seconds. What will be the volumetric flow rate internally if the pressure there is held at 10^{-1} mbar by a pump?

5.6 A pump of speed 300 l s^{-1} holds a vacuum of 2×10^{-5} mbar in a vessel to which it is connected. What is the throughput? What is the pumping

speed at a downstream location in the pumping arrangement where the pressure is 10^{-2} mbar?

5.7 Use the defining Equations 5.6 and 5.7 to prove Equations 5.8 and 5.9.

5.8 A vacuum chamber may be connected to a pump of speed 100 l s^{-1} by pipes of conductance (a) 1000, (b) 400, or (c) 100 l s^{-1}. Calculate the pumping speed at the chamber in each case.

5.9 In a pipe of diameter 50 mm, air flows at a volumetric rate of 10 l s^{-1} and a pressure of 1 mbar. What is the flow regime and is the flow turbulent or laminar?

5.10 Air flows from a region of steady upstream pressure (a) 10 mbar, (b) 1 mbar through an aperture 1 mm square into a region where the pressure is kept below 0.1 mbar by pumping. What will be the pumping speeds and throughputs in each case?

5.11 Two conductances of value 100 l s^{-1} and 80 l s^{-1} are in parallel with each other, and the combination is in series with a conductance of 180 l s^{-1}. What is the conductance overall?

5.12 Use the short pipe formula of Dushman (Equation 5.44) to estimate the molecular flow conductance of a pipe 15 cm in diameter and 75 cm long.

5.13 A component has a molecular flow conductance of 500 l s^{-1} for nitrogen. What will its conductance be for (a) hydrogen, (b) carbon dioxide?

5.14 By what factor will the molecular flow conductance of a long pipe be increased if its diameter is doubled?

5.15 From Table 5.1, the molecular flow transmission probability for a pipe whose length is equal to its diameter is 0.51, so that only about one half of the molecules that enter it pass through. What fraction will get through for a pipe with $L/D = 5$? Use this value to compute the molecular flow conductance for nitrogen gas of a pipe with entrance diameter 15 cm and 75 cm long. Compare this with the estimate of exercise 5.12 above.

5.16 The molecular flow transmission probability of a component with entrance area 4 cm^2 is 0.36. Calculate its conductance for nitrogen at (a) 295 K, (b) 600 K.

5.17 A vessel of volume 4 m^3 has to be evacuated from 1000 mbar to 1 mbar in 20 min. What pumping speed (in m^3 per hour) is required?

5.18 How long will it take for a vessel of volume 80 l connected to a pump of speed 5 l s^{-1} to be pumped from 1000 to 10 mbar? What is the time per decade?

5.19 What time constant is associated with the pumping of a vessel of volume 60 l with a pump of speed 300 l s^{-1}?

5.20 In a particular high-vacuum application, the total gas load from the chamber is 3×10^{-4} mbar l s^{-1} and the pump which is directly connected to it has a speed 150 l s^{-1}. What is the ultimate pressure?

5.21 An inventor claims to have designed a new type of high vacuum pump that has a pumping speed of 1350 l s^{-1} for nitrogen and an inlet port diameter of 10 cm. On what grounds do you doubt his claim?

Problems

5.22 A pipe of conductance 250 l s^{-1} is attached to a pump of speed 50 l s^{-1}. Calculate the ratio of the upstream to the downstream pressure and the upstream pumping speed at the pipe inlet.

5.23 Use Equations 5.40, 5.44, and 5.37, respectively, to evaluate and compare the molecular flow conductances for nitrogen of (a) a circular hole of diameter 5 cm in a thin plate, (b) a pipe 20 cm long and 5 cm in diameter that therefore has the same entry diameter as (a), and (c) a pipe of this bore 1 m long.

5.24 A vacuum chamber of volume 0.03 m^3 and internal area 0.6 m^2 is connected by a pipe of conductance 1600 l s^{-1} to a pump of speed 400 l s^{-1}. Calculate the pumping speed at the chamber. If the ultimate pressure achieved is 2×10^{-8} mbar, calculate the gas throughput and hence estimate a gassing rate per cm^2 of internal surface. If the pumping action suddenly ceases due to the closure of a valve above the pump, how much time will elapse before the pressure rises to (a) 10^{-6} (b) 10^{-5} mbar? What time constant may be associated with the pumping when it is restored?

5.25 Explain what you understand by the term "outgassing" of a surface. Define the specific outgassing rate and give a typical value for a well-precleaned surface that has been in vacuum for 1 h. Explain how this may be significantly reduced by the procedure of "baking," and why this is necessary if ultrahigh vacuum of about 10^{-10} mbar is to be achieved.

5.26 A right-angled elbow component made from tube of internal diameter 20 cm consists of two sections whose central lengths are 40 and 30 cm. Use the graphical data of Figure 5.11 to determine its transmission probability and hence molecular flow conductance. Compare this transmission probability with that of a straight pipe of the same diameter and length which from Cole's data is 0.248.

5.27 Consider the venting of a vacuum chamber, i.e., the introduction of atmospheric air to bring it back up to atmospheric pressure. This is

done by opening a small valve that is often directly connected to the ambient atmosphere. In the early stages of venting, because of the small pressure ratio, flow is choked, as discussed in Section 5.5.3. For a chamber of volume V initially at high vacuum, assume that venting is effectively through a small aperture of area A that connects it to the atmosphere at pressure p_0, and derive an expression for the time taken for the chamber pressure to rise to the value at which choked flow ceases and beyond which normal unchoked flow occurs. Show that for a chamber of volume 100 l and a venting valve of effective aperture 4 mm^2, this time is 66 s. The total venting time is roughly about twice this value. Livesey (1998) gives the formula for the total venting time and comments on the small temperature rise that occurs in venting, which is due to two opposing effects — the slowing of gas that has acquired energy in being accelerated into the system and the adiabatic cooling downstream of the aperture.

5.28 A vacuum line component has the shape of a truncated cone whose smaller (exit) diameter is half its entrance diameter. It has a molecular flow transmission probability of 0.4 (see O'Hanlon [2003], p. 43, for example). Its exit is connected to a short pipe of matching diameter whose transmission probability is 0.25. Calculate the transmission probability of the combination.

5.29 Consider the passage of molecules in molecular flow through a component in the form of a truncated cone as described in the previous question. Let subscript 1 refer to the larger diameter and 2 the smaller, so that area A_1 is greater than A_2. Transmission probabilities are, say, α_{12} and α_{21}, and the laws of physics require that $A_1 \times \alpha_{12} = A_2 \times \alpha_{21}$. For flow in both directions, draw circles to represent cosine law scattering at appropriate points on the cone surfaces in order to give a physical interpretation of this result.

6

Creating a Vacuum — Pumps

6.1 Introduction

The function of a vacuum pump is to withdraw gas from a designated volume so that the pressure is lowered to a value suitable for the purpose in hand. The variety of applications of vacuum technology is such that a very wide range of vacua has to be provided. Thus a very modest vacuum of only 100 mbar is sufficient to enable large forces based on the 900-mbar pressure difference with atmospheric pressure to be used in vacuum molding. But a pharmaceutical application may require the maintenance of 10^{-2} mbar with high flow rates of water vapor, and the vacuum coating of mirror surfaces by the condensation of metallic vapor from hot vaporization sources needs a high vacuum of less than 10^{-5} mbar to be held with large outgassing loads. An ultrahigh vacuum ~10^{-10} mbar is necessary to secure adequately clean conditions for investigations in surface science.

The types of pumps that are available to meet these needs may be usefully classified under just three headings that summarize what they do to the gas presented to them.

In **positive displacement pumps,** gas is manipulated using repetitive mechanical movements of driven parts and synchronized valve actions that displace it from the inlet to the outlet in relatively small discrete amounts at a high repetition rate and with some compression. The rotary vane pump is an example of prime importance. Others are the Roots pump, and those based on the movement of diaphragms, claws, screws, and scrolls.

In **momentum transfer pumps,** incoming gas molecules interact with either a high-velocity stream of fluid or very fast-moving solid surfaces that add a directed component to their motion and transfer them continuously to an outlet, usually at a pressure much below atmospheric. Examples are vapor jet (diffusion) pumps, drag pumps, and turbomolecular pumps.

In **capture pumps,** molecules are removed from the gas phase by being trapped on surfaces by the physical or chemical processes of condensation and adsorption. Cryogenic pumps, sublimation pumps, and ion pumps, in which the capture process is assisted by the presence of electric or magnetic

fields, are examples. In these cases there is no pump outlet and the pumped gas is stored in a condensed state.

In both displacement and transfer pumps, gas is compressed to arrive at the outlet still in a gaseous condition, and they are therefore called throughput pumps, in contrast to those whose action depends on capture and storage.

Positive displacement pumps operate mostly in the range from 1000 down to 10^{-3} mbar or a little less and usually handle gas in a continuum, fluidic state, as previously discussed. Except for rough vacuum applications where a single positive displacement pump of appropriate size may be suitable for the purpose in hand, applications requiring better vacuum normally require more than one type of pump, and a sequence of operations is needed to attain the required vacuum. In these circumstances a primary pump, defined as one that exhausts to atmospheric pressure, and very frequently a rotary pump (see Section 6.2.1), has the role of removing the air from the pumped volume, a procedure called "rough pumping" or simply "roughing." When the pressure has fallen from 1000 mbar to 10^{-1} mbar (so that 99.99% of the air initially present has been removed), a secondary pump is brought into operation. The change from the primary to the secondary pump is referred to as "crossover." Secondary pumps, as well as needing low pressure at their input in order to function, cannot exhaust to atmospheric pressure and so have to be operated in tandem with a primary pump that compresses their output to exhaust it at atmospheric pressure. By networking and valving, it is frequently arranged that this is the same primary pump as did the initial roughing, now serving in what is called a "backing" role. These matters will be exemplified later.

Pumps creating high vacuum operate in conditions of molecular flow (Kn > 1) as discussed in Chapter 5. Because of the absence of fluidic behavior and the dominance of randomizing molecule–surface collisions, there is no way in which a pump working in this regime can "suck" gas into it. Rather, when gas molecules arrive by chance, there is a mechanism that "disposes" of them by either momentum transfer to an outlet or by capture, thus preventing their return to the pumped volume. If the mechanism were 100% efficient, then all molecules entering the pump would be disposed of and, following the discussion of Section 5.6.2, the pumping speed would be equal to the conductance of the entrance, i.e., $Q = S \times p = C_A p$ and therefore $S = C_A$. For nitrogen, from Equation 5.40, $C_A = C_0 = 9.3D^2$ l s^{-1} for an aperture D cm in diameter. For $D = 10$ cm, therefore, the maximum possible pumping speed would be 930 l s^{-1} for nitrogen. Inlet ports of this size are typical of many laboratory pumps and we will see that their measured speeds for nitrogen are indeed a few hundreds of liters per second. A measure of the pumping efficiency for a particular pump is therefore the ratio of its measured pumping speed to the conductance of its inlet port, frequently a circular opening in a round flange. This ratio is called the speed factor H, and naturally has a maximum value of unity. This is also true of the Ho coefficient, which is the ratio of the measured pumping speed to the conductance of the active pumping area, which may be an annular area at the

pump inlet, for example, the inactive area being occupied by part of the pumping mechanism. The Ho coefficient therefore measures the capture and disposal probability for the pumping mechanism. Further discussion of its application and significance will be found in O'Hanlon (2003).

Unlike many domestic and industrial pumping applications in which fluid moves through pumps with relatively small pressure ratios between pump inlet and outlet, most vacuum pumping applications involve removing gas from a vessel of fixed volume to create and sustain a large pressure ratio between the exhaust at atmospheric pressure and the pumped volume. A diagram showing a schematic "pressure profile" of pressure along the gas flow pathways through a vacuum system using throughput pumping therefore resembles that of Figure 6.1. On the vertical axis, pressure has a logarithmic scale with atmospheric pressure as the upper limit. The region AB represents the substantially constant pressure of gas in quasi-equilibrium in the pumped volume and BC the fall in pressure associated with the flow of gas from this volume through a pipe to the inlet of the secondary pump S. Work is done by the secondary pump to effectively compress the gas from its inlet pressure p_C to its outlet at p_D, with compression ratio, as drawn, of 10^5. Segment DE corresponds to the high conductance path to the backing pump R that then compresses from p_E to a pressure slightly greater than atmospheric, with compression ratio 10^4, and expulsion. Overall, starting from the vessel, any given mass of gas is conveyed to the outlet with progressive compression by the pumps. There is continuity of throughput $Q = S \times p$ along the flow path, and at places where a pressure can be clearly defined and measured, at the interconnection between pumps, for example, volumetric flow rates can be deduced. As well as conveying gas from low to high pressure, the pumps also prevent gas flowing from high to low, as is its natural tendency, and thus can be thought of as providing barriers to reverse flow.

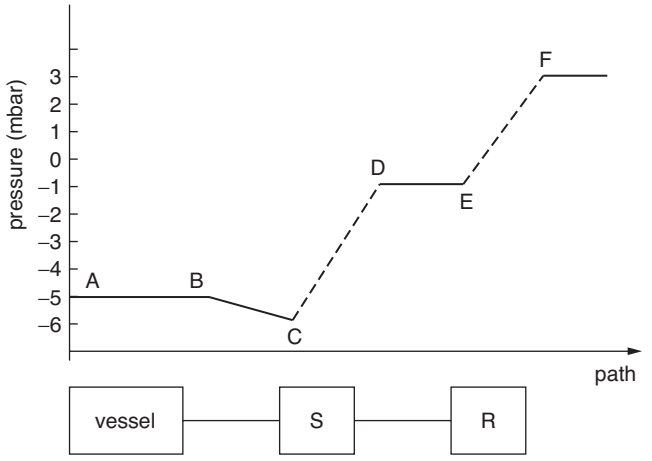

FIGURE 6.1
Pressure profile along the direction of gas flow.

The two principal measures of performance usually quoted for vacuum pumps are their volumetric pumping speed S and the best vacuum achievable, their ultimate pressure. The latter is specified with the inlet port closed by a standard test dome and may be regarded as the condition in which the effective pumping speed has fallen to zero because the pumping action is balanced by internal gas generation in the pump. The additional loads present when the pump is connected to a system cause the system to have a higher ultimate pressure. Pumping speeds S are presented as a function of inlet pressure and for many pumps are constant over a substantial region of the useful working pressure range. Depending on the operating mechanism, speeds may depend on the type of gas pumped, as will be discussed.

6.2 Positive Displacement Pumps

6.2.1 The Rotary Vane Pump

The construction of the basic form of this pump is shown in Figure 6.2. A metallic stator block immersed in oil contains a cylindrical pumping chamber in which a rotor is mounted off-center. The rotor has diametral slots containing sliding vanes that are pushed outwards by springs to make contact with the stator wall. The whole assembly is precisely machined and assembled, and the distance between the rotor and stator surfaces along the line where they are closest at T is a few hundredths of a millimeter. The movement of parts and gravity ensure that oil is drawn into the pumping chamber to cover all internal surfaces, lubricating parts that are in relative motion and providing seals by the closure of small gaps. There are thus two seals that rotate with the vanes and a stationary top seal made by the oil closing the gap at T. This seal separates the inlet side from the outlet side. The port that is connected to the vessel being pumped leads gas to the inlet of the pumping chamber. The exit from it leads to the gas outlet at the exhaust valve. The latter is usually a simple flap below the oil level designed to open when the internal pressure is somewhat greater than atmospheric, at 1200 mbar or more. The rotor is coupled directly to an electric motor that turns at about 1500 rpm. The motor has a slightly unbalanced load, but with vanes made of a lightweight plastic material, the levels of vibration induced are small. Centrifugal forces contribute to keeping the vanes in contact with the stator wall.

The pumping action is illustrated in Figure 6.3 and may be briefly described as "induction, isolation, compression, and exhaust." Referring to Figure 6.3(a), the half turn of the rotor that concludes with vane V being in the position shown induces air into the pumping chamber. The total volume available to the gas increased and so it expanded to occupy the available volume. In Figure 6.3(b), the movement of vane V past the inlet isolates the

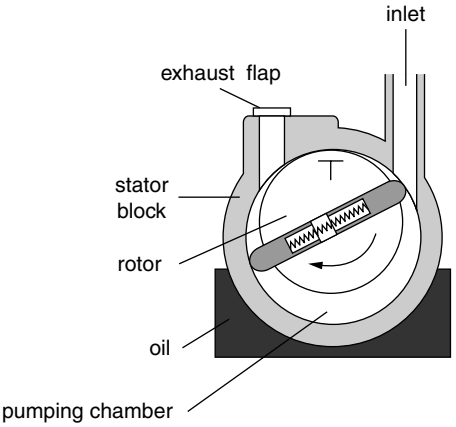

FIGURE 6.2
Structure of a typical oil-sealed rotary pump.

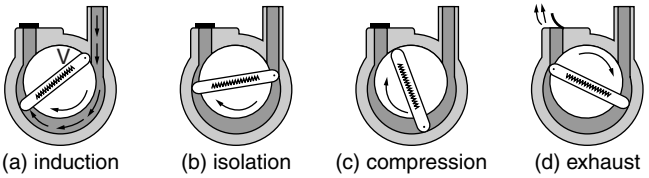

FIGURE 6.3
Pumping sequence. (From Harris, N.S., *Modern Vacuum Practice*, 2nd ed., Nigel Harris Publisher, 2001. With permission.)

induced air in a crescent-shaped volume. Further rotation, as shown in Figure 6.3(c), results in the reduction of this volume and compression of the air with a rise of pressure and temperature to a point shown in Figure 6.3(d) where the pressure is sufficient to open the exhaust valve so that the air is expelled. This sequence is completed twice per revolution. Energy is expended not only to do work against inertial and frictional forces but also to compress and move the pumped gas, and so pumps run hot, typically at about 75°C. As well as lubricating and forming the seals, the oil also has the important role, as it slowly moves through the pump, of taking away some of the heat generated in the interior. In some versions of the pump the stator exterior is not immersed in oil but is finned and air-cooled, and the oil necessary for the pump's operation is drawn from a reservoir located on top of the pump and circulated appropriately. Pumps are nowadays fitted with "anti-suckback" devices that prevent pump oil being drawn toward the vacuum chamber when they are switched off.

When air that contains water vapor is being pumped, the possibility arises that compression of the air–vapor mixture will cause the condensation of the vapor to liquid droplets. This will happen at a stage of compression

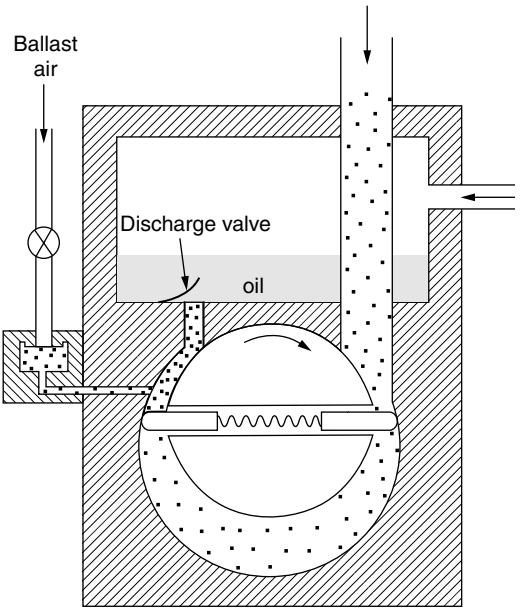

FIGURE 6.4
Gas ballasting.

when the partial pressure of water vapor in the diminishing isolated volume reaches the saturation vapor pressure at the prevailing temperature. Some of the condensate will mix with and degrade the lubricating oil with undesirable effects. Lubrication will be impaired and temperatures rise. To overcome this problem, which arises frequently in many applications, Gaede suggested the use of the so-called gas-ballasting procedure. In this arrangement, whose implementation is illustrated in Figure 6.4, atmospheric air is admitted into the post-isolation compression volume in a controlled way so that the pressure necessary to open the exhaust valve is reached and the gaseous mixture expelled before the condition for condensation is reached. The facility may be switched on and off as necessary.

The dimensions of the pumping chamber and the rotation rate n determine the gas-handling capacity of the pump. The theoretical free-air displacement, which is the volumetric flow rate when it pumps air at atmospheric pressure with no pressure differential, is just 2 nv, where v is the maximum isolated volume. But this assumes quasi-static conditions that might apply if the rotation rate were very slow, and thus ignores dynamical effects in compression and also the necessity to force open the exhaust flap. The volumetric efficiency is therefore less than 100%. Measured pumping speeds have constant values (of about 80% of the theoretical free air displacement) down to inlet pressures of about 10 mbar, below which they start to diminish, reaching

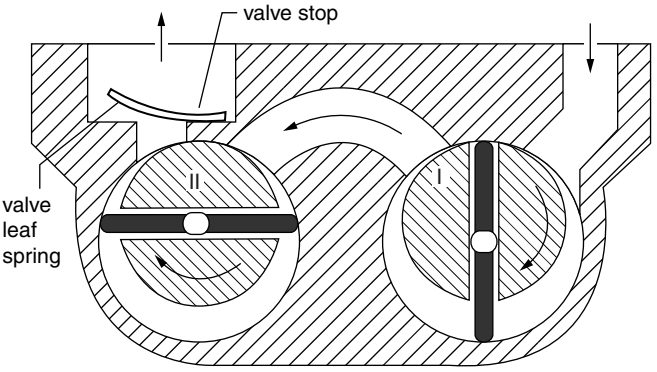

FIGURE 6.5
Arrangement for two-stage pumping. (Used with permission of Leybold, Ltd.)

zero at an ultimate pressure of less than 10^{-1} mbar typically. This is determined principally by the back leakage through the top seal T of Figure 6.2, across which the pressure differential exceeds atmospheric twice per revolution. Air dissolved in the oil is also carried across this seal and released, giving another contribution to back-flow.

To improve the ultimate pressure of the single-stage pump just described, two-stage pumps utilize two similar stages connected in series but with no intermediate exhaust valve, as shown in Figure 6.5. Both are driven by the same shaft, and at low inlet pressures the differential pressure across the top seal of the vacuum-side chamber is much reduced so that ultimate pressures better than 10^{-3} mbar may be achieved. It is also arranged that the oil flow that lubricates and seals this stage is well degassed.

At the inlet of a pump, the migration of the vapor of the hot lubricating oil or its gaseous breakdown products "upstream" against the general direction of flow will occur to some extent when gas flow into the pump ceases to be viscous and will become more significant if molecular flow conditions are reached. For a 25-mm pipe, this will be at about 10^{-3} mbar. In applications for which this would be detrimental, a foreline trap containing a sorbing material is fitted. At the pump outlet also, environmental concerns often demand the use of an oil mist filter prior to discharge from the atmosphere. The properties and choice of oils for these pumps, including these matters, are discussed by Harris (2005).

Pumping speed curves for pumps of typical laboratory size are shown in Figure 6.6 and include the curves for operation with gas ballast, which naturally differ, reflecting the changed conditions when this facility is used. While speeds of the order of 10 m^3 per hour (approximately 3 l s^{-1}) are typical for a laboratory-sized unit, these pumps are available in a wide range of sizes, with speeds up to a few hundred m^3 h^{-1}.

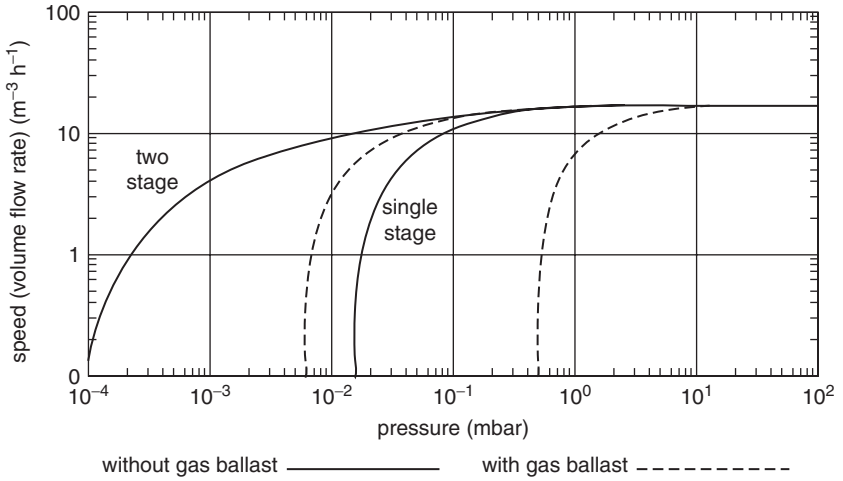

FIGURE 6.6
Typical pumping speed curves for single- and two-stage rotary pumps.

6.2.2 The Roots Pump

These pumps have high volumetric flow rates in the pressure range from 10 down to 10^{-2} mbar. The cross-sectional diagram of Figure 6.7 illustrates their construction. Two lobed rotors with an approximate "figure eight" shape rotate in opposite directions in exact synchronism driven by a motor and gearing external to the vacuum space. The rotors are enclosed in an oval-shaped stator housing with inlet and outlet ports and are of such a shape that in all positions the gaps between them and between each rotor and the stator wall are very small, typically a few tenths of a millimeter. Because there are no parts that touch there is no need for lubricants, and the rotors can turn at high speeds, typically about 3000 rpm, compensating for the fact that the gaps across which pressure differences become established cannot be perfect moving seals. Subject to the integrity of the seal between the vacuum space and the drive unit, which has to be extensively lubricated, it is thus a "dry" pump; there are no lubricants in the pumping volume whose vaporization could allow contaminants to enter the vacuum.

How gas is displaced can be seen in the sequence of diagrams of Figure 6.8. Part (c) shows the rotors in a horizontal "T" configuration that identifies the volume isolated at the pressure of the inlet. For each revolution of the two synchronized rotors, a volume equal to four times this is swept (ideally) from inlet to outlet. The ideal, zero leakage, pumping speed is therefore the number of revolutions per unit time multiplied by this volume. It is the free air displacement of the pump when there is no pressure differential Δp between inlet and outlet, and therefore no reverse flow, and is denoted S_D. These pumps, also called "lobe pumps" or "Roots blowers," were originally designed to move gas at high volumetric rates at atmospheric pressure

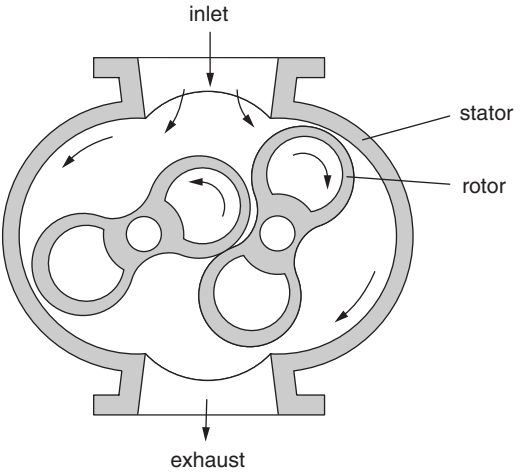

FIGURE 6.7
Geometry of the Roots pump. (From Harris, N.S., *Modern Vacuum Practice*, 2nd ed., Nigel Harris Publisher, 2001. With permission.)

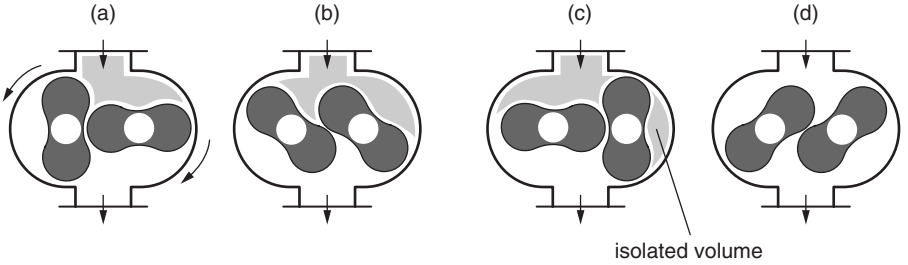

FIGURE 6.8
Pumping sequence. (From Harris, N.S., *Modern Vacuum Practice*, 2nd ed., Nigel Harris Publisher, 2001. With permission.)

against small pressure differentials, and the principle is still widely used in such nonvacuum applications.

In vacuum applications, gas is moved from the inlet at low pressure to the outlet at higher pressure, and therefore there is compression. Both this and viscous friction cause heating of the gas. The rotors and the stator housing become hot in cooling the gas, but heat is taken away externally from the stator housing rather more effectively than it is from the internal rotors and so there is differential thermal expansion and the real danger of mechanical seizure if the rotors expand excessively. To combat this, the rotors may be cooled by an internal oil flow or, with less efficiency, by cold surfaces placed in the flowing gas near the rotors on the outlet side. The effect is sufficiently large that pumps cannot run at full speed and at high pressures with high compression ratios for any reasonable length of time. The most common way

to run the pumps and to alleviate these problems is to back them with a rotary pump whose speed is typically one tenth or more of the Roots pump speed. These matters are further discussed in due course.

The compression ratio achieved by the pump, defined as the ratio p_o/p_i of outlet to inlet pressures, varies with pressure. Because of the necessary gaps at the seals, there is a backflow across them from outlet to inlet, which is driven by the pressure difference $p_o - p_i$ and which depends on the gap conductance. As we have seen in Chapter 5, under viscous conditions conductances depend on pressure and in this case will be greater, the greater the mean pressure in the seal. At lower pressures the conductance diminishes and will tend to become independent of pressure as molecular conditions are approached. A second source of back-flowing gas is desorption off the rotor surfaces at the inlet side of gases that were adsorbed at an earlier stage of each revolution at the higher pressure outlet side.

Based on the foregoing, the throughput Q of gas through the pump when the inlet pressure is p_i may be expressed as

$$Q = S_D p_i - Q_B \tag{6.1}$$

in which Q_B represents the back leakage and $S_D p_i$ is the ideal flow with zero leakage.

An important characteristic of the pump is the zero-load compression ratio K_0 and its variation with outlet pressure. It is measured as follows. The Roots pump inlet is blanked off and a rotary pump connected to its outlet to serve as a backing pump. Gauges are located at inlet and outlet and air is let into the backing line via a leak valve so that by adjusting the leak rate and if necessary throttling the backing pump the outlet pressure p_o of the Roots pump can be set at different values over a wide range. Inlet pressures are recorded as the outlet pressure is varied and compression ratios $K_0 = p_o/p_i$ are plotted against p_o, resulting in the variation shown in Figure 6.9. Typically, K_0 has a maximum value of about 40 at approximately 1 mbar and decreases to values less than 10 at both 10 and 10^{-2} mbar, reflecting dominance of the leakage back flow at higher and gas return by desorption at the lower pressures, respectively.

In the blanked-off zero-flow condition, $Q = 0$, and therefore, from Equation 6.1, $Q_B = S_D p_i$. Because S_D is a known quantity supplied by the manufacturer together with the K_0 data, this means that when the outlet pressure is p_2, Q_B can be determined as $S_D \times p_2/K_0$. This information may be used to determine the performance of the Roots pump when it is backed by a rotary or other pump, as shown in Figure 6.10, in which L and B represent the Roots and backing pump, respectively, the gas flow is a throughput Q, and the backing pressure is p_2. For the backing pump, the pumping speed S_B as a function of its inlet pressure will be known and have the general form displayed in Figure 6.6. What is to be determined is the speed of the Roots pump L as a function of its inlet pressure p_i.

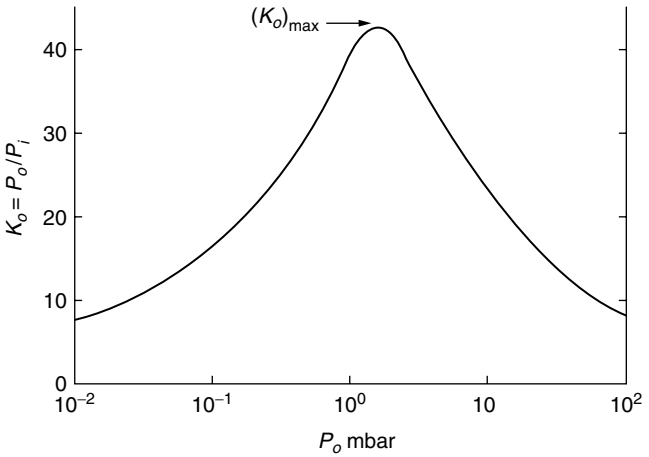

FIGURE 6.9
Typical compression ratio data for a Roots pump.

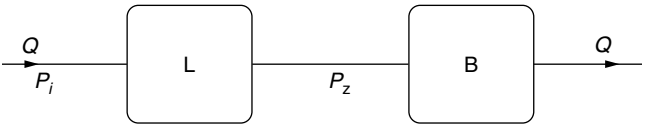

FIGURE 6.10
Roots pump and backing pump in tandem.

Considering the continuity of throughput Q and the Equation 6.1 of the Roots pump, we have

$$Q = S_{\mathrm{D}} \times p_i - S_D \times p_2/K_0 = S_{\mathrm{B}} \times p_2$$

From

$$S_{\mathrm{D}} \times p_i - S_D \times p_2/K_0 = S_{\mathrm{B}} \times p_2$$

in which, for a particular p_2, all but p_i are known, we get by rearrangement

$$p_i = p_2 \left(\frac{1}{K_0} + \frac{S_B}{S_D} \right)_{p=p_2} \tag{6.2}$$

from which, because $S = Q/p_i$ and $Q = S_{\mathrm{B}} p_2$

$$S = \frac{S_B S_D K_0}{S_D + S_B K_0} \tag{6.3}$$

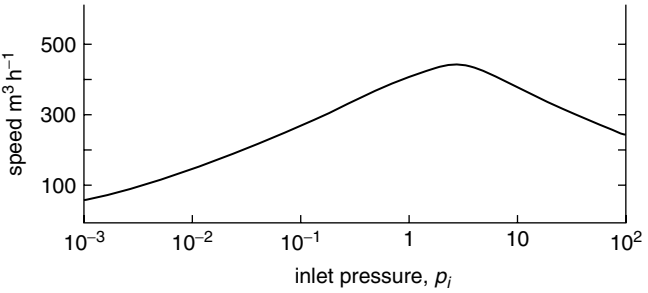

FIGURE 6.11
Pumping speed characteristic of a Roots pump backed by a rotary pump.

Thus, the performance of the Roots and rotary vane pumps in combination may be determined. Figure 6.11 shows a typical characteristic. Roots pumps are available in a wide range of sizes from various manufacturers, with speeds ranging from 10 to 1000 $m^3\ h^{-1}$.

As earlier mentioned the heat generated in these pumps may be considerable and therefore in the earliest stages of pumping from atmospheric pressure, a by-pass line may be used or alternatively a slipping clutch drive so that the Roots pump does not start pumping at full motor power until an inlet pressure of about 10 mbar has been achieved. Modern developments in motor power unit design may also allow the pump to start slowly at high pressure thus preventing overloading.

Although the Roots pump just described is a dry pump, it does, in most applications, require the support of a backing pump, frequently an oil-sealed rotary vane pump. The combined pumping characteristic gives useful performance up to atmospheric pressure.

6.2.3 Other Dry Pumps

In recent years, in response chiefly to the needs of the semiconductor processing industry, a range of dry oil-free primary pumps has evolved that will exhaust to atmospheric pressure and achieve low-pressure performance sufficient to back secondary pumps. They have essentially one of two roles: either to enable totally clean contamination-free pumping or as an alternative way of moving gas in processes that produce gritty gas-borne particulates or chemically aggressive gases, conditions that are seriously harmful to the operation of oil-sealed pumps. These matters are succinctly discussed by Troup and Dennis (1991). Notable among these dry pumps are modern versions of the diaphragm pump and those based on the screw, claw, and scroll principles.

The principle of the ***diaphragm pump*** is illustrated in Figure 6.12(a). A flexible elastomer diaphragm acting as a piston is pulled and pushed by a motor-driven reciprocating action, in conjunction with synchronous valve actions, displacing gas from inlet to outlet. Figure 6.12(b) shows a practical

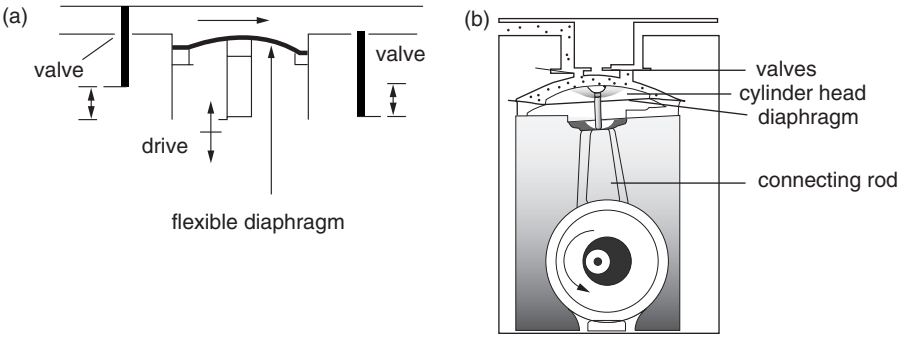

FIGURE 6.12
Diaphragm pump (a) principle and (b) practical design. (Part b, Vakuubrand GMBH from *Vacuum Solutions*, July/August 2000, Institute of Physics Publishing, Bristol. With permission.)

arrangement in which the inlet and outlet valves are leaf seals actuated by the pressure differences across them. The "dead" volume that occurs when the piston is in the uppermost position and the inlet valve is closed is minimized. A single stage will exhaust to atmospheric pressure with an inlet pressure of about 80 mbar, achieving a compression ratio of about 12. The volumetric flow rate, reflecting the limited range and speed of movement of the flexing piston, is of order m^3 per hour. When two stages are used, coupled as shown in Figure 6.13(a), ultimate pressures of about 5 mbar are achieved, as shown in the typical pumping speed curve of Figure 6.13(b). With four stages, ultimate pressures below 1 mbar are possible.

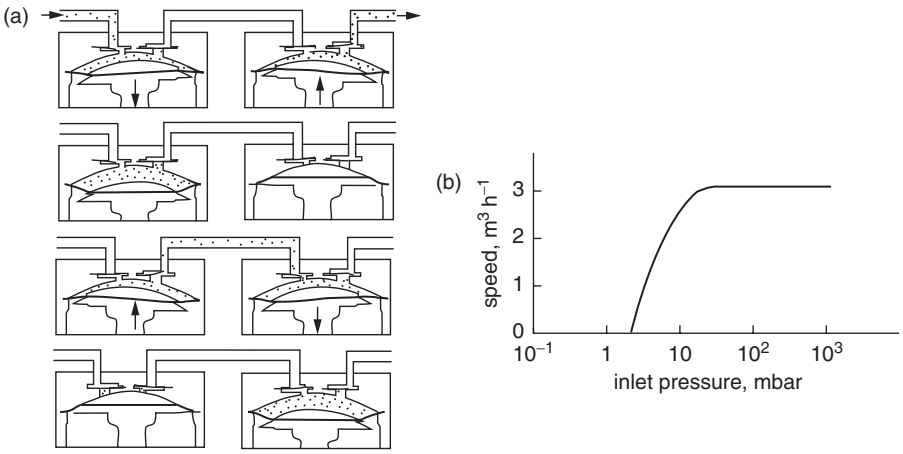

FIGURE 6.13
(a) Two-stage diaphragm pump, (b) pumping speed curve. (Part a, Vakuubrand GMBH from *Vacuum Solutions*, July/August 2000, Institute of Physics Publishing, Bristol. With permission.)

Diaphragm pumps with pumping capacity at 10 mbar inlet pressure are able to back secondary turbomolecular pumps whose outlet pressures have been raised to about this value by the addition of a molecular drag stage (see Section 6.3.3). Because a throughput at 10 mbar has a volumetric flow rate of one hundredth of that throughput at 0.1 mbar, which is the outlet pressure of a normal turbomolecular pump without a drag stage, there can be considerable-size saving in the choice of a matching primary pump. The achievement of totally clean UHV by the use of a diaphragm pump to back a turbo-drag pump has been an important recent development in contamination-free UHV pumping.

Like the Roots pump, the ***claw pump*** is based on two counter-rotating synchronized rotors inside a close-fitting stator housing. The drive unit is external to and sealed off from the pumping chamber, which is therefore dry. As shown in Figure 6.14(a), both rotors R1 and R2 have a claw C protruding outside the cylinder radius with an adjacent matching cavity inside it. As they rotate in synchronism, the claw tips run close to the stator housing and the claws move with small clearance through the cavities of the partner rotors for part of their rotation. The perspective view of Figure 6.14(b) identifies the inlet and outlet ports that are openings in opposite flat faces of the stator housing. As the rotors rotate, these ports are covered or uncovered according to the position of the rotor cavities, thus creating a self-valving arrangement. Rotor R1 controls inlet access to the pumping chamber and R2 the outlet. Let us imagine that the inlet is in the uppermost face of the stator as we view the figure, so that the outlet is below R2 in the face below.

The pumping action may be deduced by considering Figure 6.14(c). In the figure, the claws are evident and the dotted darker region represents gas coming into the pump. The cycle may be represented, starting at the upper left at representation 1, as "induction, compression, isolation, exhaust, and carryover." The representations are numbered 1 to 11, with the numbers placed adjacent to the position of the claw of rotor R1. Going around clockwise from 1 to 5 is induction; the cavity in R1 is open to the inlet and gas in the pumped volume expands to fill the increasing volume available to it. At 5, claw R1 has almost sealed off the inlet. At 6, compression of isolated gas is commencing; 7 is a further stage of compression and R1 has completed one revolution. At 9, the cavity of rotor R2 has just uncovered the outlet opening and gas is expelled. This continues and, at 10, is nearly complete. Between 11 and 1, when R1 has completed a second rotation, a small volume of gas is necessarily carried over. Between 1 and 6, gas in the lighter colored zone is undergoing the compression exhaust and carry-over portrayed in Figure 7 to Figure 11.

Rotors turn typically at 2000 rpm. Because of the carryover volume, compression ratios are limited to values of about 10, as may be guessed by inspection. Multiple stages, 3 or 4, enable inlet pressures less than 1 mbar with exhaust to atmospheric pressure, and thus compression ratios greater than 10^3. The greatest speed is achieved at about 1 mbar and has a similar

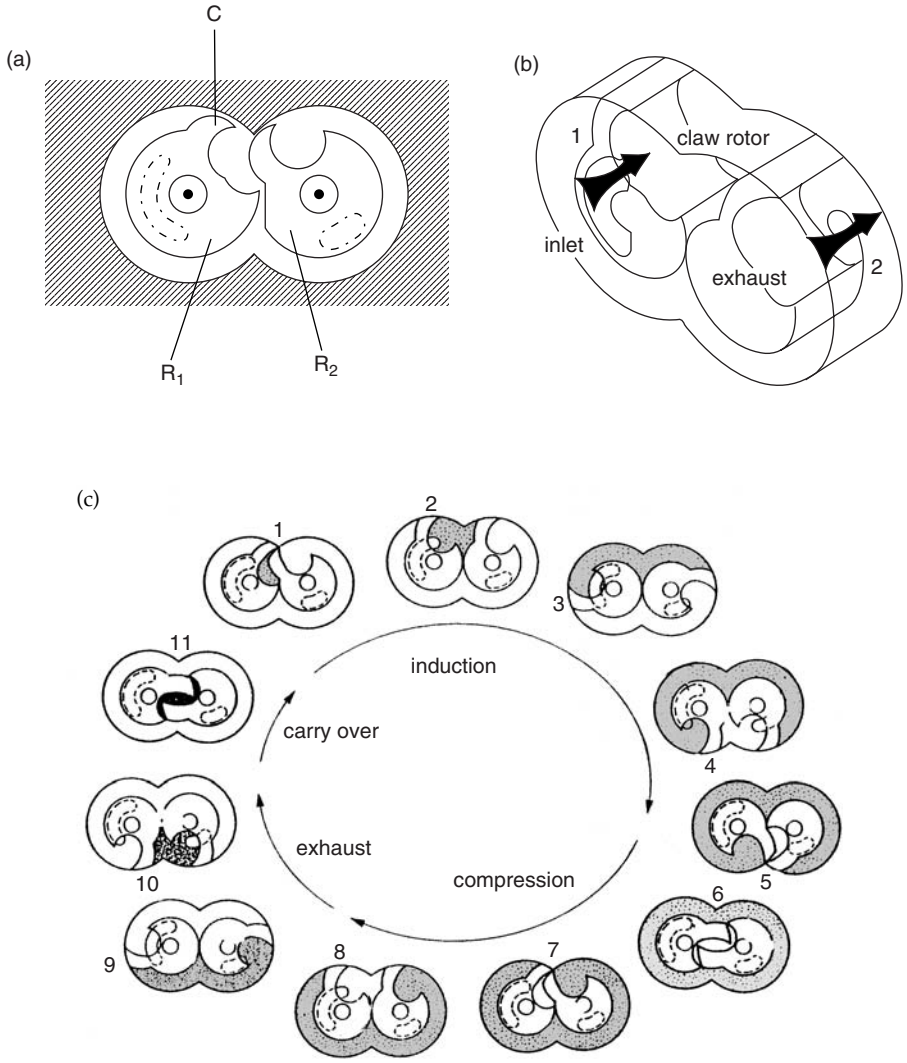

FIGURE 6.14
Claw pump: (a) geometry, (b) perspective view, (c) pumping sequence. (Used with permission of BOC Edwards High Vacuum.)

shape to that of the Roots pump. In this dry pump, of course, there must be back leakage across the small clearances that permit high rotation speeds without friction.

The action of the ***screw pump*** may be understood as follows. Imagine two enmeshed parallel screw forms with pitch $2d$ but opposite helicity as shown in Figure 6.15(a), consisting of raised parts of square section $d \times d$ on a host cylinder of diameter D. A helical channel thus runs around the screw, and for one such screw enclosed in a cylinder of diameter $(D + 2d)$, there would

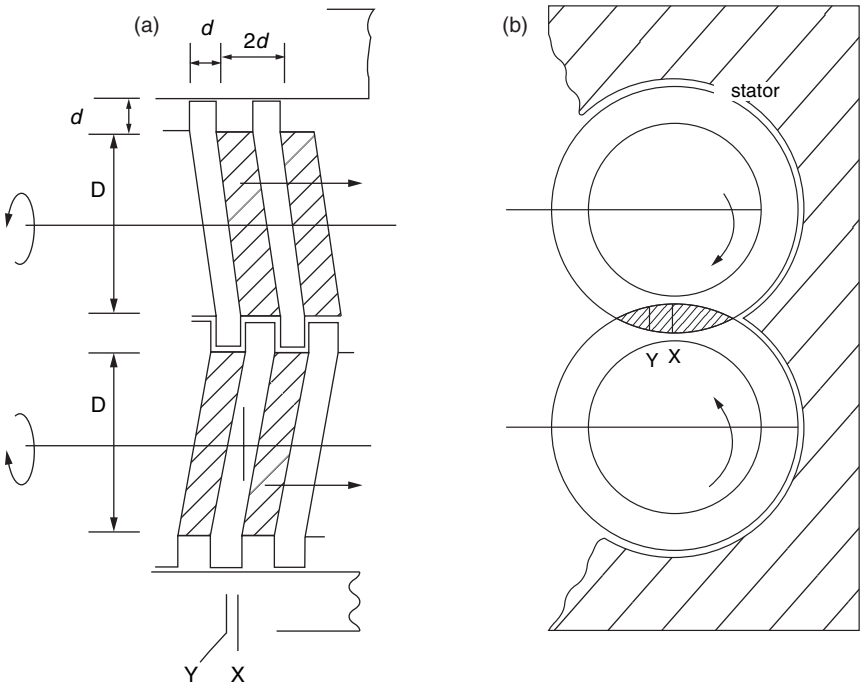

FIGURE 6.15
Principle of the screw pump. (a) Screw profile, (b) sectional view.

be a continuous helical pathway. But for the two intermeshed screws of the figure, surrounded by a close-fitting stator as suggested in the view in Figure 6.15(b), there are trapped volumes and restricted sections of channel in the two screws defined by the regions of their contact. In the shading shown in the end view, the central line represents contact between the rotors at section X, whereas those on either side correspond to shorter line contacts associated with adjacent sections, such as at Y. When the screws counter-rotate, trapped volumes advance along the axis (subject to a little leakage because clearances have to be finite), eventually to be expelled at the outlet against the prevailing pressure. Similar volumes are formed as they are drawn through the inlet opening, isolated and conveyed along the axis. Actual pumps have a more complex screw form than the simple one depicted, which is for illustrative purposes. Modern pumps designed specifically for vacuum service turn at high speeds up to 10,000 rpm because the rotating loads can be dynamically balanced and large pumping speeds are possible. They are tolerant of particulates, which tend to sweep through them, and have ultimate pressures of about 5×10^{-2} mbar. In addition, because the intermeshing surfaces do not quite touch, they can be given protective coatings that enable them, as with other types of dry pump, to be used to pump chemically aggressive gases. Most of the compression, and therefore most of the heating, occurs at

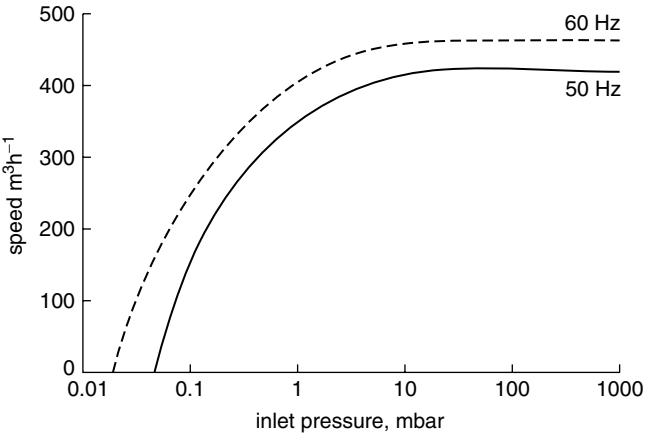

FIGURE 6.16
Pumping speed of a screw pump.

the outlet, but new developments based on screws that have a progressively reducing pitch mean that compression and heating can be distributed along the pump axis.

A typical performance curve is shown in Figure 6.16.

The ***scroll pump*** is based on the geometry of the involuted scroll. Its action may be illustrated by considering the spiral Figure 6.17(a) that can be described in *r*, θ coordinates by the equation

$$r(\vartheta) = r_0 + r_0\left(\frac{\vartheta}{2\pi}\right) \tag{6.4}$$

Note that, as drawn, the spiral starts from θ = 0 with a value r_0 to the left of the origin on the horizontal axis and θ increases clockwise. The radial distance from the center at any angle increases by the distance r_0 with each turn, and r_0 is thus the radial separation of adjacent parts. In Figure 6.17(b), a spiral of identical form but displaced by $r_0/2$ to the left and with the same convention has the equation

$$r(\vartheta) = \frac{3}{2}r_0 + r_0\left(\frac{\vartheta}{2\pi}\right) \tag{6.5}$$

and is drawn as the dashed line. If the dashed spiral is displaced by a distance $r_0/2$ to the left as in Figure 6.17(c), there are point contacts between the spirals at separation r_0 along a horizontal line, and starting from a point of contact such as A and moving clockwise, one can identify an enclosed area terminating at B that has the form of a crescent with tightening curvature in an

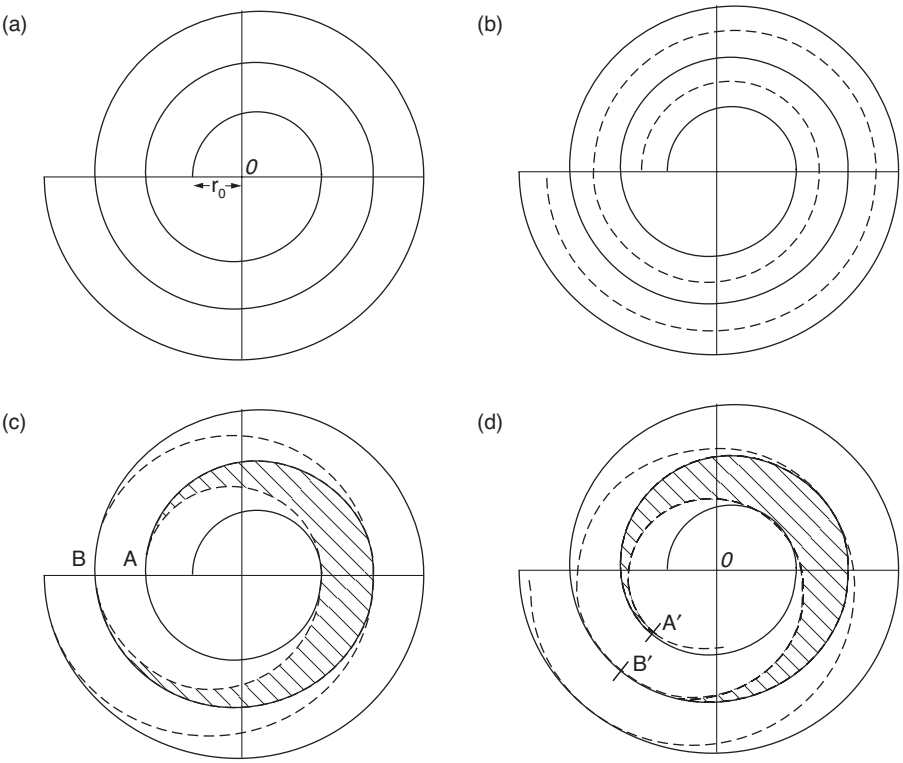

FIGURE 6.17
Geometry of the scroll pump.

anticlockwise direction. Now suppose that starting again from Figure 6.17(b) the dashed scroll had been displaced by 45° downwards and to the left, again by a distance $r_0/2$ to produce Figure 6.17(d). The points of contact are now along a line at 45° to the horizontal at separation r_0, and the area previously defined as between A and B is now between A′ and B′. It is a slightly smaller area, and the crescent shape has moved closer to the center.

The continuous spiral may be thought of as fixed and the dashed one as moveable. The change described above between Figure 6.17(c) and 6.17(d) could have been accomplished by arranging for the dashed spiral to move without change of orientation around one eighth of a circle of radius $r_0/2$ with center O. At all stages the spirals would have touched along a rotating straight line with progressive movement of the crescent shape. If the dashed spiral is made to move continuously in a circular orbit with radius $r_0/2$ about 0, that is, it does not rotate but executes a circular orbital motion (rather as one might polish a window with a circular motion of the hand), then a counterclockwise orbital motion causes the areas trapped to move inward.

This, translated into three dimensions, is the basis of the scroll pump. Two intermeshed scrolls (spirals) are constrained to have the relative motion

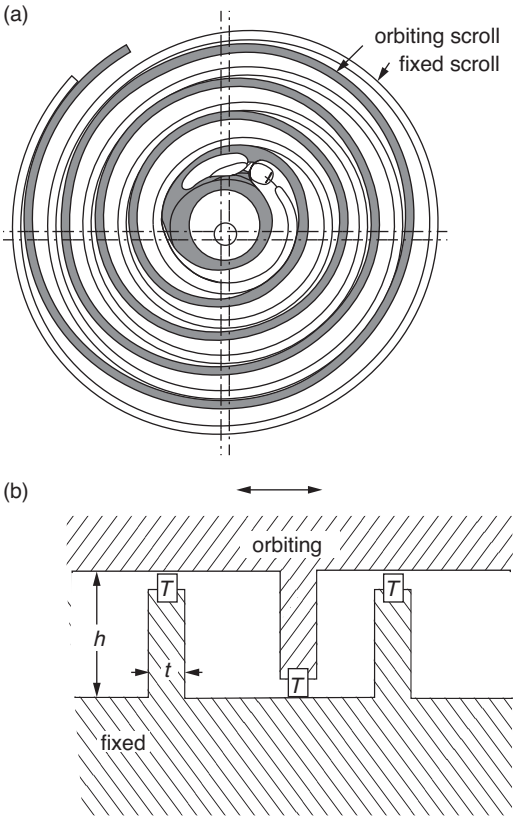

FIGURE 6.18
Scroll pump (a) geometry of scrolls, (b) section. (Part a from Su, Y., Sawada, T., Takemito, J., and Haga, S., *Vacuum*, 47, 815-818, 1996. With permission.)

described above. Each consists of a spiral wall, called the wrap, on a continuous plate. A typical arrangement is shown in Figure 6.18(a). Figure 6.18(b) shows a sectional view. A typical wall height h is 28 mm and wall thickness t = 4 mm. A tip seal, T, consisting of a Teflon strip fitting into a recess that runs along the top of each wall makes light contact with the opposing face. The minimum gap between the walls in near-contact is a small clearance of about 0.1 mm, giving an imperfect but dry seal. Counterclockwise rotation draws gas in at the outermost opening. It is isolated in a crescent-shaped pocket and then conveyed progressively towards the center, with some compression, to exit in an axial direction at the pump outlet. No exhaust valve is needed. The orbital rotation rate is typically 1450 rpm. As gas in any pocket is compressed, there is naturally some back leakage, especially across the clearance seals where the walls are at their closest. This flow may be viscous, transitional, or molecular, depending on the prevailing pressure conditions. The theoretical analysis of the pump, which is rather complicated by its

geometry, is discussed by Su et al. (1996) Typical pumps have speeds of about 10 $m^3\ h^{-1}$ and ultimate pressures of about 5×10^{-2} mbar.

6.3 Momentum Transfer Pumps

Molecular drag, turbomolecular, and vapor diffusion pumps are described in this section. In these devices, incoming gas molecules continuously acquire an enhanced velocity component in the direction of the outlet.

6.3.1 The Molecular Drag Pump

Closely related to the turbomolecular pump, in that they both depend on the interaction of gas molecules with fast-moving surfaces, the molecular drag pump was proposed by Gaede in 1913. Its development, however, was limited by the technology of the times, and was quickly eclipsed by his invention (contemporaneously with Langmuir in the U.S.) of the vapor diffusion pump, which could be manufactured more easily, and which came to dominate high-vacuum pumping for the next 70 years. Recently though, enabled by computer-controlled machining methods and improved technology for high-speed rotation, molecular drag pumps, which have some features that make them useful in particular applications, have enjoyed a resurgence of use.

The principle of Gaede's pump is the basis of modern designs and its concept is illustrated in the cross-sectional diagram of Figure 6.19. A solid cylindrical rotor R rotates at high speed inside a hollow cylindrical stator S with which it is coaxial, tending to drag gas along the channel between R and S in the direction of its motion. An intruding feature C, called a skimmer, which almost touches the rotor, presents an obstacle to circumferential flow and separates inlet and outlet ports. The channel has depth h of a few millimeters as shown and dimension w into the page. If we take reasonable values r = 50 mm and a rotation speed 30,000 rpm = 500 Hz, the surface speed of the rotor is 157 m s^{-1}, comparable with typical thermal velocities.

Although gas in a viscous condition may be moved through the pump, this is not its most effective regime of operation. Under these conditions the forward motion imparted to molecules by their interaction with the rotor surface is quickly lost in randomizing collisions with neighboring molecules. Velocity gradients develop and a pressure difference $p_2 - p_1$ between output and input is established, which is proportional to rotor speed, gas viscosity, and channel length, and inversely proportional to channel depth h. Van Atta (1965) derives this result. Under molecular conditions the pump becomes much more efficient. The prevailing gas density is now sufficiently low that the mean free path is much greater than the channel dimension h. Molecules colliding with the rotor surface leave it with a bias velocity U and travel

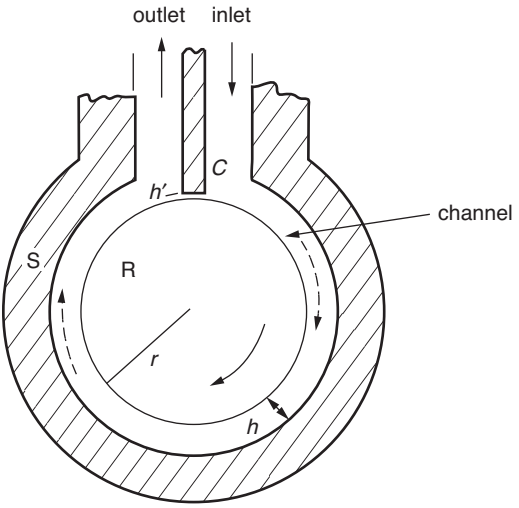

FIGURE 6.19
Geometry of a drag pump.

across to the stator as discussed in Section 4.6. The collision with the stator removes the bias and randomizes their motion, and they return to R where they again pick up the bias velocity so that their journey to the outlet continues.

The passage of molecules through the channel may be modeled in one dimension as depicted in Figure 6.20(a) and Figure 6.20(b).

A channel of width w, height h, and length L represents the pumping duct with a coordinate x along its length. It is closed on one of its faces by a high-speed plate with velocity U, representing the motion of the rotor. Given the relative dimensions, the curvature of the actual pumping channel can be ignored in a first approximation, and to simplify the analysis we will assume that $w >> h$, so that "side effects" can be ignored. At any section of the channel, half the molecules will be traveling from R to S with imposed bias velocity U, the other half will be traveling away from S having had their velocities randomized. The pumping speed or volumetric flow rate S_0 at any

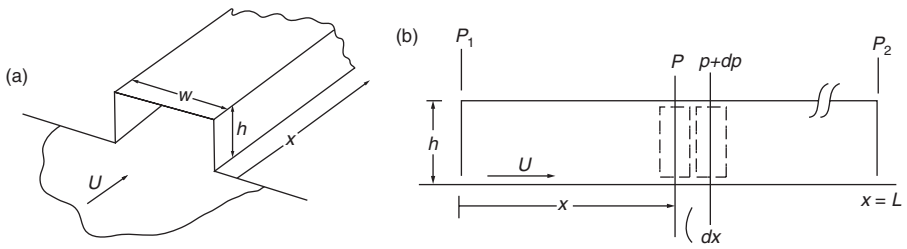

FIGURE 6.20
(a) Perspective view of drag channel, (b) diagram for analysis of flow.

section will be just $whU/2$, independent of x. (Were the whole channel translating with velocity U, then, because of surface interactions, all the gas in it would adjust to move with it, and the volumetric flow rate would be whU). At coordinates x and $x + dx$, the pressures are p and $p + dp$, and may be taken to be associated with number densities n and $n + dn$ in volumes $whdx$ that straddle the two positions symmetrically, as in Figure 6.20(b). Opposing the drag effect will be backflow due to the pressure differences developed. Defining Q as the throughput, we thus have at x

$$Q = S_0 p - C_x dp \tag{6.6}$$

where C_x is the conductance to be associated with the section of length dx at x. Now, for a channel of length L and constant cross-section, the conductance C of the whole channel is expressible (see Equation 5.36, for example) as $C = k/L$ where k is constant. Thus we may write $C_x = k/dx = CL/dx$ and so

$$Q = S_0 p - CL\frac{dp}{dx} \tag{6.7}$$

This may be rearranged as

$$\frac{dp}{p - Q/S_0} = \frac{S_0}{CL}dx$$

and integrated from $x = 0$ where $p = p_1$ to give

$$\left(p - Q/S_0\right) = \left(p_1 - Q/S_0\right)\exp\left(\frac{S_0}{CL}x\right)$$

so that for $x = L$, where $p = p_2$

$$\left(p_2 - Q/S_0\right) = \left(p_1 - Q/S_0\right)\exp\left(\frac{S_0}{C}\right) \tag{6.8}$$

This shows that the ratio S_0/C controls pump behavior. The conductance C of the channel may be taken (see Wutz, Adam, and Walcher, 1989) to be

$$C = \frac{8wh^2}{3L}\sqrt{\frac{R_0 T}{2\pi M}} \tag{6.9}$$

so that with $S_0 = whU/2$ the ratio S_0/C is

$$\frac{S_0}{C} = \frac{3LU}{16h}\sqrt{\frac{2\pi M}{R_0 T}} \tag{6.10}$$

From Equation 3.8, for the mean molecular velocity $\bar{v}$, Equation 6.10 may also be expressed as

$$\frac{S_0}{C} = \frac{3LU}{4h\bar{v}} \tag{6.11}$$

in which its dimensionless character is evident from the ratios of lengths and velocities it contains.

To interpret this analysis, consider first the steady-state zero-flow condition ($Q = 0$) that would be achieved with an outlet valve closed so that pumping was into a small isolated volume. Equation 6.8 indicates that this zero-flow compression ratio K_0, say, would be

$$K_0 = \left(\frac{p_2}{p_1}\right)_0 = \exp\left(\frac{S_0}{C}\right) \tag{6.12}$$

Inserting reasonable values $L = 250$ mm, $h = 3$ mm, $U = 150$ m s^{-1}, and 475 m s^{-1} as the mean velocity of nitrogen molecules at room temperature gives $K_0 = \exp 20$, ~ 10^8, a high compression ratio. The other extreme case is when the outlet pressure equals the inlet pressure (corresponding in a different pressure regime to the free displacement conditions of positive displacement pumps). If this condition were arranged, then from Equation 6.6, with pressure constant along the channel, the pumping speed would be just $Q/p = S_0 = whU/2$. These limiting conditions are shown in the throughput vs. pressure ratio plot in Figure 6.21(a); there is zero throughput at $K = K_0$ and maximum throughput when $K = 1$. For the usual intermediate conditions, we may write Equation 6.8 as

$$p_2 - Q/S_0 = K_0(p_1 - Q/S_0)$$

Defining S_1 as the pumping speed at the inlet so that $Q = S_1 p_1$ leads to

$$\frac{p_2}{p_1} = K_0\left(1 - \frac{S_1}{S_0}\right) + \frac{S_1}{S_0}$$

and

$$\frac{S_1}{S_0} = \frac{K_0 - p_2/p_1}{K_0 - 1} = \frac{K_0 - K}{K_0 - 1} \tag{6.13}$$

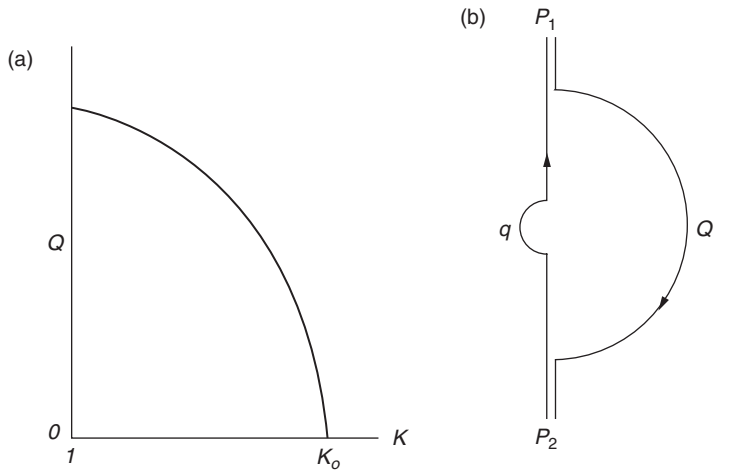

FIGURE 6.21
(a) Throughput vs. compression ratio and (b) representation of forward and reverse flow for a drag pump.

in which K is the pressure ratio p_2/p_1. Because K_0 is very much greater than unity, this may be written as

$$\frac{S_1}{S_0} = 1 - \frac{K}{K_0} \tag{6.14}$$

Evidently, when gas is being pumped, pumping speeds are less than S_0 and pressure ratios are less than K_0. Equation 6.13 enables the Q vs. K plot of Figure 6.21(a) to be completed.

The theoretically predicted high-compression ratios indicated above are not achieved in practice because theory also needs to include the gas that returns via the imperfect seal between the high and low pressure sides of the skimmer. Compared with the pumping channel, this has the same width w and smaller clearance h', but also smaller length L', through which, in addition to natural pressure-driven flow, there will be a component driven by the rotor in just the same way as in the main pumping channel. This situation is represented schematically in Figure 6.21(b). Complicating the analysis is a return flux q that will be governed by an equation similar to 6.7, with coordinate y from high to low pressure.

$$q = S'_0 p' + C' L' \frac{dp'}{dy} \tag{6.15}$$

The quantity S'_0/C' will again control the flow. The result is that, because of this counterflow and other leakages, the compression ratios achieved

are much less than predicted by Equation 6.12. However, this analysis is instructive in identifying the role of the factors involved, and its concepts are also valid in more modern developments of the basic pump. Evidently, from Equation 6.12 and Equation 6.10 for S_0/C, compression ratios will be higher for gases of higher molecular weight. Thus the compression ratio for CO_2 ($M = 44$) will be greater than that for helium ($M = 4$) by a factor $\exp\{\sqrt{44}\}/\exp 2 = 100$. Heavy hydrocarbons are pumped with a very high compression ratio.

Pumping speed is proportional to the channel depth h and the velocity U, whereas, as Equation 6.11 indicates, to achieve high compression requires large U and as large a value as possible of the ratio of L to h. Therefore, while the maximum rotation speed achievable promotes both S and K, the choice of h involves a compromise. Molecular drag pumps are of intrinsically small speed in that the entrance area through which gas enters the pump is limited, approximately wh in the Gaede pattern, with values as we have seen, typically of a few liters per second. On the other hand, this same geometry at the pump outlet means that molecular conditions are sustained up to a higher pressure before the onset of transitional and then viscous flow, so that outlet pressures are typically two orders of magnitude higher than for turbomolecular and diffusion pumps, which will be described later. This is a very useful property in the context of dry pumping and will be discussed later.

In the Holweck development of the basic design, the length of the pumping channel is increased by making it in the form of a helical groove in the inward-facing surface of the stator. Another possibility is to have multiple circular channels of the type described above adjacent to each other, with the outlet of one connected in an axial direction to become the inlet of the next, in the original Gaede pattern. In both cases, w is no longer very much greater than h. A design of the Holweck type is shown in Figure 6.22 and its pumping speed curve in Figure 6.23. The bladed rotor at the inlet improves performance by promoting flow into the drag section.

The applications for drag pumps by themselves are limited, but they have become very extensively used when incorporated into the turbomolecular pumps as a high-pressure outlet stage of the combination to form turbodrag pumps, which are discussed in the following text.

6.3.2 The Turbomolecular Pump

Like the molecular drag pump, the turbomolecular pump exploits the interaction of gas molecules with high-speed surfaces to direct them towards an outlet. But in this case the surfaces are the inclined surfaces created by machining angled slots radially inwards from the edge of a circular disc to form a bladed rotor that resembles the multiple-bladed structure of an axial flow turbine. The rotors turn at speeds sufficient to give their outer parts velocities comparable with molecular velocities. Each rotor is paired with a

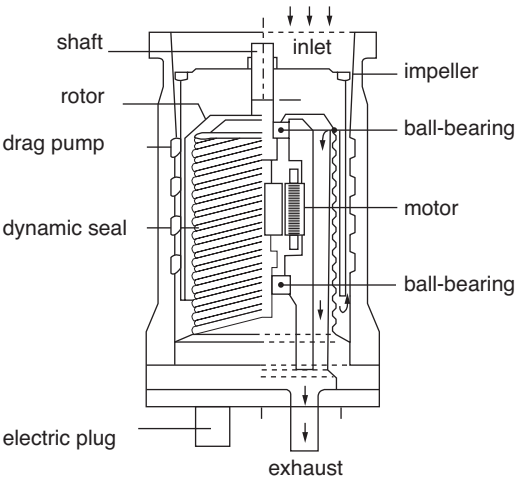

FIGURE 6.22
Structure of a drag pump. (Used with permission of Alcatel Vacuum Technology, Annecy, France.)

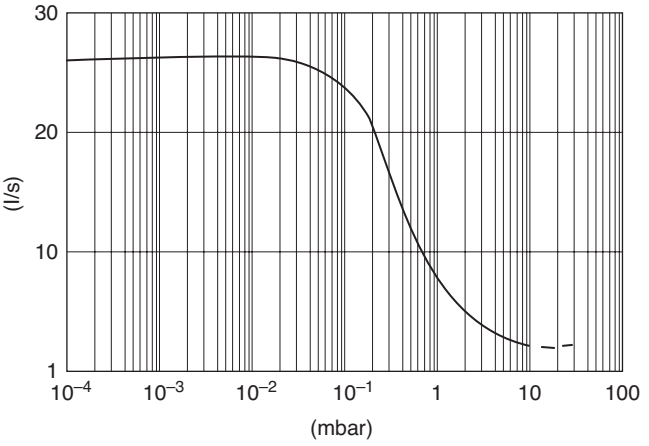

FIGURE 6.23
Pumping speed of a drag pump. (Used with permission of Alcatel Vacuum Technology, Annecy, France.)

similar set of stationary stator blades that project inwards from the wall of the cylindrical pump housing, as shown in Figure 6.24. The flow direction is axial, and although the structure resembles a turbine arrangement, it should be emphasized that the pump operates in the molecular flow regime. The blades are not aerodynamically shaped as they would be for hydrodynamic flow, but are simply flat-sided, as results from the machining process.

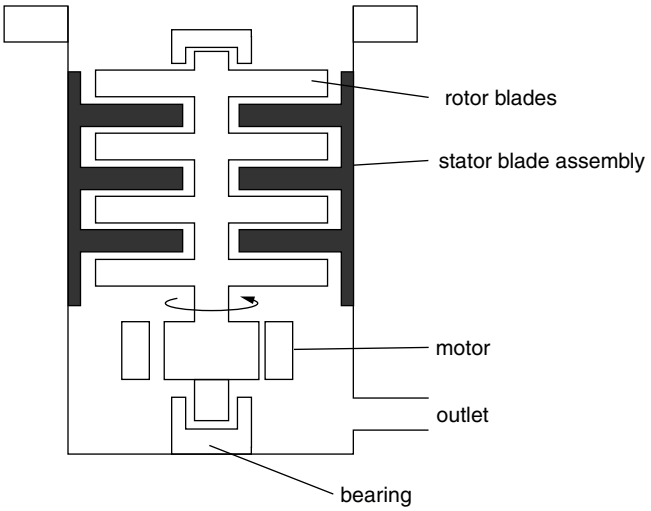

FIGURE 6.24
Principle of the turbomolecular pump.

The pump was invented by Becker in 1958 and has revolutionized high-vacuum pumping by partly displacing the diffusion pump in many of its traditional high-vacuum applications. Its design also overcame a significant problem of the drag pump that the effective pumping area in relation to the associated supporting flange was small. From a comparison of Figure 6.24 and Figure 6.22, it is evident that the turbomolecular pump, flange size for flange size, offers a significantly greater active area for pumping. Whereas pumping speeds of typical drag pumps are of the order of 10 l s^{-1}, those of comparable turbomolecular pumps are of the order 100 l s^{-1}. Another advantage it has, compared with the drag pump, is that there is no requirement for a very small gap between the high-speed rotor and stationary parts to act as a seal; the gap between the rotor blades and the housing in the turbo pump does not have that role and can be as large as 1mm, eliminating the possibility of mechanical seizure.

The essence of the pumping action is that the slanting blades, in moving at high speed through the gas, intercept more molecules on their downward-facing side than if they were stationary and tend to project them in an axial direction. This action is duplicated by the stator blades that receive molecules that have gained a velocity increment in rebounding from the rotor. The processes involved may be understood as follows. Let Figure 6.25 represent in one dimension a row of blades of separation s, length (chord) b, and inclination ϕ that move with velocity U past an associated set of stator blades with opposite inclination. We will further assume (solely to illustrate the capture process) that all molecules move with the most probable molecular velocity v_α (418 m s^{-1} for nitrogen) and that this is also the value of the blade velocity U. At any location above the blades, the distribution of

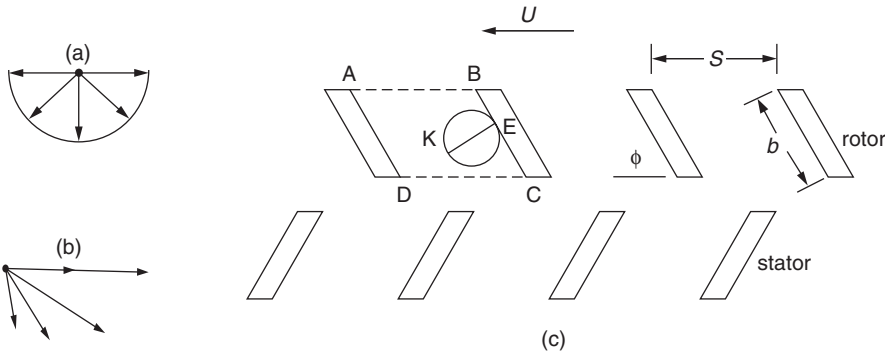

FIGURE 6.25
(a) Turbomolecular pump, (b) molecular velocity distributions, (c) blade geometry.

downward-moving molecules is isotropic and may be represented by a semihemispherical set of velocity vectors as depicted at (a) in Figure 6.25. A "cell" ABCD may be defined as a unit of the arrangement, and in the plane at the top of the blades the area AB may be subdivided into small areas, each of which has a downward flux of molecules with this velocity distribution entering it.

To visualize how blade BC intercepts molecules, we may use a standard method of mechanics and imagine it brought to rest by the addition of a velocity $-U$ to both it and the molecules in its vicinity. The distribution shown at (b) in Figure 6.25 is thus produced. Some molecules entering the cell ABCD from above will pass through to exit directly through CD. Some will impact on the surfaces AD and BC, more on the latter, and then, having interacted briefly with the surfaces, will desorb randomly. From a typical place E on BC, they will desorb according to the cosine law, represented by a sphere K, some in a downward direction, some across to AD for another surface interaction, and some upwards through AB back where they came from. This is a complicated state of affairs similar in character to that discussed in Section 5.6.3 for the transmission probability through a cylindrical tube, and in a similar way involves solid angles, probabilities that depend on the particular location on the surfaces, and the possibility of some multiple trajectories back and forth between faces BC and AD before molecules leave through either AB or CD.

Molecules from below are also incident upwards on the rotor and similar considerations apply, but the result, given the geometry and the imposed motion, is that there is a net downward flux of molecules. The incident flux is greater on BC than AD, and from a point such as E, the "downward" solid

angle is generally greatest. Franck in Wutz, Adam, and Walcher (1989) discusses these geometrical matters in more detail.

The downward urging of molecules is further reinforced and duplicated when the molecular population in the downward direction, which carries a bias due to the rotor motion, interacts with the stator blades. The rotor–stator combination, in which each plays an equal role in promoting the downward flow of molecules, can be considered as one stage of the pump. Modern pumps have typically about 10 such stages. Probabilities α_{12} and α_{21} may be assigned formally for the probabilities of passage of molecules between inlet (1) and outlet (2), with $\alpha_{12} > \alpha_{21}$ for the reasons already discussed. Defining A as the sum for all the cells of entry areas such as AB and recalling the quantity J (Equation 3.13) as the number of molecules impinging per unit area per second, the particle flow rate $\dot{N}$ through the rotor stator stage is

$$\dot{N} = \alpha_{12} J_1 A - \alpha_{21} J_2 A \tag{6.16}$$

where J_1 and J_2 are the downward and upward impingement rates associated with number densities n_1 and n_2 on the inlet and outlet sides. The probability of pumping, say, W, (which is also the speed factor H) is the ratio of the number of molecules passing through the row to the number incident on it. Thus

$$W = \frac{\dot{N}}{J_1 A} = \alpha_{12} - \alpha_{21} \frac{J_2}{J_1} \tag{6.17}$$

The ratio J_2/J_1 equals the ratio of number densities n_2/n_1, and assuming a Maxwellian distribution for the molecules exiting the rotor region, $n_2/n_1 = p_2/p_1 = K$, the compression ratio. Therefore

$$W = \alpha_{12} - \alpha_{21} K \tag{6.18}$$

and by rearrangement

$$K = \frac{\alpha_{12}}{\alpha_{21}} - \frac{W}{\alpha_{21}} \tag{6.19}$$

For $W = 0$, the zero flow condition, K takes a maximum value

$$K_0 = \frac{\alpha_{12}}{\alpha_{21}} \tag{6.20}$$

From Equation 6.18, the maximum value of W is achieved when $K = 1$ and is

$$W_{\max} = \alpha_{12} - \alpha_{21} \tag{6.21}$$

It follows from these equations that between these extremes the flow rates and pressure ratios are related by

$$\frac{W}{W_{\max}} = \frac{K_0 - K}{K_0 - 1} \tag{6.22}$$

This is the equivalent statement to that of Equation 6.13 for the drag pump. In that case it was derived in terms of the flow mechanism; in this, in terms of flow probabilities. Again, maximum flow occurs for zero compression and maximum compression for zero flow, and a plot of the type in Figure 6.21(a) will represent behavior between these limits. Evidently, from Equation 6.20 and Equation 6.21, high compression ratios are promoted by large ratios of the forward-to-reverse probabilities, whereas large flow rates call for large values of their difference.

It is of central importance, therefore, to determine the forward and reverse transmission probabilities. The problem is too complex to be solved analytically. Monte Carlo techniques were applied in a series of landmark papers by Kruger and Shapiro, most notably that of 1961. They theoretically analyzed and also measured the properties of a single rotor. Their analysis determined the dependence of the compression ratios K_0 and pumping probability $W_{\max}$ for a single row of blades on the blade speed U, blade angle ϕ, and the blade spacing-to-chord ratio s/b, which gives a measure of the "openness" of the arrangement. Blade speed U is expressed as a speed ratio $s_r = U/v_\alpha$ to the most probable molecular velocity v_α ($= \sqrt{2kT/m} = \sqrt{2R_0T/M}$ see Equation 3.5). For nitrogen its value is 418 m s^{-1}.

For $s/b = 1$, for example, K_0 increases with the speed ratio s_r in an approximately exponential way up to $s_r = 1.5$ before tending to become constant. One may write

$$K_0 \propto g(\phi)\ \exp\{U/\nu_\alpha\} \quad U/\nu_\alpha < 1.5 \tag{6.23}$$

where $g(\phi)$ represents the dependence on blade angle ϕ. K_0 increases as ϕ gets smaller and, for example, has a value 4 for $\phi = 30°$ and 8 for $\phi = 10°$ for $s_r = 1$. $W_{\max}$ increases approximately linearly with s_r up to $s_r = 1$ before tending to become independent of s_r or even decreasing. Under the same conditions ($s_r = 1$, $s/b = 1$), $W_{\max}$ has a value of about 0.5 for ϕ in the range 20° to 40°, but is less than 0.3 for $\phi = 10°$. These trends of K_0 and $W_{\max}$ with ϕ are true, with slightly altered values, for other s/b ratios in the range 0.5 to 1.5. Evidently, large blade angles promote pumping speed, and smaller ones compression. The reader may consult O'Hanlon (2003) or the original Kruger and Shapiro article for greater detail.

In currently available pumps, blades of radius about 50 mm turn at 60,000 rpm = 1000 Hz so that peripheral rotor speeds are approximately 100π = 314 m s^{-1}. For nitrogen, therefore, $s_r = 314/418 = 0.75$, and for the conditions

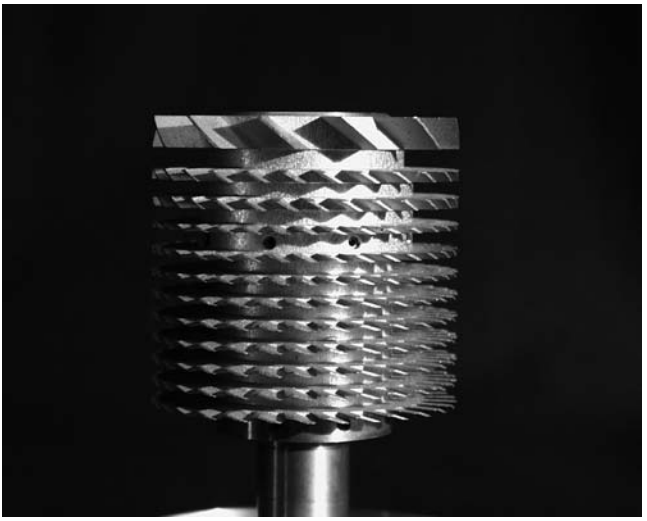

FIGURE 6.26
Rotor blade structure.

$s/b = 1$ and blade angle 30°, $K_0 = 2.3$ and $W_{max} = 0.3$. The compression ratio for a single rotor/stator stage for nitrogen is thus $2.3^2 = 5.3$. For ten stages in series, the compression ratio would be 2×10^7. Of course, analysis and design are refined to account for the changes in s_r and s/b with radial position along a rotor, and in practice only the faster-moving outer section is bladed. Flow and compression of the whole multistage assembly are optimized by having high-angle blades of more open aspect at the input to achieve high pumping speed. The stages near the output have blade angles of about 10° to generate the greatest compression. Each stage sustains a pressure difference, with gas becoming progressively more compressed so that the volumetric flow rate diminishes towards the outlet. Figure 6.26 illustrates the way in which the blade angles change for a rotor stack of radius 30 mm designed to operate at 90,000 rpm.

How compression ratios K_0 vary with the type of gas being pumped may be deduced from Equation 6.23. The most probable velocity depends inversely on the square root of the molar mass M, and therefore, for a given value of U, the velocity ratio s_r is proportional to $\sqrt{M}$. Hence

$$K_0 \propto \exp\sqrt{M} \tag{6.24}$$

If we take nitrogen as the gas of reference and consider relative molecular masses M_r

$$\frac{K_0}{(K_0)_{N2}} = \frac{\exp\sqrt{M_r}}{\exp\sqrt{28}} = \frac{\exp\sqrt{M_r}}{199} \tag{6.25}$$

For hydrogen, $M_r = 2$ and $\exp\sqrt{2} = 4.1$, so that for one rotor the compression ratio is approximately only about 2% of that for nitrogen. Compounded through many stages this leads to compression ratios that are many orders of magnitude smaller. Typical values achieved in a modern pump for nitrogen, helium, and hydrogen are 10^9, 10^5, and 10^3, respectively. Values as low as this for hydrogen are especially significant, as will be described later. Conversely, for heavy molecules such as hydrocarbons, compression ratios are very high, so that backstreaming of such molecules from the vapors of backing pump oils does not occur.

Pumping speeds tend not to vary significantly for different gases. Because the dependence of W_{max} on $s_r = U/v_a$ is approximately linear for $s_r = 1.5$, then for a given operating rotor speed U,

$$W_{max} \propto 1/v_a \propto \sqrt{M} \tag{6.26}$$

Hence pumping probabilities are smaller for the lighter gases, partly because a greater proportion of faster-moving molecules will pass between pumping blades and not acquire any velocity enhancement. However, the pumping speed of a gas depends on the product WJ of the probability of pumping and the arriving flux. From Equation 3.23, the flux J of the gas at the pump inlet is $J = n\bar{v}/4 = pN_A/\sqrt{2\pi MR_0T}$, and so for different gases at the same pressure, it is greater for lighter gases, varying as $1/\sqrt{M}$. In the product WJ, therefore, M does not appear, and pumping speed does not depend on gas type within the limits of the approximation. It may be shown that with blade area A, the volumetric pumping speed is $S = W\bar{v}\,A/4$, from which the independence of S on M may be deduced directly.

Measured speeds vs. inlet pressure are shown for a typical pump in Figure 6.27. Speeds are constant at low pressure and are slightly less for helium and hydrogen than for nitrogen, but this depends on details of design and the reverse can be true. Typically, a pump with an inlet flange having an opening 100 mm in diameter has a pumping speed of 250 l s^{-1} for nitrogen, and hence an overall speed factor S/C_A equal to $250/930 = 0.27$. This figure is a measure of the efficiency of the "disposal mechanism" referred to in the introduction. (The active pumping blade area will be rather less than the area of the opening, and so the Ho coefficient is greater than 0.27.) From Figure 6.27 pumping speeds start to decrease for inlet pressures greater than about 10^{-3} mbar, falling to about 10% of maximum at 0.1 mbar. This is due to loss of efficiency as the rear blades come into transitional and then viscous flow conditions, with slowing of the rotation rate due to friction. In Figure 6.27 (dashed line) throughput $Q = Sp$ is plotted vs. pressure, just for nitrogen. In the region of constant speed, it increases linearly with pressure and then reaches a roughly constant maximum as the speed falls off.

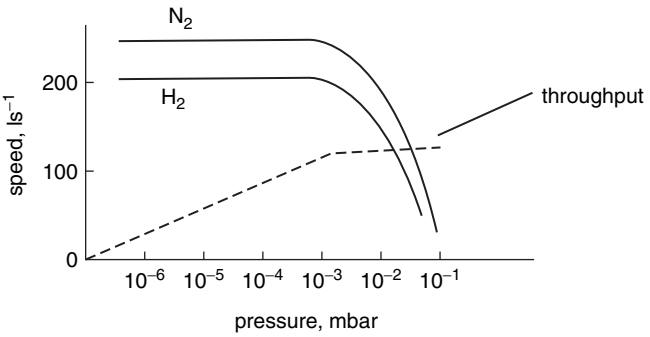

FIGURE 6.27
Pumping speeds of a turbo pump, and throughput curve for nitrogen (dashed line).

Figure 6.28 shows how the zero flow compression ratio K_0 depends on the foreline pressure provided by the backing pump, falling below its maximum value at pressures above 10^{-1} mbar for nitrogen (10^{-2} mbar for hydrogen and helium) and becoming essentially zero between 10^{-1} and 1 mbar. These data are taken by blanking the pump inlet and adjusting the pressure in the backing line while also measuring the inlet pressure, rather as described for the Roots pump in Section 6.2.

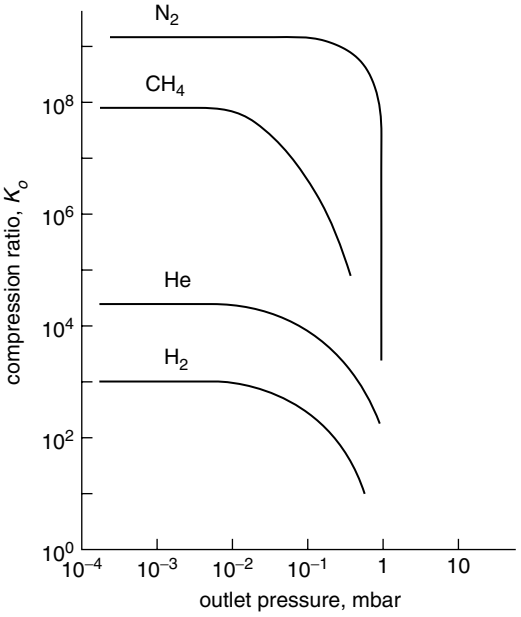

FIGURE 6.28
Zero-flow compression ratio vs. backing pressure.

It is important to consider the matching of a turbomolecular pump to its backing pump via the foreline. The criterion usually taken is that when the inlet pressure to the turbomolecular pump is such that the throughput has reached its maximum value, the foreline pressure should be low enough to keep the blades at the foreline side in molecular flow. As the data of Figure 6.27 show, maximum throughput is achieved at an inlet pressure p_1 of about 10^{-3} mbar, and a backing pressure $p_2 = 10^{-1}$ mbar in the foreline sustains molecular flow. Because throughput $Q = S \times p$ is continuous, $S \times p_1 = S_B \times p_2$ where S and S_B are the speeds of the turbo and backing pumps, respectively. Thus a backing pump of speed

$$S_B = \left(\frac{p_1}{p_2}\right) S = \frac{10^{-3}}{10^{-1}} S = \frac{S}{100}$$

is indicated. The ratio S/S_B is called the "staging ratio," and although 100 in this example, is frequently made smaller than this, perhaps 50 or less, by choice of a larger backing pump, for reasons connected with the pumping of hydrogen, which are discussed below.

Returning to Equation 6.22, taken now to describe the whole pump, and considering pumping speeds S rather than pumping probabilities, leads to

$$\frac{S}{S_{\max}} = \frac{K_0 - K}{K_0 - 1}$$

For the large K_0 values encountered in this pump, this may be written as

$$\frac{S}{S_{\max}} = 1 - \frac{K}{K_0} \tag{6.27}$$

Pumping speeds will be close to the maximum value for $K << K_0$, and only when compression ratios increase to approach K_0 will they reduce to become zero for $K = K_0$. For nitrogen and similarly relatively heavy gases, compression ratios of 10^8 or more mean that speeds are close to the maximum for K up to about 10^7, i.e., with p_2 held at say, 10^{-1} mbar by the backing pump, speeds are retained down to pressures of 10^{-8} mbar and less. This is not the case for hydrogen which, as we have seen, has K_0 values of about 1000 and sometimes rather less. For $K_0 = 500$, pumping speeds for hydrogen will start to diminish significantly at a much earlier stage and be a tenth of the maximum at a K value of only 250.

The bearings that define the location of the shaft that carries the rotor blades may be of oil- or grease-lubricated type, or magnetic, or a combination of the two, with a magnetic bearing nearer to the high-vacuum side. In the case of lubricated bearings, concerns arise about the possibility of hydrocarbon contamination of the pumped volume. However, as previously mentioned, the

compression ratio for hydrocarbons at normal operating speeds is so high that no backstreaming of vapors occurs, and provided that the rotation speed does not fall below about 60% of maximum, this remains true. It is prudent, therefore, having switched off the power to the motor, to vent the system at speeds not less than this. Venting is carried out not from the foreline but from either the high vacuum side or at a location incorporated in the turbo itself that injects gas at an intermediate level in the rotor sequence. For similar reasons, it is not wise to leave a non-operational turbo pump with lubricated bearings under vacuum. Pumps with wholly magnetic bearings use a combination of the fields of permanent magnets and electromagnets to levitate the rotor assembly and control the motion. They are free from contamination concerns and therefore ultraclean, though relatively expensive. In their venting procedure, care must be exercised not to switch off the levitating facility until the rotation has completely stopped. Hablanian (1997) discusses design, performance, and operational matters in detail.

6.3.3 The Turbo-Drag Pump

To add a drag stage after the turbomolecular blades on a common drive shaft is a natural evolution in pump design and a very desirable one for reasons discussed in Section 6.3.1. It extends performance at the upper end of the pressure range by two orders of magnitude, and therefore enables the use of roughing pumps of considerably smaller volumetric capacity, particularly dry pumps. Most manufacturers have added this feature to their designs so as to create turbo-drag pumps, and some now make these exclusively. The drag geometry is of either the Holweck or the Gaede pattern.

6.3.4 The Vapor Jet (Diffusion) Pump

In this device, molecules of the gas to be pumped collide with heavy molecules traveling at high speed in a directed vapor jet and are given a biased motion towards the pump's outlet. Although a vapor jet provides the basic pumping mechanism, it was for many years described as a diffusion pump because the pumping action is initiated by the diffusion of gas molecules into the vapor jet, and this term is still widely used. The pump was invented by Gaede in 1913 with mercury as the working fluid, further developed by Langmuir, and first used with an oil charge by Burch in 1929. It has served as a mainstay of high-vacuum pumping ever since and has been developed to a very high level of performance. It remains a widely used and robust workhorse today, though in many applications requiring a clean environment free of possible oil vapor contamination, it has been displaced by the more expensive turbomolecular pump already described. Such clean environments and ultrahigh vacuum are achievable with the vapor jet pump, though special precautions, as will be described, are then necessary.

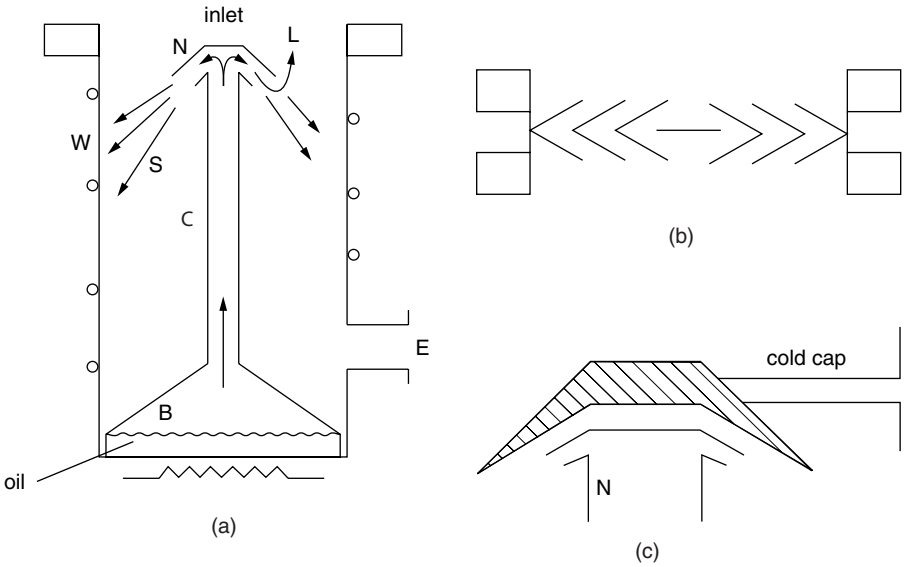

FIGURE 6.29
(a) Single-stage vapor jet pump, (b) chevron baffle, and (c) cold cap.

The working principle of the pump can be described with reference to the single-stage version illustrated in Figure 6.29(a). An oil with low vapor pressure and consisting of molecules of high molecular weight, typically 400–500 amu, is heated in a boiler B so that its vapor exerts a pressure of 1–2 mbar. This vapor fills the chimney C and is expelled through an annular nozzle N at its top into a region maintained at lower pressure, 10^{-1} mbar or less, by a rotary backing pump connected to the outlet E. The expansion through the nozzle causes the jet created to be supersonic, with the form of a thickening skirt S that extends downwards and outwards to the inner wall of the water-cooled pump casing W, usually made of stainless steel. Here the vapor condenses back to liquid, which runs down the wall under the influence of gravity and reenters the boiler, to be heated and vaporized again. The oil thus undergoes a continuous cyclic process.

At the pump inlet, if the gas pressure is sufficiently low that its behavior is molecular rather than fluidic, gas molecules that drift into the jet are likely to have collisions with its fast-moving oil molecules, be given additional downward momentum, and move through it to the lower part of the pump. They emerge from the jet into the region downstream that is at a higher pressure than the inlet and are removed from there by the backing pump. As well as moving the gas in a downstream direction and producing compression, the jet forms a seal between the pump inlet and its outlet at the higher backing pressure. The reason that gaseous conditions at the inlet cannot be fluidic is that if the gas is too dense, it breaks up the jet and destroys the pumping action, causing the pump to stall. Therefore, prior to a vapor jet pump being brought into action, a primary pump has to be used

to lower the pressure at its inlet to about 0.1 mbar, approaching molecular conditions. The vapor jet pump can then lower the pressure in a vessel connected to its inlet by many orders of magnitude, compressing the gas to backing pressure from inlet pressures that fall as pumping progresses.

The sealing action mentioned above that prevents backflow from the outlet at higher pressure to the inlet exists only if the backing pump keeps the outlet pressure below a critical value referred to as the critical backing (or "foreline") pressure, which is typically a few tenths of the vapor pressure in the boiler that drives the jet. Otherwise, the jet action is overwhelmed from the outlet side and the pumping action ceases. In normal operation, backing pressures are usually much less than this critical value, and it is only at crossover, defined in Section 6.1 and exemplified in Section 8.6, or in situations in which continuous high gas loads are being handled, that the pump and its backing pump operate near this condition.

The performance of the vapor jet pump depends on a number of factors. The molecular flow conductance of the annular region at the plane of the nozzle determines the arrival rate of molecules at the jet stream. Their diffusion into the jet and acquisition of increased downward momentum depend on the dimensions of the stream, and the density, mass, and velocity of the oil molecules in it. If the jet is too dense, the gas will not penetrate into it substantially; if not dense enough, too few of the gas molecules will acquire the necessary additional momentum to create adequate compression in the gas below the jet. The properties of the jet depend on the design of the nozzle and the properties of the oil, particularly its vapor pressure, viscosity, and stability at the high operating temperature, which is typically about 200°C. Not all the gas molecules that enter the jet are sent to the outlet and pumped away. Some collisions within it will result in gas molecules leaving in an upward direction against the general direction of flow, so the probability of interaction and onward transmission is less than unity. Ho coefficients are typically 0.3–0.4, but may be as high as 0.5 in the best modern designs. The pumping action and the analysis of it are quite complex. Over many years of study, analysis has progressed to a highly refined state. Hablanian's text (1997) gives a good introduction to these matters.

One unavoidable feature of the performance of the nozzle is that oil molecules escape from the jet, some in an upward direction towards the vacuum chamber, as indicated at L in Figure 6.29. This is because of interactions at the exit of the nozzle and scattering between molecules in the stream whose motion following expansion still has a random thermal component superposed on the directed hydrodynamic component. Those that escape upwards present a problem, which is described as backstreaming. Its magnitude is such that unless preventive measures are taken, the oil charge in the pump can be seriously depleted and the chamber grossly contaminated in relatively short operational times.

In order to reduce the effect, baffles may be located in the pump inlet above the nozzle to intercept the backstreaming flux. A typical design of a chevron baffle is shown in Figure 6.29(b). It allows no line of sight up to the vacuum

chamber for backstreaming molecules. They thus impinge on it and are trapped by condensation, accumulate as condensed liquid, and drip back into the pump to eventually find their way back to the boiler. The price to be paid for this preventive action is that the conductance for gas into the pump, and therefore the pumping speed above the baffle, is reduced, typically by a factor of about 0.3. A design feature that does not entirely eliminate but significantly reduces the problem at source is the fitting of a cold cap above the top cap of the nozzle assembly as shown in Figure 6.29(c). Attached to the water-cooled pump body in such a way that it is kept much cooler than the adjacent top cap, it intercepts, condenses, and ultimately returns to the boiler a considerable fraction, 95% or more, of the backstreaming flux. Another mechanism by which oil molecules may travel upwards to reach the high vacuum above the inlet is by condensation of some of the backstreaming flux on nearby pump walls, diffusion along them, and subsequent reevaporation in unfavorable directions. This is called back-migration and is reduced by the water cooling of the walls. Backstreaming and back-migration may be eliminated by well-designed cold traps, to be discussed later.

The single-stage version of the pump described above serves to explain its working principle and the main features of its operation. In practice, most pumps are multistage devices, with three or more jet-producing nozzles, as shown in Figure 6.30. The jet assembly is driven by the boiler pressure of about 2 mbar, and there is progressive compression of the pumped gas through the stages. The top nozzle has an associated annular open area that is relatively large to maximize gas entry and a jet that is designed to have a high capture probability with a low compression to minimize reverse flow. In the subsequent stages, denser jets produce more compression into smaller volumes and raise the outlet pressure to a level matched to the backing pump. Some multistage pumps have a final horizontal ejector stage into the backing line connection with an associated cooled baffle to counter fluid loss. Such a stage raises the critical backing pressure. Values generally are in the range of 0.3 to 1 mbar, and depend on nozzle design, power input, and the oil used. Fractionating pumps are designed with separate concentric chimneys and fluid-guiding channels on the base of the boiler so that vapors and the more volatile fractions of oil mixtures are led preferentially to the nozzles of the lower-stage jets. By reducing the presence of highly volatile components in the top jet, this leads to an improvement in the ultimate pressure achieved.

Generally, oil that has backstreamed to condense on baffles or other surfaces in the region of the inlet will exert a vapor pressure that contributes to and may limit the ultimate pressure achievable. Many of the oils used have vapor pressures of about 10^{-6} mbar at room temperature and the more expensive ones 10^{-8} mbar or less. Water cooling is arranged to hold surfaces near the inlet close to room temperature or sometimes less if baffles are refrigerated. One aspect of this problem is that at the high boiler temperature, some oil may break down into more volatile fractions.

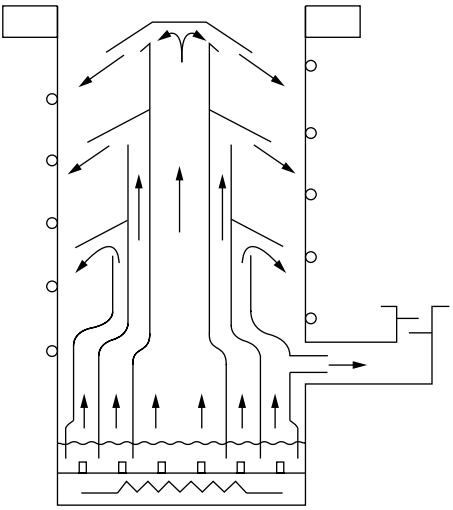

FIGURE 6.30
Typical multistage pump.

The pumping speed curve of a typical diffusion pump for nitrogen is shown in Figure 6.31(a) in which volumetric speed on a linear scale is plotted against pressure on a logarithmic scale. The region of constant speed up to about 10^{-3} mbar is where the jet has a pressure-independent capture coefficient so that the pumping rate is the product of this and the molecular arrival rate at the entrance area, determined by the particular geometry and Equation 3.13. Pumping speeds are roughly proportional to the area of the inlet port. The value given of 250 l s^{-1} would be typical of a pump with an inlet port 100 mm (4 in.) in diameter. Pumps with a larger inlet port 150 mm in diameter are widely used in laboratory practice and have speeds of 600–700 l s^{-1}. For very big installations, pumps with throat diameters up to 1 m and pumping speeds of many thousands of liters per second are manufactured.

Above an inlet pressure of about 10^{-3} mbar, pumping speed begins to diminish and at 10^{-2} mbar has fallen to a fraction of its maximum value. The associated throughput, being the product of speed and pressure, increases to reach a maximum at a pressure corresponding to point Q_{max} and becomes approximately constant, as shown by the dashed curve in Figure 6.31(a). What is happening in this condition is that the jet system is becoming overloaded. The number of molecules that can be pumped, given the power supplied to it from the boiler, has reached a maximum.

At the lowest pressures, although the capture efficiency of the jet is undiminished so that the horizontal characteristic that represents it may be extended back to zero pressure, at some inlet pressure value that depends on the oil used, the reverse flow of backstreaming oil molecules balances the pumped flow, so that the net speed is zero. The curves shown are for DC705, a silicone oil, and Santavac 5, a polyphenyl ether, which is much more

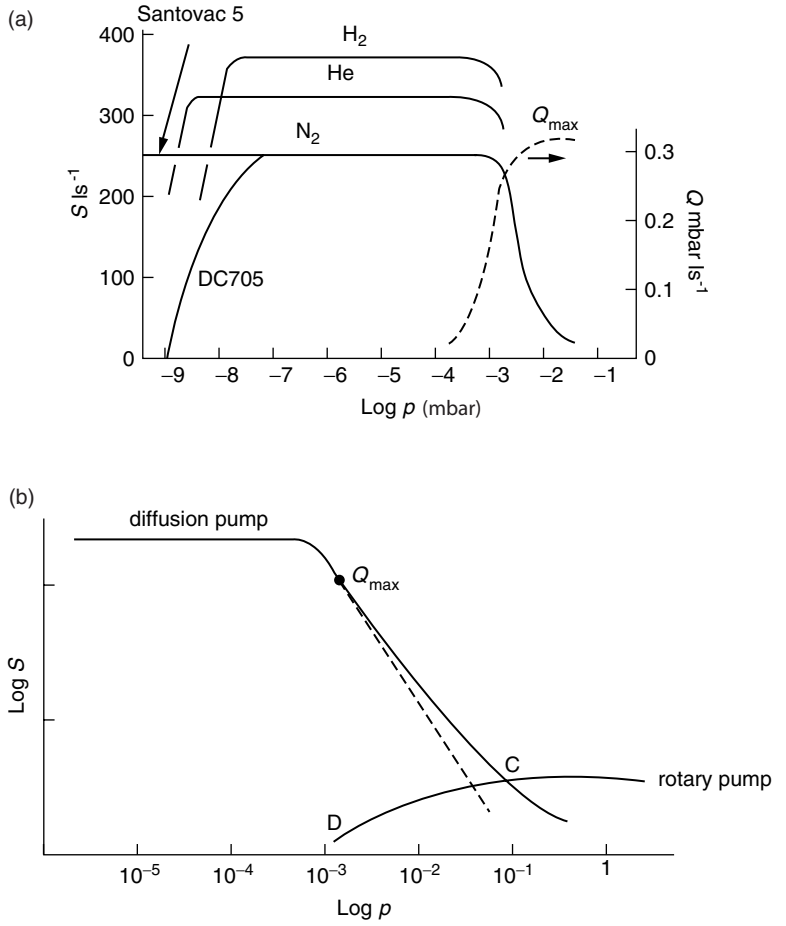

FIGURE 6.31
(a) Pumping speed curves and (b) speed and throughput curves for vapor jet and backing pumps.

expensive. Operation at the lowest pressures is possible with the oils that have the least vapor pressures at room temperature, but other considerations, such as oxidation resistance and how stable their backstreamed vapor is to charged particle beams in the pumped vacuum chamber, may have an overriding importance, depending on the application. Harris (2005) or O'Hanlon (2003) may be consulted to pursue this further.

For gases such as argon, carbon monoxide, and for water vapor, pumping speeds are comparable with the values for nitrogen. For light gases such as hydrogen and helium, one might expect, because of their much higher molecular flow rate into the pump (see Equation 5.38, for example), that pumping speeds would be dramatically larger than they are for nitrogen. This is not so, however, as the labeled curve for hydrogen demonstrates. Typically,

pumping speeds for hydrogen and helium are just 50 and 30% greater than for nitrogen, respectively. This reflects reduced efficiency of downward momentum transfer in the jet for less massive molecules. Another aspect of the curve for these gases is the reduction in the pumping speed that occurs at inlet pressures below about 10^{-8} mbar. This is due to the limited values of the maximum compression ratio achievable with these gases, about 10^6 in the most favorable cases, but often less. Again, this is attributable to dynamics within the jet. Thus, when the ratio of outlet to inlet pressures begins to exceed 10^6, the backflow increases disproportionately and the pumping speed starts to fall. In these circumstances, the pressure of hydrogen or helium at the inlet is equal to the outlet pressure divided by the compression ratio. Such compression ratio limitations occur for all gases, but with discharge pressures into the backing line of 10^{-2} mbar or less when pumping at the lowest inlet pressures, the maximum compression ratio with heavier gases, usually $\times 10^8$ or more, is sufficient to have the region of constant speed extend back to pressures of 10^{-10} mbar and lower, so that in normal operation this limiting behavior is not encountered.

In Figure 6.31(b) in which pumping speed is plotted on a logarithmic scale, pumping speed curves for a vapor jet pump and a backing pump are shown, together with the plot of the constant throughput Q_{max} vs. pressure for the vapor jet pump. This will be used in Section 8.6 to illustrate the matching of the two pumps in a practical example.

Vapor jet pumps are used in many high-vacuum applications; TV tube manufacture and vacuum coating (see Chapter 9) are just two examples. Their reliability, size range, versatility, and general robustness make their operation very economic. Unlike turbomolecular pumps, they do not contain any moving machined parts and the comparative simplicity of their basic design, engineering, and power provision is reflected in their lower cost. In many applications, the small amount of backstreaming that occurs in a well-designed pump fitted with a cold cap and with appropriate choice of oil is either not important or acceptable. When the reduction or elimination of backstreaming is important, the fitting of baffles and traps above the pump has to be considered.

As noted earlier, the role of a baffle is to intercept backstreaming oil vapor and return it to the boiler, and this occurs with a significant reduction in overall pumping speed. It is generally arranged that a baffle is water cooled so that its temperature is a little below ambient. In some cases they may be cooled to further below ambient, for example, by using the Peltier effect, to give increased effectiveness against backstreaming and improved pump ultimate pressure. If the temperature is so low as to immobilize condensate and prevent its eventual return to the boiler, the baffle becomes a trap. With Santovac 5®, a polyphenyl ether as the pumping fluid, it is possible to achieve pump ultimate pressures of 10^{-9} mbar or less using only water-cooled baffles.

As distinct from a baffle, a trap is intended to capture and immobilize backstreaming vapors in a condensed state, at temperatures sufficiently low that the condensed species exert negligible vapor pressure. Traps designed

for attachment above vapor jet pumps allow no direct line of sight paths between inlet and outlet in a similar way to baffles, but contain a liquid nitrogen reservoir centrally located in their structure whose outer wall is the trapping surface on which all passing molecules impinge. As well as acting as effective barriers to backstreaming molecules, the cryogenically chilled surfaces have a very high pumping speed for water vapor from the pumped vacuum chamber. Such a trap, mounted above a baffle and on a pump that uses high-grade pumping fluid and with provision for maintaining the level of liquid nitrogen, can reliably provide ultraclean pumping at ultrahigh vacuum of suitably baked systems for extended experimental times, months or longer, for lengthy investigations in surface science. Pump, baffle, and trap selection at the design stage must take account of the reduction of pumping speed due to the presence of the added components, which might typically reduce it by a factor of 5 or 6. The authoritative text of Hablanian already referred to may be consulted for detailed discussion of all operational aspects of the vapor jet pump.

6.4 Capture Pumps

Capture pumps work by trapping and retaining molecules in a condensed state. They, therefore, have a finite capacity and require regeneration when it is reached. Pumps exploiting capture at low temperatures will be described, and then those based on gettering, which include sublimation and ion pumps.

6.4.1 The Sorption Pump

The sorption pump captures gas by physical adsorption onto the surfaces of porous molecular sieve, chilled to the temperature of liquid nitrogen. It has limited capacity, but is clean, contamination-free, and useful as a "one-shot" roughing pump. Frequently, two are used in sequence in ion-pumped systems to bring the system to a pressure of less than 10^{-3} mbar at which the ion pump can be started. A typical simple design is shown in Figure 6.32. A cylindrical stainless steel can C has a reentrant lower section N that defines an annular region A holding zeolite pellets Z. This is immersed in liquid nitrogen in a polystyrene container P. A tube T enables the liquid to fill space N. An isolation valve (not shown) is incorporated in the pumping line to the system and a simple spring-loaded valve S allows release of pumped gases to the atmosphere during subsequent regeneration.

The molecular sieve Z, which does the pumping, consists of small pellets of synthetic zeolites (alkali metal aluminosilicates) from which the water of hydration has been driven out by heating at 350°C so as to leave an open network of voids to which there is easy access and in which gaseous molecules

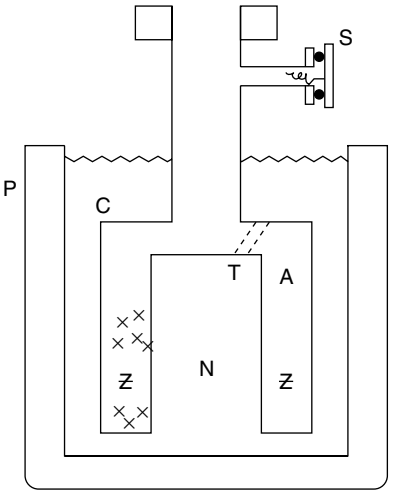

FIGURE 6.32
Cryogenic sorption pump.

may become trapped. If molecules are too big, they cannot, of course, enter the structure; if too small, they may enter it but would be unable to make adequately strong bonds to their surroundings to become trapped. For the frequently used Linde 5A sieve, the average pore size is 0.5 nm. By analogy with a sponge, one may expect materials in such a form to possess a large area of accessible surface. Delchar (1993) discusses the various sieves in detail. A crude model (Problem 4.9) shows that one cubic centimeter of this material has an effective surface area of about 600 m^2. It is this huge surface area per unit volume that, when chilled, allows a large number of molecules to be physically adsorbed. Gases such as nitrogen, oxygen, and carbon dioxide are very effectively pumped by incorporation and retention in the chilled structure, but not helium or neon. A charge of about 1 kg has a capacity of 10^5 mbar liter. Water vapor is pumped very effectively but can only be released by a regeneration process that involves heating to 250°C for several hours, whereas the other pumped gases desorb at room temperature. Periodically, such high-temperature regeneration is necessary to restore capacity.

Because of the poor thermal conductivity of zeolites and the small area of contact between them and the container wall, it takes about 20 min to thoroughly chill the pump and bring it to working temperature. Used in sequence, two such pumps can reduce the pressure in a typical 100 l volume to 10^{-3} mbar, but the procedure is critical. The first pump is used to reduce the pressure to about 1 mbar and then its valve quickly closed, so that much of the neon (which is not pumped) in the original volume is transported to accumulate at a location where it is isolated from the rest of the system, thereby decreasing considerably the quantity of inert gas to be handled by the second pump. Alternatively, if only one sorption pump is available, a mechanical pump can be used to pump down to 0.1 mbar, safely above the

limit for back-diffusion of pump oil vapors into the system, before the sorption pump is used.

6.4.2 The Cryopump

In these pumps, gaseous molecules are captured either by condensation or, for certain gases, by adsorption on surfaces held at very low temperature, processes called cryocondensation and cryosorption, respectively. For both economic and practical reasons, earlier designs that used liquid nitrogen (77 K) and liquid helium (4.2 K) to cool condensing surfaces directly are nowadays rarely used except in large installations. Rather, laboratory cryopumps incorporate reliable self-contained refrigerator units to achieve the low temperatures required. They are clean and oil-free. The modern form of pump is widely used in both high and ultrahigh vacuum applications, subject to a restricted ability to pump hydrogen and helium. It is a secondary pump, designed for use at inlet pressures of less than about 10^{-3} mbar for which flow conditions are molecular. Continuous operation at higher pressures is not feasible; the rapidly increasing deposit thickness and its poor thermal conduction would cause the temperature of the condensing surface to rise to a value at which evaporation and condensation were comparable, so that effective pumping action ceases. Being a capture pump, it naturally requires periodic regeneration.

Central to the physics of the condensation process is the dependence of the vapor pressure of a substance on its temperature. The equilibrium vapor pressures of some gases commonly encountered in vacuum practice are shown in Figure 6.33.

Any point on these curves gives the temperature and pressure at which the condensed phase of a substance and its saturated vapor are in equilibrium. Thus, familiarly, the vapor pressures of liquid nitrogen at 77 K and

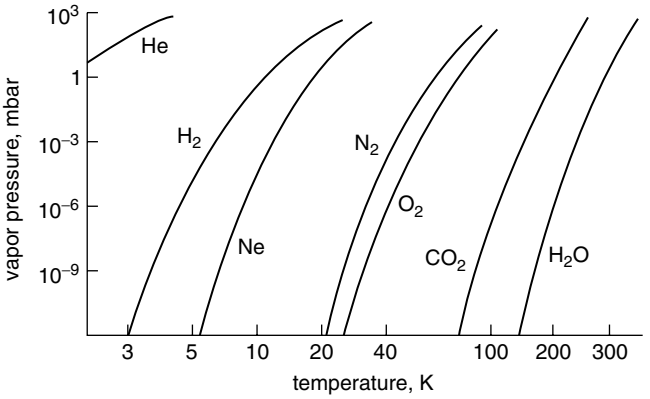

FIGURE 6.33
Vapor pressures of common gases.

water at 373 K, their boiling points, are 1013 mbar. Note that at the temperature of liquid nitrogen, the vapor pressure of water has diminished to be vanishingly small and that of carbon dioxide is 4×10^{-8} mbar.

Net condensation on a surface at a given temperature occurs when the partial pressure of a substance in the vapor phase is greater than its saturation vapor pressure at that temperature and can be related to arrival and departure fluxes as discussed in Chapter 3. For example, in an enclosure imagined to contain just gaseous CO_2 at room temperature, a small area of surface maintained at 77 K will pump CO_2 and reduce the pressure to a value close to 4×10^{-8} mbar, and a state in which the arrival and departure fluxes will be equal. For the pumping of water vapor by a surface at this temperature, capture would be complete because its vapor pressure is essentially zero. In such circumstances the pumping speed of the cold surface would be equal to the conductance of an aperture of the same area, as discussed earlier. For water vapor using Equation 5.37 and Equation 5.38, this gives a pumping speed of $11.8 \times (28/18)^{1/2} = 14.7$ l s^{-1} per cm^2. Because of the efficiency of the condensation process (the probability of condensation onto very cold surfaces is close to unity), pumping speeds in cryogenic condensation are potentially high. Unlike most pumping situations, temperature is not a constant and care has to be exercised in analyzing flow. It is best treated in terms of particle flow rates, as will be discussed later.

It is evident from Figure 6.33 that for the purposes of cryocondensation, gases fall into three groups. There are those like water vapor and higher hydrocarbons with negligible vapor pressure at temperatures of about 100 K. A second group, such as N_2, O_2, and CO_2, are such that pumping will be complete at condensing temperatures of 10–20 K. A third group, helium, hydrogen, and neon, have substantial vapor pressures even below 10 K, and for these, cryocondensation is not feasible. These last can, however, be pumped effectively by cryosorption on activated charcoal at temperature 10–15 K. Although the adsorbing surface saturates at just one adsorbed monolayer, the porous structure, as is the case in the sorption pump discussed in the previous section, presents an extremely large area for pumping, hundreds of m^2 per gram, so that large capacities of order 100 sccm are possible.

The design of a common type of cryopump is shown schematically in Figure 6.34(a). A two-stage cryogenic refrigerator R abstracts heat to keep a first-stage "cold head" F, a massive copper block, at a temperature of 60–80 K, and a second cold head S at 10–15 K. Attached to and cooled by F is a cylindrical thermal shield and enclosure E made of copper, completed at the top by an annular louvered structure A, which allows the inlet of gas but prevents an optical line of sight for thermal radiation to the inner stage attached to S. The walls of the enclosure are nickel plated on the outside so as to have low emissivity and high reflectivity for ambient temperature thermal radiation from the pump housing H, and they are blackened on the inside so as to adsorb the radiation that passes through the chevron, thus preventing its reflection to S. The louver and shield, because of their physical

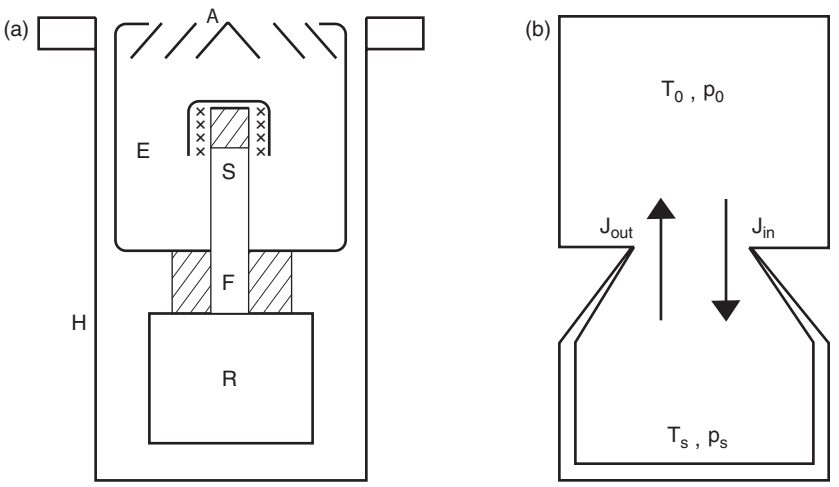

FIGURE 6.34
(a) Structure of a cryopump. (b) Model of chamber and pump for analysis.

and thermal attachment to the copper block F, are held at temperature 60–80 K, depending on the thermal load received. The louver acts as the pump for incident water vapor. Noncondensible species at this temperature, for example, nitrogen, hydrogen, and helium, pass through it into the enclosure. Water vapor that negotiates the louver without collision will be condensed on the shield inner wall.

Attached to the second cold head S is an inverted cup-shaped piece made of copper, on the underside of which activated charcoal granules are glued. In some designs this structure is more complex. It is maintained at temperature 10–15 K and so will condense gases such as nitrogen, oxygen, and argon on its upper surface, leaving ideally just hydrogen, helium, and neon, which can only be pumped by cryosorption, to find their way to the activated charcoal underneath. The structure and temperature distribution in the pump is thus that water vapor is pumped by the louver at ~80 K, nitrogen and others by cryocondensation on the surfaces at ~10 K, and hydrogen, helium, and neon by cryosorption in locations not accessed by others. At the temperatures prevailing, the deposits develop as solid "ices" of their species, and thickening deposits present an extra thermal resistance in the heat-flow pathway to the primary sources of cooling, the cold heads F and S.

The action of the two-stage refrigerator that cools F and S may be understood in terms of the single-stage device described below. It utilizes the Gifford–McMahon refrigeration cycle which is a zero-work expansion cycle with regenerative heat exchange. Figure 6.35(a) shows the design schematically. A cylindrical housing H with a copper cold head K contains a cylindrical piston-like structure called the displacer D that moves up and down at about 1 Hz, driven by an external motor. A seal S separates the upper (cold) region of variable volume, V2, and the lower warm region, V1. The

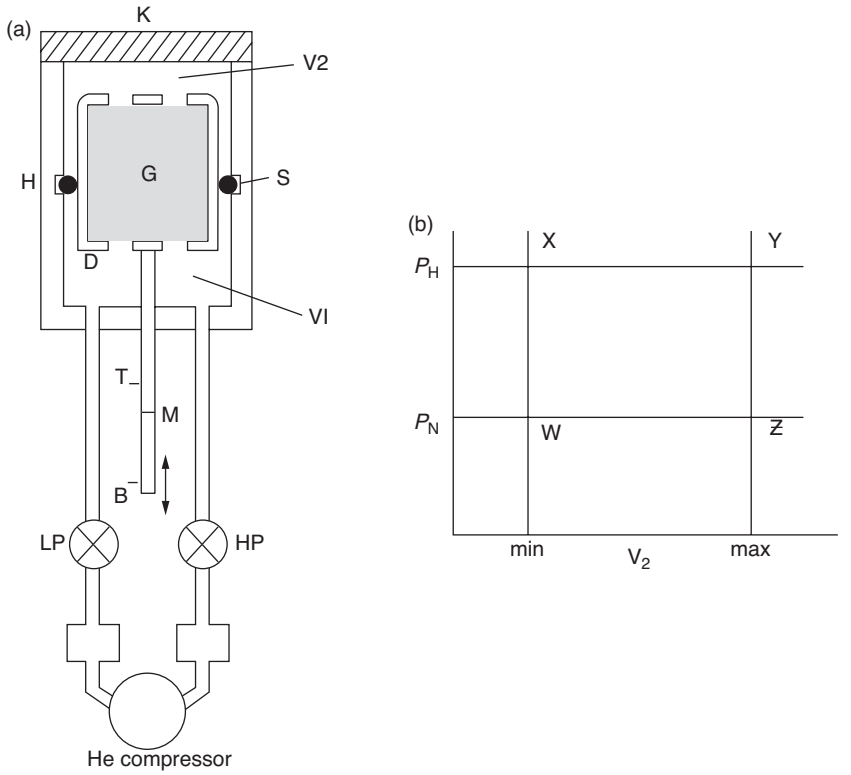

FIGURE 6.35
(a) Principle of Gifford–McMahon refrigerator, (b) working cycle.

displacer, made of plastic with poor thermal conductivity, contains regenerating material G in the form of a matrix of compressed layers of very fine copper mesh. Its function is to provide thermal capacity with large surface area in order to make heat exchange as efficient as possible. Gas can easily pass through the regenerator matrix so that the pressures are essentially the same on both sides of it as the displacer moves between its upper position (marker M at T), where the cold volume V2 is at a minimum, and the lowest position (M at B), where volume V1 is a minimum. The volume V1 + V2 stays constant, and the equality of pressures means that negligible work is done by or on the gas in moving the displacer up and down. At the warm end of the housing are valved connecting pipes to ballast volumes of a helium compressor. High-pressure helium gas is the working fluid of the refrigerator. The inlet valve marked HP controls flow into it from the compressor at about 20 bar (300 psig). The outlet valve LP is the exhaust to the lower pressure side of the compressor at about 6 bar. The opening and closing of valves is synchronized with the movement of the displacer.

The Gifford–McMahon cycle that creates the cooling is represented by the rectangle WXYZ in the diagram of pressure vs. V2, the "cold" volume, in

Figure 6.35(b). Let us imagine the device to be in state W on the diagram. Marker M is at T, V2 at a minimum, and the pressure throughout at the lower value P_N. Valve LP closes and HP opens, so helium flows from the compressor at pressure P_H to fill the whole volume, passing through the regenerator. V2 is at its minimum value and state X is reached. With HP still open and LP closed, the displacer is then moved downwards to its lowest position. The regenerator moves through the helium and V2 increases. A small amount of cooling of the helium occurs by heat exchange with the matrix, causing a further small amount of high pressure helium to be withdrawn from the compressor. Thus Y is reached where V2 is a maximum. Valve HP is now closed and LP opened. There is a drop in pressure, and hence temperature, as gas at high pressure is extracted from the device from the compressor. Some passes through the regenerator and cools it, and cold head K is cooled. Stage Z is reached and Y to Z is the "cooling stroke" of the cycle. Finally, with valve LP still open, the displacer is moved to minimize the cold volume V2. The regenerator, in moving through the gas, is cooled and the bulk of the gas passes to volume V1, at which point the cycle can start again.

By using two such interconnected stages with a common drive, cold heads F and S in Figure 6.35 are maintained at temperatures 10–15 and 60–80 K. In practice, in order to achieve vibration isolation and at no cost to thermal efficiency, the compressor is set up as a remote unit connected to the refrigerator R in the pump body by flexible lines. The cooling capacities of the first and second stages are about 15 and 3 W, respectively. The time to reach temperature is 1–3 h. The temperature of the second-stage cold head is sensed by a hydrogen vapor pressure thermometer and monitored in the control unit.

Thermal loads on the refrigerator need to be considered and are of two types, arising from thermal radiation and molecule–surface interactions, particularly those that result in condensation. Thermal radiation incident on a surface at temperature T_S that faces surroundings at room temperature T_0 K, say, is σT_0^4 per unit area, where σ = Stefan's constant. Assuming a "grey" surface, i.e., one whose emissivity at all wave lengths is a constant fraction ε of that of a perfect black body at the same temperature, then the amount of radiation adsorbed per unit area is $\varepsilon\sigma T_0^4$, while radiation emitted is $\varepsilon\sigma T_S^4$, so that the net amount absorbed per unit area is $\varepsilon\sigma(T_0^4 - T_S^4)$. Taking a worst case ε =1, with $\sigma = 5.67 \times 10^{-8}$, T = 295 K and T_S = 80 K leads to a power per unit area of 430 Wm^{-2} or 43 mW per cm^2.

For a pump with entrance aperture D = 10 cm, a typical small laboratory pump, the total area of the enclosure E (louver plus shield) that intercepts room temperature radiation from the vacuum chamber and the pump housing is ~500 cm^2. Were all these surfaces technically black (ε = 1), the load of about 20 W would be too high for the refrigerator's first stage. But because the exterior shield and chevron are nickel-plated, these surfaces are highly reflecting, with emissivity ε =0.03 so that loads are reduced by factors of 10 to an acceptable value. Once the louver becomes ice coated, its good reflectivity is

lost and there is a higher load on the pump inlet area, but because the access of gas to the sides of the enclosure is restricted, the low emissivity of other parts of the enclosure is not seriously compromised.

Thermal radiation from this enclosure E, at 70 K, say, on average, is a load on the second stage parts at 10–15 K. The thermal flux onto these colder parts may be estimated in a similar way to be 1.36 W/m^{-2} = 0.14 mW/cm^2, which leads to an acceptably small load on the 10–15 K cold head. The necessity for an intermediate enclosure to shield the coldest stage from room temperature radiation is evident.

Compared with the above loads, the thermal loads due to condensation are quite small and may be demonstrated as follows. Let us suppose (reasonably, as we shall see) that the measured pumping speed for nitrogen gas is 500 l s^{-1}. Nitrogen finds its way into the pump to be condensed on the 10–15 K surfaces. At a chamber pressure of, say, 10^{-5} mbar, the number of molecules being pumped per second is the number contained in a volume of 500 l at the prevailing pressure. Because a mole contains $N_A = 6 \times 10^{23}$ molecules and occupies 22.4 l at standard pressure and temperature, the number of molecules being pumped is

$$N_A\left(\frac{500}{22.4}\right)\left(\frac{273}{295}\right)\left(\frac{10^{-5}}{1013}\right) = 2.05 \times 10^{-7} N_A = 1.2 \times 10^{17}$$

Now, the kinetic energy content of a mole of diatomic molecules such as nitrogen is, according to the kinetic theory, $(5/2)R_0T$ per mole, amounting to 6.13 kJ at 295 K. To a good approximation, all of this energy will be extracted in condensation at 10–15 K. Therefore, the heat load presented to condensing surfaces in one second is (6.13×10^3) (2.05×10^{-7}) = 6.3 mW, over an area of ~50 cm^2, and therefore ~0.1 mW per cm^2, an acceptable load. Of course, part of this load will be deposited anyway on the inner surface of the 70 K enclosure with which molecules interact first. Thermal loads due to gas condensation even at a relatively high inlet pressure are therefore smaller than radiation loads.

Consider now a simple model for the pumping of a particular gas, as depicted in Figure 6.34(b). The upper part represents the vacuum chamber at room temperature T_0 containing gas at pressure p_0. An aperture of area A is the entrance to the cryopump, assumed to be at a uniform low temperature T_S, in which molecules exist at pressure p_S in equilibrium with the condensed phase on the cryopump walls. Arrows represent fluxes J_{IN} and J_{OUT} into and out of the pump, associated with p_0 and p_S, respectively. The net inward particle flux is

$$\dot{N} = AJ_{IN} - AJ_{OUT}$$

Substituting for J's from Equation 3.23

$$\dot{N} = \frac{Ap_0}{\sqrt{2\pi mkT_0}} - \frac{Ap_S}{\sqrt{2\pi mkT_S}}$$

Were $p_S = 0$, then $\dot{N}$ would have a maximum value $\dot{N}_{max}$ equal to the first term so that all incident gas would be pumped. Therefore

$$\frac{\dot{N}}{\dot{N}_{max}} = 1 - \frac{p_S}{p_0}\sqrt{\frac{T_0}{T_S}} \tag{6.28}$$

At the ultimate pressure, $\dot{N}$ becomes zero and in the chamber the value of p_0 will have fallen to p_u given by

$$p_u = p_S\sqrt{\frac{T_0}{T_S}} \tag{6.29}$$

This is simply the thermal transpiration effect discussed in Section 3.11. For $T_S = 15$ K, the ultimate pressure in the chamber would be higher than that in the pump by a factor $(295/15)^{1/2} = 4.4$. Substituting in Equation 6.28 gives

$$\frac{\dot{N}}{\dot{N}_{max}} = 1 - \frac{p_u}{p_0}$$

It may be straightforwardly shown that $\dot{N}_{max}$ is related to aperture conductance C_A and the maximum possible pumping speed S_{max}, and so we may write

$$\frac{S}{S_{max}} = 1 - \frac{p_u}{p_0} \tag{6.30}$$

When $p_0 >> p_u$, the pumping speed has its maximum value, and it is not significantly reduced until the pressure approaches its ultimate value.

This model is rather idealized and oversimplified by the assumption that all molecules arriving at condensing surfaces stick to them. More sophisticated analysis is presented in van Atta (1965) and Roth (1990). Nevertheless, it describes the impingement and capture of water vapor on the entrance louver quite reasonably. Measured pumping speeds for water are close to the theoretical maximum. For example, a pump with inlet opening of diameter $D = 10$ cm has a measured speed 1100 l s^{-1} for water vapor, which is close to the theoretical value $930(28/18)^{1/2} = 1160$ l s^{-1}. Other gases have to reach the interior pumping surfaces and are impeded by the louver and interior geometry. There is, therefore, a transmission probability of less than

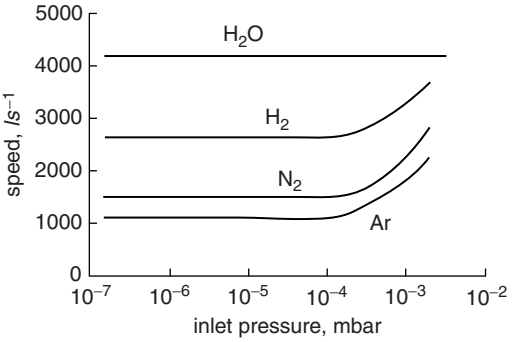

FIGURE 6.36
Pumping speeds of a cryopump for various gases.

100% that they reach the pumping surfaces at 10 K, and on account of this and the efficiency of the capture process, speeds are reduced to well below the maximum values of the entrance conductances.

Figure 6.36 shows measured pumping speeds vs. pressure for a pump with an inlet diameter D = 20 cm. It should be stressed that such data are characteristic of a pump that has been freshly regenerated (see the following text). The speeds are 1500 l s^{-1} for nitrogen, 1200 l s^{-1} for argon, 2500 l s^{-1} for hydrogen, and 4000 l s^{-1} for water vapor. It is interesting to take the ratios of the measured speeds for these gases to the pre-louver entrance conductances, which are $9.3 \times 20^2 = 3720$ l s^{-1} for nitrogen, $3720 \times (28/18)^{1/2} = 4338$ l s^{-1} for water, and $3720 \times (28/2)^{1/2} = 13900$ l s^{-1} for hydrogen. The overall capture efficiency, or capture coefficient, is over 90% for water, as previously noted, about 40% for nitrogen, and only about 16% for hydrogen. The pumping speed for the cryosorbed gases decreases as adsorption sites are used up and the adsorbed population increases with the associated increase in p_S. This is particularly important in the case of hydrogen, which is a natural constituent in the outgassing of vacuum chamber walls, whereas the others are not. Except for water vapor for which the speed is constant, speeds are (unusually) higher at the upper end of the operating range. This is because access to the interior at these pressures is promoted by the increased conductance values at pressures that are transitional compared with the constant values associated with molecular flow.

As indicated in Equation 6.29, ultimate pressures are controlled by the pressure in equilibrium with the condensed phase. For species such as nitrogen, for example, for which at 10 K $p_S = 10^{-14}$ mbar, this is well below 10^{-11} mbar. But for cryosorbed species such as hydrogen, p_u depends on p_S, the pressure in equilibrium with the adsorbed phase. This depends on the amount adsorbed, and also quite sensitively on temperature. For 50% saturation at 15 K, it is ~10^{-9} mbar, and ~10^{-5} mbar at 20 K. In the pumping of typical mixtures, it is the pumping of cryosorbed gases that determines the ultimate pressures achieved.

As the pumped gases accumulate as condensed solids on the first-stage louver and inner second-stage parts, cryosorbing surfaces become saturated, and the point is reached where performance deteriorates to a level that is not acceptable. The pump then requires regeneration. The loss of performance may be indicated as a marked drop in pumping speed or an excessively long time for pump down. The loss of performance will depend very much on the circumstances of use and particularly the gas mixture pumped. Associated with it may be the accumulation of condensate layers up to ~cm in thickness, partly blocking the louver. The capacity of a pump for a given gas is defined as the quantity in mbar × liter units that can be pumped before the speed for that gas falls to half its original value. For example, the 20-cm pump discussed above might have a capacity of 400,000 mbar l for nitrogen and therefore be able to pump at a constant pressure of 1×10^{-5} mbar with speed 1500 l s^{-1} for a time $(4 \times 10^5)/(1500 \times 10^{-5}) = 2.7 \times 10^7$ s, which is about 10 months. Average pressures will determine the period of useful service. Capacities may be alternatively specified in slm or sccm (see Section 5.2).

In principle, regeneration consists of closing the high-vacuum isolation valve, switching off the refrigerator, and allowing the pump to warm up so that trapped gases are expelled via a suitable exhaust valve. A safety valve is also incorporated because the rapid warming of condensates, if not properly vented, could lead to large internal pressures and an explosion hazard. In practice, quite sophisticated procedures are used, first purging with warm clean dry nitrogen at atmospheric pressure in a way programmed in the control unit that reduces the time to accomplish regeneration and protects the sensitive cryosorbing elements from contamination. Sometimes partial regeneration is carried out for only the second stage condensates that may have become saturated prematurely, depending on the conditions of use. These important practical and operational matters are discussed in various texts, notably Hablanian (1997) and Bigelow (1994), and in manufacturers' instructions. Once regeneration is completed, a rough pumping connection, ideally to a dry pump, enables low vacuum to be achieved, prior to restarting the refrigerator, activating the pump, and opening the high vacuum valve to resume high-vacuum operation. If this roughing is carried out with an oil-sealed rotary pump, it should be terminated at about 1 mbar so that oil vapors cannot reach the sensitive sorbing surfaces by back-diffusion.

Reconnection of the pump to the vacuum chamber is made only when the chamber pressure has been reduced by primary pumping to a value referred to as the "crossover pressure" p_X. It is associated with the pump's crossover capacity, denoted by c_X. This is the maximum quantity of gas, expressed in mbar l, which can be suddenly presented to the pump as an impulsive load without overloading the second stage so that its temperature rises above 20 K. If that happens, previously cryopumped gases desorb disastrously. For a pump of the size we have been discussing, c_X is about 200 mbar l. The crossover pressure, therefore, depends on the chamber volume, v, say and $p_X = c_X/v$. For a vacuum chamber of volume 100 l, p_X will be 2 mbar. On reaching this chamber pressure, the high vacuum valve to the already

activated cryopump can be safely opened. The relatively high crossover pressures, allied with a dry pump to rough both pump and chamber, offer an attractive oil-free pumping combination. These pumps are available in a range of sizes, the largest having speeds of order 10,000 l s^{-1}.

6.4.3 Water Vapor Pumps

Arising from the need to enhance the pumping of water vapor in many high-vacuum applications, and the existence of refrigeration technology that allows the maintenance of cryopumping surfaces without the use of liquid nitrogen, a design of cryogenic pump has emerged specifically to pump water vapor. Times to pump down to high vacuum are considerably shortened, thereby making production processes more efficient.

Essentially, a condensing plate (cryopanel) of large surface area is attached to a refrigerated cold head maintained at a controlled temperature of about 110 K, sufficient to condense water vapor completely with a pumping speed, as we have seen, of 14.7 l s^{-1} per cm^2. In one design, the "inline" pattern, the cryopanel in the form of a louver or chevron, is located inside a collar that, with its externally attached refrigerator, can be mounted on top of the inlet to a turbo or other secondary pump. For a collar of 10 cm bore, the pumping speed is 1100 l s^{-1} and the conductance for nitrogen is 450 l s^{-1}. In another design, a suitably shaped cryopanel is located inside the vacuum chamber supported and regulated by attachment to a flange. These can be particularly useful in the context of argon sputtering at relatively high pressures, where the ability to selectively pump water vapor and prevent its contributing to the discharge can be critical.

6.4.4 The Titanium Sublimation Pump

This pump is used in an auxiliary role to augment pumping in some high- and ultrahigh vacuum applications. In essence, it works by capturing molecules in strongly bound chemisorbed states on freshly deposited surfaces of titanium film, forming stable surface compounds such as titanium hydride, oxide, nitride, etc. This type of capture process is called "gettering" and is discussed in detail by Ferrario (1998). When freshly deposited, titanium has a highly reactive surface, gettering most gases except helium, argon, and other chemically inert species, including methane. In one widely used design, titanium sublimes from the surface of a filament made of Ti85%Mo15% alloy. The filament is about 20 cm long and 2 mm in diameter, and is resistively heated by passing a current ~40 A through it. In contrast to those of the pure metal, alloy filaments retain their mechanical stability at temperature and during cyclic use. The molybdenum exerts negligible vapor pressure at the temperatures used, about 1000°C. Usually three filaments are mounted on a flange, as illustrated in Figure 6.37. Current to the supporting electrodes is supplied via feedthroughs in the flange. The

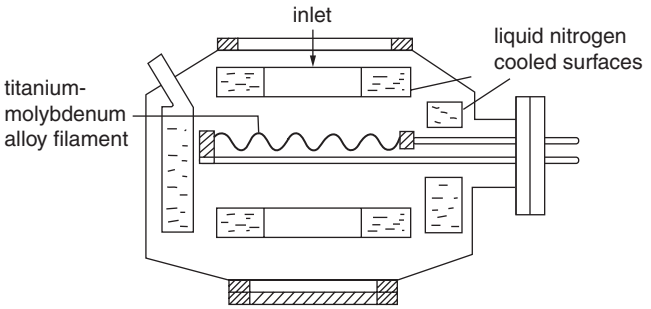

FIGURE 6.37
Titanium sublimation pump. (From Chambers, A., Fitch, R.K., and Halliday, B.S., *Basic Vacuum Technology*, 2nd ed., Institute of Physics Publishing, Bristol, 1998. With permission.)

filaments get thinner with use and eventually fail, so spare capacity is useful, though filaments have to be degassed prior to use.

Ideally, the surfaces onto which titanium is deposited are chilled to improve the capture rate, and they also provide a shield to prevent it from reaching inappropriate parts of the system. Table 6.1 shows measured speeds, in the useful unit l s^{-1} per cm^2, for various gases in circumstances of direct gas access with no conductance limitation.

Operation does not normally need to be continuous for long periods. The highest pressure for continuous operation is about 10^{-5} mbar. For pressures higher than this, the gas arrival rate exceeds the rate of supply of titanium to provide gettering sites on the depositing surface, which is determined and limited by the rate of sublimation from the filament onto the surrounding pumping surfaces. At pressures 10^{-6} mbar and less, fresh surface is created by depositing titanium only intermittently to replenish the pumping speed when it is required as the deposited layer becomes saturated. Freshly deposited surface has pumping speeds of the order of those given in the table. At low pressures, the time to saturation is longer and replenishment is necessary only at long intervals. The power supply and controller enable deposition periods to be timed. TSPs, as they are usually called, are very useful in dealing with anticipated large gas loads at higher pressures to reduce the time to reach working pressure in a process, and also in retaining UHV conditions at 10^{-10} mbar and less. Their operation is discussed in detail by Welch (1991).

TABLE 6.1

Pumping Speeds in l s^{-1} per cm^2 for Various Gases

Gas	O_2	N_2	H_2	CO	H_2O
Room temperature, 295 K	9	4	3	8	3
Chilled, at 77 K	11	10	10	9	14

6.4.5 Nonevaporable Getter (NEG) Pumps

Developed by SAES Getters S.p.A of Milan, NEG pumps are so called because, unlike the TSP, the active gettering surface is generated on the surface of the material itself, rather than by evaporating it from a source onto other surfaces. The working materials are metallic alloys of zirconium with other metals. Two of particular importance are Zr70V25Fe5, with trade name St 707, and St 101 of composition Zr84Al16. They getter all the active gases and are especially efficient in pumping hydrogen and its isotopes. Inert gases are not pumped, and methane and hydrocarbons only when operating at elevated temperatures. Getter St 707 is the more widely used; St 101 has enhanced performance for hydrogen but has to be operated at higher temperatures.

The alloys are prepared in powdered form with particle size 50 μm (1/20 mm) so as to present a large effective area and are bonded onto both sides of thin metallic sheet, usually constantan, to support them. This permits their configuration into various shapes. Heat is necessary for their activation and operation. In strip form, this can be supplied as resistive heating, but more frequently it is by radiative heating from an adjacent heater. Figure 6.38 illustrates (a) a strip and (b) a cartridge configuration in which a sheet multiply folded in concertina fashion is placed around a central heater. The cartridge can be mounted on a flange to protrude directly into the vacuum space or in a flanged cylindrical housing that is attached to the vacuum chamber as an appendage pump. The cartridge shape presents a relatively large area with good gas access for pumping.

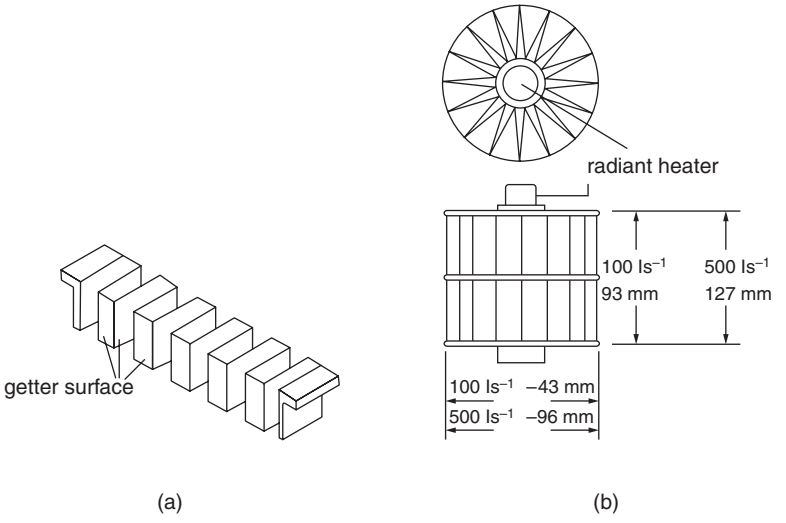

FIGURE 6.38
(a) Strip, (b) cartridge configuration of an NEG pump. (From Chambers, A., Fitch, R.K., and Halliday, B.S., *Basic Vacuum Technology*, 2nd ed., Institute of Physics Publishing, Bristol, 1998. With permission.)

As supplied, getters are naturally passivated by oxide and carbide layers due to contact with the atmosphere. But once activated by heating for about 1 h in a vacuum of about 10^{-5} mbar at 350–450°C (St 707), fresh metallic gettering surface is exposed. In the course of this activation, the oxides and carbides are dissolved and diffuse into the bulk. If now cooled to lower temperature or even room temperature, the fresh surface pumps all active gases such as N_2, O_2, CO, and CO_2 by first dissociating them and then trapping them in chemisorbed states as surface nitrides, carbides, etc. By maintaining an elevated temperature, gases are continually pumped at the surface while products diffuse into the bulk. Eventually, surfaces saturate and have to be activated again. Saturation is reached sooner, the lower the operating temperature. Reactivation is necessary when the pumping of active gases causes surface saturation, or whenever it has been necessary to vent the vacuum chamber to air. Repeated reactivations are possible until the getter has ingested so much that its capacity (see the following text) is reached, at which point the cartridge has to be replaced.

The pumping mechanism for hydrogen is different. Molecular hydrogen first dissociates and then the hydrogen atoms rapidly diffuse into the bulk to form a solid solution. The process is reversible, principally because of the ease with which H atoms can diffuse even at room temperature. Figure 6.39(a) shows some thermodynamic data. The quantity dissolved in the bulk is associated with a unique pressure of hydrogen in the adjacent gas phase. The equilibrium conditions are strongly temperature dependent. For example, for a hydrogen concentration in the bulk of 1 mbar l per gram at 400°C, the pressure is 10^{-4} mbar, though at room temperature it is less than 10^{-16} mbar. When the temperature of the getter is such that the dissolved hydrogen it contains has an associated equilibrium pressure less than the partial pressure of hydrogen in the vacuum chamber, pumping of hydrogen will occur. If these conditions are sustained long enough, the bulk concentration will build up until the equilibrium pressure is reached, in which state the pumping would have reduced to zero. From Figure 6.39 one deduces that if the getter were reactivated at 400°C after saturation due to active gas pumping and had previously pumped say 0.1 mbar l of hydrogen, this gas would be expelled until another equilibrium at lower partial pressure was reached.

There are approximately 3×10^{19} molecules in 1 mbar l and this corresponds to about 0.1% of the number of atoms in 1 g of the host alloy. At 10 mbar l per gram of hydrogen, the embrittlement limit is approached, beyond which stresses in the material cause it to break down mechanically and flake. Water vapor and hydrocarbons dissociate to be trapped as surface oxides, etc., while hydrogen is taken into the bulk.

Figure 6.39(b) shows pumping speed curves for different gases at 280°C and room temperature of a cartridge pump (SAES SORB-AC GP200) that has a 10-cm diameter inlet to its housing. It contains 170 g of getter alloy that presents an area 0.45 m^2. Pumping speeds are plotted vs. the sorbed quantity of gas in torr liter (= 1.33 mbar l). For hydrogen, the pumping speeds are about 500 l s^{-1} at both temperatures, indicating that the limitation on

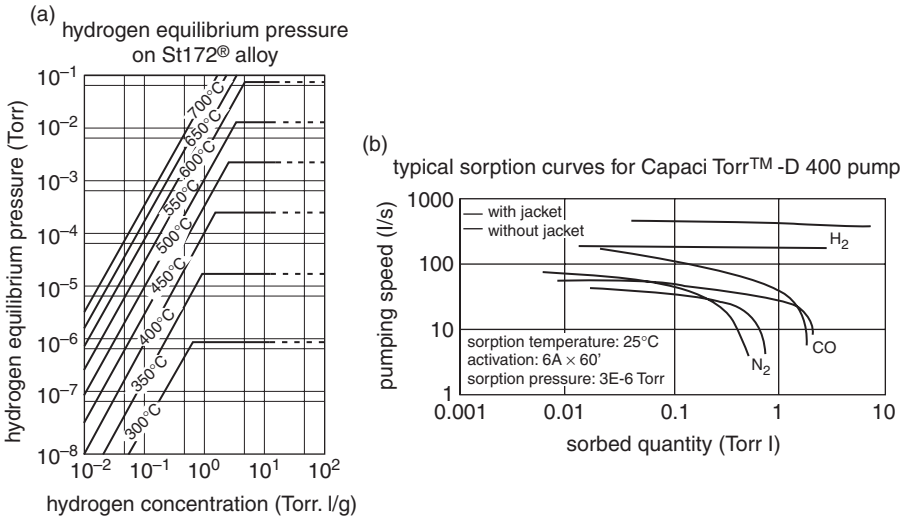

FIGURE 6.39
NEG pump: (a) thermodynamic data, (b) typical pumping curves. (Used with permission of SAES Getters, S.p.A., Milan, Italy.)

pumping speed is due to factors other than the temperature, such as access. The limiting capacity for room temperature operation is 1700 torr l. For the chemisorbed CO, the speed is initially 180 l s^{-1}, diminishing less rapidly at the higher temperature for the reasons discussed. While operation at elevated temperatures favors the diffusion of active gases into the bulk and prolongs the period of this useful pumping prior to reactivation, it also means that the equilibrium hydrogen pressure will be higher. Operation at elevated temperatures is preferred for large general gas loads at high vacuum, but to pump hydrogen in UHV conditions requires that getter St 707 be operated at room temperature. The 0.45 m^2 = 4500 cm^2 of active surface assuming 5×10^{14} sorbed molecules per cm^2 implies saturation for room temperature operation at 2.2×10^{20} molecules equivalent to $(2.2 \times 10^{20}/3 \times 10^{19}) = 7$ mbar l approximately, consistently with the data of Figure 6.39(b). The total capacity for active gases is of the order of 2000 mbar l so that many reactivations are possible. For continuous elevated temperature operation at 10^{-6} mbar with average speed 200 l s^{-1}, a crude calculation suggests that the lifetime would be $2000/(200 \times 10^{-6}) = 10^7$ s, which is several months. With uninterrupted use in UHV conditions, reactivation may not be necessary for even longer periods. A regeneration treatment to remove hydrogen is possible when its capacity has been reached and if the accumulated load of irreversibly pumped gases allows. The number of times that the pump can be reactivated for active gases is limited because the pumping of these is not reversible. Eventually, the getter material has to be replaced.

Because they can be configured in various ways, there are many applications for these versatile pumps. A notable example has been to provide

continuous linear pumping for the 28-km vacuum ring of the LEP, now being rebuilt as the large hadron collider. They are widely used to boost the pumping of hydrogen in turbo-, cryo-, and ion-pumped HV/UHV systems. One design of the ion pump has an incorporated NEG element for just this purpose. They are an equally attractive pump for UHV installations that are rarely vented because they can pump at room temperature and therefore with zero power for long periods. Also, they provide a means of keeping inert gases free of active gas contamination. In pellet form, they assist in maintaining the static vacuum in cathode ray tubes.

6.4.6 The Sputter-Ion Pump

Ion pumping effects were evident in the first half of the 20th century in devices such as thyratrons, which depended on electrical discharges in gases. The detrimental and annoying phenomenon of "clean-up" occurred as ionized gas molecules, with enhanced chemical reactivity compared with their neutral state, were captured by cathode surfaces, causing a reduction of pressure. Similar effects had been noted by vacuum practitioners as a complicating feature in the operation of ionization gauges. It was not until about 1950, however, that the effects involved were pursued with the specific aim of making pumping devices. The history of their development is discussed by Weston (1985). Ion pumps, better described as ion-getter pumps because they exploit ionization and titanium gettering, were developed in a number of forms with various operating principles. Of these, the sputter-ion pump has acquired a dominating position and is so called because, in addition to purely ionic pumping mechanism (to be described), part of its pumping action arises from the laying down of freshly deposited titanium gettering film by the process of sputtering. This is the ejection of a number of atoms from a surface due to the impact of a single energetic particle such as an ion accelerated in an electric field, as discussed in Chapter 4.

The basic element of a sputter-ion pump is the Penning cell, in which a self-sustaining "Penning discharge" is set up. The cell, depicted schematically in Figure 6.40, is a diode consisting of a short cylindrical tube anode A, made of thin stainless steel, and two cathode plates K made of titanium whose planes are perpendicular to the cylinder axis. The cell is quite small. Typically, the anode cylinder is 15 mm in diameter and 20 mm long, with the cathodes separated from it by 4 mm. A strong magnetic field B ~0.15 T from permanent magnets external to the stainless steel pump casing E is directed parallel to the cylinder axis. The cathodes are earthed and the anode is at a potential 3–7 kV.

The discharge set up in this cell has a complex character that has been widely investigated and which is described in detail by Henning (1998). The discharge is initiated by the creation of an electron due to a cosmic ray interaction somewhere in the cell, or field emission at a cathode. Its acceleration to the anode is prevented by the presence of the strong magnetic field that causes it, on account of the $v \times B$ force, to spiral around a magnetic field

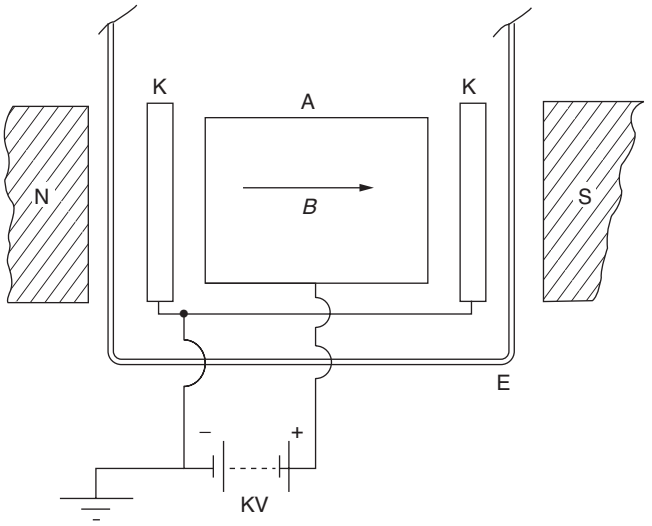

FIGURE 6.40
Sputter-ion pump cell.

line. The outcome of a cascade of ionizing collisions, each of which yields a positive ion and an electron, is that a dense electron cloud becomes trapped in the anode cylinder, confined by the magnetic field and restrained axially by the cathodes at ground potential. The electrons follow long-lasting spiral trajectories around field lines back and forth between the cathodes, and it is this, in contrast with the rapid exit of partner ions from the region, that causes them to dominate the electrical behavior in the cell. What happens to the ions will be discussed shortly. The presence of the large negative space charge causes the potential along the cylinder axis in the central region between the cathodes to be slightly below cathode potential. As a result, the electric field within the cylinder is *radial* and therefore perpendicular to the magnetic field. Outside the cylindrical space charge region, the electric field becomes increasingly axial (and therefore reflecting for electrons) as the cathode is approached. The spiraling back and forth motion of electrons between the cathodes is accompanied by a rotation of the charge cloud about the central axis due to what is called $E \times B$ drift, a feature of charge motion in perpendicular E and B fields. The rotation rate has been measured to be about 70 MHz. It constitutes a circulating "ring current" of about 1 A, in which the density of electrons is very high. This rotating cloud is the principal characteristic of the Penning discharge. Once established, if the neutral background of molecules could be removed, it would persist as a stable entity. The ring current is substantially independent of pressure from 10^{-4} mbar down to very low pressures. Above 10^{-4} mbar it is converted to a glow discharge that can extend outside the cell.

The high electron density in the cloud means that there are many ionizing collisions of electrons with the neutral gas molecules in the cell, each

producing a positive ion and a new electron. While one electron remains in the discharge, the excess electron, urged outward by the electric field but constrained by the effect of the *B* field, diffuses slowly out of the discharge to reach the anode and contribute to the discharge current. The positive ion, on the other hand, being relatively massive, is little affected by the magnetic field and accelerates under the influence of the electric field to strike the cathode, falling through a large potential difference. The impact of ions causes titanium to be sputtered in all directions and deposited on the interior of the anode cylinder and the opposite cathode. In the outer region of the cathode it builds up with time, whereas the sputtering causes the central region to be eroded, forming a depression. The freshly deposited titanium surfaces pump the neutral molecules of active gases such as N_2, CO, etc., by chemisorption as in the TSP. The pumping is enhanced if molecules have been dissociated or otherwise excited in the discharge. This is one of the principal pumping mechanisms of the cell.

Another mechanism is true ion pumping in which some of the energetic incident ions penetrate deeply into the cathodes to become buried as stable trapped compounds. However, depending on their location, some may be later released by resputtering. This is particularly the case for argon atoms that are only weakly bound. The principal pumping mechanism for them, in pumps that have been run for a while so that burial and resputtering have come into balance, has been shown to be their rebound as neutrals from the cathode with appreciable energy and burial elsewhere after a trajectory uninfluenced by the fields.

A third distinct mechanism, principally at the anode surface, is the simple burial of gas atoms or molecules, briefly resident there, under titanium atoms that are deposited on top of them. A schematic representation of these mechanisms is given in Figure 6.41. Gas molecules that are captured in the titanium deposit on the interior anode surface stay trapped because this region is not subjected to ion bombardment, whereas the cathodes are, with the effects previously noted.

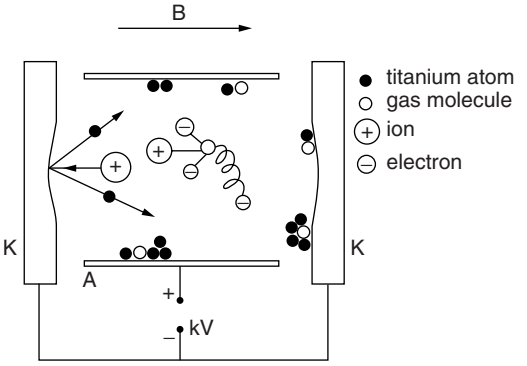

FIGURE 6.41
Pumping mechanisms in diode sputter-ion pump.

The pumping speed of a single cell even for active gases is only about 1 l s^{-1}, and so working pumps contain many such cells in an array, with arrays distributed in pockets of the stainless steel pump casing. The protrusions allow economic use of both poles of the external magnets. The total ion current is roughly proportional to the molecular number density and hence gas pressure. This means that the pump is self-regulated; the amount of sputtering is the amount needed to deal with the gas load, and the metered pump current is a guide to the pressure.

The sputter-ion pump will pump all types of gases. Of the active gases, the pumping of hydrogen differs from that of others such as N_2 and CO. Hydrogen ions do not, because of their small mass, cause much sputtering of the cathode. On the other hand, once dissociated, hydrogen atoms are small and mobile and penetrate very deeply into it by diffusion, and this is their principal pumping mechanism. The cathode bulk has a very large capacity for hydrogen, and when, as is frequent, other gases are being pumped, the sputtered film deposited on the anode also serves similarly as a good diffusion matrix for arriving hydrogen. As a result, pumping speeds for hydrogen are about double those for air. Hydrocarbons, such as normally stable methane, are pumped by being broken down into more active fragments in the discharge and then gettered, with speeds comparable to that for air.

Of the inert gases, helium is surprisingly well pumped, it is thought, because of its small size that enables easy diffusion into the bulk. It has speeds ~20% that of air. Argon, because it has a small speed — only 1–2% that of air — may present certain problems in prolonged use under high vacuum at pressures ~10^{-6} mbar. When it is a working constituent of the vacuum chamber residual gas, or, if there are tiny leaks to the atmosphere (where its partial pressure is about 10 mbar), it may slowly accumulate in the system to the point where it cannot be stably retained. Argon instability (see Welch, 1991) is the term used to describe the ensuing periodic pressure bursts up to about 10^{-4} mbar that occur, followed by recovery. There are designs of pumps that avoid this potential problem. One exploits the "neutral rebound" mechanism referred to earlier. At some loss to general pumping speed, one of the cathodes is made of tantalum instead of titanium. Because of its much higher atomic mass (Ta — 181 amu, Ti — 48 amu) the probability and energy of rebound for argon atoms (40 amu) is much greater. The energetic neutrals are safely buried in the anode and the result is an argon pumping speed ~20% of that for air. The other, more favored, strategy is to use slotted cathodes, as shown in Figure 6.42(a). Argon ions incident at grazing incidence on the slot sides may be reflected as neutrals without great loss of kinetic energy and therefore penetrating power, and the enhanced sputtering that occurs at small angles causes burial under a greater number of deposited layers at the bottom of the slots.

Another design that markedly improves argon pumping is the so-called triode configuration of the pump, shown in Figure 6.42(b), in which the cathode has an open grid structure. The anode cylinder is now at earth

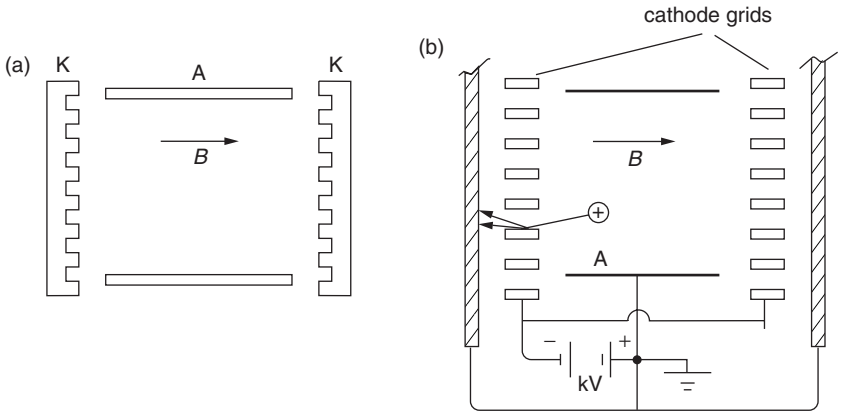

FIGURE 6.42
(a) Slotted cathodes, (b) structure of triode pump.

potential, as is the pump casing, and the power supply provides negative kV for the cathode. In spite of its open structure, the grid still provides a continuous two-dimensional region at cathode potential. Ions passing through the grid will not reach the casing, effectively an auxiliary anode, because of the reversed electric field, but will oscillate back and forth before eventually reaching the cathode. But ions grazing the sides of the cathode grid will sputter Ti onto the pump casing with much greater efficiency than for normal incidence sputtering, to form non-ion-bombarded gettering surfaces in addition to that of the anode cylinder. Argon is pumped efficiently at about 20% of the air speed because the ions neutralized by glancing incidence at the grid retain much more of their kinetic energy than when rebounding backwards in the diode, and so proceed to the pump casing beyond to be very effectively buried in the surface of the casing.

Figure 6.43 shows air pumping speed curves for both a diode and a triode pump with 10-cm inlet diameters, for comparison with other pumps, though many pumps are larger with a 15-cm inlet. Pumping speeds are a maximum

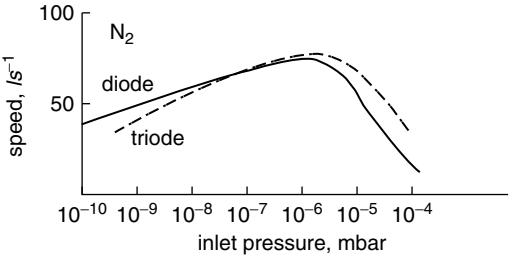

FIGURE 6.43
Pumping speed curves for diode and triode pumps.

in the region of 10^{-5} to 10^{-7} mbar, falling off at lower pressure. Because of this and because the pumps operate at a maximum pressure of about 10^{-4} mbar, the throughputs they can handle are low compared with turbomolecular or diffusion pumps of the same speed. However, when brought down to operating pressure by contamination-free rough pumping, they provide scrupulously clean ultrahigh vacuum environments. The practicalities of their UHV use are informatively dealt with by Singleton (1998) and Bigelow (1994) giving a useful general discussion of the pumps' operation.

7

Measuring a Vacuum

7.1 Introduction

Although the devices developed to measure vacuum are frequently referred to as pressure gauges, in many applications of vacuum technology the use of the concept of pressure as a force per unit area is not relevant. In just a few low-vacuum applications, such as vacuum molding, pressure is a quantity of explicit importance because they exploit the mechanical force created by the difference between atmospheric pressure and a pressure of about 100 mbar. And, of course, all vacuum vessels operating at essentially zero internal pressure must be constructed to withstand the large forces placed on them by the external atmosphere. But for most applications of vacuum, it is n, the number density of molecules in the residual gas, that makes a vacuum suitable for its intended purpose by determining, for example, the scattering of accelerated particles out of the beam in a high-energy particle physics experiment, the mean free path λ in a vacuum-coating application, or the residual gas impact rate J on surfaces in a surface science experiment. We know, however, that n and p are related by the simple equation $p = nkT$, and with the understanding that a fixed temperature of reference is taken, room temperature, 295 K, the familiar and convenient pressure p serves as a measure of vacuum over its whole range.

It will be evident from the discussion of previous chapters that the vacua to be measured span a range of at least 15 orders of magnitude, from pressures just below atmospheric down to 10^{-12} mbar or less. No single type of sensing device can cover this range, and as one might anticipate, those that measure force directly are restricted to higher pressures. Determination of lower pressures requires the measurement of other phenomena in the vacuum that depend, ideally in linear fashion, on the number density n. Thus, in particular ranges of vacuum, the effects of thermal conduction, viscosity/drag and ionization are exploited, and in essence n is measured and p inferred because of its proportionality to n. Such methods are described as indirect, and calibration is usually required. It is worth stressing the enormous change of the n value associated with the pressure range from

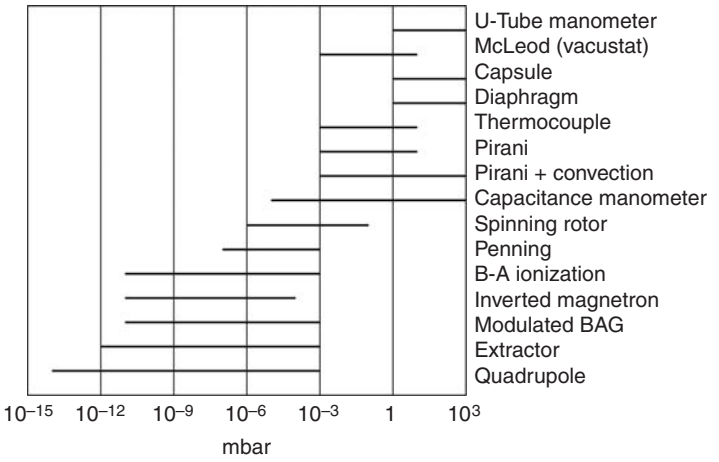

FIGURE 7.1
Range of operation of various gauges.

atmospheric to 10^{-10} mbar, from 10^{19} down to 10^{6} molecules per cm^3, respectively, as may be deduced from Table 3.1.

The direct and indirect gauges to be described and their range of operation are listed in Figure 7.1 and are restricted to those that are widely used and commercially available. A more comprehensive survey of the wide range of types developed is given, for example, in Roth (1990). Because the operation of most vacuum systems requires the use of a primary pump to rough from atmosphere and a secondary pump to attain the higher degrees of vacuum, more than just one type of gauge is usually needed to span the range of operation, monitoring the pump-down process and the ultimate state achieved.

For gauges that measure force directly, the reading does not depend on the identity of the gas. For equal molecular number densities n of different gases, the force exerted has the same value $p = nkT$. But for indirect gauges exploiting a property that depends on n, such as gas ionization, for example, the magnitude of the response for equal number density will vary from gas to gas because some gases are more easily ionized than others. Thus, an inert gas such as helium with its closed electronic shell structure is more difficult to ionize than methane, which is very easily ionized. These gas dependencies and how they are allowed for will be discussed later in the context of particular gauges. Nitrogen is usually taken as the gas of reference in terms of which scales are calibrated.

It is very frequently the case that the gas whose pressure is to be measured is a mixture of different gases and the considerations related to those mentioned earlier apply. Gauges that measure force directly record the total pressure of such a mixture correctly because, in molecular terms, the contributions due to the surface impacts of the various species are proportional to their individual number densities n, each with the same constant of

proportionality, *kT*. In a sense, therefore, the gauge responds to the summed partial pressures, without of course distinguishing them. For indirect gauges, it is necessary in principle to allow for the different sensitivity of their mechanism to the different components of the mixture, but this composition is not known unless a partial pressure analyzer is available. In practice, the total pressure would need to be known with the accuracy implied rather rarely, and in many applications these refinements may be justifiably ignored, though as discussed at the end of Section 7.3.1., there are exceptions, and they have an important safety aspect. For many commonly encountered residual gas components in high vacuum, the relative sensitivities are not markedly different from that for nitrogen, and there are in any case other more basic device-imposed limitations on the accuracy achievable.

Knowing the composition of a residual gas mixture, so as to be aware of its potential for chemical activity, for example, is an important characteristic of the vacuum in many diverse applications and, in some, is arguably more important than knowing the total pressure. Residual gas analyzers, also called partial pressure analyzers, provide another measure to characterize the vacuum and will be described at the end of this chapter.

With the exception of the simple but rarely used U-tube manometer and McLeod gauge, most gauges require calibration because they do not measure pressure directly. Calibration may be done at the time of manufacture, but may drift with time. In addition to concerns about accuracy, stability, and other operational matters to be discussed for various gauges, the optimum location and mounting of a gauge is important. Ideally, they should be located as close as possible to where the pressure is required and mounted vertically so that debris cannot fall into them. This is also the position in which they will have been calibrated.

Gauges should also be mounted and used in such a way as will have minimum disturbance on the system and will be able to sample it appropriately. For example, in Figure 7.2, continuum flow at relatively high pressure through a pipe with an elbow is measured by two identical gauges A and B placed close together but with different orientations to the flow. A typical streamline of the flow is XYZ, and the effective location of the actual pressure sensor is within the gauge heads at S. Gauge A measures the true static pressure at Y because, once steady state is achieved, there will be no pressure

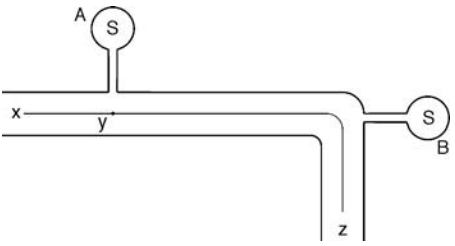

FIGURE 7.2
Pressure measurement at different locations.

drop between S and Y. Gauge B, on the other hand, will record a higher pressure including a dynamic component because, essentially, force supplied by gas at a higher pressure has to stop the movement of gas towards the gauge and make it turn the corner. In practice, for example, for a high flow rate of 1.6 l s^{-1} in a 25-mm bore pipe, gauge B's reading was 33 mbar and A's 30 mbar. Gauge B is therefore not properly located for measuring the static pressure in the flowing gas. In the context of molecular flow, gauges should be located so that they sample gas isotropically, as far as this is possible. If a gauge is located too close to a pump and especially if it has sight of the pump entrance, then the sampled flux will be reduced below its quasi-equilibrium value. Equally, if a gauge that is openly located within a vacuum chamber receives directed flux due to beaming of gas injected into the system, its reading will be too high and not characteristic of a random isotropic distribution.

Gauges based on pressure-induced effects, thermal properties, viscous drag, and ionization are described in the following text in that order, which accords roughly with their place in the spectrum of measurement.

7.2 Gauges Measuring Pressure Directly

In the U-tube manometer and the McLeod gauge, pressure in a gas is opposed by a liquid surface acting as a piston; in other devices, pressure causes a well-characterized deformation of a solid sensor, which is converted to a pressure reading.

7.2.1 U-Tube Manometers

Though nowadays rarely used, the simple U-tube manometer depicted in Figure 7.3(a) affords a simple and direct measure of pressure differences. It consists of transparent tubing, usually glass, that contains a liquid of known density ρ. The pressure difference $p_1 - p_2$ is measured by the difference in height h between the meniscus levels in the two columns, according to the equation

$$p_1 - p_2 = \rho\, g\, h \tag{7.1}$$

For work of the highest accuracy, the local value of g, the acceleration due to gravity, and the precise density of the liquid at the prevailing temperature would be employed. With one limb sealed at vacuum as shown in Figure 7.1 (b), and mercury (relative density 13.6) as the sensing liquid, so that only mercury vapor exerting a vapor pressure of approximately 10^{-3} mbar at room temperature exists in the closed volume, it is possible to measure from

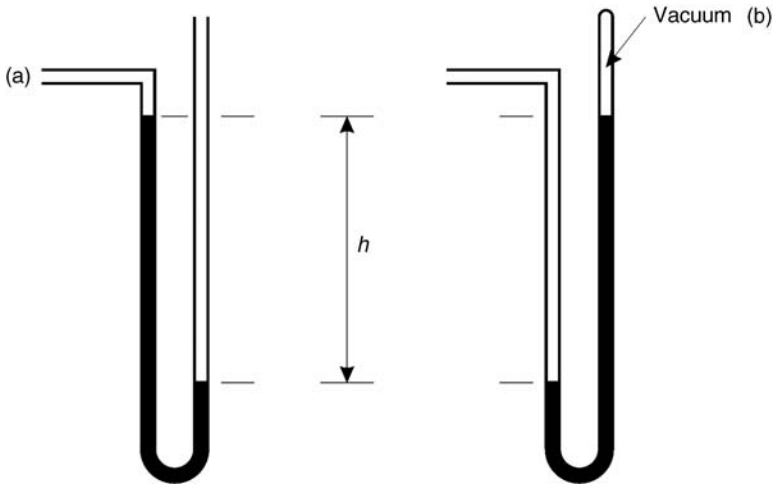

FIGURE 7.3
(a) and (b): U-tube manometer configurations.

atmospheric pressure ~1000 down to about 1 mbar in the other limb. In ordinary use, the accuracy of measuring h by eye against a scale graduated in millimeters would be limited to a millimeter or a little less. The transport of mercury vapor toward the measured vacuum has to be prevented by the use of a liquid nitrogen cooled trap, adding to the inconvenience of the procedure. Occasionally, the configuration of Figure 7.1(a) may be useful for measuring differential pressures. For example, if these were of the order of 10 mbar ($\equiv$ 7.6 mmHg), silicon oil with density approximately 1 g per cubic centimeter and negligible vapor pressure could be used, giving h = 13.6 × 7.6 = 103 mm and reasonable sensitivity against a scale graduated in millimeters.

7.2.2 The McLeod Gauge

The McLeod gauge was widely used until the early 1980s as a standard for calibration purposes and was of considerable importance in early vacuum science. It is now obsolete, but its principle of operation remains fundamental and of academic interest, and a few remain in use. In essence a known volume of gas at the low pressure of the system is isolated and compressed to a higher measurable pressure and volume. From the ratio of the initial and compressed volumes, the initial system pressure is deduced by Boyle's law. The basic elements of the device, constructed in glass, are shown in Figure 7.4. The left-hand limb consists of a large bulb extended at its top by a closed vertical fine capillary tube L and having a total volume V above the junction at level X. The right-hand limb is connected at the top to the vacuum system but divides in part so that there is a capillary section R identical in

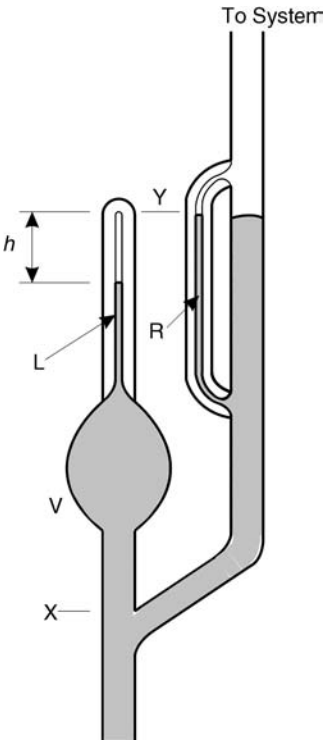

FIGURE 7.4
Schematic diagram of the McLeod gauge.

bore to that of L and located alongside it. Below the junction at X is a reservoir of liquid mercury with the means attached to raise or lower the mercury level.

To make a measurement, the reservoir is pressurized in a controlled way so that the mercury level rises. As it reaches level X, a volume V is isolated at the prevailing system pressure p. As the mercury level is raised further to eventually reach level Y at the point in R opposite the top of the closed capillary in L, the mercury in L has compressed the isolated gas to occupy a length h, which is also the difference in mercury levels in the two capillaries. Thus, if A is the cross-sectional area of the capillaries, by Boyle's law,

$$pV = (h + p)Ah$$

In this equation, p is expressed as a length equivalent to the pressure of a column of mercury. Compression ratios are typically of the order of a thousand so that $p << h$ and therefore with sufficient accuracy

$$p = h^2A/V \tag{7.2}$$

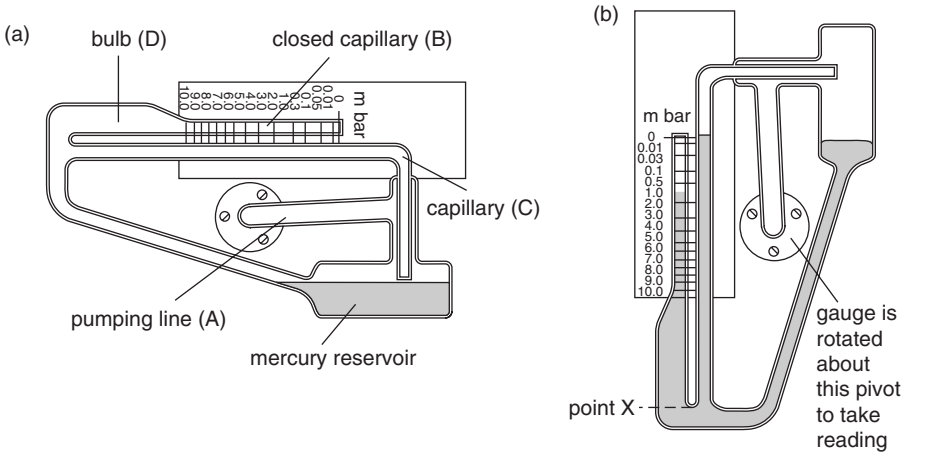

FIGURE 7.5
Vacustat (From Harris, N.S., *Modern Vacuum Practice*, 2nd ed., Nigel Harris Publisher, 2001, and BOC Edwards. With permission.)

The nonlinear scale located alongside capillary L is calibrated directly in pressure units. Despite its attractively simple principle, which yields an absolute pressure value, in practice, the gauge had to be very well constructed and maintained, and required considerable skill in its operation for good accuracy to be achieved. For example, the accuracy of measuring h is critical especially at lower pressures, and any vapors contained in the sampled gas whose saturation vapor pressure is exceeded in the course of the compression will condense, rendering the reading invalid. Typical designs for high vacuum spanned the measurement range from 10^{-2} to less than 10^{-5} torr (mmHg), their natural unit in the context of mercury columns. The many operational concerns relating to the gauge are discussed by Leck (1989).

In spite of its obsolescence as a device for measuring high vacuum, a useful variant designed for the range of 10 to 10^{-3} mbar is still available commercially and very useful in routine laboratory calibrations of other gauges. Shown in Figure 7.5, it is a more compact device than the traditional type described above, is portable, and employs a simple swiveling action to change from the sampling to the isolation and measuring condition. Of course, because of health and safety regulations in force, devices containing mercury, which is toxic, are less acceptable than hitherto.

7.2.3 Gauges Depending on Deformation of a Sensor

These gauges depend on the deformation within elastic limits induced by pressure changes applied to a sensor. There are various types. The relatively small movement created is magnified mechanically so as to move a pointer over a scale. Deflections vary linearly with pressure, and usually therefore, so do the scales.

In the ***Bourdon gauge***, a hollow metallic tube of oval cross section and shaped in the form of a U or a C is closed at one end and connected to vacuum at the other. Excess internal pressure tends to straighten out such a shape, and vacuum pressures therefore have the opposite effect. The reference pressure is that of the atmosphere. The movement of the closed end is converted to a pointer movement across a dial. In its simpler forms this serves as a rough vacuum indicator over the range of 1000 to 10 mbar.

The ***capsule gauge*** is a simple and robust mechanical device that measures pressures from 1000 down to 1 mbar by sensing and magnifying mechanically the small changes in dimension of a capsule that expands as the pressure surrounding it is reduced. The capsule consists of two thin-walled corrugated diaphragms welded together at their edges and hermetically sealed, and is located inside a leak-tight housing that attaches to the vacuum system. An internal lever mechanism moves the pointer over a scale within the housing. Because of their construction, they are not affected by changes in atmospheric pressure and are approximately linear in response. They are supplied by a number of manufacturers and available for ranges such as 0–25 and 0–100 as well as 0–1000 mbar, with accuracies quoted as a few percent of their full-scale deflection. They are rugged and immediate indicators of the vacuum they measure, although the possibility of the mechanism becoming impaired due to corrosion by hostile gases or other reasons may be a concern.

In ***diaphragm gauges***, the pressure difference across a single circular corrugated diaphragm causes a deflection that is coupled with magnification to a pointer and scale. In an ingenious version of the device described by Wutz et al. (1989) and manufactured by the Leybold Company, the reference pressure on the side remote from the pressure being sensed is sealed off at a negligible value of 10^{-4} mbar. Its operating range is from 1000 down to 1 mbar, with an expanded scale in the lower range so that 1–20, 20–100, and 100–1000 mbar occupy comparable lengths of a generous 280° scale. This is accomplished by having a large, very sensitive diaphragm and a set of concentric circular stops that render it progressively less sensitive at higher pressures. The accuracy quoted is ±1 mbar between 1 to 10 mbar and ± 10% of reading between 10 and 1000 mbar.

A series of handheld manometers that utilize a miniaturized ***strain gauge*** is manufactured by Sifam Instruments under the trade name "Digitron." They are portable, easily attached to host systems, and can be supplied to measure either differential or absolute pressures. Their sensor signal is electronically processed and has a 4-digit LCD output display. One type spans the range of 0–2000 mbar with good linearity and accuracy of order 1%.

7.2.4 The Capacitance Diaphragm Gauge

When the flexure of a diaphragm is sensed electrically rather than mechanically, greater sensitivity and precision are possible. The basic principle

employed is that the pressure-induced displacement of a metallic diaphragm with respect to a fixed partner electrode close to it causes a change in their mutual capacitance, which, by appropriate signal conditioning, is converted to a pressure reading.

The flexure of a thin-plane circular diaphragm due to a pressure difference depends on how its edge is supported and whether or not it is pre-tensioned. Two possibilities are shown in Figure 7.6 for diaphragms of radius R across which there is a pressure difference Δp. In Figure 7.6(a), the edge is rigidly clamped by forces perpendicular to the plane. For $\Delta p = 0$, there is no stress in the diaphragm and no deformation. In Figure 7.6(b), however, the edge clamping is also made to exert a large outwardly directed tension T per unit length of the rim so that the diaphragm is radially stressed. These deformations are analyzed in Prescott (1961). The result for the case illustrated in Figure 7.6(a) is given in a convenient form in *Kempe's Engineers Year-book* (2001). The maximum displacement x_0 at the center is

$$x_0 = \frac{3}{16}\frac{\left(1-\sigma^2\right)R^4}{E\,t^3}\Delta p \tag{7.3}$$

where t is the thickness, E is Young's elastic modulus, and σ is Poisson's ratio. Between the center $r = 0$ and the edge at $r = R$, the displacement x decreases with r as $(R^2 - r^2)^2$. For the highly pretensioned diaphragm of case (b), the displacement x depends on r as $(R^2 - r^2)$ and the central displacement x_0 is

$$x_0 = \frac{R^2}{4T}\Delta p \tag{7.4}$$

Also, as proved in Feather (1961), the fundamental frequency f of vibration of such a tensioned diaphragm made of material of density ρ is

$$f = \frac{0.7655}{2R}\sqrt{\frac{T}{\rho t}} \tag{7.5}$$

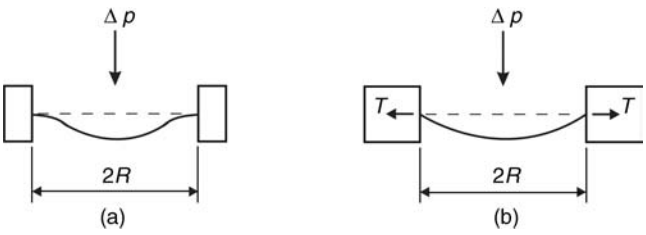

FIGURE 7.6
(a) and (b): Deflection of diaphragms due to a pressure difference.

Exploiting diaphragm deflections to measure a vacuum by their capacitative coupling to a reference electrode was proposed by Alpert, Matland, and McCoubrey in 1951. Such devices have subsequently been very intensively developed and have come to be known as capacitance manometers, or more usually, capacitance diaphragm gauges. They are instruments of superior reliability and accuracy in which advances in engineering materials and in electronics have been exploited. Their early evolution is discussed by Sullivan (1985).

As a simple model of one type of this device, consider the arrangement of Figure 7.7 in which a tensioned diaphragm TT of radius R and thickness t separates two regions at different pressure, the left-hand side at high vacuum ~10^{-7} mbar, essentially zero for mechanical purposes, and the other side at a pressure p. In this mode, the device measures absolute pressure. In a differential mode, the left-hand-side pressure would be held at some reference value other than zero and Δp would be sensed. A fixed electrode E of area A is located close to the diaphragm, and pressure p causes bowing towards it, altering the capacitance. If $p = 0$, the undistorted diaphragm is parallel to E at a distance X so that, discounting edge effects, the two approximate a parallel plate capacitor with capacitance $C = \in_0 A/X$, in accordance with the well-known formula. Otherwise, for a finite pressure, regarding the central part of the diaphragm as being approximately flat and displaced according to Equation 7.4, the capacitance is increased to

$$C = \frac{\in_0 A}{X - x_0} \tag{7.6}$$

The sensitivity and measuring range will depend on the stiffness of the diaphragm. For a device with a nominal range of 0–1000 mbar, in order to produce a reasonable change of capacitance, the high pressure must move

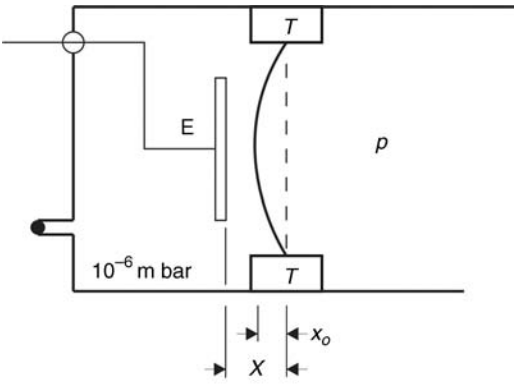

FIGURE 7.7
Simple form of diaphragm gauge.

the diaphragm close to electrode E without touching it, and a relatively stiff diaphragm is required. For a version designed for, say, 10 mbar maximum, a weaker diaphragm will be appropriate. In practice, electrode E is positioned to act as an overpressure stop at approximately 20% above nominal full scale. Its presence is particularly important for more sensitive diaphragms, for which the recommended practice when they are not in use is to isolate such gauge heads by valving them off at low pressure, so that exposure to atmospheric pressure does not force the diaphragm onto the electrode. For a diaphragm of any sensitivity, the lower end of the measuring range at zero deflection is nominal zero, and as we shall see, this is where sensitivity is least.

Let us investigate the orders of magnitude of the various quantities involved in the device. In typical designs, diaphragms are 50 mm in diameter, made of inconel (density 8.5×10^3 kgm^{-3}) and tensioned radially such that their fundamental resonant frequency is about 3 kHz. Putting these numbers into Equation 7.5 yields a ratio of 3×10^8 Nm^{-2} for the ratio T/t — the greater the diaphragm thickness t, the greater the outward tension needed around its rim for the same resonant frequency. Substituting $R = 25$ mm and $T = 3 \times 10^8\ t$ in Equation 7.4 with $\Delta p = p$ gives

$$x_0 = 4.8 \times 10^{-13}\ p/t \tag{7.7}$$

The maximum displacement x_0 corresponding to the full pressure difference must be a little less than X, which is typically 0.2 mm. Therefore, taking x_0 to be 0.16 mm, so that there is another 25% displacement to the overpressure stop, and putting this value in Equation 7.7 with $p = 10^5$ Pa (1000 mbar), gives $t = 3 \times 10^{-4}$ m = 0.3 mm. The associated value of T will be 9×10^4 Nm^{-1}, incidentally, which is quite considerable.

Now consider electrical matters. Taking $A = 1$ cm$^2 = 10^{-4}$ m^2, with $X = 0.2$ mm and approximating $\varepsilon_0 = 8.85 \times 10^{-12}$ Fm^{-1} as 10^{-11} Fm$^{-1} = 10$ pFm^{-1}, the capacitance between electrode and diaphragm is $C = \varepsilon_0 A/X = 5$ pF. When bowed at the position corresponding to a pressure of 1000 mbar, $X - x_0 = 0.04$ mm, and from Equation 7.6 the capacitance is increased to 25 pF. These are small easily measured capacitances, but what is important is the change of capacitance with pressure and its measurability. Setting the capacitor separation $X - x_0 = s$ so that $C = \varepsilon_0 A/s$ and differentiating gives $dC/ds = -\varepsilon_0 A/s^2$ so that, at the lower end of the pressure scale, nominal zero, for which $s = X = 0.2$ mm, the rate of change of capacitance with separation is least, with value 2.5×10^4 pF m^{-1}. The increment $\delta C = (dC/ds) \times \delta s$, associated with a displacement $\delta s = 1.6 \times 10^{-7}$ m that corresponds to a pressure of 1 mbar, is therefore about 4×10^{-3} pF, and thus $\delta C/C \sim 10^{-3}$. The minimum capacitance change that could be detected in this example corresponds to a displacement of about 10^{-9} m and therefore pressure $\sim 10^{-2}$ mbar. The limits of sensitivity are determined by signal-to-noise considerations in circuitry and thermal effects in the transducer.

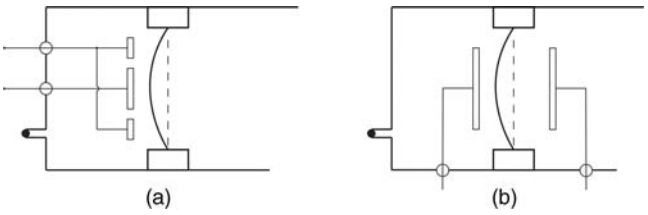

FIGURE 7.8
(a) and (b): Alternative configurations.

In practice the electrode configurations used are slightly more complex than that of Figure 7.7. A frequently encountered arrangement is shown in Figure 7.8(a). It exploits the capacitances between the diaphragm and two static electrodes in the form of a central disc and a concentric annulus mounted in a "bull's eye" arrangement on an insulating ceramic support. The two capacitances form part of an ac bridge circuit, driven at a frequency ~10 kHz, which is balanced at $p = 0$. When the diaphragm deflects under pressure, the central capacitance increases more than the outer, unbalancing the bridge to create a signal that is processed to give an output DC level proportional to the pressure. In the arrangement of Figure 7.8(b), there are again two static electrodes, but the diaphragm is located symmetrically between them. Equal capacitances to the diaphragm at zero deflection allow the balancing of an ac bridge at $p = 0$ in a similar way to that of arrangement (a), and in this case the out-of-balance signal arising from deflection is approximately proportional to pressure. As depicted, both these arrangements are in the absolute mode in which a reference vacuum ~10^{-7} mbar is maintained in the left-hand reference side by activating a getter after sealing off at high vacuum. Both are adaptable to a differential mode if a reference pressure other than zero is established in the left-hand volume. Arrangements of the type in Figure 7.8 (a) have the advantage that the sensing electrodes are isolated by the diaphragm from the vacuum being measured, so that the dielectric properties of the gas involved do not affect the measurement. This is not true for devices of type shown in Figure 7.8 (b), where the small dielectric effects may be a concern for work of the highest accuracy. On the other hand, the symmetric structure tends to give good stability against zero drift caused by temperature changes. These problems arise because, even though the thermal expansion coefficients of ceramic supporting electrodes and the diaphragm are as closely matched as possible, some differential thermal expansion occurs when ambient temperature changes. This effect is quantified by manufacturers and can be allowed for. In some designs, electronic means are used to compensate for temperature-induced change. A more direct, but more expensive, method is to hold the whole sensing assembly at a uniform, controlled, elevated temperature, typically 45°C. Zero checking and adjustment is an essential part of operation. In the absolute mode, a pressure at least one order smaller than the lowest pressure that a particular device can measure must be supplied at the measuring side for this purpose.

The general considerations presented above concerning flexure and capacitance mean that by a suitable choice of diaphragm stiffness various ranges of vacuum are accessible. Instruments are available for maximum pressures of 1000, 100, 10, 1, and 0.1 mbar with any instrument spanning four decades of measurement so that at the upper end instruments span the range 1000 to 0.1 mbar while the most sensitive operate from 0.1 down to 10^{-5} mbar. Typical accuracies are better than ±1% in the upper three decades of operation, with reduced accuracy at lower pressures. For the more stable instruments with sensor heads that operate above room temperature, corrections for thermal transpiration are necessary for the most accurate work. These corrections are discussed by Nash and Thompson (1983). Otherwise these devices measure total pressure independently of the type of gas. Such is the stability and accuracy of the more expensive models that they are used as transfer standards (see Section 7.6)

7.3 Gauges Depending on Thermal Conduction

For gases in a continuum state, as was discussed in Sections 3.9 and 3.11, thermal conductivity does not depend on pressure. But in a given geometrical configuration, as gas pressure falls so that conditions become transitional and eventually molecular, thermal conductivity progressively diminishes, thus providing a property that can be exploited for pressure measurement. The range of vacuum spanned by instruments based on this phenomenon is from about 100 to a little less than 10^{-3} mbar.

Consider a hot surface, that of a wire, say, to anticipate the structure of a device, at a temperature T_W. The molecular mechanism by which it loses heat to its cooler surroundings at temperature T_S is that incident molecules arrive at the hot surface with mean velocity associated with the temperature T_S, accommodate to it thermally to some degree, gain energy, and then depart from it traveling faster, on average, than when they arrived. Energy is thus removed from the hot surface to its surroundings. By analogy with Equation 4.2, an energy accommodation coefficient α may be defined to express the extent of thermal accommodation:

$$\alpha = \frac{T_d - T_s}{T_w - T_s} \tag{7.8}$$

where T_d is the temperature equivalent to the average velocity of the departing molecules. If accommodation has been complete, $T_d = T_w$, and the cooling is maximum. The rate of heat loss due to conduction is the difference between the rates at which molecules bring energy to the surface and take it away. The impact rate of molecules on a surface was shown in

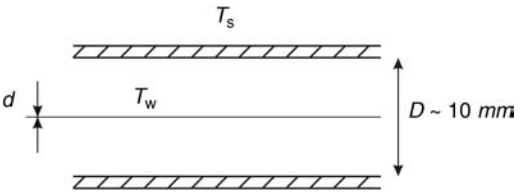

FIGURE 7.9
Geometry for thermal conductivity gauge.

Section 3.4 to be $J = n\bar{v}/4$ (Equation 3.13), also expressible as $p / \sqrt{2\pi mkT}$ (Equation 3.22). Molecules have a range of speeds determined by the Maxwell–Boltzmann distribution, and as may be shown by carrying out the integration suggested in Question 4.13, at temperature T the average energy incident on the surface per molecule is $2kT$. (It is not correct to argue that each arriving molecule has the average energy $(3/2)kT$ and then multiply this by J, because faster molecules make a larger contribution to the counting process that determines J.)

Figure 7.9 shows the usual cylindrical geometry of this type of gauge. A fine tungsten wire filament of diameter d ~0.01 mm is heated by the passage of electric current to be at temperature T_w. Surrounding it and located coaxially is a cylindrical surface of diameter D ~10 mm at room temperature T_s. The wire surface, though typically 100 K hotter than its surroundings, presents a very small area to the gas relative to that of the surrounding cylinder. Thus, when gaseous conditions are such that the mean free path becomes rather greater than d but still less than D, molecules are in a transitional state and collide dominantly with themselves and with the cylinder wall so that their temperature may be taken to be T_s. This becomes accurately true as conditions become molecular with mean free path comparable with or greater than D. Thus the impact rate on all surfaces, and that on the hot wire, will be J as given earlier by Equation 3.22, with $T = T_s$. The energy arriving per unit area in one second is therefore $J \times 2kT_s$. Similarly, the energy taken away will be the larger quantity $J \times 2kT_d$. Therefore, the rate of loss of energy due to this conduction mechanism is, using Equation 7.8,

$$\dot{Q} = J \times 2k\left(T_d - T_s\right) = J \times 2k\alpha\left(T_w - T_s\right)$$

Substituting $J = p/\sqrt{2\pi mkT_S}$ and rearranging gives

$$\dot{Q} = \sqrt{\frac{2k}{\pi mT_s}}\alpha\left(T_w - T_s\right)p$$

which can also be written as

$$\dot{Q} = \sqrt{\frac{2R_0}{\pi M T_s}}\alpha\left(T_w - T_s\right)p \tag{7.9}$$

This analysis assumes a monatomic gas such as helium whose thermal energy is purely translational. A more general treatment accounts for energy in other modes for gases that consist of diatomic or more complex molecules. The result is (see Kennard, 1938) that the formula of Equation 7.9 is multiplied by a factor $(\gamma + 1)/4(\gamma - 1)$ where γ is the ratio of principal specific heats ($\gamma = 5/3$ and 7/5, respectively, for monatomic and diatomic gases near room temperature). However, the essential information is in Equation 7.9. The heat loss, in addition to being directly proportional to the temperature difference and to the pressure p (fundamentally to the number density), also depends on gas type through M and on its accommodation coefficient α with the heated surface. Typically, α values for commonly encountered gases on tungsten, taken from Wutz et al. (1989), are ~0.8, though for helium it is about 0.4. The inverse dependence on $\sqrt{M}$ means that conduction is greater for the lighter species, a consequence of their greater number of visits to and energy transactions with the wire per unit time. Thus, at a given pressure, the heat loss depends on the identity of the gas.

Equation 7.9 applies at lower pressures such that λ is significantly larger than the wire diameter d. As may be deduced from Table 3.1, λ is comparable with $d = 0.01$ mm at a pressure of about 1 mbar. At this and higher pressures, as λ becomes less than d, molecular collisions in the gas close to the hot wire raise the gas temperature and cause the heat loss from the wire per molecule to be reduced. With further increase of pressure, continuum conditions are approached in which the thermal conductivity of the gas ceases to depend on pressure, although at higher pressures the heat loss continues to increase slightly as bulk convective heat flow become possible when gas adjacent to the wire is heated by conduction, expands, becomes bouyant, and transports energy away upwards.

When these phenomena are exploited in a pressure gauge, other sources of heat loss, though relatively small, have to be considered. Thus, there will be heat lost by radiation and some by conduction from the ends of the hot filament through its supports. For a constant wire temperature, these losses do not depend on pressure. Figure 7.10 shows schematically how heat loss varies with pressure.

An equation that governs the working of any gauge based on these principles is therefore

$$\sqrt{\frac{2R_0}{\pi M T_s}}a\left(T_w - T_s\right) \times p + \dot{Q}_{rad} + \dot{Q}_{end} = e \times i \tag{7.10}$$

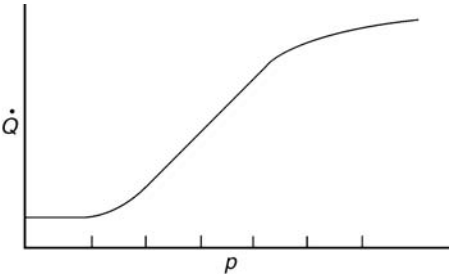

FIGURE 7.10
Variation of heat loss with pressure.

where e and i are the voltage across and current through the filament that supply the power to maintain the temperature and balance the various heat losses.

7.3.1 The Pirani Gauge

This gauge was developed by Pirani (1906) in the context of understanding the physics involved in the manufacture of light bulbs. The basic arrangement is that the filament, electrical resistance R_F, is incorporated into a Wheatstone bridge circuit as shown in Figure 7.11. The DC power source of voltage V sends current through branches $(R_F + R_1)$ and $(R_C + R_2)$. The current through the former maintains the filament with resistance R_F at a designated temperature of about 100–150° above room temperature. The branch $(R_C + R_2)$ contains either a "dummy" filament R_C isolated and encapsulated at high vacuum alongside the sensing resistor, or a compensating resistor, to allow for changes in ambient temperature that would cause bridge unbalance. A detector D senses out-of-balance current. The simplest way in which the bridge can be used, and the way it was first used historically, is with a constant driving voltage V. The bridge is balanced at high vacuum where the heat loss due to gaseous conduction is negligible in comparison with the small heat loss due to thermal radiation and end losses. These determine a threshold level typically a little less than 10^{-3} mbar, below which the gauge is not sensitive. At higher pressures there is appreciable conduction through the gas and the filament is cooled according to Equation 7.9. The cooling causes a lowering in filament temperature and therefore resistance, unbalancing the bridge whose output signal gives a measure of pressure above the threshold of sensitivity. Filaments are made of pure refractory materials such as tungsten, whose electrical resistance depends strongly on temperature. In this mode of operation the upper limit of sensitivity is about 1 mbar.

In modern forms of the instrument, it is arranged so that the power source drives the filament either with *constant current* so that changes in resistance are measured as the pressure is varied or in the *constant temperature* mode in which case the power to the filament is measured at constant resistance.

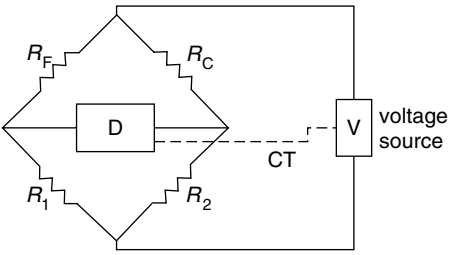

FIGURE 7.11
Circuit for Pirani gauge. (From Chambers, A., Fitch, R.K., and Halliday, B.S., *Basic Vacuum Technology*, 2nd ed., Institute of Physics Publishing, Bristol, 1998. With permission.)

In this mode, the output signal is electronically processed and used to control the driving voltage V, as suggested by the dotted link in the figure, so that as the pressure changes, the bridge stays balanced with R_F unchanged. Thus, if pressure increases, there is increased energy loss to the gas and a larger current I is supplied so that the power I^2R_F maintains the wire temperature and therefore R_F at a constant value. The pressure value displayed is derived electronically from the power provided. On instruments with analogue display, the range from 10^{-3} to 10 mbar is displayed on a generous 230° scale that becomes "squashed" at both ends due to the nonlinear overall response, greatest sensitivity being between 10^{-2} and 1 mbar, and roughly linear reflecting loss according to Equation 7.9 in the lower region of sensitivity. Because the temperature is constant, the instrument's response to pressure changes is rapid and can be used to initiate switching operations at predetermined pressure values.

With sophisticated electronic control, specified gauge head orientation to exploit convection losses, and digital output display, the range can be extended up to 1000 mbar, though the resolution is coarse in this region with 100 mbar increments, typically. Leck (1989) and Wutz et al. (1989) describe the theory of the instrument and its operational modes and circuitry in detail.

The Pirani gauge measures the total pressure for gases and vapors and, as noted earlier, its response can vary considerably from gas to gas. Thus, care should be taken when interpreting the gauge readings, especially if argon pressures are being monitored, when an indicated pressure of order 10 mbar may in fact be a pressure of order atmospheric and present an overpressure hazard. In use, where they frequently monitor gas mixtures, they are of limited accuracy. Their function in practice is usually to indicate if pressure is in the lower, middle, or upper part of a decade. The complexities of the heat transfer processes at surfaces and their dependence on the state of the filament are such that, even for a single known gas, pressures cannot be reliably predicted by analysis and so calibrations are necessary. The gauges are usually calibrated for nitrogen or dry air, and manufacturers supply response data for other common gases, as shown, for example, in Figure 7.12.

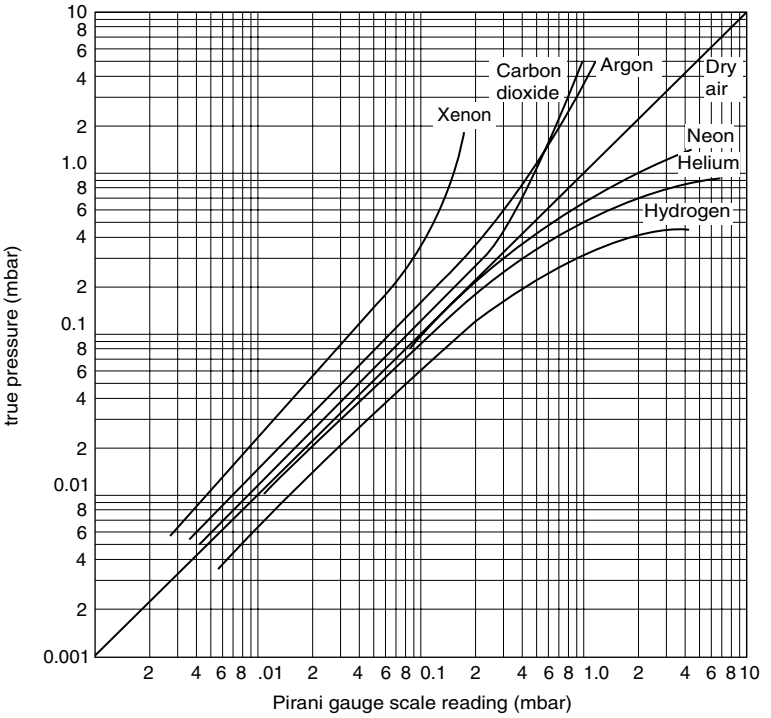

FIGURE 7.12
Calibration curve for a Pirani gauge. (Used with permission of BOC Edwards High Vacuum.)

7.3.2 The Thermocouple Gauge

This gauge operates on the same filament-cooling principles as the Pirani gauge, but in this case the changes in temperature of the filament due to pressure changes are measured directly by a very fine thermocouple attached to its center, as shown in Figure 7.13. The thermocouple voltage is processed to give a pressure reading. Its working range is typically from 5 down to 1×10^{-3} mbar.

Although they are of limited accuracy, both Pirani and thermocouple gauges are robust, reliable, and relatively inexpensive. They play an important role in a majority of vacuum systems to monitor, for example, the

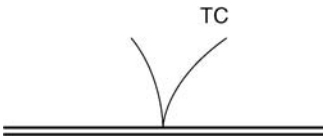

FIGURE 7.13
Thermocouple gauge.

performance of the rotary or sorption pump in primary pumping from atmospheric pressure, and in continuous monitoring of backing line pressure for secondary high-vacuum pumps.

7.4 The Spinning Rotor Gauge (SRG)

First suggested by Beams et al. (1962) and subsequently intensively developed by Fremerey and coworkers (1982), this gauge is now a highly accurate and sophisticated, though expensive, instrument. It measures the slowing down by molecular drag of the spinning motion of a magnetically levitated steel sphere, and operates over a range from about 0.1 down to 10^{-7} mbar. The subject of molecular drag was introduced in Section 4.6

A schematic diagram of the SRG is shown in Figure 7.14. The rotor is a steel ball 4.5 mm in diameter, magnetically levitated inside a horizontal stainless steel tube, using permanent magnets and current-generated magnetic fields. In addition, there are drive and pick-up coils. The tube is closed at one end and the open end is attached to a flange and samples the vacuum. In use, the driving magnetic field is rotated and the ball accelerated to about 400–415 Hz. The driving field is then switched off and the reduction in the rotation rate of the ball, caused by the decelerating effects of gas-induced friction at its surface, is measured over a period of time by pickup coils that sense the ball's rotating magnetic moment. Digital electronics in the control unit computes the slowing rate and hence the pressure, which is updated every few seconds using an average measuring period of 10 s.

The deceleration may be analyzed as follows. Consider, as in Figure 7.15, a sphere of mass M_S and radius R levitated and spinning with angular velocity ω about the axis shown. Its moment of inertia for this motion is $I = (2/5)M_S R^2$. It is located in gas at pressure p whose molecules have number density n and bombard it at an impingement rate J.

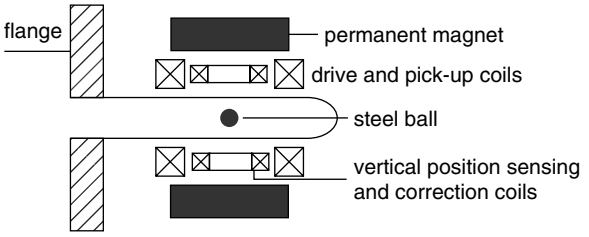

FIGURE 7.14
Schematic diagram of the spinning rotor gauge. (From Chambers, A., Fitch, R.K., and Halliday, B.S., *Basic Vacuum Technology*, 2nd ed., Institute of Physics Publishing, Bristol, 1998. With permission.)

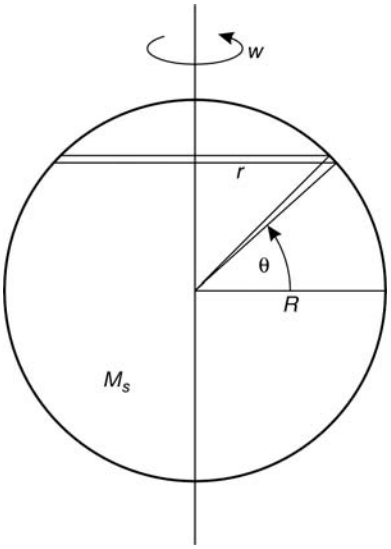

FIGURE 7.15
Geometry for analyzing deceleration of sphere.

The *average* direction of arrival of molecules from the gas onto the surface of the sphere at any part of its surface is perpendicular to it. Let us assume that the molecules stay on the surface, if only briefly. In acquiring the circulating motion of the surface at any point, the molecules exert a force on the surface which, by Newton's third law, is equal and opposite to the one that gives them this new motion and causes a slowing torque about the axis of rotation. In desorbing at a later time, the average direction of departure from the surface is perpendicular to it in the rotating reference frame of the sphere, so there is no net average torque associated with departure; all the influence on the motion of the sphere can be regarded for the purpose of this analysis as taking place at the time of molecular arrivals.

Consider molecular impacts on an elemental ring at angle θ with radius $r = R \sin\theta$. It presents an area $2\pi rRd\theta$ to the gas. Molecules of mass m arriving at this area are given a sideways velocity $r\omega = R\sin\theta\omega$ in a direction transverse to the average perpendicular direction of arrival. Therefore, the momentum change per second in the direction of motion of the surface, and by Newton's second law, the force on this element of area is

$$J \times 2\pi rRd\theta \times m\ r\omega$$

Multiplying by a further r to give the torque about the axis and substituting for r leads to

$$\text{torque} = 2\pi mJ\omega R^4 \sin^3\theta\ d\theta$$

and integrating this from $\theta = 0$ to π gives the total torque as $(8\pi/3)\omega JR^4m$.

The angular momentum of the spinning sphere is $I\omega$, and again applying Newton's second law in the form "torque = rate of change of angular momentum," gives

$$\frac{8\pi}{3}\omega JR^4m = \frac{2}{5}M_SR^2\,\dot{\omega}$$

where the fractional slowing rate, defined with a negative sign so that the quantity itself is positive, is

$$\left(-\frac{\dot{\omega}}{\omega}\right) = \frac{20}{3}\frac{\pi JmR^2}{M_S}$$

Substituting for J from Equation (3.23), expressing the mass M_S of the sphere in terms of its density ρ and diameter $d = 2R$, and simplifying, gives the fractional deceleration rate as

$$\left(-\frac{\dot{\omega}}{\omega}\right) = \frac{10}{\rho d}\sqrt{\frac{M}{2\pi R_0T}}p$$

Evidently, the slowing effect due to drag depends on the molar mass M of the gas involved, as well as on the pressure (strictly on the number density). Rearranging,

$$p = \frac{\rho d}{10}\sqrt{\frac{2\pi R_0T}{M}}\left(-\frac{\dot{\omega}}{\omega}\right) \tag{7.11}$$

Two refinements of this equation are needed in order to determine pressure from the measured slowing rate. The first is related to the efficiency of the momentum transfer when molecules impinge on the surface of the ball, and the second is related to the existence of other sources of deceleration.

If the momentum exchange between molecules and the surface is incomplete, for example, if some incoming molecules are not captured by the surface and so retain part or all of their momentum in a direction parallel to the surface, then the slowing effect is less. This is formally accounted for by inserting a coefficient σ the tangential momentum accommodation coefficient, which is the fraction of the incident parallel momentum absorbed, into the denominator of the above equation, thus correcting pressure values deduced, which would otherwise be too low. To determine the σ value for a particular ball, the device is used to measure a known pressure generated by independent procedures in special equipment, such as is available at national standards laboratories. It may vary a little from gas to gas. Surprisingly, values

of σ may be slightly greater than unity, by a few percent. This is a consequence of the roughness of the ball surface on a microscopic scale, which means that molecular velocity components perpendicular to the average plane of the surface play a role in exerting force parallel to it, just as a ball dropped onto a slope exerts a force on it that has a horizontal component. Measured σ values are generally close to unity, and for all but the most accurate work, as Comsa et al. (1980) have demonstrated, may be assumed to be unity.

The second refinement to Equation 7.11 arises because there is a small braking effect of electromagnetic origin due to induced eddy-currents. It contributes what is called the "residual drag" term to the measured slowing rate. By its nature it is independent of both the pressure and identity of the gas being measured, but it adds a constant extra contribution to the deceleration rate and therefore is equivalent in its effect to an added constant pressure *P*, which has to be subtracted from the value deduced from the raw deceleration data to get the true pressure. The value of *P* is determined by measuring the deceleration in a vacuum ~10^{-8} mbar or less in which the gas drag contribution is negligible. Typically, *P* has a value of a few ×10^{-6} mbar.

The result of the refinements discussed above is that Equation 7.11 is modified to

$$p = \frac{\rho d}{10\sigma}\sqrt{\frac{2\pi R_0 T}{M}}\left(-\frac{\dot{\omega}}{\omega}\right) - P \tag{7.12}$$

The values of σ and *P* are entered into the program of the controller, together with other fixed quantities and the appropriate *M*, which may be that of a pure gas or vapor, or a composite value for a mixture.

The normal linear operating range of the SRG is from about 10^{-1} to 10^{-7} mbar with an accuracy and reproducibility in the readings of order 1%. Because of its potential as an accurate and stable gauge, its operation has been investigated in great detail. Peacock (1998) discusses operational concerns and good practice. Frequent checking of the residual drag term is advised as this can vary slightly. For example, because the levitated sphere is quite thermally isolated, it will warm up a little with use due to the dissipation of eddy-currents, and in doing so, expand slightly. This implies a small increase in its moment of inertia with time, and so a small change in the dynamical properties. The effect can be quantified.

The use of the gauge at nonmolecular pressures above 0.1 mbar is possible, but response becomes less sensitive and nonlinear in the transitional and viscous ranges. This upper range of the instrument is discussed by Lindenau and Fremerey (1991). The lower level of sensitivity, in normal use about 10^{-7} mbar, is principally determined by the stability of the residual drag term *P*. As Equation 7.12 demonstrates, when the effect of gas drag becomes smaller than that associated with the *P* term, the problem is that of determining the small difference between two comparable numbers.

The SRG samples the vacuum passively without producing thermal, chemical, or pumping effects and is bakeable to 400°C. Because of its potential accuracy and stability, it is used as a transfer standard. In a system containing a calibrated SRG, because of its range and reliability, it can be used to calibrate other gauges that overlap its range, such as ionization gauges and capacitance manometers at its lower and upper limits, respectively.

7.5 Ionization Gauges

The routine measurement of high and ultrahigh vacuum is possible only using gauges of this type. Gauges whose action is based on mechanical, viscous, or thermal effects, with the exception of the expensive spinning rotor gauge, have inadequate sensitivity at pressures below about 10^{-3} mbar. Ionization gauges are therefore important and essential components of vacuum systems that operate in these regions. They are of two basic types. In the hot cathode ionization gauge, a current of electrons released thermionically from a filament is accelerated to ionize some of the sampled gas molecules, creating positive ions. The rate of ionization is proportional to the number density of molecules and hence the pressure. The current of ions thus generated is collected and affords a measure of pressure. In the cold cathode type, the current that sustains a magnetically confined discharge is similarly created in proportion to the molecular number density.

Because the probability of ionization, and hence the associated current, depends on the identity of a gas, it is convenient to have a gas of reference in terms of whose ionization the pressure is expressed. This is usually nitrogen. In addition, the complexity of the processes involved, even with a single species of gas, means that an accurate calculation from first principles of number density from ionization current is not possible, and therefore it is essential to calibrate such gauges against a standard. This may be done by procedures carried out at the time of manufacture and can be later checked with a spinning rotor gauge if one is available.

7.5.1 The Hot Cathode Gauge

Fundamental to the operation of this gauge is how the probability of ionization depends on the energy of the electron with which a gas molecule interacts. Figure 7.16, based on the work of Tate and Smith (1932), illustrates how the ionization depends on electron energy for a number of common gases. The probability is expressed per centimeter length of electron path in the gas at a pressure of 1 mbar and temperature 273 K. There is no ionization below threshhold ionization potentials of typically about 15 eV, but then ionization increases rapidly to reach a maximum at about 100 eV, thereafter

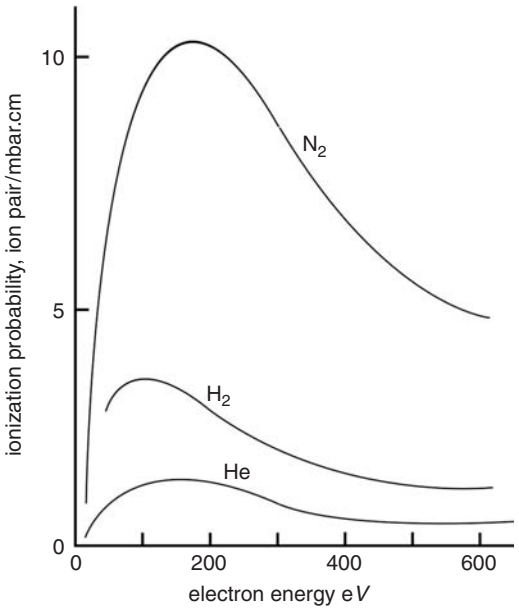

FIGURE 7.16
Ionization cross sections.

decreasing gradually. This is because, at higher energies, the ionizing impulse delivered to the molecule by the electric field of the electron is reduced as its interaction time with the target molecule diminishes. Note that the shape of the curve is similar for most gases.

Although the curves relate to gases at a particular pressure and temperature and therefore with a particular number density, they may also be taken, if divided by the product of that number density and the unit of length, to represent the ionization cross sections σ as a function of electron energy. Consider a small elemental volume of length l and constant cross section occupied by molecules with number density n through which an electron current i_e, distributed uniformly over the cross section, passes in the direction of its length. The number of ions generated per second i_+ is given by

$$i_+ = nl\sigma i_e \tag{7.13}$$

Remembering that $n = p/kT$, one may write

$$i_+ = Kpi_e \tag{7.14}$$

where $K = l\sigma/kT$. The quantity K, being the ion current per mbar per unit of electron current, is a measure of sensitivity. In the geometry of this simple, short, one-dimensional model element, it is proportional to electron path length and ionization cross section. In practical device geometries, the aim

is to increase the electron path length in order to enhance the ionization probability prior to capture at the electron-collecting electrode. The configuration of the electrodes, potential distributions, and consequent electron trajectories determine effective average values of electron path length. The ionization probability is optimized, given the range of gases likely to be encountered, by choosing the potential difference that accelerates the electrons to be about 150 V. For a given gas, a gauge therefore has a K value as defined by Equation 7.14, which depends on its electrode configuration and operating potentials. Rearranging 7.14 gives the equation that governs the gauge:

$$p = \frac{1}{K}\frac{i_+}{i_e} \tag{7.15}$$

K is usually called the gauge constant. How its value depends on the identity of the gas will be described later. The gauge constant may vary considerably even between gauges of nominally identical design because of manufacturing tolerances.

Early ionization gauges were three-electrode devices based on the thermionic triode valve. This was the active device of electronic technology before the invention of the semiconductor triode that we know as the transistor. This early pattern of gauge is usually referred to as the normal triode. Such a gauge, with cylindrical symmetry, is represented in Figure 7.17. Although now largely superseded by a simple variant to be described, it is the best with which to explain how ionization gauges work.

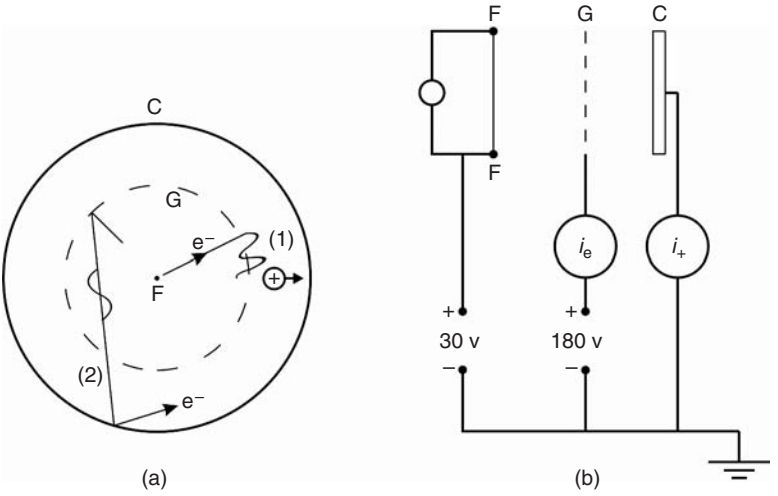

FIGURE 7.17
Ionization gauge, conventional triode pattern (a) geometry, and (b) potentials.

Part (a) shows, in cross section, an axial filament F, a concentric open-grid structure G surrounding it, and an electrode C, the ion collector, which encloses them both. Part (b) shows schematically the potentials on the electrodes, the source which heats the filament FF to release electrons by thermionic emission at potential 30 V, and current-measuring arrangements. The open grid sets the potential at this radial distance to be 180 V, creating an inwardly directed radial electric field between G and F that accelerates electrons from the filament through a potential difference of 150 V. Between grid G and the grounded collector C, the outwardly directed electric field will accelerate any positive ions generated towards C but will have a retarding effect on those electrons, the majority, that pass through the open grid structure. Even those that lose no energy will be able to penetrate only to positions where the potential is 30 V, very roughly 1/6 of the G to C distance away from C, before their radial velocity is reversed. The electrons therefore oscillate back and forth in the region of the grid before eventually reaching it and being registered in the filament-grid circuit as the filament emission current. This is represented by the trajectory labeled (1) in Figure 7.17(a), which, because of the oscillations, is much longer than the radial dimension in which the electrons are constrained. In the course of such lengthened trajectories, the electrons ionize neutral molecules of the background gas to produce, in each ionizing collision, a positive molecular ion and an additional electron. The ions are accelerated to the collector where they are registered as an ion current; the electrons eventually find their way to the grid. Operated with a constant emission current i_e, typically 1 mA, the ion current measured is proportional to the pressure, as previously discussed. Referring to Equation 7.14, a typical value of K for nitrogen is about 10 per mbar, so that an emission current of 1 mA at a pressure of 10^{-6} mbar creates a small ion current of 10 nA.

This normal triode gauge has a linear response proportional to pressure over a range from 10^{-3} down to about 10^{-7} mbar, as shown in Figure 7.18(a), in which the pressure indicated by the measured ion current according to Equation 7.15 is plotted vs. the true pressure, a slope of unity implying that calibration has been carried out. The upper limit is determined by the onset of space charge effects and multiple electron-molecule collisions that cause the efficiency of ionization to be reduced with the nonlinearity indicated. In addition, operation at higher pressures tends to quickly shorten the life of filaments, frequently made of tungsten.

The lower limit is caused by a small pressure-independent residual current that, below a true pressure of about 10^{-7} mbar, masks the true ion current, as indicated by the deviation labeled (NT) from the line of linear response. This residual current i_r is due to causes other than ionization within the gas. The effects contributing to it, whose magnitudes are negligible at higher pressures, are described below. The measured collector current i_+ and its relationship to the pressure is therefore more accurately described by modifying Equation 7.14 to be

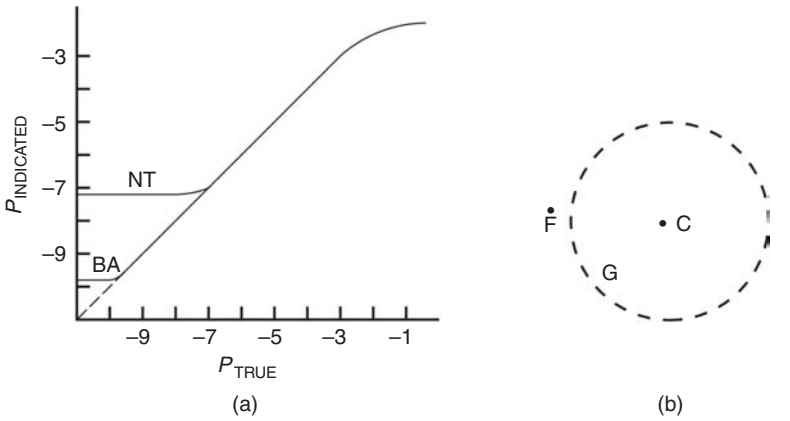

FIGURE 7.18
(a) Ionization gauge response and (b) Bayard–Alpert (BAG) configuration.

$$i_+ = Kpi_e + i_r$$

The dominant contribution to i_r is generated as follows. Some of the energy of the electrons accelerated from F that strike the grid G is dissipated in the creation of soft x-rays (energy 180 eV and less). These x-rays are intercepted by the relatively large enclosing collector surface where they may release electrons by the photoelectric effect. An electron leaving the collector and a positive ion arriving cannot be distinguished in the measuring circuit, and the electron current will add to the measured collector current. This process is represented schematically by sequence (2) in Figure 7.17(a). In a given gauge, it is independent of pressure and gas type, though it does, of course, depend on the electron emission current. The phenomenon is described as the "x-ray effect" and the pressure value at which the ion current and residual current are equal is called the "x-ray limit." Until it was recognized and dealt with in the 1950s, this effect had delayed progress in the measurement of vacua that were in many cases certainly better than those indicated by the normal triode working at its lower limit.

The x-ray limit was considerably reduced in magnitude, and the useful range of the gauge extended at the lower end in an inverted triode construction proposed by Bayard and Alpert (1950) and now known as the Bayard–Alpert gauge (BAG). It has the configuration shown in Figure 7.18(b). The open grid structure of the normal triode is retained, but the electron-emitting filament is outside it and the ion collector, in the form of a thin wire, is located on the central axis. The potentials are similar to those in the normal triode, but because of the much smaller surface area of the collector, the flux of soft x-rays intercepted by it and the associated photoelectron current is greatly reduced by a factor of 100 or more. This simple strategy dramatically lowers the x-ray limit and extends the range of the BAG down to about 10^{-10} mbar, as indicated by the curve (BA) in Figure

7.18(a). Its useful range is, therefore, from 10^{-3} mbar, down through the high vacuum and deeply into the ultrahigh vacuum region. The BAG has become the hot cathode gauge universally adopted in ultrahigh vacuum practice. Weston (1985) discusses the electron trajectories in this gauge, which are very different from those in the normal triode. Another feature of this geometry is that, because of the logarithmic way in which potential varies between the grid and the central collector wire, ionization is promoted in a large fraction of the volume inside the grid.

The electrode potentials and electron emission currents in ionization gauges are regulated electronically by an associated control unit, usually with digital display of the pressure derived from the ion current, although the older analogue patterns that display current directly on a meter have some advantages when investigating minute pressure changes. Emission currents may be set at 100 μA, 1 mA, or 10 mA by increase of the filament heating current. The highest value is appropriate at the lowest pressures in order to maximize the ion current.

As already noted, gauges are calibrated for nitrogen, but sensitivities depend on the type of gas because of differing ionization cross sections and also because of differences in other characteristics such as the efficiency of ion collection. It is possible to make allowance for measurement in gases other than nitrogen for which a gauge is calibrated by introducing a relative sensitivity R defined as $R = K_{\text{gas}}/K_{\text{N2}}$. The true pressure p is then related to the indicated nitrogen pressure p_{ind} by

$$p = \frac{1}{K_{gas}}\left(\frac{i_+}{i_e}\right) = \frac{1}{RK_{N2}}\left(\frac{i_+}{i_e}\right) = \frac{p_{ind}}{R} \tag{7.16}$$

Some representative values of R are given in Table 7.1, abstracted from data quoted in O'Hanlon (2003) and Leck (1989). Thus, when used in pure helium, the true pressure might be about six times the indicated pressure, reflecting that helium is more difficult to ionize than nitrogen, as evident also in Figure 7.12.

It should be stressed that these R values will give only an approximate correction to an indicated pressure because of the variability of gauges and because the measured K values on which they are based are average values evaluated from data on various gauge designs. Leck (1989) and Tilford (1992) discuss these matters. If more reliable operation in any gas is necessary,

TABLE 7.1

Values of the Relative Sensitivity R for Some Gases and Water Vapor

Hydrogen	Helium	H_2O	Nitrogen	CO	Oxygen	Methane
0.4	0.17	0.9	1	0.9	0.86	1.6

individual gauge calibration is advised. Measurement of the ion current at two distinct pressures well above the lower gauge limit, separated by at least an order of magnitude, and verified with a spinning rotor gauge, would enable K to be estimated. For many purposes, however, with a typical residual gas mixture of components whose sensitivity is close to that of nitrogen, corrections that would in any case require a knowledge of the gas composition would not be justified. Also, in many cases, the purpose of the gauge is to give an indication of the general state of the vacuum system, and it is not necessary to know the pressure value precisely. Rather, with the incorporation of a residual gas analyzer (Section 7.7) into the system, it is of greater value to know the general level of the vacuum and the relative proportions of its constituent gases.

The accuracy of ionization gauges is a complex matter that has been extensively investigated and which has many aspects. Ideally, one would like an instrument, gauge plus controller, that responds linearly with nitrogen pressure and, having been once accurately calibrated against a standard, retains that performance. Linearity is reliably achieved, but K values change with time and use as structures relax slightly and electrode surface conditions alter. In addition, it is not usually feasible in practice for manufacturers to calibrate ionization gauges individually, so at best, calibration of samples from a batch may be carried out and uniformity of construction within small tolerances relied upon to give a gauge of reasonable accuracy. It is generally thought that hot cathode gauge readings (in nitrogen) are likely to be within ± 50% of the true value, so that the linear response line of Figure 7.18(a) may differ in slope appreciably from that drawn. The many factors involved have been thoroughly investigated by a number of workers, and the publications from NIST by Tilford and his coworkers are notable, with the discussion given by Peacock (1998) forming a good starting point.

The problem of distinguishing between true ion current and the residual current i_r is not insurmountable and was addressed by the modulated BAG technique proposed by Redhead (1960). In a simple application of the technique, the modulator is an additional electrode in the form of a straight wire located inside and close to the grid parallel to the gauge axis, and alternately held at either grid potential or ground potential. In the former case, the ion current is scarcely affected, and the collector current i_1 may be taken to be

$$i_1 = i_+ + i_r$$

But in the latter case with modulator potential at ground, the electric field within the grid structure is considerably modified and the current at the collector changed to be

$$i_2 = (1 - k)i_+ + i_r$$

where it is assumed, most simply, that i_r is unaffected by the different modulator potentials and that k is a constant. From these equations, $k = (i_1 - i_2)/i_+$, and k may be determined by operation at higher pressures where i_+ is little different from i_1 so that $k = (i_1 - i_2)/i_1$. Values are typically about 0.6. In operating at the lower pressures using the modulation technique, knowledge of k then enables the ionic component of i_+ to be determined and a corrected value of pressure deduced. By this procedure, and refinements of it, the lower limit of the BAG may be reduced by an order of magnitude or more to pressures of 10^{-11} mbar and less. As Singleton (2001) notes, it is unfortunate that this modulation feature is not more widely available in commercial gauges because although originally proposed to deal with the x-ray problem, the availability of a modulator electrode and the freedom to vary its potential is, in addition, an extra tool in the evaluation of other contributions to the residual current, which become particularly important at the lowest pressures. One of these, whose effects may be evident at higher pressures too if grids have not been adequately outgassed (see the following text), is that arising from electron-stimulated desorption.

Electron-stimulated desorption, or ESD as it is known, is the release of molecules adsorbed on surfaces due to electron impact. The disruptive interaction of the incident electron is such that some of the desorbing molecules are ionized. Thus, in the ionization gauge, the impact of energetic electrons on the grid may cause the release of ions by ESD into the grid-collector space. Following acceleration and collection, these ions constitute a current that is not a true gas-phase current in its origin, and which, like the x-ray contribution to the collector current, cannot be distinguished from the true current in the basic BA gauge. This will be further discussed later.

One of the practical necessities for the operation of BAG gauges at pressures below 10^{-7} mbar is to outgas them. Gases are adsorbed on the electrodes as a result of venting of the vacuum system to gases, frequently air, at atmospheric pressure. When pumping is started again, gases that have been adsorbed on the filament are rapidly thermally desorbed when, having reached 10^{-5} mbar, the gauge is switched on. It is wise at this stage to run both of the filaments that are usually provided on a gauge head in case the spare needs to be activated later. Outgassing of the grid and collector requires a special procedure and is accomplished by electron bombardment heating, with the controller supplying high voltages to them so as to raise their temperatures to much higher values than are characteristic of normal operation. Adsorbed gases are thus driven off, though they will be taken up again to some degree at the normal running temperature, depending to an extent on the much-reduced background pressure and composition. To get gauges ready for operation in ultrahigh vacuum usually requires outgassing procedures to be carried out for times of order an hour. Singleton (2001) discusses these and other practical procedures in detail.

Gauges of the "nude" type are attached to a flange so as to protrude more or less directly into the vacuum space. Gauges housed in a tubular envelope, usually made of glass, are necessarily connected to the main chamber by a

pipe. The connection of such gauges to the system should be via a pipe of high conductance because gas flow in either direction will cause a pressure difference between the indicated pressure at the gauge and the main volume, so that the indicated pressure is not quite that of the system. The difference may be of either sign, depending on whether there is a net outflow or inflow of gas. The former would occur in the operation of a gauge that had not been adequately outgassed, and the latter due to pumping. Gauges possess a small pumping speed, of the order 0.1 l s^{-1} due to two distinct effects. Ionic pumping occurs when ionized gas species are trapped at active sites on the collector surface rather than being returned to the gas phase as neutral molecules. It is therefore greatest on a freshly outgassed surface and will diminish as the adsorbed coverage increases to reach an equilibrium value at the prevailing pressure. It also occurs by trapping and burial of energetic ions onto electrode surfaces. The other pumping effect is chemical due to reactions with impinging gas molecules that take place on the surface of the heated filament. Tungsten filaments operate at temperatures as high as 2100 K that are sufficient to cause dissociation of water, hydrocarbons, hydrogen, and other molecular species. Filaments with a lower work-function such as thoria-coated iridium operate at substantially lower temperatures and correspondingly have a smaller pumping action. They also have the advantage of being more resistant to burnout if accidentally exposed to higher pressures, though they have been found to be less stable, and so tungsten filaments are still preferred in many cases. These nonpassive attributes of hot cathode gauges are potentially most serious in small-vacuum systems, particularly if, as is sometimes the case, a tubulated gauge is attached to a volume not substantially bigger than itself. For a nude gauge in a large system with matching pumping speed, pumping effects will be negligible, but the role of the hot filament in affecting the residual gas composition may be a concern.

The drive to develop gauges that will reliably measure pressures of 10^{-11} mbar and less, what is now known as the XHV region, has a long history and is described in the article of Peacock (1998) to which reference has already been made. The classic text of Redhead et al. (1968), recently reissued, presents an account of developments to its date, many still relevant and significant. In more recent reviews (1987), (1993), and (1998), Redhead has discussed progress in the subject, which remains an active field of development.

Reducing the contributions to the collector current of (1) electrons released by x-ray impingement and (2) ions desorbed by electron impact are the two principal problems that face the designer of gauges to operate in the region below 10^{-11} mbar. To a lesser extent, outgassing from hot surfaces is also a concern. Although modulation techniques as previously described enable the magnitude of the x-ray effect in the modulated BAG to be estimated, allowing pressure estimates down to ~10^{-12} mbar, uncertainties increase, and other gauge geometries have been developed to try to eliminate them. These are known as ***extractor gauges*** because they are designed to extract the ions generated in the gas phase to a remote collecting electrode that is shielded

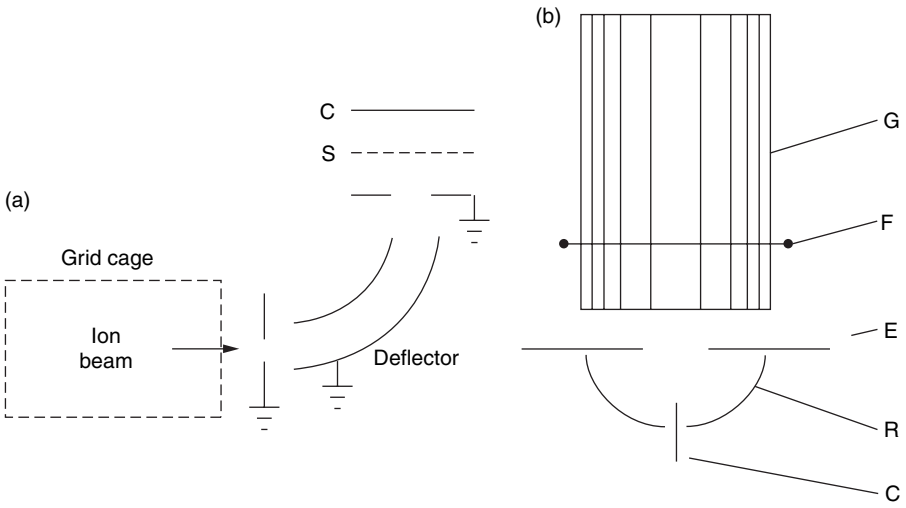

FIGURE 7.19
(a) Helmer pattern, (b) Leybold extractor gauge. (From Chambers, A., Fitch, R.K., and Halliday, B.S., *Basic Vacuum Technology*, 2nd ed., Institute of Physics Publishing, Bristol, 1998. With permission.)

from the x-rays produced at the grid. The basic Helmer and Hayward (1966) design that has come to be known for obvious reasons as the "bent-beam" gauge is shown in Figure 7.19(a). The ions are extracted from the ionizing volume inside the grid and are electrostatically deflected through 90° onto the ion collector (C). A suppressor grid (S) placed in front of the collector reduces any secondary or photoelectric emission from the collector. The electron and ion ballistics are rather complicated. Its low pressure limit is about 10^{-12} mbar. It has been further developed in a number of ways that are reviewed by Redhead (1993). These gauges also give some reduction of ESD effects because, in the process of ion extraction, the energy filtering properties associated with the beam-bending electric fields tend to favor the passage of the "proper" gas-phase ions from the center of the ionization volume.

A pattern of extractor gauge due to Beeck and Reich (1972) and manufactured by the Leybold Company has been available commercially for some time and is frequently quoted as a reliable gauge of reference by workers developing other devices in the low UHV region. Its principle is shown in Figure 7.19(b). A ring filament F at +100 V potential surrounds an open grid structure G at +220 V. Grounded shield E hides the small collector C from grid generated x-rays because of its small aperture and, at the same time, positive ions are attracted to it and through the aperture into the collection region. The small collector wire C is surrounded by a reflector R at +205 V potential that intercepts most ESD ions from the grid, whereas those formed within the grid are directed to C because they originate from a region of

somewhat different potential than the grid itself, due to the effects of the space charge. Measurements down to 10^{-12} mbar are possible.

The field of XHV total pressure measurements is in active state of development. For example, Watanabe (1993) reported a sophisticated design of ***ion spectroscopy gauge*** incorporating a hemispherical energy analyzer that enables separation of the ESD and gas-phase ions. The x-ray limit of this gauge is reported as 2.5×10^{-13} mbar and by an optimal choice of materials and heat conduction paths, temperature rises are restricted so that outgassing is reduced to a very low level.

In a type of hot cathode ionization gauge developed by Akimachi et al. (1997) and called the ***axial symmetric transmission gauge,*** the ion source and ion detector, which is a secondary electron multiplier, are separated by an energy analyzer of the Bessel-box type. The design prevents direct axial line of sight from the ion source region to the detector, as well as acting as an energy analyzer. Gas-phase ions are separated from ESD ions because the energy filter collects the former from a restricted volume of the grid cage where the potential is about 30 V different from the grid voltage. Calibration involves an especially developed conductance modulation method. The sensitivity is 0.23 per mbar for H_2 from 10^{-12} up to 10^{-6} mbar, and pressure measurements down to 3×10^{-14} mbar are possible.

7.5.2 Cold Cathode Gauges

In these gauges the ion current associated with a magnetically confined self-sustaining discharge is the basis of the pressure measurement. Of a number of types, the Penning gauge (10^{-2} to 10^{-7} mbar) and the inverted magnetron gauge (10^{-4} to 10^{-11} mbar) are the most widely used. The electrons that initiate the discharge come from chance events such as ionization induced by natural radioactivity or field emission at electrode surfaces. The absence of an electron-emitting filament has a number of advantages.

The Penning gauge is based on the Penning discharge described in some detail in Section 6.4 in connection with ion pumps, though the electrode arrangements and voltage used may differ slightly. One pattern of gauge is shown in Figure 7.20. An anode A in the form of a short tube and typically at a potential of 3 kV is located symmetrically between earthed circular cathode plates K within an earthed gauge enclosure. A strong magnetic field ~0.1 T supplied by an external magnet is directed parallel to the anode axis. The natural tendency of electrons to be accelerated directly towards the anode is prevented by the strong magnetic field, around whose field lines they spiral due to the $v \times B$ force, resulting in much longer trajectories back and forth between the cathodes and increased probability of ionization. The spiral radius $r = mv/eB$ is typically ~0.1 mm. A steady discharge is established, dominated by electrons, in which the presence of electronic space charge causes the potential along the axis to be near ground so that the electric field over a substantial volume is radial and therefore perpendicular to the

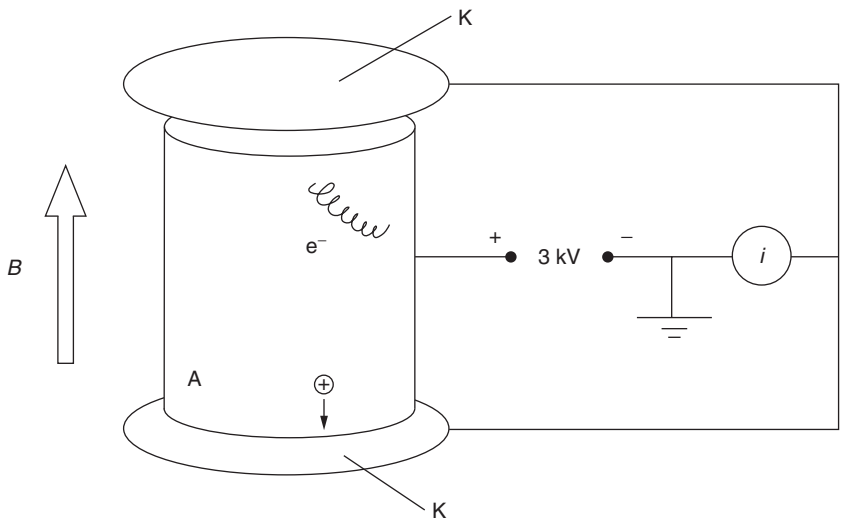

FIGURE 7.20
Penning gauge.

magnetic field. Positive ions produced by ionization of the background gas are relatively unaffected by the magnetic field, travel directly to the cathode, and are rapidly removed from the discharge; the resulting electrons drift slowly outward to the anode. The ion current affords a measure of pressure. The rate of ionization depends on the molecular density and the ion current and is roughly proportional to pressure in the range $< 10^{-6}$ to 10^{-3} mbar. At the upper end of the range, a glow discharge becomes dominant; at the lower limit, the discharge extinguishes because there is insufficient ionization to sustain it. A typical sensitivity is 10 A per mbar for nitrogen, so that the ion current is 1 mA at 10^{-4} mbar; such currents are easily measured directly and there is no need for amplification. The response of the gauge is gas dependent and the discharge somewhat unstable, so that good accuracy is not claimed for these gauges. They are sometimes referred to as "order of magnitude" gauges and thought of as pressure indicators. Nevertheless, they are robust and reliable and provide an adequate measure of the state of a vacuum in many laboratory and industrial applications.

The ***inverted magnetron*** cold cathode gauge, due to Hobson and Redhead (1958), operates down to much lower pressures in the UHV range. The construction of a typical device is shown schematically in Figure 7.21(a). A strong external magnet produces an axial magnetic field in a cylindrical electrode geometry in which the outer cylinder is a cathode and a central axial rod forms an anode A at a potential of several kV. An auxiliary electrode G at cathode potential attracts the field emission currents rather than cathode C that collects the ion current, thus avoiding a spurious addition to the true ion current. The electric field is radial and perpendicular to the magnetic field. As in the Penning gauge, the crossed fields lengthen electron paths

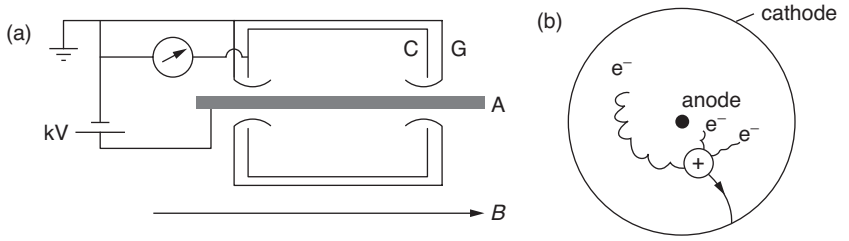

FIGURE 7.21
Inverted magnetron gauge: (a) construction, (b) discharge geometry. (From Chambers, A., Fitch, R.K., and Halliday, B.S., *Basic Vacuum Technology,* 2nd ed., Institute of Physics Publishing, Bristol, 1998. With permission.)

that take the form of cycloidal hops in a general circumferential direction as shown in Figure 7.21(b). A stable discharge builds up in which the current depends on pressure. The ions produced, being much heavier than electrons, follow simpler trajectories to the cathode. The current–pressure relationship is $i = kp^n$, with n in the range of 1.05 to 1.25, depending on the particular design; control units are designed to translate this current into a displayed pressure value. Typically k has a value of a few A mbar^{-1} at 10^{-6} mbar, taking n as unity at this pressure, and gauges may be used down to 10^{-10} mbar or less, when currents become very small and concerns about spurious leakage currents arise. The time for the discharge to become established can become rather long at lower pressures (typically 30 s at 10^{-9} mbar and 2 min at 2×10^{-10} mbar), and so if circumstances allow, the gauges are started at high pressure.

As with the Penning gauge, there are pumping effects of the order of 0.1 l s^{-1} or up to 1 l s^{-1} for active gases. They may be noticed, particularly in small volumes, by the changes of pressure that occur on a partner gauge, if one is available, when the IMG is switched off. Operation for prolonged periods at higher pressures ~ 10^{-5} mbar or higher is discouraged because of substantial sputtering of electrode materials onto insulators that may lead to spurious leakage currents and hence false readings. Also, insulating surface coatings that may be deposited onto conducting electrode surfaces due to breakdown of hydrocarbons can alter the electrical conditions of the discharge and hence the calibration. Because the power consumed in normal operation is so small, outgassing is minimal. Any x-ray effects are proportional to the current and scale with pressure and so do not present a problem. Because it has no hot filament to influence the gas composition by promoting chemical reaction, the IMG arguably intrudes less on the vacuum measurement than does the thermionic gauge.

The relative merits of hot and cold cathode gauges have been examined in detail by Peacock et al. (1991). More recently, Kendall and Drubetsky (1997) have evaluated the performance of a number of cold cathode gauges for UHV work and make the case that they offer comparable operating accuracies to those available using well-operated thermionic gauges. The cold

cathode gauges in various forms continue to be a very useful device in high and ultrahigh vacuum. They have a number of very special applications such as when, for example, the visible radiation from a hot filament cannot be tolerated.

7.6 The Calibration of Vacuum Gauges

The principles that underlie the calibration of vacuum gauges were carefully described by Steckelmacher (1987). Its importance has increased in recent years because of the precision required in specifying pressures for a number of scientific and technological processes, particularly in the fabrication of semiconductors. Hinkle and Uttaro (1996) have discussed the methodology and performance of standards for both vacuum and gas flow measurements suitable for the production environment.

There are basically three different procedures by which the readings of a gauge can be related to an accurately known pressure, and thus calibration achieved. They are (1) a comparison of the gauge with an absolute gauge, (2) attachment of the gauge to a purpose-designed vacuum system in which known pressures can be generated, and (3) comparison of the gauge with a reference gauge that has already been calibrated, and which thus acts as a "transfer standard." Once calibrated by one of these procedures, it may be possible by adjustment of the instrument's processing and display system to make it read accurate values.

By absolute standard in procedure (1) is meant a gauge that measures pressure in terms of other more fundamental quantities that can be determined with high accuracy. Absolute liquid manometers can be used as a standard from atmospheric pressure to about 0.1 mbar and to as low as 10^{-5} mbar if optical interference techniques are used. The McLeod gauge, now obsolete, served as an absolute device for many years, but it was very difficult to achieve accurate and reliable results. Manometers and other primary devices are established at national standards laboratories such as NIST in the U.S., NPL in the U.K., and PTB in Germany. Extraordinary care and thoroughgoing procedures are necessary to measure a pressure absolutely with high accuracy; the simplicity of Equation 7.1, for example, when brought into application, is rich in minutiae, all of which have to be attended to, and which are informatively discussed by Tilford (1992).

Procedure (2), generating a known pressure, may be accomplished by either static or dynamic methods, depending on the pressure range. The essence of the static method is to use Boyle's law to determine a final pressure that is the outcome of isothermally expanding a small volume of gas, initially at a relatively high known pressure, by a known volume ratio. A schematic diagram of an arrangement for a three-stage series expansion, is given in Figure 7.22.

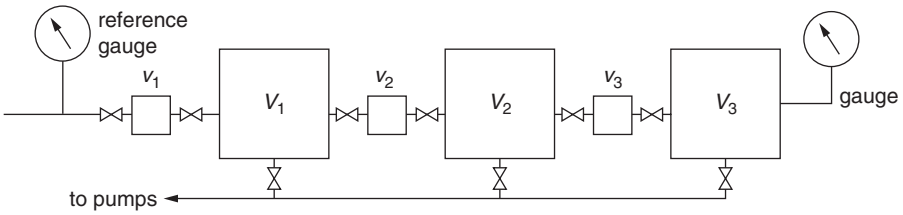

FIGURE 7.22
Schematic diagram of the series expansion technique. (From Chambers, A., Fitch, R.K., and Halliday, B.S., *Basic Vacuum Technology*, 2nd ed., Institute of Physics Publishing, Bristol, 1998. With permission.)

Large volumes V_1, V_2, and V_3, individually connected via valves to a high vacuum pump, are connected in series by small volumes v_1, v_2, and v_3. The gauge under test and a reference pressure gauge, a manometer of high accuracy, are attached as shown. With all volumes initially evacuated and isolated, an inert gas such as nitrogen or argon is admitted into v_1 at a known pressure p_1 and isolated. It is expanded isothermally by opening connections into V_2 and v_2. The gas in v_2 is now isolated and then expanded by opening connections to the next two volumes, and so on. The final pressure, p_3, is given by the equation

$$p_3 = p_1 \times \frac{v_1}{v_1 + V_1 + v_2} \times \frac{v_2}{v_2 + V_3 + v_3} \times \frac{v_3}{v_3 + V_3} \tag{7.17}$$

Thus, if the volume ratios are known, p_3 may be determined from the initial p_1. For work of the highest accuracy, Equation 7.17 may be refined to allow for the deviation from ideal gas behavior expressed in the virial Equation 2.2 of Chapter 2. Depending on the gauges to be calibrated, pressures may be generated in a range from 10 down to 10^{-7} mbar, with uncertainties of order 0.1%. For example, a generated pressure of 10^{-6} mbar will have uncertainty ±0.2%, and higher pressures less. The sequence of procedures presents a good example of error analysis in which the various uncertainties accumulate. Among the experimental concerns, particularly when generating the lower pressures, is the outgassing of the vessel walls, since Equation 7.17 assumes of course a fixed mass of gas. Thus prior baking is appropriate. Equally however, freshly baked clean surfaces may pump by adsorption, so careful attention to these matters is necessary. Adequately low outgassing is achieved when the rate of pressure rise with the system isolated from the pumps is sufficiently small for the purpose in hand. Expansions using inert gases such as argon are less problematic than those using chemically active gases such as oxygen. Increasingly, the calibration of SRGs in gases of industrial importance is necessary, and so this remains an active field. Jitschin et al. (1990) discuss the method and experimental concerns in detail.

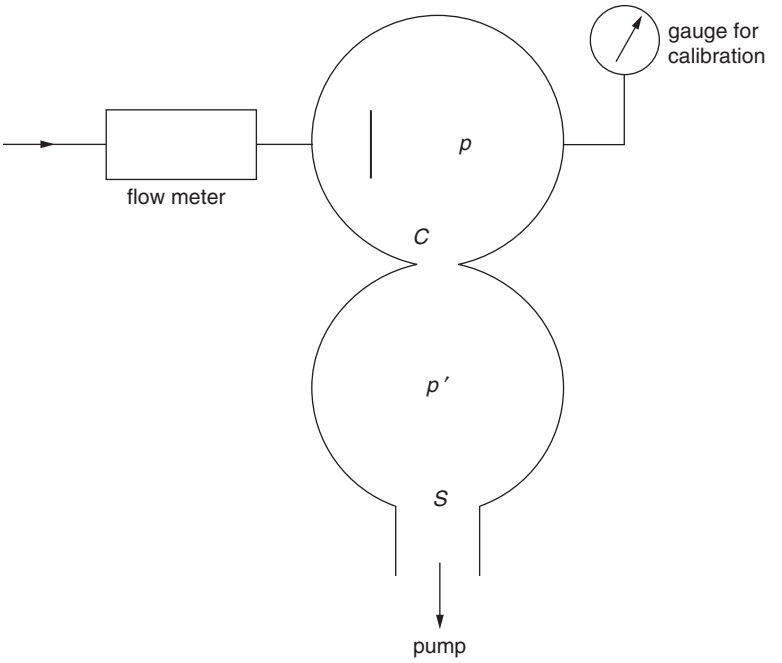

FIGURE 7.23
Schematic diagram of the orifice flow technique.

The dynamic method of generating a known pressure is used at lower pressures. The technique is referred to as "dynamic expansion," or more frequently and informatively as "orifice flow." Figure 7.23 is a schematic diagram of a typical arrangement used.

Two spherical volumes are connected via an orifice of known conductance. To the upper one, referred to as the calibration volume, is attached the gauge to be calibrated, and gas is led into it through a calibrated flowmeter, the flow rate being controlled by a fine-leak valve. The lower volume is connected to a pump of large speed S. The pressures involved are sufficiently low that flow through the system is molecular and the shape of the volumes promotes randomization of the velocity distribution. Suppose that the pressures in the calibration volume and the lower pumped volume are p and p', respectively, in a steady state for which the throughput measured on the flowmeter is Q. If the conductance of the orifice is C, then using Equation 5.7

$$Q = C\,(p - p')$$

But also by Equation 5.6, $Q = S \times p'$, so that the unknown pressure p is determined by the equation

$$p = Q\,(1/C + 1/S)$$

The value of C for an aperture of area A in an ideally thin (zero thickness) plate is known exactly and given by Equation 5.38. Because plates have finite thickness, a refinement of this result with a Clausing transmission probability factor is possible. Interestingly, the conductance of a circular hole in a plate of finite thickness fabricated by spherical grinding from both sides is exactly calculable (Edwards and Gilles, 1966), and this has been exploited in recent applications of the method (Butler et al., 1999). The value of C for a practical aperture 1 cm in diameter will, from Equation 5.40, be about 10 l s^{-1} for nitrogen. The pumping speed S is of order several hundred l s^{-1}, so that the second term in the bracket makes a relatively small contribution to the result. Q may be measured in several ways. In so-called constant pressure flowmeters, a constant pressure is held in the gas upstream of the fine leak valve that regulates flow into the calibration volume, while the volume upstream is reduced by the inward movement of a piston at an appropriate steady rate. Then $Q = p(dV/dt)$. In constant volume devices, on the other hand, the equation $Q = V(dp/dt)$ describes the flow into the calibration volume from a fixed volume V in which pressure falls with time at a controlled constant rate (dp/dt).

The dynamic expansion method can be used to generate pressures from 10^{-6} down to 10^{-10} mbar. Compared with the static expansion method, there are fewer concerns about the effects of adsorption and desorption, and generating a set of known pressures can be accomplished more rapidly. The interested reader might consult the papers of Poulter (1977) and Tilford et al. (1988) for more information.

Procedure (3), calibration against a transfer standard, requires that the gauge serving as the transfer standard, as well as having high accuracy, be sufficiently stable that its accuracy and calibration are retained with long-term use. Its calibration will have been carried out by either procedures (1) or (2). The most widely used transfer gauges are the capacitance diaphragm and the spinning rotor gauges. Care has to taken, of course, that the gauge to be calibrated and the standard be located in a vacuum system in such a way that they both sense the same pressure.

7.7 Partial Pressure Gauges

In an increasing number of the applications of vacuum, it is necessary to know not only the total pressure in a vacuum system, as measured by one of the gauges previously described, but also the composition of what is usually the mixture of gases, rather than just a single species, that constitutes the background residual gas in a system. In terms of matters introduced in Chapter 2, partial pressures are sought. To determine the constituents of a mixture, ionization by electron impact is exploited, as in the ionization gauge, with the addition of means to sort out the various ion types thus produced and relate the ion currents to the quantities of gases present.

7.7.1 General

Essentially, a partial pressure analyzer, more usually referred to as a residual gas analyzer (RGA), consists of three parts, as depicted schematically in Figure 7.24(a). An ion source samples the gas in the system and produces a representative set of ions. These are sent to a mass analyzer that sorts them out according to their mass-to-charge ratios m/e and acts as a filter, selecting ions of a particular m/e from those present to be passed on to a detector, where they are counted, giving a measure of how many molecules of the type from which they originate are present in the gas sampled. By varying the selected m/e value so as to scan the range possible, the detector output can be arranged to display a "mass spectrum" of ion current vs. the mass number associated with the particular species, e.g., 32 amu for oxygen. This simple conceptual scheme assumes that molecules are just singly ionized, which is not strictly true as there are small proportions of multiply ionized species, but this and other effects such as the dissociation of some molecules into fragments by the electron beam are accounted for in the ensuing analysis of the raw data. Figure 7.24(b) illustrates schematically the mass spectrum of a mixture of helium and water vapor.

The variation of ionization efficiency and other important instrumental factors with ion type means that partial pressures are not simply related to the magnitudes of their associated ion currents, and so, as is the case with ionization gauges, if high accuracies are sought, very careful individual calibration may be needed to refine the calibration built into the instrument by the manufacturer. For many applications, however, it is sufficient to have an indication of the general proportions of a mixture as revealed by the displayed spectrum and to know that certain species are below the level of detection or that the amount of others is below some threshold critical for the purpose in hand.

The RGA does not have the range or precision of the traditional form of mass spectrometer, exemplified by the Dempster mass spectrograph for the accurate measurement of isotopic masses and abundances, and similar instruments designed for use in analytical laboratories. These are capable of accuracies as high as six significant figures in mass determination, may have a range up to

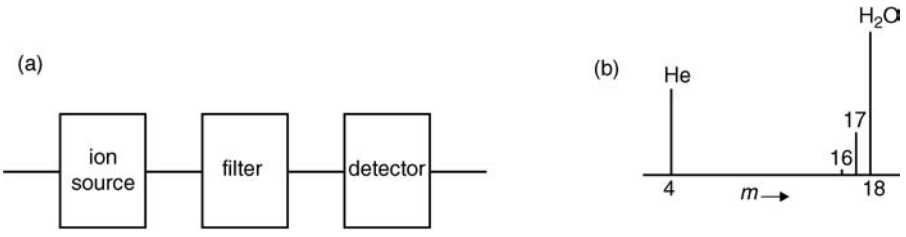

FIGURE 7.24
(a) Essential components of an RGA, and (b) simple mass spectrum.

several thousand amu, and have to be physically large. The requirements of an RGA, however, are different. It has to be physically small so that it can be simply incorporated into a vacuum system, and usually it is desirable that the gas spectrum, from 1 to 100 amu typically, be measured repeatedly at a reasonably fast rate, say, once every few seconds, in order to monitor change. High sensitivity is thus called for so that, as the instrument scans through the mass spectrum, the collection time for measured currents can be adequately short. Signal-to-noise considerations thus arise; ion currents must be detectable above the background electronic noise of the measuring system.

Intimately related to its sensitivity is an instrument's resolution — its ability to distinguish masses of different value. Sensitivity and resolution are in direct competition because increasing sensitivity is automatically allied with a coarsening of resolution. This is true for all instruments of this general type. For example, an optical prism spectrometer used to examine light from a white light source disperses it spatially in angle into its constituent colors. Increasing the width of the entrance slit that is imaged at the output through the telescope produces a brighter image but one to which more wavelengths contribute. In the context of mass spectrometry, an analogous situation is that where the directions of travel of ions sorted according to their mass are focused and spread across a linear region in which the size of a suitably located aperture determines those that contribute to the measured current. Such is the case in the magnetic sector type of analyzer, and similar sensitivity/resolution issues exist for instruments based on different principles, including the quadrupole mass analyzer to be described next.

The resolution achieved may be specified in terms of a resolving power defined as $M/\Delta M$, where M is the mass number of a peak in the spectrum and ΔM is its width measured at some specified fraction of the peak height, either at 50% (then referred to as FWHM — full width at half maximum) or at 5%. The latter is very widely adopted and is recommended by the American Vacuum Society. Figure 7.25 illustrates how it gives rise to "10%

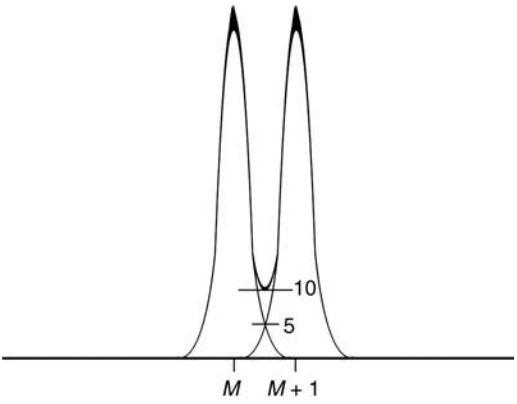

FIGURE 7.25
Resolution of adjacent mass peaks.

valley resolution" for adjacent peaks with masses M and $M + 1$ of the same height. The valley between the peaks is deep and the masses are easily distinguished, though one can imagine that if the peak width were substantially broadened, resolution would be lost. With this, 5% definition resolution is also adequate for peak heights that are quite dissimilar. Because, as noted, resolution and sensitivity are interdependent, quoted values of resolving power should indicate the associated sensitivity. Such is the need for high sensitivity that the ability to resolve masses differing by one mass unit is quite acceptable in practice. Thus, in the oft-quoted and important matter of distinguishing nitrogen and carbon monoxide that have molecular masses 28.0061 and 27.9949 amu, respectively, both are sensed as mass 28, and it is their characteristic fragmentation patterns (to be described later) that allow them to be separately identified.

In practice, typical values of sensitivity achieved are a few $\times 10^{-4}$ A per mbar for nitrogen, usually taken as the gas of reference. Although ion sources use electron impact to create the ions, the value of electron current is not quoted because the two are not proportional to each other on account of space charge effects that are exploited in the operation of ion sources. The sensitivity value quoted above is two orders less than comparable ionization gauge sensitivities — 10 A per mbar (for 1 mA of electron current) because of restrictions imposed by the deflecting fields and apertures associated with the ion collection process. Sensing a partial pressure of 10^{-13} mbar, corresponding to a component that is roughly 0.1% of a vacuum of 10^{-10} mbar, therefore involves measuring a current of about 10^{-17} A, and requires electronic amplification. Many designs have mass range 1–100 amu, which is adequate for many applications, but instruments with ranges up to 200 or 300 amu are also available.

7.7.2 The Quadrupole Instrument

The variety of analyzer types developed in the period 1950–1970 is summarized in Roth (1990). The instrument based on the quadrupole mass filter has come to occupy a position of almost total dominance at the present time and will therefore be described in some detail, but the principles of magnetic sector and time-of-flight instruments that have special fields of application will also be briefly described.

As shown in Figure 7.26(a) and (b), the quadrupole mass filter consists of four parallel cylindrical conducting rods set in a square array and connected electrically so that opposite pairs are at the same potential, which is the sum of a DC component U and an alternating component $V \cos \omega t$ at a radio frequency ω of a few MHz. At any instant t, the potentials are $(U + V \cos \omega t)$ on one pair and $-(U + V \cos \omega t)$ on the other, as may be deduced from the circuit arrangement. Ions are injected down the longitudinal axis z. The rods are typically about 100 mm long and 6 mm in diameter, with each a distance r_0 of about 3 mm from the central axis so that the assembly can fit

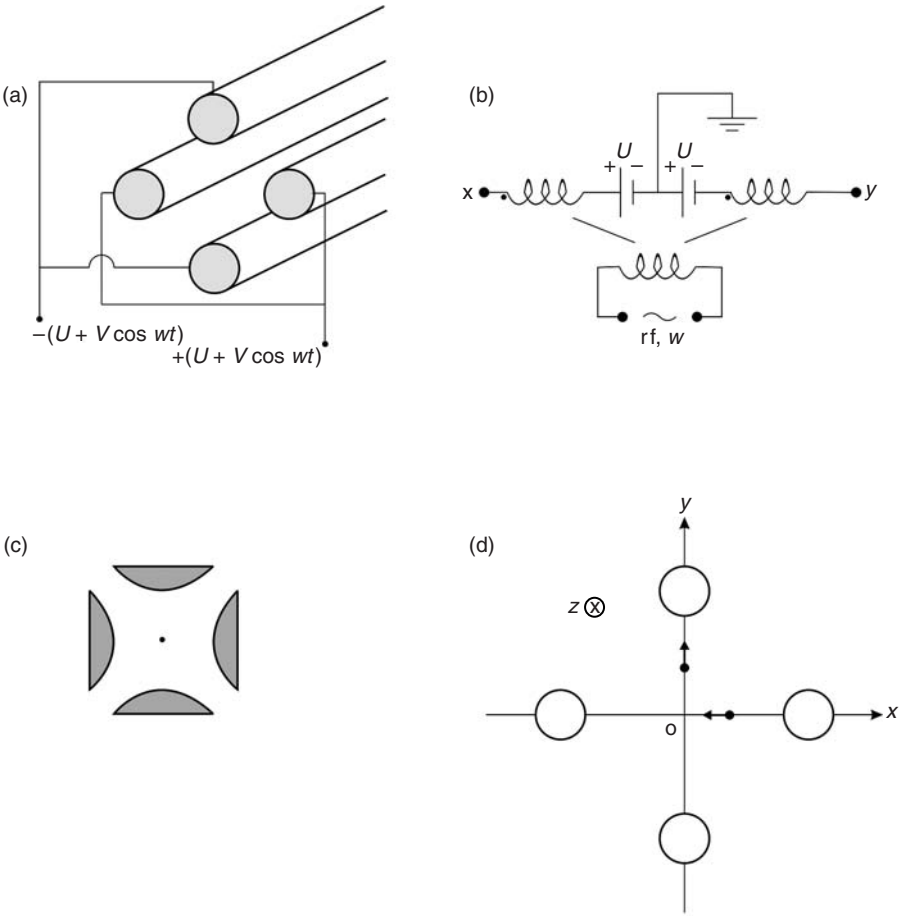

FIGURE 7.26
(a) Perspective view, (b) circuit arrangement, (c) ideal section, (d) force vectors.

compactly into a tubular volume of about 35-mm outside diameter attached to a standard 70-mm flange. Ideally, the rods have the hyperbolic cross section suggested in Figure 7.26(c), and analysis is based on this geometry, but circular rods chosen to have radius 1.16 r_0 are used because they are much cheaper to produce and assemble. The electric field produced is similar in character.

In the region near the z axis the potential ϕ is

$$\phi = \left(U + V\cos\omega t\right)\frac{\left(x^2 - y^2\right)}{r_0^2} \tag{7.18}$$

The associated electric fields are $E_x = -d\phi/dx$ and $E_y = -d\phi/dy$. Writing the spatially independent part $(U + V \cos \omega t) = \phi(t)$ for simplicity, we have therefore, in the plane containing the x and z axes,

$$E = E_x = -\frac{2x}{r_0^2}\phi(t) \tag{7.19}$$

while in the yz plane

$$E = E_y = \frac{2y}{r_0^2}\phi(t) \tag{7.20}$$

Differentiation shows that at a general point x,y,z the total field E has components E_x and E_y that are also given by these expressions. There is no E_z at any point because the potential does not depend on z. Evidently, if $\phi(t)$ is positive, the force eE_x on a positive ion of charge $+e$ at a displacement x in the xz plane is directed towards the z axis, tending to promote stability and transmission down the axis, whereas for displacements in the y direction the opposite is true; ions are urged outward away from the axis towards the electrodes. This is suggested by the arrows representing force vectors in Figure 7.21(d). The quantity $\phi(t) = (U + V \cos \omega t)$ is always positive if $V < U$, but if $V > U$, which is the case in practice, it becomes negative for a part of each cycle, reaching a negative value $U - V$ at its lowest, and if V becomes very much greater than U, the negative part occupies almost one half of the cycle. As will be explained below, it is the consequent reversal of the force vectors for part of each cycle that is the basis of the quadrupole's filtering action, allowing transmission down the axial direction of a restricted set of ions with particular m/e value, while those with lower and higher values are driven away from the central region in trajectories that lead to neutralization and collection at the quadrupole rods. One may extend the interpretation of force vectors in Figure 7.26(d) to include their time dependence: as drawn, they can be thought of as the vectors corresponding to the maximum value $U + V$ of the potential, and each will diminish and reverse once per cycle, reaching a peak negative value corresponding to $U - V$.

The equations of motion of a positive ion of mass m and charge e in the field above are, by Newton's second law,

$$\ddot{x} + \frac{2e}{m}(U + V\cos\omega t)\frac{x}{r_0^2} = 0 \tag{7.21}$$

and

$$\ddot{y} - \frac{2e}{m}(U + V\cos\omega t)\frac{y}{r_0^2} = 0 \tag{7.22}$$

with

$$\ddot{z} = 0 \tag{7.23}$$

From this last equation we deduce that the initial axial velocity component with which an ion enters the region remains unaltered. The equations in x and y are a form of the Mathieu equations whose solutions are complex and have to be obtained numerically. For this purpose, they are cast into a form involving dimensionless parameters $(2eV/mw^2r_0^2)$ and $(4eU/mw^2r_0^2)$. Dawson (1976) discusses the quadrupole extensively, and Leck (1989) gives a very detailed interpretation of the solutions. A helpful nonmathematical discussion of quadrupole operation in terms of the stability of trajectories in the potential field, originally given by Mosharrafa (1970), is presented in O'Hanlon (2003), and followed below using force rather than potential.

The solutions of the equations show that ions injected close to the axis and making small angles with it follow trajectories that are a composite of independent motions in the x and y directions. They are oscillatory in character on account of the cos ωt term and generally have increasing amplitude that leads to their loss from the system. These are referred to as unstable solutions. However, depending on the values of the dimensionless parameters, determined by U, V and ω, there are stable solutions corresponding to trajectories that although oscillatory, stay within the space defined by the rods to reach the exit for subsequent detection. The existence of these stable solutions may be qualitatively understood as follows. Remembering that the amplitude V of the alternating component of the applied voltage is greater than the DC component U, for an ion traveling in the z direction, that part of its motion controlled by the electric fields in the x direction will tend to be stable while $\phi(t)$ is positive and unstable when it is negative. The destabilizing influence when $\phi(t) < 0$ will have the greatest effect on the lighter ions of smaller m/e, which after several cycles can be ejected to the electrodes, while heavier ions, feeling the same force but with greater inertia, will be less perturbed and stay in a stable transmitting trajectory. The pair of rods along the x coordinate axis that carries the positive U potential therefore acts as a high pass filter. At the same time, the motion controlled by the field in the y direction will tend to be unstable while $\phi(t)$ is positive, but stable and transmitting when it is negative. Stability and transmission are now achieved for lighter ions on which the field has more influence; for heavier ions, the stabilizing influence is insufficient to prevent their loss to the electrodes. This pair of rods on the y axis carrying the $-U$ potential therefore acts as a low-pass filter. Taken in combination, the two pairs act as a band pass filter, low and high mass ions outside the band being rejected, with limits determined by the values of U, V, and ω. How these determine the filtering action is shown in Figure 7.27.

Stability and transmission through to the detector for an ion of mass-to-charge ratio m_1/e and associated mass number M_1 occurs only when the

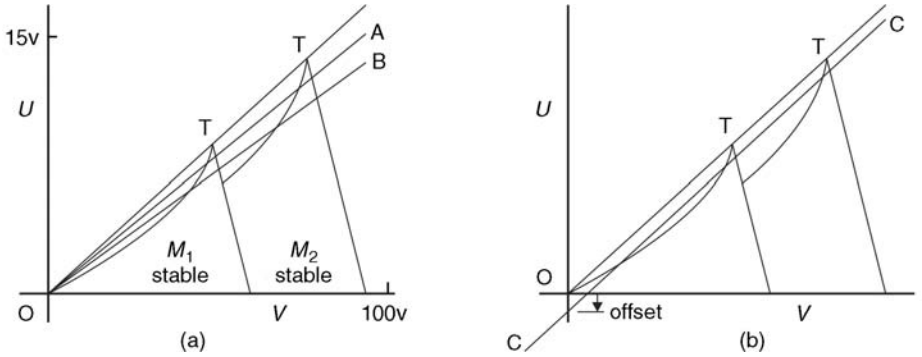

FIGURE 7.27
(a) Stability plots and scanning options, (b) scanning with constant resolution.

values of U and V lie within the indicated triangular area of the U, V plane; values outside it correspond to instability. Each value of M has a similar "stability triangle," and so there is an overlapping family of them, although for clarity, only two are fully shown, those for M_1 and M_2, with M_2 rather greater than M_1. Importantly, the topmost points T for each M lie on a line of constant slope U/V in the plot. Scanning through the mass range is accomplished by increasing the U and V values (at fixed ω) so that U/V stays constant. Following a line of slope OA to just undercut the top of T results in high resolution. Scanning along line OB of smaller slope will clearly reduce resolution while increasing the transmitted current. Notice that U/V is about 0.16. The resolving power $M/\Delta M$ depends only on this ratio and so is constant, resolution ΔM varying with mass, as may be deduced by inspection. It is more convenient in practice to have constant resolution ΔM through the spectrum. This is accomplished by having a constant small negative offset on the value of U so as to scan along a line CC parallel to OA. Driven electronically in this way, the quadrupole filter selects M values across the range with constant resolution, less than 1 amu. With increasing m, this implies a decrease in sensitivity for which allowance has to be made. These matters and other refinements are discussed by Ellefson (1998).

Returning to Figure 7.24(a), the filter described above has to be matched to an ion source at its input and a detector at the output. The quadrupole instrument usually has an axial ion source whose design is shown schematically in Figure 7.28(a). From a linear filament F, electrons are accelerated through a potential difference of about 70 V into an open cylindrical grid structure G in which they travel in long trajectories before collection, ionizing the residual gas. A repeller plate R behind and a few volts negative with respect to the filament assists the projection of electrons towards the grid region. The grid is coaxial with the quadrupole filter and is held at a small potential positive relative to zero potential on its axis. Between the grid and quadrupole are apertures E. Because of the distribution of electronic space charge inside the grid, positive ions tend move to the central axis, from

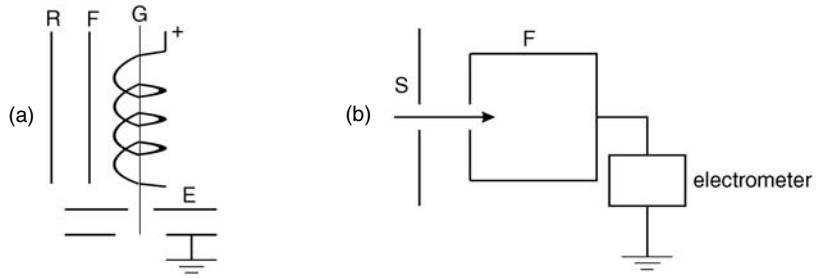

FIGURE 7.28
(a) Axial ion source, and (b) Faraday cup collector.

where they are extracted axially and roughly focused by apertures E carrying appropriate potentials and sent into the quadrupole filter with energies of a few eV. As previously noted, the transmitting properties of the filter at the selected *m/e* values do not depend on the *z*-component of the ion velocity at entry. The operation of the electron-emitting filament, usually of thoriated iridium, limits the upper operating pressure of the instrument to 10^{-4} mbar.

The transmitted ion current may be measured directly by a simple Faraday cup arrangement or by more sophisticated means that involve its multiplication to give greater sensitivity. The principle of the former is shown in Figure 7.28(b). Positive ions enter the cup-shaped electrode F and are neutralized at its surface, inducing an equal electron current flow in the circuit that is measured by the electrometer. A suppressor electrode S adjacent to the cup entrance helps to repel back into the cup any secondary electrons generated by the initial ion impact, whose escape otherwise would cause a spurious additional increment in the indicated ion current. The input resistance ~10^{11} Ω and stray capacitance ~pF associated with the electrometer circuitry result in a response time-constant $RC \sim 0.1$ s. The currents range from about 0.1 μA at the upper pressure limit down to the smallest that can be detected above noise levels, typically about 10^{-14} A, that correspond to minimum detectable pressures of about 10^{-9} mbar, or a little less.

Greater sensitivity is achieved by having a multiplier device into which the ion current is deflected when Faraday cup detection becomes inadequate. All devices exploit secondary electron emission but in different geometries. The "venetian blind" type with discrete dynodes has tended to be replaced by the channeltron type illustrated in Figure 7.29. Microchannel plate structures may also be used. In the widely used channeltron, a highly resistive thin film whose surface has a high secondary emission yield coats the interior of a glass horn shape and sustains a potential difference of 2 kV from a high voltage source. Potential increases from entrance to exit. The ion beam is accelerated onto the entrance area where each ion produces several electrons. From their point of departure, these electrons are accelerated further into the horn, and on impact each produces several more so that there is a developing cascade of electrons that eventually reaches the output, where the current is collected by an electrode measured by an electrometer as

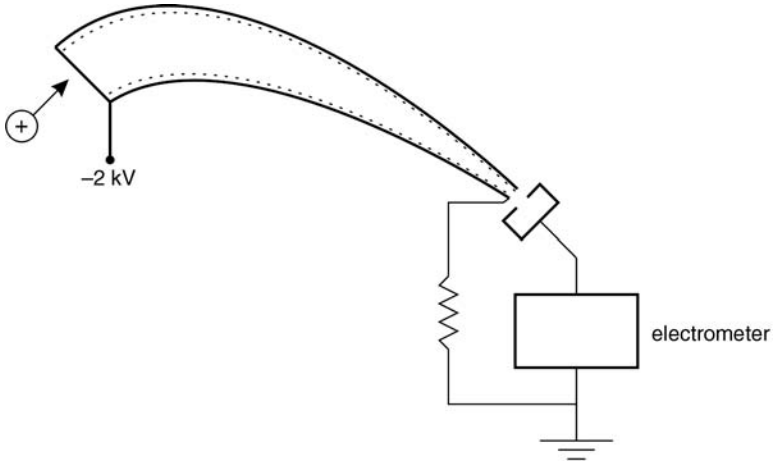

FIGURE 7.29
Channeltron multiplier.

before. The current multiplication achieved may be as high as $\times 10^6$. The sensitivity of the electrometer need not be as high as in the Faraday cup detector and lower input resistance is therefore possible so that the time constant is much reduced, enabling fast scans. Minimum detectable partial pressures are typically about 10^{-14} mbar.

7.7.3 Magnetic Sector and Time-of-Flight Analyzers

The field of residual gas analysis is dominated by the quadrupole instrument described above. Instruments based on the magnetic sector principle of mass separation no longer have a large role in this field, but the principle is still used in leak detectors that selectively sense helium gas, and so it is appropriate to describe it briefly. A magnetic analyzer with a 180° sector is shown schematically in Figure 7.30.

An electron beam from a thermionic filament F is accelerated through an aperture into the ion source box (ISB) and travels across it causing some ionization before exiting through an aperture in the opposite side to a collecting electrode, X. The positive ions are extracted from the ion source via an exit slot with the assistance of a small repeller potential on R and are then accelerated through a potential difference V into a region of uniform magnetic field B, whose direction is normal to the plane of the figure. This field and the path-defining slits S1, S2, S3, and S4 form the analyzer. Entrance slit S1 is at earth potential, and box ISB is at positive potential V. Ions of mass m and charge e injected into the magnetic field will travel in a circular path of radius R defined by the slits if their velocity v is such that the force Bev due to their motion perpendicular to the field provides a centripetal acceleration v^2/R. Thus, it is required that $Bev = mv^2/R$. It follows

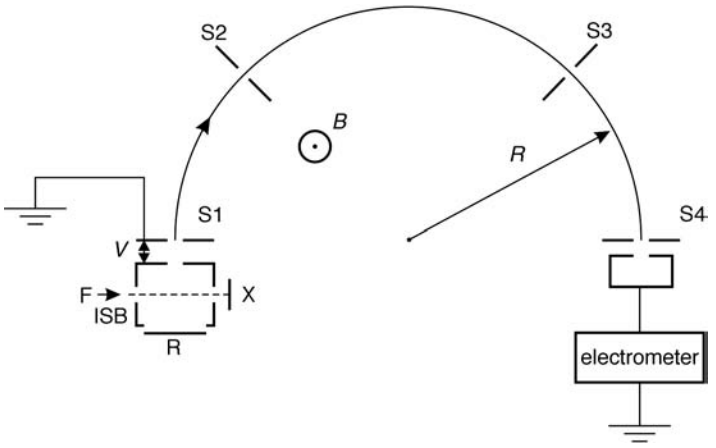

FIGURE 7.30
Schematic diagram of a 180° magnetic sector spectrometer.

that $R = mv/eB$ so that with fixed R and B the analyzer sorts ions by their momentum mv. The velocity of the ion is determined by the potential difference V as $\frac{1}{2}mv^2 = eV$ whence

$$\frac{m}{e} = \frac{R^2B^2}{2V}$$

Ions with a particular value of m/e will pass through all four slits to reach the detector, a Faraday cup (F), if the value of V through which they are accelerated is given by this equation. Thus the mass spectrum can be scanned by varying the ion accelerating voltage V, and m is inversely proportional to V. Magnetic sector analyzers also have a focusing property, so that ions of the same energy that diverge slightly from the central direction converge to a focus at the exit slit.

The ion source described is known as the Nier type after its inventor. The ion beam it produces is a "ribbon" whose elongated cross section is aligned parallel to the input slit (in the figure this is not shown; in fact, line FX should be perpendicular to the page). Slit widths control resolution and sensitivity. Resolution increases with decrease of slit width, but sensitivity decreases, so a compromise has to be made in design. In practice, small devices for vacuum application have an ion deflection radius of about 1 cm and fixed magnetic field of about 0.2 T.

Another rather specialized type of analyzer is that based on sorting by time-of-flight, TOF, and shown schematically in Figure 7.31. A set of ions created locally by electron beam ionization or other means such as laser desorption off a surface is subjected to a brief voltage pulse 100 V that draws them into an electric field in which they are accelerated through a potential

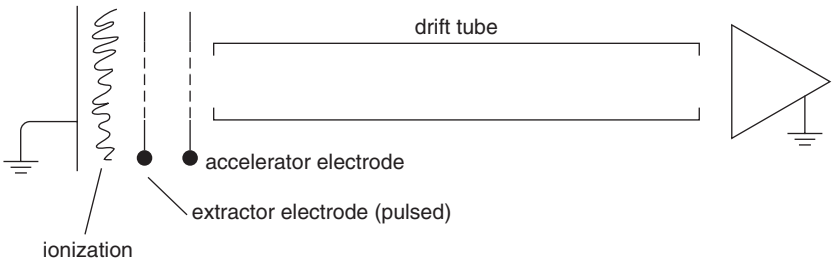

FIGURE 7.31
Time-of-flight analyzer.

of 3 kV and then injected into the field-free space of a long drift tube. Down this they travel with constant velocity to a detecting multiplier. All ions acquire the same energy zeV and so, equating this to $(1/2)mu^2$, various drift velocities $u = \sqrt{2zeV/m}$ are present corresponding to the range of mz/e values. Including an integer $z = 1, 2,\ldots$, allows for ions that are multiply-ionized. The time of flight t down a tube with source-to-detector distance L is

$$t = \frac{L}{u} = \left(\frac{m}{ze}\right)^{\frac{1}{2}} L/(2V)^{\frac{1}{2}}$$

A display of the multiplier output signal vs. time elapsed from the injection therefore presents a spectrum of ions spread out according to their mass/charge ratio. Because t is proportional to $\sqrt{m}$, the scale is not uniform. The times of flight down a 1-m tube following acceleration through 3 kV for singly ionized masses 2 (H_2^+), 28 (N_2^+), and 100 are 1.9, 7, and 13 μs, respectively, so repeated pulsing at an appropriate rate covers the above range of elapsed times. For each value of the mass/charge ratio, the signal will have finite width because of the small spatial extent of the region from which the ions are extracted and the consequent slight variation in the accelerating potential experienced. As mass increases, the separation in time Δt of adjacent mass peaks, which depends on the factor $\sqrt{m+1} - \sqrt{m}$), decreases, and so resolution at high mass values becomes problematic, though an ion–optical technique using a device called a "reflectron" helps to address the problem. Inexpensive fast, modern electronics and computing power have led to a renewal of interest in this TOF technique that is described in more detail in the article by Ellefson already noted.

7.7.4 Analysis and Display

Whatever type of mass analyzer is used, the spectrum of ions presented by the ion source for analysis has to be carefully interpreted in order to deduce

the composition of the system gas that has been sampled. A principal feature of the ionization that occurs in the source is that, in addition to producing ions by the loss of one electron from the parent molecule, the energy of the ionizing electrons is sufficient to break apart a fraction of the parent molecules into charged fragments, each with a distinct value of the ratio m/e and therefore a corresponding peak in the spectrum. Thus, for each molecular species, the analyzer gives a characteristic "cracking pattern." These are quantified by quoting the percentage of each fragment signal relative to that of the singly ionized parent molecule taken as 100. For example, for CO_2, as well as CO_2^+ at mass 44, there are fragment peaks at mass 28, (11%), mass 16, (8%), and mass 12 (6%), due to CO^+, O^+, and C^+, respectively, with some other very much smaller peaks from other causes. Nitrous oxide, N_2O, also has mass 44 but a different cracking pattern with fragments at masses 30, 28, 16, and 14. Water vapor, an important constituent of the residual gas in many vacuum systems, gives fragment peaks at masses 16 and 17 due to O^+ and OH^+, with typically 1 and 23%, respectively, of the amplitude of the main peak at mass 18. The relative proportions of fragments in a cracking pattern are not invariant. They depend on the design of the analyzer, the energy of the electron beam that produces the ions, and may vary with operating conditions, particularly the state of multiplier surfaces that may slowly degrade with use. The fragments are, of course, created in the ion source and are not present in the system's residual gas.

There are two other aspects of the ionization process that are important. One is the creation of multiply charged ions and the other is the existence of isotopes. For example, the spectrum of pure argon has two peaks — one at mass 40 from singly ionized A^+ and another with relative magnitude 11% at mass 20 arising from the A^{++} ion with mass/charge ratio $m/2e$. Mass value 20 would be truly present from singly ionized neon, in which case there is also an isotope at mass 22 with relative abundance 9%. Most elements have naturally occurring isotopes, but in many cases one is dominant and the existence of the others insignificant for most purposes. Thus, the nitrogen atom exists as two isotopes of mass 14 amu (99.63%) and 15 amu (0.63%) so that there are nitrogen molecules with mass 28, 29, or 30, but with mass 28 strongly dominant. In fact, in the mass spectrum of molecular nitrogen, the only substantial subsidiary peak is 7% at mass 14 arising from N_2^{++}, N^+. However, in some cases, isotopes occur in comparable proportions. Xenon has five such isotopes in the mass range 129–136. Chlorine is a well-known example, with isotopes of mass 35 and 37 in the ratio 3:1. Silicon has isotopes at mass 28, 29, and 30 in the proportions 92:5:3. The isotopic abundances, together with the cracking patterns of various gases and vapors, are listed in O'Hanlon (2003).

The distinction referred to earlier between nitrogen and carbon monoxide, both frequently occurring together, is made on the basis of their cracking patterns. Both give signals at mass 28 = (2 × 14) and (12 + 16), but nitrogen has an associated peak at mass 14 as noted, while for CO there are peaks at masses 12 and 16 from C_{12}^+ and O_{16}^+ with magnitudes approximately 3.5 and 1.5%, respectively, of the mass 28 peak.

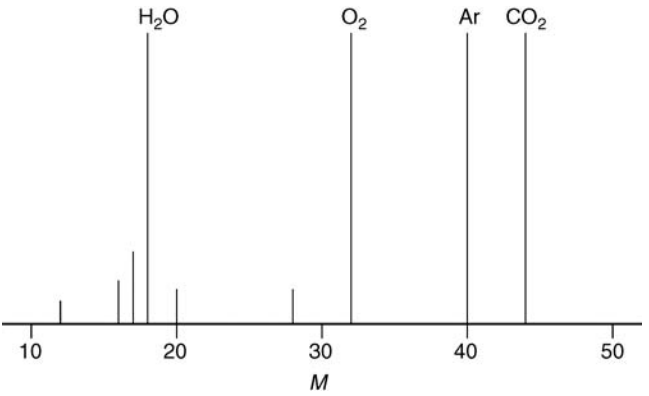

FIGURE 7.32
Hypothetical spectrum.

When, as is frequently the case, there are two or more gases present, each will contribute its own characteristic cracking pattern, and there may be some overlapping of signals, so that the proportions of each gas can be determined from the relevant peak heights only if their cracking patterns are known either from data compilations or, ideally, because instruments differ in their detailed sensitivities, from previous calibrations of the instrument in use. Figure 7.32 shows a hypothetical spectrum of a simple mixture of argon, water vapor, oxygen, and carbon dioxide in roughly equal proportions and assuming comparable sensitivities to each. There are some overlapping peaks. If approached as an unknown spectrum, the characteristic and frequently occurring water vapor cluster is evident; carbon dioxide, argon, and oxygen are suggested by the strong parent peaks at 44, 40, and 32, the argon being corroborated by the minor peak at 20. The peak at 28 might be nitrogen, but the absence of a peak at 14 suggests it may not be; 28 is likely to be a fragment from CO_2, with minor contributions to the O^+ peak at mass 16. To identify the constituents in a mixture of known total pressure and estimate their general proportions, so that rough estimates of the partial pressures could be made, is adequate for many purposes. This qualitative analysis involves some detective work, a familiarity with cracking patterns, and a trial-and-error approach, frequently with some expectations of residual gas composition based on prior knowledge of how the system is pumped and operated. The simple cracking pattern charts supplied by some of the equipment manufacturers are easily used and very helpful in practical situations.

Quantitative analysis to give reasonably accurate values of partial pressures is much more problematical and requires that instrumental performance be thoroughly characterized so that the relative sensitivity to different gases and vapors is known. If a spectrum can be interpreted in terms of constituents that do not produce overlapping peaks, then an estimate of the partial pressures may be relatively straightforward. But when there are

overlapping peaks, the fact that they arise from constituents, some or all of which initially at least may be unidentified and in unknown proportions, makes the task of unraveling the detail quite complex. For a thorough and practical introduction to the problems that arise in analyzing spectra, the reader should consult O'Hanlon (2003), an excellent resource. Residual gas spectra of representative vacuum systems are to be found in a number of texts, but notably in the AVS monograph of Drinkwine and Lichtman (1995).

Returning to the quadrupole instrument, its operation and use, it is fortunate that its control by computer software makes it particularly versatile. The high sensitivity and fast scans possible with electron multiplier detection mean that data can be gathered and processed at a high rate. Manufacturers supply the software for controlling the instrument. Basically this drives it, acquires data from it, stores and processes that information, and presents it on screen as required. There are various modes of operation. In the analogue mode, the raw data acquired of ion currents at different m/e values, as the spectrum is scanned through the mass range, is displayed. The peak shapes are evident, and resolution, sensitivity, and the range scanned can be adjusted to suit the purpose in hand. The display of peak height vs. mass number may be linear or logarithmic with a range of up to six decades. The logarithmic scale can be very useful in indicating the presence of very small signals, though in revealing many peaks by boosting weak signals the logarithmic display can at first sight sometimes present information overload; switching to a linear or reduced logarithmic scale reveals a simpler display that emphasizes only the main features present. In the bar-graph mode this information is processed by the software to give a cleaner-looking display than the corresponding analogue mode, allowing easier recognition of groups of features.

The library mode contains the stored data and spectra of a wide range of gases and vapors that can be compared, using a split-screen facility, with an acquired spectrum under investigation and be used quantitatively in sophisticated analyses. A particularly valuable mode is that which displays selected peaks vs. time. The magnitude of up to ten mass peaks can be monitored and displayed, each with a different color trace and a time scale whose span across the screen, minutes or hours, is set to suit the purpose in hand. The correlation or lack of it between the signals is very informative in helping to identify the source of some operational problems. This mode has many applications, for example, in following the changes to gas composition that might be caused by local outgassing when an evaporation source is switched on, or changes occurring during an overnight baking period. In the leak detector mode, the partial pressure of helium may be monitored to detect its passage into the vacuum through leaks when it is presented on the atmospheric side of the vacuum wall.

Quadrupole instruments cannot be operated at pressures greater than 10^{-4} mbar because of their thermionic filaments. It is, however, a frequent industrial need to do gas analysis at higher pressures than this. Sometimes this problem can be solved by housing a quadrupole in an auxiliary, separately

pumped high-vacuum chamber and taking a continuous sample of gas into it from the system at higher pressure via a suitably designed connection of known conductance for the various gases sampled. Gas analysis is thereby accomplished while keeping the instrument under its proper operating conditions. (See Question 8.5.)

It is arguable that an RGA is an essential part of a vacuum system even if one is not concerned to have a detailed analysis of the residual gases, because the residual gas spectrum may give important and otherwise unknown information about the state of the system. The information may not always be welcome, but it may not be irrelevant! That an RGA can be used to detect leaks, as will be discussed in Chapter 10, is also advantageous.

8

Illustrative Examples and Representative Laboratory Systems

8.1 Introduction

In the early sections of this chapter, four examples of flow and pumping calculations are given. They are intended to illustrate the use of the information and methods of the previous chapters in analyzing practical situations. In the later sections, some representative small laboratory pumping systems are described and analyzed, with account taken of their size and purpose, the choice of pumps, and some details of their operation. It is hoped that, as a result, readers may be better equipped to approach and analyze the systems they encounter, with increased recognition and understanding of their form and purpose. The treatment given is not intended as a guide to the design of systems. For that purpose the chapter on high-vacuum system design by W. Schwarz (1998) forms a good general starting point; a comprehensive survey of systems for various applications with thoroughgoing descriptions of their configuration and operation is given in O'Hanlon (2003). In operating vacuum systems, it is of the utmost importance to do so safely, and to follow the recommendations given in manufacturers' literature. Chapter 16 of Harris (2005) may be consulted as a general guide.

8.2 Pumping Water Vapor

In the first part of this example the conditions assumed are slightly artificial but, nevertheless, some useful procedures are illustrated that are relevant in later, more realistic applications. Consider the vacuum system shown schematically in Figure 8.1(a). It consists of a cubical vessel of side 30 cm containing air at atmospheric pressure connected via an isolation valve to a pump of speed $S^* = 2\ \mathrm{l\ s^{-1}}$. Suppose that the internal walls of the vessel are covered with a film of water 0.1 mm thick. The contained air will be saturated

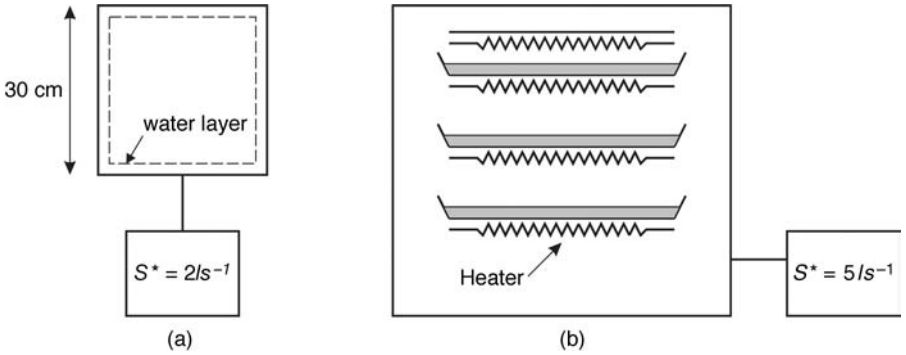

FIGURE 8.1
(a) Pumping water vapor and (b) vacuum drying.

with water vapor at room temperature, taken to be 22°C = 295 K, exerting a partial pressure equal to the vapor pressure of water, which is 23 mbar = 2300 Pa at this temperature, and in equilibrium with the liquid film. How long will it take to pump the water away?

Let us assume that the vessel wall and the film on it are sufficiently good thermal conductors that the energy needed to supply the latent heat for the forced evaporation of the water can be conducted to the surface quickly enough to maintain the evaporating surface at 295 K. (This is the rather unrealistic assumption referred to earlier.)

The volume to be pumped is $(30)^3$ cm^3 = 0.027 m^3 = 27 l and the time constant V/S^*, therefore, 13.5 seconds. The film of water covers an area $6 \times (30)^2 = 5400$ cm^2 = 0.54 m^2 with thickness 0.1 mm and therefore has a volume 5.4×10^{-5} m^3. Because the density of water is 1000 kg m^{-3}, the mass of water is 5.4×10^{-2} kg = 54 g, which is $^{54}/_{18}$ = 3 mol. By assumption, this is pumped away at room temperature, so we need to know what volume is occupied by this mass at a pressure of 23 mbar. From Equation 2.1, this volume is

$$\frac{n_M R_0 T}{p} = \frac{3 \times 8.3 \times 295}{2300} = 3.2 \text{ m}^3 = 3200 \text{ liter}$$

Making the working assumption that the liquid water will not start to evaporate at a significant rate until the pressure of air in the volume has fallen to 23 mbar, we need to determine how long it will take to pump down to 23 mbar. Calling this time t_1, from Equation 5.53 it will be $t_1 = (V/S) \times \ln(1000/23) = 13.5 \times 2.3 \times 1.64 = 51$ seconds, less than a minute. Having reached this condition, the pump

that has a speed of just 2 l s^{-1} would require 3200/2 = 1600 seconds or 27 minutes more to pump away the water vapor at 23 mbar. As vapor is pumped away, it is replaced by fresh evaporation from the surface. After all the water vapor had been pumped away, the pressure would start to fall again towards the ultimate pressure. It takes almost half an hour, therefore, to remove the water in the system.

It is interesting to consider briefly the quantities in this problem in terms of the kinetics of evaporation discussed in Chapter 3, where Equation 3.26 gives the rate of uninhibited free evaporation into vacuum. Under the assumption that the temperature of the evaporating surface could be maintained at 295 K and substituting p_e = 2300 Pa, M = 0.018 kg, and T = 295 K gives a mass flow rate of 2.49 kg m^{-2} s^{-1}, which is quite considerable and implies that a water film 0.1 mm thick would evaporate in less than one tenth of a second. Thus, although the intrinsic rate of evaporation from the surface, controlled by thermal energy and the breaking of bonds, is very rapid indeed, in this example a strong limitation on the rate at which water vapor can be removed is imposed by the available pumping speed. By contrast, in the context of thin film deposition, evaporating metallic copper, say, from a source at a temperature of 1550 K with vapor pressure 10^{-2} mbar (1 Pa) in a high vacuum of 10^{-6} mbar or less, the evaporation rates are close to those predicted by Equation 3.26.

The principal assumption in the argument above was that the water surface could be maintained at room temperature by the arrival of heat from below to supply the necessary latent heat of evaporation. A more realistic analysis is possible in the case of an actual vacuum-drying application, as depicted schematically in Figure 8.1(b), in which a wet product — for example, a water-bearing paste in a pharmaceutical application — is mounted on heated trays in the vacuum system and may also be heated by thermal radiation. The latent heat of vaporization of water is 2.5 MJ kg^{-1}. We will suppose that the energy input into the trays is 100 W = 100 J s^{-1} and that the pumping speed is 18 m^3 h^{-1} = 5 l s^{-1}. The rate at which water is forced to evaporate is $(100/2.5 \times 10^6) = 4 \times 10^{-5}$ kg s^{-1}. By Equation 5.4, this corresponds to a throughput $Q = (R_0T/M)\dot{W} = (8.3 \times 295/0.018) \times 4 \times 10^{-5} = 5.4$ Pa m^3 s^{-1} = 54 mbar l s^{-1}. From the fundamental pumping Equation 5.6, the pressure created by the pump will be $p = Q/S = 54/5 = 10.8$ mbar. The temperature at which water exerts this vapor pressure may be read off from the data of Figure 2.4 and is 281 K = 8°C. The steady condition achieved, in which the heat supplied is removed by evaporation, is such that the temperature of the evaporating surfaces falls from the initial room temperature to 8°C, at which the rate of heat supply and removal are in balance. This condition persists until the water is pumped away, and the time taken could be estimated if the mass to be pumped were specified. For a pump of twice the speed, 10 l s^{-1}, the halved pressure of 5.4 mbar would correspond to a further lowering of temperature to just below the freezing point of water and the surfaces would release vapor by sublimation. With a fixed energy input rate, pumping harder over the surface is associated with a lowered surface

temperature; the mass flow rate is determined by the power input, and greater volumetric pumping speed implies lower pressure and therefore reduced temperature. These are the conditions sought in vacuum freeze drying, which is carried out, depending on the sensitivity of the material being dried to temperature, at temperatures as low as –30°C, with appropriate provision of power input and pumping speed. The thermal design of vacuum drying arrangements to supply vaporization or sublimation energy at the desired temperature and rate is not trivial.

Returning to the first slightly artificial example of pumping the wet walls of a vacuum system, it may be noted that if water has to be pumped from walls that are wet, for some probably accidental reason, or from imperfectly dried internal components, the forced evaporation causes a fall in temperature because the rate of energy supply from the surroundings is usually limited, and the consequent lowering of vapor pressure considerably extends the time to pump it dry.

8.3 Pumping through a Long Pipe

Consider, as shown in Figure 8.2, a vessel of volume 0.1 m^3 = 100 l containing air at an atmospheric pressure of 1000 mbar connected to a small two-stage rotary pump of speed S^* = 1 l s^{-1} by a pipe 1 m long with internal diameter 1 cm. Suppose that it is to be pumped down to a pressure 0.1 mbar. If the pump were connected directly to the vessel, the time constant for pumping would be V/S^* = 100 s and the time per decade 230 s. Pumping through 4 decades down to 0.1 mbar would, therefore, according to the discussion of Section 5.7, take just over 15 min. Our purpose in this example is to determine the effect of inserting the connecting pipe.

With a pipe of diameter 1 cm, our expectation is that flow will be in the continuum regime for most of the pressure range. Equation 5.1 indicates that

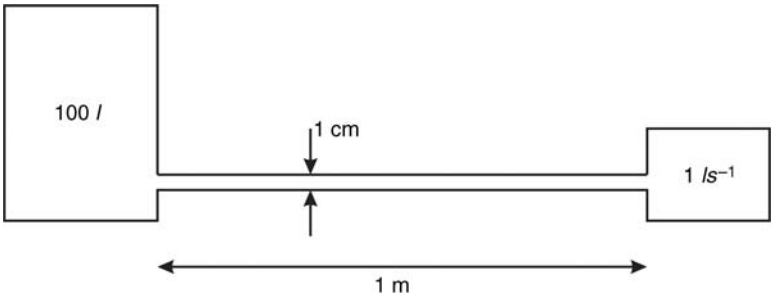

FIGURE 8.2
Pumping through a long pipe.

flow is continuum for pressures greater than 0.64 mbar. At the lower limit of 0.1 mbar, therefore, flow will have just entered the transitional state. The conductance of the pipe in the continuum regime, assuming that flow is viscous laminar, depends on the mean pressure and is given by Equation 5.34. Equation 5.35 expresses it in practical units as $C = 133\ (D^4/L)\ \bar{p}$ l s^{-1}, with L and D in cm and $\bar{p}$ in mbar. Writing p_U as the upstream pressure at the vessel and p_D the downstream pressure at the pump, $\bar{p} = (p_U + p_D)/2$. It will be useful for manipulative purposes to write $C = \beta\ \bar{p}$. In this case $\beta = 1.33$.

Consider the gas flow from the vessel to the pump. Following the methods introduced in Section 5.3, the throughput Q at any stage of the pumping process is

$$Q = Sp_U = S^*p_D = C(p_U - p_D) \tag{8.1}$$

where S is the pumping speed at the vessel. Substituting for C, we have

$$Q = Sp_U = S^*\,p_D = \frac{\beta}{2}\left(p_U^2 - p_D^2\right) \tag{8.2}$$

Examining this set of equalities, we see that, of the quantities involved, only S^* and β are known, and they are constant in time. The upstream and downstream pressures, known to be 1000 mbar at the start of pumping, will decrease as pumping proceeds, with $p_D < p_U$, in a way that has to be determined. Throughput $Q = S^*p_D$ will fall accordingly. The upstream pumping speed S and pressure p_U are determined to the extent that their product is Q.

The pressure dependence of C in this regime complicates the problem of determining the pumping speed at the vessel. From Equation 8.1, introducing a pressure ratio $K_p = p_U/p_D$,

$$\frac{S^*}{C} = \left(\frac{p_U}{p_D} - 1\right) = K_p - 1$$

and therefore

$$K_p = \frac{S^* + C}{C}$$

Because $S = S^*/K_p$, we get for the pumping speed at the vessel

$$S = \frac{S^*C}{S^* + C}$$

This is the result of Equation 5.11 obtained in a slightly different way. If C were constant and known, S would be directly determined, constant, and the general analysis presented in Section 5.7 could be followed to find how pressure fell with time.

However, because C is not constant, a different approach is necessary. We proceed as follows. From Equation 8.2

$$S^* p_D = \frac{\beta}{2} p_U^2 \left(1 - \frac{p_D^2}{p_U^2}\right) = \frac{\beta}{2} p_U^2 \left(1 - \frac{1}{K_p^2}\right)$$

and by rearrangement

$$\frac{2S^*}{\beta p_U} = \frac{K_p^2 - 1}{K_p}$$

Now S^* and β are known, but p_U is not. However, if we insert representative values of p_U and write $(2S^*/\beta p_U) = \alpha$ for convenience, we can solve the resulting quadratic equation

$$K_p^2 - aK_p - 1 = 0$$

and for each value of p_U determine K_p, and then the downstream pressure p_D and associated throughput $Q = S^* p_D$. With Q now known, the upstream pumping speed at the vessel can be determined for each of the chosen p_U values. Thus, solving the quadratic and restoring physical quantities gives

$$K_p = \frac{S^*}{\beta p_U} + \sqrt{1 + \left(\frac{S^*}{\beta p_U^2}\right)^2} \tag{8.3}$$

This analysis has been carried out for representative upstream pressure values in the range 1000 down to 0.1 mbar, and the results are listed in Table 8.1. Pressure ratios K_p are given, then downstream pressures, pressure differences, throughputs, and upstream pumping speeds. Also given are values of Q/D.

From the column of downstream pressures, one notices that at higher pressures they differ very little from the upstream pressures, reflecting that the conductance values are very high. For example, at an upstream pressure 100 mbar, the mean pressure has essentially that value and $C = 1.33\ \bar{p} = 133$ l s^{-1}, still higher than the pump speed of 1 l s^{-1} by a factor of more than 100. Alongside each p_D value is $\Delta p = (p_U - p_D)$ and $\Delta p/p_U$, expressed as a percentage. Their variation emphasizes that until lower pressures are reached, the pressure drop is only a small fraction of the upstream pressure. Throughputs fall as indicated and the derived upstream pumping speed at the vessel, on account of the high conductance, does not fall significantly below the speed

TABLE 8.1

Results of Flow Analysis in Pipe with $L = 1$ m, $D = 1$ cm, $S^* = 1$ l s^{-1}, and $\boldsymbol{\beta} = 1.33$

p_U mbar	$S^*/\beta p_U$ $= 3/4\ p_U$	K_p	$p_D = p_U/K_p$ mbar	$p_U - p_D$ Δp mbar	$\Delta p/p_U$ %	$Q = S^* p_D$ mbar l s^{-1}	$S = Q/p_U$ l s^{-1}	Q/D mbar mm
1000	0.00075	1.00075	999.3	0.7	0	999.3	1	100
300	0.0025	1.0025	299.3	0.7	0.2	299.3	1	30
100	0.0075	1.0075	99.3	0.7	0.7	99.3	0.99	9.9
10	0.075	1.078	9.3	0.7	7	9.3	0.93	0.93
1	0.75	2	0.5	0.5	50	0.5	0.5	0.05
0.1	7.5	15.1	0.007	0.093	93	0.007	0.07	0.007

of the pump until pressures of about 1 mbar are reached. At this pressure and below, decreasing conductance causes relatively larger pressure drops across the pipe. Conductance has become comparable with pump speed, and the lower parts of Figure 5.3 have been reached. Fortuitously, the numbers for $p = 1$ mbar enable an easy check on the self-consistency of the analysis. Thus, with $p_D = 0.5$ mbar, the mean pressure is 0.75 mbar and the conductance therefore $1.33 \times 0.75 = 1$ l s^{-1}. Whence $Q = 1 \times 0.5 = 0.5$ mbar l s^{-1}, in agreement with the tabulated value.

At $p_U = 0.1$ mbar, with associated downstream pressure 7×10^{-3} mbar, the conductance is $1.33 \times 0.053 = 0.07$ l s^{-1} and the pumping speed at the vessel is markedly reduced. Also, the rotary pump will be approaching the region where its speed starts to fall, and flow is becoming transitional.

The column with Q/D is included to test whether flow is turbulent or not according to the criterion of Equation 5.5, that is $Q/D > 24$ mbar mm for turbulence. Thus, for upstream pressures greater than approximately 300 mbar, turbulent flow is indicated and the analysis is inappropriate. The throughput will be rather less than indicated and the time to pump the first decade rather more than $2.3 \times V/S = 230$ seconds. Various authors, Livesey (1998) for example, give appropriate throughput expressions for turbulent flow that could be used to get a more accurate analysis for this region. It may be checked that for $p_U = 200$ mbar, conditions have become viscous laminar. For the next decade 100–10 mbar, the upstream speed is effectively equal to the pump speed and this is true for most of the next decade. The pumping time from 100 to 1 mbar could, therefore, be taken as the time for two decades, 460 seconds, with a true value a little higher than this. Below 1 mbar the analysis becomes increasingly approximate because flow becomes transitional, and close to the pump, increasingly so.

In summary, therefore, the analysis just presented shows that even for pumping through a long narrow pipe for which reasonable estimates of conductance are possible, care has to be exercised in interpretating the data. At pressures down to 1 mbar, notwithstanding the early turbulence, the conductance is so high that the pumping speed at the vessel is virtually the same as that of the pump, and the pumping time for the first three decades will be only slightly greater than 3×230 seconds. At the lowest pressures, conditions enter the transitional regime, gassing starts to become significant,

and the rate of fall of pressure slows. It must be expected, therefore, that the time needed to reduce the pressure from 1 to 0.1 mbar will be rather greater than that taken for the earlier decades.

8.4 Differential Pumping in an Electron-Optical Application

I am indebted to Malcolm Baker, formerly of BOC Edwards High Vacuum International, for the instructive example of an actual design shown in Figure 8.3. The differential pumping in this particular application differs from that in the general multi-pumped configuration described in Section 5.8 because it is achieved with a single pump and an appropriate arrangement of pumping pathways in the system.

The configuration is shown in Figure 8.3. It is a system for carrying out electron microscopy of biological materials and the sample chamber is at a relatively high pressure. An electron beam necessarily produced and conveyed through a region of high vacuum passes through a small hole H into the sample chamber, and so there is a flow of gas in a direction opposite to that of the electron beam, with a known throughput $Q = 0.002$ mbar l s^{-1}. Vacuum chambers 1 and 2 are separated by a small tube with conductance C_1, and pumped by connection to the two limbs of a symmetric T-shaped pipe arrangement that is connected to a pump of speed 100 l s^{-1}. The pipe above the pump is 63 mm in diameter and 30 cm long with conductance C_4. The two limbs to which it is connected are identical and also have a bore of

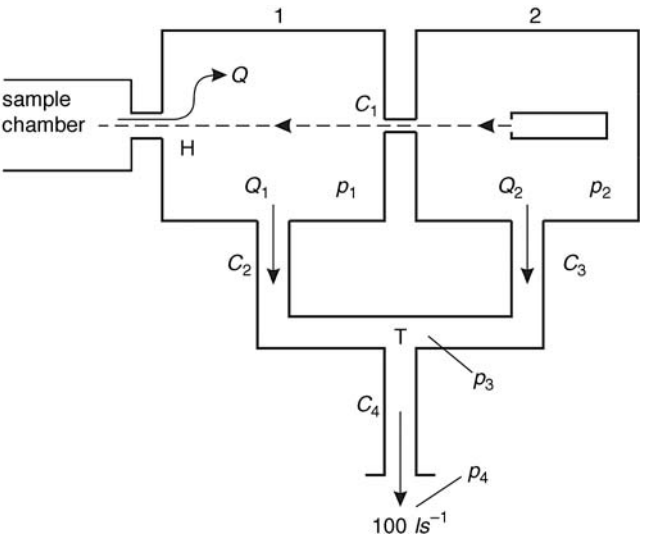

FIGURE 8.3
Differential pumping.

63 mm, but are 50 cm long, each with a right-angle elbow to connect them to chambers 1 and 2. These pipes with elbows have molecular flow conductances C_2 and C_3 as labeled, which are equal.

The arrangement is designed to allow the electron gun to operate at low pressure in chamber 2, while most of the gas entering chamber 1 from the sample chamber is pumped away through the left-hand limb. Outgassing loads from the wall are ignorable compared with the primary gas load from the sample chamber and conditions are molecular throughout. Pressures and pumping speeds p_1, S_1 and p_2, S_2 in chambers 1 and 2 are to be determined, as are pressures p_3 at the T junction and p_4 at the pump. The incoming throughput Q divides into throughputs Q_1 and Q_2 as shown, which recombine at the junction of the three pipes in the T at pressure p_3. In using p_3 in the Formula 5.7 ($Q = C\Delta p$), it should be borne in mind, as discussed in Section 5.6.6, that the conductance formulas are derived on the assumption that pipes terminate in volumes where pressures are clearly defined, as are the pressures in chambers 1 and 2, but this is not the case for p_3, so that a slight uncertainty is introduced into the calculations.

All the pipes involved are technically short, so the use of the straight pipe formula of Equation 5.44 may be used to calculate conductances C_1 to C_4. Inserting numerical values gives $C_1 = 5.6 \times 10^{-3}$ l s^{-1}, $C_4 = 81$ l s^{-1}, and $C_2 = C_3 = 53$ l s^{-1}, to two significant figures. In calculating the conductance for pipes that contain a right-angled elbow, the effective length is sometimes increased by adding a length equal to the diameter to the actual length to allow approximately for the effect of the bend. In this case the conductances C_2 and C_3 would then be reduced to 47 l s^{-1}. However, this procedure overcompensates for the effect of the bend. A comparison of transmission probabilities of a pipe with a right-angled bend and a pipe of the same center length, using graphical data of Figure 5.11, shows that the molecular flow conductance of the pipe with the elbow is only about 4% less than that of the straight pipe. This is readily understood when the randomizing nature of the molecule–surface collisions that determine flow in this regime is considered. Accordingly, it seems appropriate to reduce conductances C_2 and C_3 by 4% from 53 to just 51 l s^{-1}. A right-angled bend is a less serious obstruction in molecular flow than in viscous flow.

The analysis proceeds as follows, based on the throughput and conductance Equation 5.6 and Equation 5.7, $Q = S \times p$ and $Q = C(p_u - p_d)$, respectively. The gas from the sample chamber with throughput Q at the input divides and then recombines, reaching the pump of speed $S = 100$ l s^{-1} where the pressure is p_4. Because the throughput here must be the same as that at the input, $p_4 = Q/S = 0.002/100 = 2 \times 10^{-5}$ mbar. Because Q, C_4, and now p_4 are known, the conductance equation can be used to determine p_3. Thus, $p_3 - p_4 = Q/C = 0.002/81 = 2.47 \times 10^{-5}$ mbar so that $p_3 = 4.5 \times 10^{-5}$ mbar. Conductance C_1 is four orders of magnitude smaller than C_2 and correspondingly the throughput Q_1 down the left-hand limb is much greater than Q_2 through C_1. Therefore, we can take $Q_1 = Q = 0.002$ mbar l s^{-1} and use the conductance equation to determine p_1. Thus $p_1 - p_3 = Q_1/C_2 = 0.002/51 = 3.9 \times 10^{-5}$ mbar, whence $p_1 = 8.4 \times 10^{-5}$ mbar.

This is the pressure in vacuum chamber 1, an adequate value to allow the passage of the electron beam without undue scattering.

Conductances C_1 and C_3, through which Q_2 flows, are separated by the large volume in which the pressure is p_2, and therefore their combined conductance is accurately given by the series formula for just two conductances, in Equation 5.10. This gives $C = (C_1 \times C_3)/(C_1 + C_3) = 5.6 \times 10^{-3}$ l s^{-1} showing that the small conductance C_1 of the short narrow tube totally dominates the combination. Thus $Q_2 = C(p_1 - p_3) = 5.6 \times 10^{-3}$ $(8.4 \times 10^{-5}$ $4.5 \times 10^{-5}) = 2.2 \times 10^{-7}$ mbar l s^{-1}. Because Q_2 is so small, the pressure p_2 in chamber 2 will be essentially the same as p_3, i.e., $p_2 = 4.5 \times 10^{-5}$ mbar. Thus, finally, the pumping speed in chamber 2, the volumetric flow rate of gas at the prevailing pressure, is $S_2 = Q_2/p_2 = 4.9 \times 10^{-3}$ l s^{-1}, which is very small, as expected.

In summary, pressures and flows are determined using the two basic equations of pumping. The pressure drop from chamber 1 to chamber 2 is the same as that from chamber 1 to the T junction, with most of the gas flowing by this latter low impedance, high conductance path, and a very small flow through the high impedance path that connects the two chambers.

8.5 High Flow Rates at Low and Medium Vacuum with Roots and Rotary Pumps in Combination

There frequently arises, usually in industrial processing applications, the need to pump gas at a pressure in the approximate range 10^{-3} to 10 mbar at high volumetric flow rates of the order of 100 m^3 h^{-1}. This corresponds to about 28 l s^{-1}, and is a substantial flow rate. In this context, mass flow rates are usually specified, either in kg h^{-1} or standard liters per minute (SLM). With pumping speeds expressed in m^3 h^{-1}, a natural unit of throughput is then the mbar m^3 h^{-1}. Two examples illustrate the general orders of magnitude encountered. First, consider a mass flow rate of 1 SLM at 22°C (295 K) at a pressure of 0.2 mbar. One standard liter SL at 22°C occupies (295/273) = 1.08 l, and in pV measure the quantity of gas is 1013 × 1.08 = 1095 mbar l, or 1.095 mbar m^3. One SLM is, therefore, a throughput of 1.095 mbar m^3 per minute and, by extension, 60 × 1.095 = 65.7 mbar m^3 h^{-1}. The volumetric flow rate (pumping speed) required at a pressure 0.2 mbar is, therefore, $S = Q/p = 65.7/0.2 = 328$ m^3 h^{-1}. Secondly, consider a mass flow of 5 kg h^{-1} of oxygen at a pressure of 10 mbar for which the volumetric flow rate has to be determined. By Equation 5.4 — $Q = (R_0T/M)W$ — we get $Q = [(8.314 \times 295)/0.032] \times 5 = 3.83 \times 10^5$ Pa m^3 h^{-1} = 3.83×10^3 mbar m^3 h^{-1}. In this case the required pumping speed is $Q/p = 383$ m^3 h^{-1}.

For this pressure range, pumping speeds of the above order and higher are usually achieved by the combination of a Roots pump and a rotary pump, as described in Section 6.2.2. Both the above examples call for a pumping speed of about 400 m^3 h^{-1}. Manufacturers specify the performance of the

TABLE 8.2

Analysis of a Roots/Rotary Pump Combination

p_B mbar	100	10	4	2	1	4×10^{-1}	10^{-1}	10^{-2}	10^{-3}
K_0	7.7	26	38	42	36	26	14	7.4	7
S_B m^3 h^{-1}	74	74	74	73	72	62	50	28	6
$S_D/K_0S_B = y$	0.88	0.26	0.18	0.14	0.19	0.31	0.71	2.4	12
$1/(1+y) = S/S_D$	0.53	0.79	0.85	0.88	0.84	0.64	0.58	0.31	0.08
S m^3 h^{-1}	265	395	425	440	420	320	290	150	38
p_i mbar	30	2	7×10^{-1}	3×10^{-1}	2×10^{-1}	8×10^{-2}	2×10^{-2}	2×10^{-3}	2×10^{-4}

combination by data that include a plot of the type shown in Figure 6.11. It is instructive to assemble the data for this plot using the analysis presented in Section 6.2.2. The recommendation for these applications would be to have a Roots pump with theoretical displacement S_D = 500 m^3 h^{-1}, backed by a rotary pump of speed about 75 m^3 h^{-1}, as analysis will show. Referring to Section 6.2.2, the zero flow compression ratios K_0 for the Roots pump are taken from Figure 6.9, and two equations are to be exploited: Equation 6.3, which expresses the speed S in terms of S_D, K_0, and the speed of the backing pump S_B at various backing pump pressures p_B, and the equation $Sp_i = S_Bp_B$ that enables the inlet pressure p_i to be determined once S is known.

The ingredients and results of calculation are shown in Table 8.2. The top row lists selected pressures across the range of interest, the second the corresponding K_0 values for the Roots pump, and the third the backing pump speeds S_B at those pressures. While K_0 peaks at 2 mbar in the middle of the range for the reasons discussed, S_B decreases progressively below about 1 mbar. In the fourth row, the quantity $S_D/K_0S_B = y$, say, is plotted, and next $1/(1 + y) = S/S_D$, which is essentially an efficiency expressing the ratio of the achieved speed to the theoretical displacement. Thus, in the next row, S is computed, and last, the corresponding inlet pressure.

From the data in the last two rows, a plot such as Figure 6.11 can be made of speed vs. inlet pressure for the combination. For the pressure range of 10 down to less than 1 mbar, the pumping speed available is 400 m^3 h^{-1} or more, and the requirements of the two flow examples above would be met by this combination. Notice that while the backing pump with inlet pressure p_B exhausts to atmospheric pressure with relatively high compression, the pressure drop across the Roots pump is quite small. The value of the so-called staging ratio, S_D/S_B = 500/75 = 6.6, is fairly typical.

8.6 System for Vacuum Coating at High Vacuum, 10^{-6} mbar

The system described is typical of many small vacuum coating plants to be found in university or industrial laboratories. As depicted schematically in Figure 8.4, it consists of a cylindrical stainless steel vessel of diameter

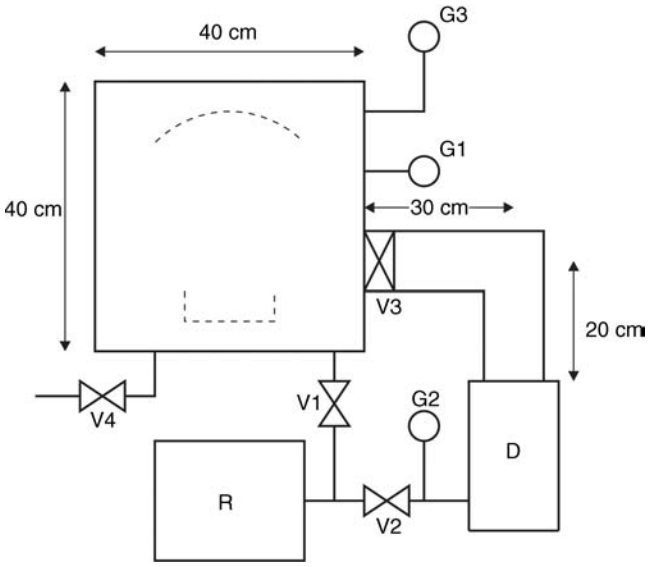

FIGURE 8.4
System for vacuum coating at 10^6 mbar.

40 cm and height 40 cm, with internal fixturing (shown dotted) that consists principally of evaporation sources on the base plate and the items to be coated supported from the lid of the vessel, elastomer sealed, which is raised to allow access and reloading. An alternative, and slightly more convenient, arrangement is to have a cubical vessel with front-door access. In either case the vessel would be equipped with ports for the attachment of monitoring devices and, especially, windows to allow inspection of the interior, with associated shutters to prevent their being coated with evaporant. Gauging, G1, G2, and G3, to be discussed later, is provided, and valves V1 to V4. The use of the base plate to support internal fixtures means that pumping access is via the side wall.

Operation at high vacuum requires primary and secondary pumps, as discussed in Section 6.1. In this example, the secondary pump is a vapor jet (diffusion) pump, still widely used for high vacuum, and the primary pump is a rotary pump. Matters of concern are to hold an adequate vacuum under the process load that occurs when the evaporant sources are operating; the ultimate pressure of the system; and how long it takes to achieve a condition in which the process can be operated. Let us consider the specification of system and the choice of component values. The analysis is inexact, principally because outgassing rates cannot be accurately predicted. In these circumstances, calculations using approximations with the occasional incorporation of safety factors are frequently adequate.

The volume of the cylindrical vessel $v = (\pi D^2/4)h$ is quickly calculated to be 0.05 m^3 = 50 l, to two significant figures, with internal area $(2\pi D^2/4 + \pi Dh)$ = 0.75 m^2, which we increase to 0.8 m^2 = 8000 cm^2 to allow for the surface

area of internal fixtures. With the constraint of side exit from the vessel to the vacuum pump, a pipe with a right-angled elbow is necessary to connect to the diffusion pump that has to be vertical, and we will suppose that the center length of the pipe is 50 cm, 30 horizontally and 20 vertically. We will assume that the system is to be cycled regularly up to atmosphere and down to vacuum, so that the expense of a high-conductance gate valve V3 to isolate the vessel from the diffusion pump is justified. Had it been possible to connect the diffusion pump directly below the vessel to the base plate, a type of pump with an integrated butterfly swing valve could have been chosen to fulfill this function. With the remote pump, V3 is correctly located to minimize the vessel volume for primary pumping. We will specify that a high vacuum with a working pressure between 1 and 3×10^{-6} mbar be achieved in a time of less than 2 h, and that at this stage the process be switched on and the system pressure not rise above 2×10^{-5} mbar when handling a typical process load Q_P. This requirement would be set not by mean free path considerations for the evaporant trajectory from source to target, but because of the undesirable incorporation of residual gas molecules into the growing deposit, which can be calculated with a worst-case assumption using Equation 3.22. Operating procedures will be described later.

To pump down the 50 l volume from atmospheric pressure with a primary pump, we will initially assume a rotary pump with a speed of 2 l s^{-1} at the vessel. Thus the time constant of primary pumping would be $v/S = 25$ s and the time per decade $2.3 \times 25 = 57.5$ s, essentially 1 min. The time to reach 10^{-1} mbar or a little less, at which the secondary diffusion pump could take over, is therefore less than 5 min, which seems quite adequate at this stage of the design. Considering high vacuum pumping, we will take the value of the 1-h specific gassing rate (Section 4.5) to be 5×10^{-8} mbar l s^{-1}/cm^2 of surface, so that the 1-h gassing load from the walls and fixtures will be $Q_G = 8000 \times 5 \times 10^{-8} = 4 \times 10^{-4}$ mbar l s^{-1}.

A pumping speed of 200 l s^{-1} at the vessel, using Equation 5.6, is predicted to give a pressure $p = Q/S = 4 \times 10^{-4}/200 = 2 \times 10^{-6}$ mbar after one hour, and if a $1/t$ outgassing law (Equation 4.10, Section 4.5) is assumed, 1×10^{-6} mbar after two hours, and 2×10^{-7} mbar after ten hours, though deviations from this relationship will be such as to give a slower rate of fall of gassing and hence pressure. Times longer than this extend beyond the relevance of the design, though the pressure would not be likely to fall much further given the trend of the outgassing and the likelihood that the pump, with normal oils and no special trapping, would be approaching its ultimate pressure. With this 200 l s^{-1} nominal pumping speed, therefore, the pressure and time requirements are satisfied, and the additional process gas load that would raise the pressure to 2×10^{-5} mbar from the 1×10^{-6} mbar reached after two hours is $Q_P = S \times (20–1) \times 10^{-6} = 200 \times 19 \times 10^{-6} = 2 \times 10^{-3}$ mbar l s^{-1}.

A speed of 200 l s^{-1} has to be accomplished using a remote pump and, therefore, the conductance of the connecting pipe has to be considered. Manufacturers' catalogs show that pumps with a nominal 100-mm (4-in.) inlet diameter typically have speeds S^* for nitrogen of about 300 l s^{-1}, which

is greater than we need. However, a rough calculation of the conductance of a matching 100-mm diameter pipe using the formula of Equation 5.44 ($C = 12.4D^3/L[1 + 4D/3L]$) yields $C = 196$ l s^{-1}, whence by Equation 5.11 the speed S at the vessel will be $S^*C/(S^* + C) = (300 \times 196)/(300 + 196) = 118$ l s^{-1}, which is inadequate. The next greater size of pump has a nominal 150-mm (6-in.) inlet and typical pumping speed 650 l s^{-1}. A matching pipe with length 50 cm and diameter 150 mm has a conductance 598 l s^{-1}. The gate valve specified earlier is taken to be part of the pipe because, when open, it has a very high conductance and ideally is just a continuation of pipe in which it is located, as separate calculations allowing for its exit conductance would verify. As noted earlier in Section 5.6.4, Formula 5.44 will overestimate the conductance, particularly at this L/D ratio of $50/15 = 3.33$, and so it is legitimate to correct it. Using the aperture conductance formula of Equation 5.40 with $D = 15$ cm gives $C_A = 2090$ l s^{-1}, and interpolating the transmission probability data gives $\alpha = 0.26$ for $L/D = 3.33$, whence for the pipe $C = 543$ l s^{-1}. In addition, the pipe contains an elbow, so it is legitimate to reduce the conductance further by 4%, as discussed in Section 8.4, to get a value 522 l s^{-1}. Using Equation 5.11, the pumping speed at the vessel is then $(650 \times 522)/(650 + 522) = 290$ l s^{-1}.

This choice of pump, therefore, gives a speed at the vessel with a good margin of safety. With the outgassing conditions assumed, the pressure after one hour of pumping would be estimated as $4 \times 10^{-4}/290 = 1.4 \times 10^{-6}$ mbar, and the process load that could be handled, $Q_P = S \times (20\text{-}1.4) \times 10^{-6} = 5.4 \times 10^{-3}$ mbar l s^{-1}.

Returning to the matter of the size of the rotary pump, provisionally taken as 2 l s^{-1} in order to estimate a time for rough pumping, it is important to note that this is not the criterion for selecting its size. What is critical is the condition at crossover, when the gate valve is opened to allow the diffusion pump to take over the pumping and when the diffusion pump has its highest load. At crossover the pumping pathway is changed from the roughing line to pass through the diffusion pump and backing line (foreline) to move to the rotary pump. As Figure 6.31(b) shows, the characteristic pumping curves of the diffusion and rotary pumps necessarily overlap. For the diffusion pump, the falling pumping speed in its roughly constant throughput (overload) region is such that the speed is a fraction of its constant, maximum high-vacuum value at about 10^{-2} mbar, and at this pressure the speed of a two-stage rotary pump may have fallen a little below its rated value at higher pressure. Manufacturers frequently specify a maximum throughput Q_{max} for a given diffusion pump, and its value at any pressure can be computed directly from the speed–pressure curve as the product of the two, shown as the dashed line in Figure 6.31(a). When the diffusion pump throughput is Q_{max}, the speed of the rotary pump must be sufficient to keep the backing-line pressure below its critical value, typically between 0.5 and 1 mbar. Because the throughput at the diffusion pump is equal to that in the backing line we have, when it is a maximum, $Q_{max} = S_B p_B$, where S_B and p_B are the pumping speed and pressure at the diffusion pump exit

into the backing line. In this example the values of Q_{max} and p_B may be taken as 1.3 mbar l s^{-1} and 0.8 mbar, indicating that S_B should be not less than $1.3/0.8 = 1.6$ l s^{-1}. Rotary pumps of this speed have matching connecting pipes of nominal 25 mm (1-in.) bore, and it is necessary to know if the connection will involve appreciable loss of pumping speed at the lower pressures. Following the methods of Section 8.3, one may estimate that for a pipe of 25 mm bore taken to be 50 cm long, for an upstream pressure of 5×10^{-2} mbar, the upstream speed S_B in this context with an assumed pump of speed $S^* = 2$ l s^{-1} is reduced by 25% to 1.5 l s^{-1}. However, this relatively small reduction is also allied with the reduced intrinsic speed of the pump as this pressure region is entered, and therefore it is prudent to increase the specification of the rotary pump speed by about 50%. This would raise the pump speed requirement to about 3 l s^{-1} or 10.8 m^3 h^{-1}, which is consistent with the manufacturer's recommendation. The pump-down time for roughing would be correspondingly shortened, with time-constant v/S now 17 seconds. Interestingly, when the system is in its designed high-vacuum condition, the throughput $Q_P = 5.4 \times 10^{-3}$ mbar l s^{-1} is quite small compared with $Q_{max} = 1.3$ mbar l s^{-1}.

The change from primary to secondary pumping at crossover should be accomplished as rapidly as possible. This critical transition is extensively discussed by Hablanian (1997). In Figure 6.31(b), examination of the region where the two pump characteristics cross shows that if pumping is transferred to the diffusion pump at too high a pressure, significantly above that of the crossing point C, there is a sudden reduction in pumping speed with the diffusion pump forced to operate in a region of very low speed and therefore an associated fall in pump-down rate and possible temporary increase of vessel pressure. On the other hand, if the rotary pump is allowed the time to pump the vessel to a pressure such that, on changeover, the diffusion pump is working close to its maximum speed in region D, the roughing line pressure will have fallen to a value of the order of 10^{-3} mbar that encourages back-migration of rotary pump oil vapors into the vessel because of the onset of molecular flow conditions. Best practice is, therefore, to aim to changeover at conditions corresponding to point C, generally at about 0.1 mbar or a little less, when the adjustment to the changed pumping is accomplished in a few seconds and full speed of the diffusion pump quickly reached.

Turning to operational matters in detail for such a system and referring to Figure 8.4, one may identify four valves: the roughing line valve V1 in the line that connects the vessel directly to the rotary pump, the backing (or foreline) valve V2 between the rotary and diffusion pumps, the high vacuum valve V3 between the vessel and diffusion pump inlet, and a venting valve V4 to let air back into the vessel. An ionization gauge G3 of either the Penning or thermionic type will measure high vacuum in the vessel, and gauges G2 and G1, frequently Pirani gauges, will measure the backing line and vessel pressures, respectively. There are three distinct activities involved in using the system: starting it from cold, operating it cyclically, and shutting it down.

Starting the system from cold, with pumps off, all valves closed, and the vessel at atmospheric pressure, one would switch on the rotary pump, open V2 to evacuate the diffusion pump, switch on its heater, and turn on its cooling water. After 20 to 30 minutes, the diffusion pump will be operational. After about 20 minutes it is permissible to close V2 and open V1 so that the rotary pump exhausts the vessel, as monitored on gauge G1. While this is happening, the diffusion pump exhausts into that part of the backing line upstream of valve V2, a volume that is usually adequate to prevent the pressure rising above the critical backing pressure, as monitored on G2, but if it does rise too much, V1 can be temporarily closed and V2 reopened to restore the situation. When the vessel is pumped down to 10^{-1} mbar or a little less, V1 is closed, V2 opened and after a second or two to ensure a low backing line pressure, the high-vacuum valve V3 is slowly opened to create the pumping pathway through the diffusion pump, backing line and rotary pump, and eventual exhaust. The slow opening of V3 prevents a large impulse of gas that would cause undesirable backstreaming of diffusion pump oil vapor. As it is opened, the backing line pressure rises briefly as the pump takes the initial maximum load, and the vessel pressure rapidly falls. The reading of G1 falls below the bottom of its scale and the ionization gauge may soon be switched on to watch the progress of the vessel to high vacuum. The vessel pressure falls rapidly in response to the increase of pumping speed, from just a few liters per second to of the order 100. The load being dealt with is dominantly the outgassing load, and systems of this sort typically reach middle 10^{-5} mbar pressures in about ten minutes.

When the system is in regular cyclical use — that is, the vessel being taken from atmospheric pressure down to high vacuum, the high-vacuum process carried out, and being brought up to atmosphere again for venting — the procedure for taking the vessel from high vacuum to atmosphere is as follows: Gauge G3 is switched off and the high-vacuum valve V3 closed. V1 is already closed, and so V4 may be opened to vent the vessel to atmosphere for extraction of the product of the process and reloading. Getting high vacuum again in the vessel then involves closing V4, closing V2 to temporarily isolate the diffusion pump, opening V1, and exhausting the vessel down to 10^{-1} mbar or a little less, at which point V1 can be closed, and V2 and V3 opened in sequence as before, G3 switched on, and the vessel monitored to high vacuum. With repeated cycling, careful use, and venting procedures that minimize the exposure to atmosphere, outgassing rates are likely to fall somewhat so that pump-down times improve.

Shutting down the system and leaving it under vacuum involves closing V3, switching off the power to the diffusion pump (though not yet the cooling water!), waiting for about 20 min as the pump cools, then closing V2 (V1 is already closed), and finally switching off the rotary pump. The rotary pump will auto-vent itself to leave the vessel, diffusion pump, and lines under vacuum, and thus promote rapid restarting when the system is next used.

8.7 High and Ultrahigh Vacuum Pumping with Turbomolecular Pumps

As earlier noted, turbomolecular pumps have substantially displaced diffusion pumps for many high-vacuum applications. One reason for this, in spite of their greater expense, is the absence of backstreaming hydrocarbon vapors to the vessel when they are operated correctly. The potential source of these vapors is the oil or grease used in the rotor bearings and the primary pump. In the example of a diffusion-pumped high-vacuum system described in the previous section and subject to the side entry constraint that necessitates the use of a connecting pipe, a turbomolecular pump of the same inlet flange size, nominal 150 mm, would have comparable pumping speed and do the same job, though operating procedures would differ a little and the criteria for backing pump selection would be different, as was discussed in Section 6.3.2. So as to exemplify these matters, let us consider a vessel of the same size as that discussed in the previous section, with volume 50 l, surface area 0.8 m^2, and similar assumed outgassing properties, but suppose that it can be pumped directly by a turbomolecular pump because it can be attached via a matching high-conductance gate valve V3 at the base plate. This arrangement is shown in Figure 8.5 and includes valves V1, V2, and V4 as before so that the vessel can be cycled regularly. (In an alternative configuration of a turbo-pumped system where frequent cycling is not a requirement, and in very small systems, a roughing line and associated valve are dispensed with because the vessel can be roughed through the turbo pump and roughing can be substantially completed by the time the turbo pump is up to speed.)

As in the previous example, a gassing load $Q_G = 4 \times 10^{-4}$ mbar l s^{-1} is estimated after one hour, at which time a pumping speed of 200 l s^{-1} at the vessel is predicted to give a pressure $p = Q/S = 4 \times 10^{-4}/200 = 2 \times 10^{-6}$ mbar, and 2×10^{-7} mbar after ten hours, assuming a $1/t$ outgassing law. As noted in Section 6.3.2, pumping speeds for a nominal inlet diameter of 100 mm are about 250 l s^{-1} for nitrogen, so that taking account of the conductance of the gate valve, typically 1700 l s^{-1} for this diameter, which includes an exit correction, the speed at the vessel will be $(250 \times 1700)/(250 + 1700) = 218$ l s^{-1}, slightly exceeding the requirement. With no interconnecting piping, therefore, the necessary speed is achieved with a smaller secondary pump.

According to the discussion of Section 6.3.2, the primary backing pump size for a staging ratio of 50 will be 5 l s^{-1} (18 m^3 h^{-1}), rather greater than that of the diffusion pumped example and a value that will promote fast roughing. In general, for gas loads containing a substantial proportion of hydrogen, and depending on hydrogen compression ratios of the particular turbo pump, staging ratios of 20 or less may be indicated, leading to the specification of a primary pump of proportionately higher speed. These matters are discussed in detail by O'Hanlon (2003).

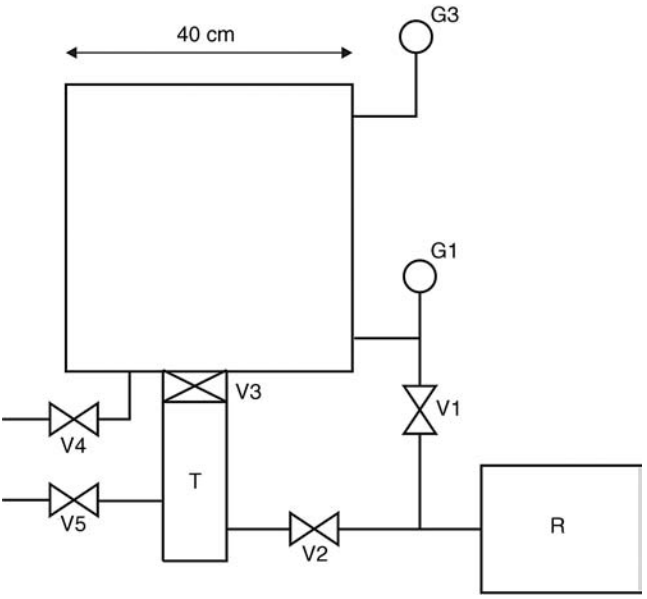

FIGURE 8.5
Turbo-pumped high-vacuum system.

In the operation of turbo pumped systems configured as in Figure 8.5, starting up from a situation in which the system has been left in a fully vented state (a desirable state of affairs because the migration of hydrocarbon vapors is inhibited) and with all valves closed, V2 would be opened, and the rotary pump and the turbo drive unit switched on so that the turbo blades start to spin. As with the diffusion pump, V2 may soon be closed and V1 opened to rough out the vessel. Roughing needs to proceed through three decades down to about 1 mbar at which point, if the turbo has reached close to maximum rotor speed, V1 can be closed, V2 reopened, and then gate valve V3 opened to cross over from the roughing to turbo pumping pathway. Because roughing is carried out with an oil-sealed pump, it is important not to let the roughing line pressure fall below 1 mbar in order to avoid oil vapor migration to the vessel. If the vessel pressure has fallen to this value before the turbo pump is up to speed, V1 is closed and one waits.

In cyclic operation, say, at approximately 2-hour intervals, after the vessel has reached a working pressure of about 10^{-6} mbar and processing is complete, gauge G3 can be switched off, valve V3 closed, and the vessel vented with appropriate safety arrangements, via V4, preferably to dry nitrogen to help reduce water vapor adsorption and subsequent pump-down time, while the turbo continues to be pumped through V2, switched back to idling speed. With product extraction and reloading complete, V2 is closed to temporarily isolate the turbo while V1 is opened to rough down again to 1 mbar, then V1 can be closed, the pump raised to full speed, and V2 and V3 opened as before. Subsequently, the initial rate of fall in pressure is rapid

because of the much-increased pumping speed. The volume gas still left at 1 mbar is quickly pumped away, and as pressure falls further, the outgassing load comes to dominate the pumping, gradually falling with a corresponding fall in pressure until the operating condition of about 10^{-6} mbar is reached. If allowed to fall further in extended time — overnight, for example — a lower ultimate pressure of about 10^{-7} mbar may be reached in which the outgassing load, principally water vapor, is essentially constant and in balance with the pumping speed. Left for even longer, the pressure in such a system might fall to a mid-10^{-8} mbar level, determined by a balance between pumping speed and an essentially constant gassing rate, in which the outgassing from elastomer seals had assumed more importance.

In closing down the system, V3 and V1 would be closed and the turbo rotor drive switched off so that rotor deceleration commenced. V2 would be closed so that the rotary pump could be switched off and auto-vented. After about ten minutes, and before the rotor reached half speed, the turbo pump would be vented to dry nitrogen at atmospheric pressure via valve V5 at a slow rate, occupying a minute or so, thus bringing the rotors to rest and preventing any oil or grease vapors reaching the upper area of the pump from where they might subsequently, under vacuum, migrate upwards towards the vessel. Venting at full speed is inadvisable because of the initial shock and forces on the bearings. In pumps with magnetic bearings, the magnetic levitation should not be switched off until the rotors have to come to rest. In unvalved turbo-pumped systems, many of the operational concerns for starting and stopping are similar. In modern systems, the operation of gauges, pumps, and valves in the correct sequence and at appropriate times may be carried out by an automatic control system.

For a turbomolecular pump that includes a drag stage, the outlet pressure is raised from about 0.1 to of the order of 10 mbar, so that a dry backing pump may be used. The combination of a turbo-drag pump with magnetic bearings backed by a dry diaphragm pump, for example, secures complete freedom from hydrocarbon contamination. The appropriate dry pump for a given application may be considerably smaller than the corresponding traditional pump because the same throughput is handled at higher pressure with a proportionately smaller volumetric pumping speed.

Turbomolecular pumps may be used for ultrahigh vacuum. Figure 8.6 shows a simplified schematic diagram of an ultrahigh vacuum evaporator used by the author and his colleagues for the growth of multilayered magnetic thin films. The stainless steel vessel is basically cylindrical in form, 40 cm in diameter and 40 cm high, with numerous ports entering it from the side and the lid to allow sample manipulation and deposition, and product monitoring by electron, ion, and magneto-optical probes, as well as gauging and residual gas analysis. A turbo-pumped airlock allows sample exchange without the necessity for venting of the main vessel. All seals are of the copper Conflat type except that of the hoistable lid, which is a well-outgassed viton ring of nominal 400 mm diameter. The vessel is supported on a lower cylinder that houses a titanium sublimation pump and furnishes

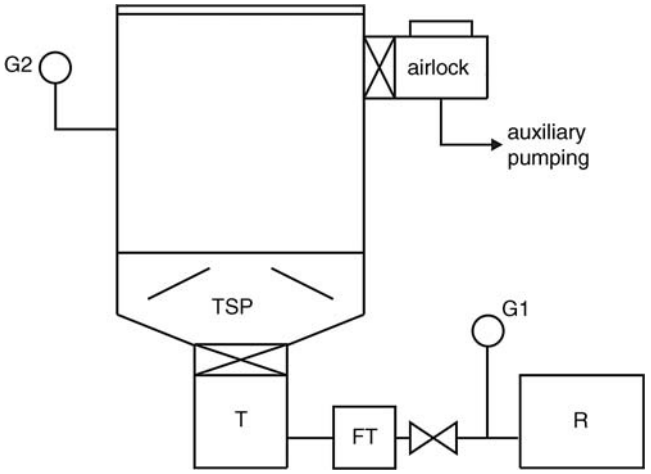

FIGURE 8.6
Ultrahigh vacuum with turbomolecular pumping.

a high-conductance pathway with reducing diameter to a 150-mm flange where a matching gate valve to a 300 l s^{-1} turbo pump with oiled bearings is attached, backed by a rotary pump with foreline trap. The total system volume is approximately 85 l. In normal operation, the system is recharged with evaporant materials at approximately monthly intervals, so initial rough pumping is through the turbo rather than a separate roughing line. The whole system above the turbo at its base is baked at 180°C by a combination of heater jackets and radiation in a detachable upper oven. After baking for periods of 24 to 36 hours and cooling back to room temperature, the system reaches a base pressure of about 2×10^{-9} mbar that is further lowered to $3–4 \times 10^{-10}$ mbar by the intermittent use of the sublimation pump as operation of gas generating processes demands. Both airlock and main chamber venting is to dry nitrogen.

The total area of internal surfaces, which includes coils, numerous shields, and a number of devices with considerable surface area, is approximately 1.2 m^2. If the servicing period with vessel vented and lid raised is prolonged, the advantage of venting to dry nitrogen is lost as atmospheric air permeates the system. In these circumstances the subsequent pump-down to high vacuum of about 2×10^{-6} mbar, dominantly water vapor, takes about ten hours, so that with 300 l s^{-1} pumping speed, the mean overall specific gassing rate is calculated to be $Q/A = Sp/A = 300 \times 2 \times 10^{-6}/1.2 \times 10^4 = 5 \times 10^{-8}$ mbar l s^{-1}, which is relatively large. The subsequent bakeout to reach the lower 10^{-9} mbar pressures indicates that the gassing rate is reduced by three orders of magnitude. Hydrogen is then the dominant residual gas, together with some carbon monoxide and very little water vapor.

8.8 Ion-Pumped Ultrahigh-Vacuum System

Figure 8.7 shows schematically a small ion-pumped system, essentially a UHV work station, with a stainless steel vessel designed to allow various UHV investigations — for example, the outgassing of stepper motors. As such, the vessel has numerous ports that allow for residual gas analysis, pressure measurement by both inverted magnetron and BA gauges above and below a variable conductance, which is manipulated by feedthroughs for mechanical movement, and others for electrical power input and signal output. The vessel, built on modular collars, typically has a volume of about 20 l and has copper Conflat seals throughout. It sits directly on top of the 150-mm bore inlet flange of a 400 l s^{-1} ion pump. No gate valve is necessary because investigations are of relatively long duration.

The vessel plus ion pump are roughed through a line that contains an all-metal isolation valve V1 close to the vessel and leads to a manifold that connects to a cryosorption pump C isolated by valve V2 and a relatively large single stage rotary pump R isolated by V3. A Pirani gauge P monitors roughing pressure and valve V4 allows venting to dry nitrogen. Bakeout ovens, shown by the dotted outline, allow heating of both vessel and pump and the line up to and including V1. Vacua of less than 10^{-10} mbar are obtained.

The starting procedure from atmospheric pressure with all valves closed and the sorption pump chilled is to open V1, switch on the rotary pump, open V3, and pump down the line, vessel, and ion pump volume just as far as 1 mbar indicated on P, and then close V3. This takes only a short time but

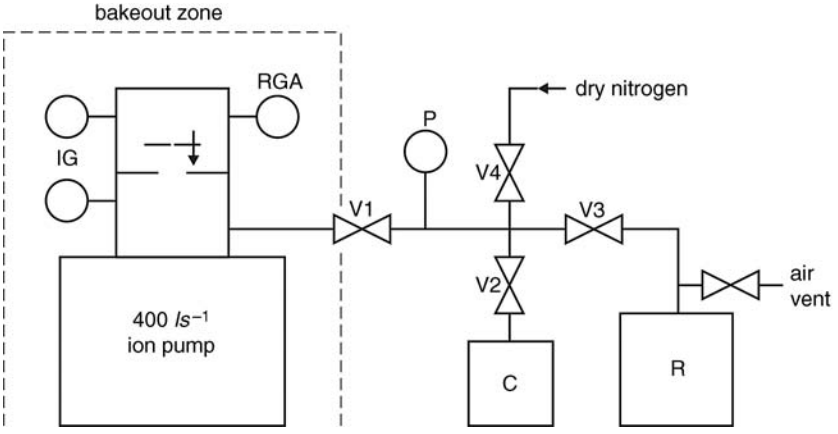

FIGURE 8.7
Ion-pumped UHV system.

removes a large volume of system gas so that the single sorption pump will have only a small load. Stopping at 1 mbar ensures that no hydrocarbon vapors from the rotary pump enter the system. V2 is then opened and the sorption pump reduces the pressure to 10^{-3} mbar or a little less in about 20 minutes, at which time the ion pump is switched on. Initial loading means that the ion pump takes a little while, perhaps a half hour, to reach full voltage, but as soon as one is sure that early instabilities are past, valve V1 can be closed to isolate the main system and V2 closed. Leaving a vacuum in the manifold helps to prevent oxidation of the low vacuum side of valve V1 when it is subsequently part of the bakeout zone.

The ion pump takes the system pressure to about 10^{-6} mbar in a few hours, at which stage the bakeout procedures are implemented. Heating the system for about 15 h or more at 220–250°C achieves the necessary outgassing. The temperature is regulated initially by the pressure, that rises in response to the early outgassing to a set point of 5×10^{-6} mbar, at which value heating is temporarily discontinued until pressure falls again below the set point. Eventually, the constant set temperature is reached and held constant, now by thermostatic control of heating from a suitably placed temperature sensor, and outgassing proceeds. With thorough outgassing, the pressure at constant temperature begins to diminish as gas sources become depleted. When heating is discontinued, the system cools, assisted by motorized fans, and the final outgassed state of the system is reached. It is prudent to degas gauges at the end of the bakeout cycle when the system walls will be hot. Easier starting of ion pumps is possible if they can be roughed to lower initial pressures ~10^{-6} mbar with a turbo pump.

Problems

8.1 A vacuum vessel of volume 9 l is connected by a pipe 1 m long with bore 12.5 mm to a single-stage rotary pump of speed 1.8 $m^3\ h^{-1}$ = 0.5 l s^{-1}. Estimate the time to pump down from 1000 to 10^{-2} mbar, assuming that in the decade from 10^{-1} to 10^{-2} mbar the pumping speed is falling off and may be crudely represented as having one half of its value at higher pressures. Treat the continuously varying conductance of the pipe as having a mean value appropriate to the mean pressure in each decade, ignore early turbulence and gassing at the lowest pressures, and estimate times per decade.

8.2 A high-vacuum system has a base pressure of 1.0×10^{-6} mbar measured on a reliable thermionic ionization gauge and is equipped with a fine-leak valve to allow the controlled ingress of helium. For a particular purpose, it is required to create a partial pressure of

helium of 5.0×10^{-6} mbar in the residual gas. On letting helium into the system, to what value should the indicated pressure be allowed to rise in order to achieve this? Take the relative sensitivity of the gauge for helium to be 0.17. What would be the partial pressure of helium for an indicated pressure of 6.0×10^{-6} mbar?

8.3 In the context of a molecular drag investigation, a circular disc of radius R and thickness t is magnetically levitated in high vacuum and spins with angular velocity ω about an axis through its center and perpendicular to its plane. The residual gas consists of molecules of mass m and is at temperature T and pressure p. Assuming that the tangential momentum accommodation coefficient is unity, show that the retarding torque on the disc due to molecular drag is

$$\sqrt{\frac{m\pi}{2kT}}\omega\left(R^4 + 2R^3 t\right)p$$

8.4 A pipe of relatively small bore with conductance C for molecular flow is closed at one end, and the pumping speed at the open end, where it is connected to a large volume, is S. Analysis (e.g., Lewin, G., *Fundamentals of Vacuum Science and Technology*, McGraw-Hill, New York, 1965) shows that the steady-state pressure distribution along the pipe is given by

$$p(x) = p(0) + \frac{Q_w}{CL^2}\left(Lx - x^2/2\right)$$

where the coordinate x is measured from the open end $x = 0$ and Q_w is the total gas load due to outgassing, assumed uniform, of the interior wall of the pipe. Express $p(0)$ in terms of S and show that the pressure drop Δp between $x = 0$ and $x = L$ is $\Delta p = Q_w/2C$. Sketch the form of the variation of p with x and show that the mean pressure in the pipe is

$$\overline{p} = p(0) + Q_w/3C = p(0) + \tfrac{2}{3}\Delta p$$

Deduce from this result by solving an appropriate quadratic equation that, to measure the mean pressure, a gauge (of an ideal, nonintrusive sort) might be placed at $x = 0.42L$.

Note that the pumping speed S determines $p(0)$ but not the pressure distribution along the pipe. This pressure distribution is encountered in distributed systems such as particle accelerator tubes in which long pipe sections are pumped by remote pumps.

8.5 In order to analyze a flowing gas mixture at a total pressure of 10^{-1} mbar into its constituents, it is proposed to continuously extract a small sample from the mixture into an adjacent vacuum chamber at a pressure of 10^{-5} mbar, in which a residual gas analyzer can be safely operated. Extraction is to be accomplished via a circular aperture of 0.4 mm diameter in a very thin plate under conditions of molecular flow. Verify from Table 3.1 that flow will be dominantly molecular in character and determine the throughput into the analyzing region of nitrogen at a pressure 10^{-1} mbar. What pumping speed must be provided to maintain 10^{-5} mbar in this region? Assuming that the analyzer is accurately calibrated and has the same sensitivity for all the gases encountered, by what factor should its indicated reading for argon be modified in order to deduce the true partial pressure of argon?

9

Applications in Industrial Products and Scientific Instruments

9.1 Introduction

The purpose of this chapter is to demonstrate the role of vacua in a selection of technological products and scientific investigations. It is a rather arbitrary selection. The first four topics, vacuum-insulated panels, vacuum deposition of thin film coatings, microelectronic fabrication, and flat panel displays, are about products and processes that underlie technological accomplishments in the world around us. The subsequent five, the very large telescope, synchrotron radiation sources, controlled nuclear fusion, gravity wave detection, and lastly, the LEP and LHC in particle physics, are exciting scientific activities whose pursuit would not be possible without the expert knowledge and skills that vacuum scientists, engineers, and designers bring to each of them. In each of the nine examples, I have tried to introduce the subject and deal with the basic underlying physics in a reasonably thorough way before going on to outline the aspects of vacuum technology that are involved.

9.2 Vacuum-Insulated Panels

The principle of the Dewar vessel, a thermally insulating container that exploits the poor conduction of heat through rarefied gas, and which is usually referred to domestically as a "thermos flask," is well known. As depicted in Figure 9.1(a) the container has double walls made of either glass or thin stainless steel that enclose a vacuum of about 10^{-4} mbar. At this pressure the number density of molecules in the residual gas is such that thermal conduction through it between the outer wall at ambient temperature and the inner at a different temperature determined by the contents, hot or cold, is negligible. Radiative heat transfer, partly reduced by silvering of the vacuum-facing walls in the case of the glass structure, and thermal

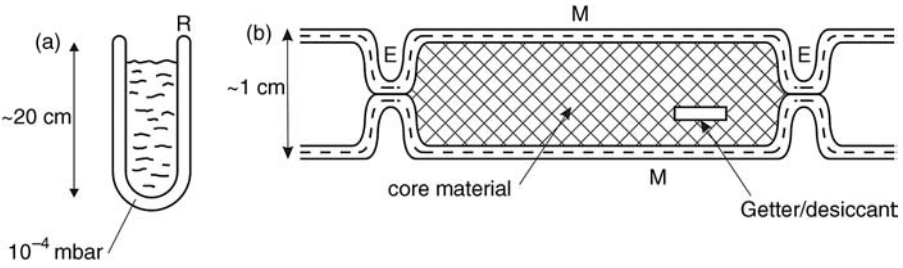

FIGURE 9.1
(a) Structure of Dewar vessel, (b) vacuum-insulated panel.

conduction in the solid wall down the temperature gradient established from inside to outside via the rim R, determine the relatively small and acceptable actual heat loss that makes the device so practically useful.

Unfortunately, the concept of the Dewar vessel is not easily extendable to geometries that do not have its cylindrical symmetry, in which the rim R plays a critical role in the mechanical integrity of the structure, holding inner to outer with no extra mechanical supports and enabling the vacuum enclosure. Also with this symmetry, long coaxial cryogenic delivery lines are possible with an inner delivery pipe at low temperature separated from an outer by an annular vacuum sheath with minimal mechanical and thermal connection between the two and sufficient mechanical strength to prevent collapse. It would be highly desirable to have extended "two-dimensional" sheets with the same highly insulating properties as the Dewar vessel, forming a sort of "vacuum sandwich." But the colossal force that would have to be sustained to prevent such a plane-sided structure being crushed — 10^5 Pa = 10^5 Nm^{-2} = approximately 10^4 kg weight or 10 tonne per m^2 — makes the simplicity achieved in the Dewar design impossible. Regularly spaced struts to maintain the separation of the two walls with adequate mechanical strength would also supply thermal pathways that would short circuit the insulation. Also, ideally, such a structure should be robust and to some degree flexible. One imagines that the familiar material "bubble-wrap," used for mechanical protection in packing, and itself a tolerably good thermal insulator because of its cellular structure that limits convection, would be much more so if the individual bubbles could be evacuated.

Vacuum-insulated panels, or VIPs as they are known, though not having the simplicity or thermal efficiency of the basic Dewar concept, nevertheless provide thermal insulation over extensive areas, which is a significant improvement over that of conventional methods. Thus, the thermal conductivities of VIPs are lower by a factor of about five than those of fiberglass blankets or solidified polystyrene foams. In various forms, they have been made for several decades for special applications such as thermal protection of biological samples in flights of the Space Shuttle and in the transportation of vaccines. They remain relatively expensive and so are not yet widely used in everyday applications, but as their development continues and concerns

grow to improve thermal insulation in diverse applications, their use will become increasingly economical and important.

A representative design is shown in Figure 9.1(b). Panels are typically some tens of centimeters square in lateral extent and consist of thin outer membranes M, ideally impermeable to gas and moisture, that define the vacuum envelope, sealed around the edges at E. This envelope contains a solid porous core material throughout which a vacuum is maintained to inhibit thermal conduction. It also has the role of supporting the envelope by maintaining the structural integrity against the large crushing force of the external atmosphere. The aim is to create an open porous structure with evacuated cavities of the largest size consistent with adequate mechanical strength of the solid component and minimal thermal conduction through it when a temperature difference exists between opposite sides of the panel. There is necessarily a region of higher conduction through the panel's surrounding edge E, and practicable membranes that are required to be robust and flexible to some degree cannot be made of glass or metal, the ideal impermeable materials. Plastic laminates have therefore been developed in which a thin aluminum foil (to act as a permeation barrier) is sandwiched between plastic layers that can be heat-sealed at the join E. While permeation through the membranes and their edge join is thus minimized, the membrane material, together with the core material, will outgas to some extent, and so provision has to be made to try to keep the installed vacuum close to its initial value of a few $\times 10^{-2}$ mbar. This is done, depending on the choice of core material, by the inclusion of appropriate getters and desiccants within the envelope. Once the getter is exhausted so that the pressure starts to rise, thermal performance is increasingly degraded. Nevertheless, useful lifetimes can be as long as tens of years.

Well-established core materials, suitable for a number of applications, are mineral powders, fiberglass, and silica, but the preparatory treatments required prior to their encapsulation and evacuation are expensive. Recently developed core materials include open-celled solidified polystyrene foams that are relatively inexpensive and easily evacuated, though they outgas to an extent that calls for matching getters, and as a consequence, lifetimes are relatively short. Newer core materials based on forms of silica gel have the advantage of being self-gettering and long lasting, but are rather more expensive to make.

The thermal conductivities achieved in VIPs are typically about 0.005 W/mK, which compares favorably with 0.041 W/mK for glass wool and 0.035 W/mK for cellular polystyrene, both widely used as thermal insulation. These of-order fivefold improvements in insulating properties are technically and economically significant to users. The materials science issues, the various aspects of thermal and mechanical design, and problems of economic production involved in the manufacture of VIPs for a variety of applications make it a challenging and active area of development. Progress may be monitored by consulting the Web site of the Vacuum Insulation Association that represents the manufacturers involved — www.vacuuminsulate.org.

9.3 Vacuum Deposition of Thin Film Coatings

The number of applications of thin film coating by vacuum processing is very large and the applications diverse. Many manufactured products depend on the associated technology. Some of the more evident ones are the aluminum reflecting coatings on, for example, the non-simple shapes of modern car headlights, the aluminized surfaces of audio CD and DVD video discs, and the cobalt/chromium layers that store information magnetically on tapes and discs. Copper alloys deposited on plastic and metal parts produce decorative gold-like effects. Less visible, by design, is the coated optics on binoculars consisting, most simply, of a quarter-wavelength thickness of transparent material of appropriate refractive index to eliminate the natural reflection from the objective lenses and improve the instrument's transmission. In the context of environmental energy concerns, vacuum-deposited coatings are features of solar panels and architectural glass, for which coatings with high reflectivity and low emissivity reduce the thermal load on the cooling plant in summer and heat loss by thermal radiation in winter. In engineering applications, special coatings of such materials as titanium nitride and carbide protect surfaces against wear and reduce friction. The decorative and protective aspects of a metallic chromium coating on mild steel are combined in the minting of coinage. Thus, a very large range of materials — elements, alloys, and compounds — is coated onto surfaces of many sorts, shapes, and sizes.

Coating is done by methods that fall into one or the other of two basic categories: physical vapor deposition, PVD, or chemical vapor deposition, CVD. In the former, the source of the material to be deposited as a coating is a piece of the material itself, which is either thermally vaporized or sputtered so that its atoms/molecules traverse the vacuum space to be deposited on designated receiving surfaces. In the case of CVD, the depositing material is produced in the proximity of the intended receiving surface by a chemical reaction in a hot flowing mixture of suitable gases. Thus, as will be described in more detail later, the reaction between gaseous silane SiH_4 and oxygen at about 400°C can be used to deposit silicon dioxide onto silicon.

In the PVD processes that employ evaporation, a number of heating techniques are possible. A relatively simple though not highly controllable method is that of resistive heating in which the source material is contained in a crucible of a refractory metal such as molybdenum through which electric current is passed to raise the temperature. High-temperature reactions between crucible and contents restrict the use of this method. A more widely used and controllable method is that of electron-beam heating. Typically, an electron beam with energy of order several keV is magnetically or electrostatically focused, usually the former, to deposit tens of watts of power onto the surface of the target source material, supported in a water-cooled hearth. For metallic source materials, it is frequently necessary that they be

molten liquid in order to achieve the necessary vapor pressure, and the size of the focused beam spot being smaller than the charge means that the material may form its own crucible, that part of in contact with the supporting hearth remaining solid and relatively cool. Quite refractory materials can be evaporated in this way.

The vacua needed for these applications are typically about 10^{-5} mbar or less, usually obtained by either diffusion or turbomolecular pumps, and sufficient to allow free passage of the evaporated product across the vacuum to the receiving surfaces without appreciable scattering. Appropriately shaped apertures in plates are frequently used to define the path of the evaporant to its intended target and intercept that which would otherwise produce inconvenient coatings elsewhere. Auxiliary devices used to assist the processing are automatically operated shutters to interrupt the evaporant streams, especially if multiple coatings of different materials from more than one source are to be produced in sequence, and quartz crystal rate meters. These monitor evaporation rates by detecting the small downward shift in natural frequency of oscillation of a small thin quartz plate located in such a way that one of its faces becomes loaded with material received from the source, thus increasing the inertia. The changing frequency signal can be processed to control the power supplied to the evaporation source and hence the evaporation rate, as well as to enable the developing deposit thickness to be determined. Small electron-beam-heated evaporation sources are also designed for use in ultrahigh vacuum in research applications, with particular attention paid to minimizing their outgassing, so that the conditions of substrate surface cleanliness are preserved.

The alternative PVD method is based on the phenomenon of sputtering, introduced in Section 4.6. Variants and sophisticated developments of the basic DC diode technique are very extensively used. This basic DC diode configuration is shown schematically in Figure 9.2(a). The substrate to be coated is connected as the anode and held at earth potential, while the target made of the source material is made the cathode and held at a distance of a few centimeters away. The cathode potential is set at –2 kV with a power supply. Following evacuation to high vacuum, argon gas is let into the system to raise the pressure to about 1 mbar. The application of the cathode potential then initiates a glow discharge in which argon ions bombard the cathode with energy sufficient to sputter target atoms, mostly neutrals, towards the substrate where they arrive with energy of a few eV, having been slowed by scattering as they pass through the neutral argon atoms and the plasma. Deposits adhere well because of the high incident energy compared with the energies associated with thermal evaporation, which are typically ~0.2 eV. The discharge is maintained by secondary electrons released by the ion bombardment of the cathode, which enter the plasma to cause further ionization of argon neutrals, eventually finding their way to the anode. Deposition rates are fairly slow because of the scattering of the sputtered atoms, and the technique cannot be used for insulating cathode materials because charge buildup would reduce the diode potential and stop the discharge.

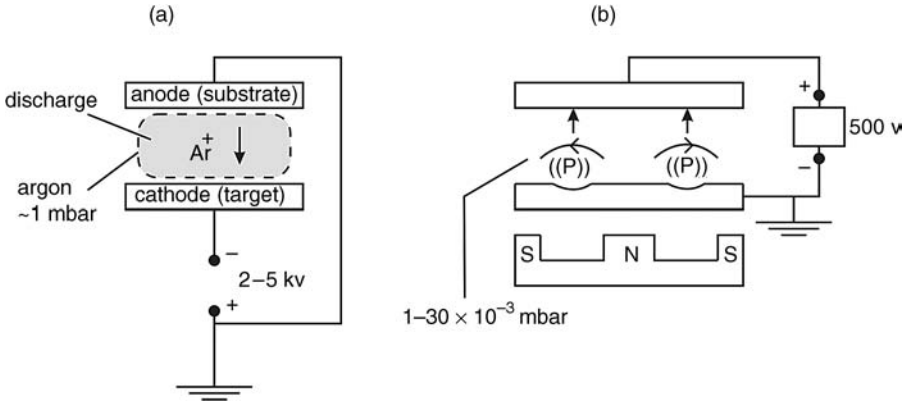

FIGURE 9.2
(a) DC diode sputtering, (b) DC magnetron sputtering.

The DC diode technique is little used nowadays, but it served as the starting point for a number of developments, including AC (RF) diode sputtering, which enables the sputtering of dielectrics, and DC and RF techniques based on the triode configuration and the magnetron. The principle of the very widely used DC magnetron sputtering technique is illustrated in Figure 9.2(b). The cathode target material to be sputtered is located above permanent magnets that create magnetic fields in the directions shown, so that the electric and magnetic fields are crossed. A DC voltage of typically –500 V on the cathode in argon pressures of $1 - 30 \times 10^{-3}$ mbar initiates a discharge in which, under the influence of the $ev \times B$ force, electrons spiral back and forth around the magnetic field lines and tend to drift circumferentially so as to be localized in the regions of high plasma density P. The result is an increased probability of argon ionization and hence sputtering as ions from this region are accelerated into the cathode, and because of the lower argon pressure compared with simple diode sputtering, fewer of the sputtered atoms are scattered out of their intended paths to the anode substrate.

A distinctly different approach to the exploitation of the sputtering process is that of ion-beam sputtering, which does not involve having an ion-generating plasma in contact with the material to be sputtered. In this technique, ions are extracted from a remotely generated plasma or by other means and formed into a beam that is directed onto the target material, from where sputtered atoms travel to the receiving substrate. That the ion beam energies, currents, and beam cross-sectional dimensions, as well as the identity of the ion can be designed to serve particular applications, and that the method can be operated in very high vacuum, make this a very powerful and widely used technique. An elegant example of its use is shown in Figure 9.3 in which targets of different materials mounted on a carousel C are sequentially presented to the ion gun G, with sputtered products received at the substrate S. The auxiliary ion gun I shown in the figure, whose beam is directed at the substrate, may be used to improve the quality of deposits by ion

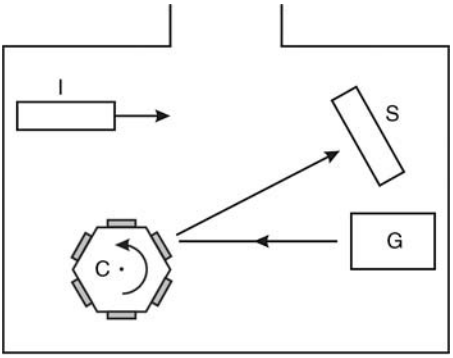

FIGURE 9.3
Sequential ion-beam sputtering.

bombardment of the substrate both before and during deposition, a procedure that is employed in a number of other deposition techniques and referred to as IBAD — ion-beam-assisted deposition. Ion-beam technology is discussed by Powers (1998). The various sputtering techniques are described by Westwood (2003).

Significant advantages of sputtering are that the source material does not have to be heated, so that sources can be operated in orientations unavailable with molten sources, and that alloys are more reliably deposited. This is because, for the deposition an alloy AB, for example, preferential evaporation of the higher vapor pressure component is likely to occur with thermal excitation, whereas in sputtering, a self-correcting mechanism tends to operate; if an excess of A is sputtered at some instant, then the surface is enriched in B for subsequent sputtering, and so on.

Chemical vapor deposition (CVD) is quite different in concept from these PVD techniques, depending, as was mentioned earlier, on chemical reactions in the gas phase. It is possible, using CVD and its variants, to deposit an extraordinarily large range of substances including many metals, alloys, graphite, diamond, carbides, oxides, nitrides, silicides, and other compounds. While CVD can be carried out at atmospheric pressure, low-pressure LPCVD at pressures in the range 0.1–10 mbar is widely employed for the growth of deposits of high structural quality, adhesion, and uniformity. To briefly illustrate what is possible, a few of the reactions that can be exploited, together with the temperature at which they have to be carried out, are listed below, where bracketed (g) and (s) signify the gaseous and deposited solid states, respectively.

1. $SiH_4(g) \rightarrow Si\ (s) + 2H_2\ (g)$ .....650°C
2. $WF_6(g) + 3H_2(g) \rightarrow W(s) + 6HF(g)$ .....300°C
3. $SiH_4(g) + O_2(g) \rightarrow SiO_2(s) + 2H_2(g)$ .....450°C
4. $BF_3(g) + H_2O(g) \rightarrow BN(s) + 3HF(g)$ .....1100°C

Precursor materials that are not naturally gaseous are generated as vapors by preheating of their volatile solids. A particularly useful feature of CVD methods is that materials that are difficult to deposit by thermal means because of their very high melting temperatures can be produced at much lower temperature than would be required for their evaporation. In reaction (2), for example, tungsten (T_M = 3387°C) is produced in a reaction at 300°C. But one notes that, particularly in the case of reactions (2) and (4), some dangerous gaseous products have to be removed by the pumps and safely disposed of by appropriate treatment at their exhaust. This is a feature of many reactions employed in CVD. Aggressive, toxic, and flammable gases, and in some cases flow-borne grit and dust generated as solids in the reactions but not deposited, are potentially hazardous to operator health as well as to equipment, and have to be dealt with. In some cases the ingredients as well as the products are similarly problematic in character.

The control of the flow rates of the precursor gases, sometimes diluted in carrier gas, and the temperature of the reacting gases and the receiving substrate surface, are critical in what is a very complex multistep physicochemical process. A schematic diagram of a system to implement reaction (3) is shown in Figure 9.4.

Mass flow controllers (MFC) regulate the flow of silane and oxygen into the heated reactor vessel in which sets of silicon wafers to be coated with SiO_2 are loaded. Pressure is sensed by a capacitance diaphragm gauge and gases are pumped by a robust combination of a Roots pump backed by a dry pump. Pressure, temperature, and gas flow conditions must be stably controlled and reproducible from load to load. In order that this can be accomplished, it is necessary for the pumping speed to be variable, either by the use of an adjustable throttle valve or an electronically controlled motor speed for the Roots pump. Flow controllers are discussed by Hinkle (1998).

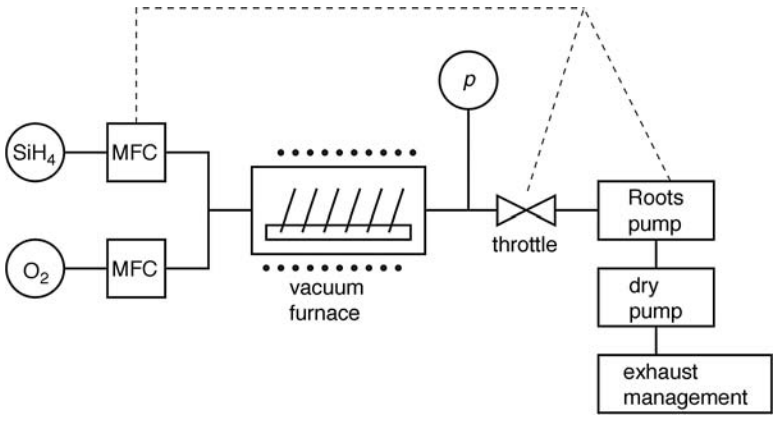

FIGURE 9.4
System for chemical vapor deposition.

Variants of the chemical vapor deposition process are PECVD, plasma enhanced CVD in which plasma excitation enables reactions to proceed at lower temperatures, and MOCVD, widely used in the electronics industry, that utilizes the vapor of metal–organic compounds as precursors, to deposit aluminum or gallium arsenide, for example.

9.4 Microelectronic Fabrication

The impact of modern semiconductor microtechnology on a range of human activities associated with work, living, and leisure is staggering in its magnitude and extent. Mobile phones, home computers, automobile electronics systems, and avionics are just a few examples of newly created technologies or existing ones whose capabilities have been dramatically enhanced. The integrated circuit (IC) is the miniaturized device that has allowed these developments. It consists of numerous individual transistors, resistors, capacitors, diodes, and logic gates, etc., depending on its design function, interconnected and packaged on a small piece of monocrystalline semiconductor just a few millimeters square and 0.5 mm thick, to form a single monolithic structure, a "chip."

Silicon is the most widely used semiconductor material, and typically several million individual circuits can be made on a 5-mm square. The texts by Morgan and Board (1994) and Jaeger (2002) are good introductions to this rather complex subject, discussing not only the physics of the devices and the fabrication processes but also the factors related to the economics of production. As will be explained below, processing is carried out simultaneously at a large number of equivalent sites evenly distributed over the surface area of a thin silicon wafer. This defines sections of it, called "dice" that will each carry a complete IC when processing is complete. Vacuum technology plays an important role in many of the stages that are involved in the manufacturing process.

In structure, an IC consists of a pattern of doped areas in the subsurface of the host silicon with patterns of overlying insulating, conducting, and semiconducting regions and interconnections built on top to a height of a few microns. The technique of lithography used to make these patterns is of central importance in the fabrication process. Photolithography, which shares some principles with the "wet chemistry" of traditional photography, has been widely used for many years and is illustrated in a simple example by the sequence (a) to (d) of Figure 9.5. This shows how windows may be cut in a thick oxide overlayer on silicon by selecting parts of it for removal in a subsequent etching process, to give the result shown in Figure 9.5(d). This is done by creating areas that will resist the etching process. Thus in Figure 9.5(a), the silicon oxide has been covered with a complete layer R of "positive photoresist," a polymeric material, shown by the lighter shading.

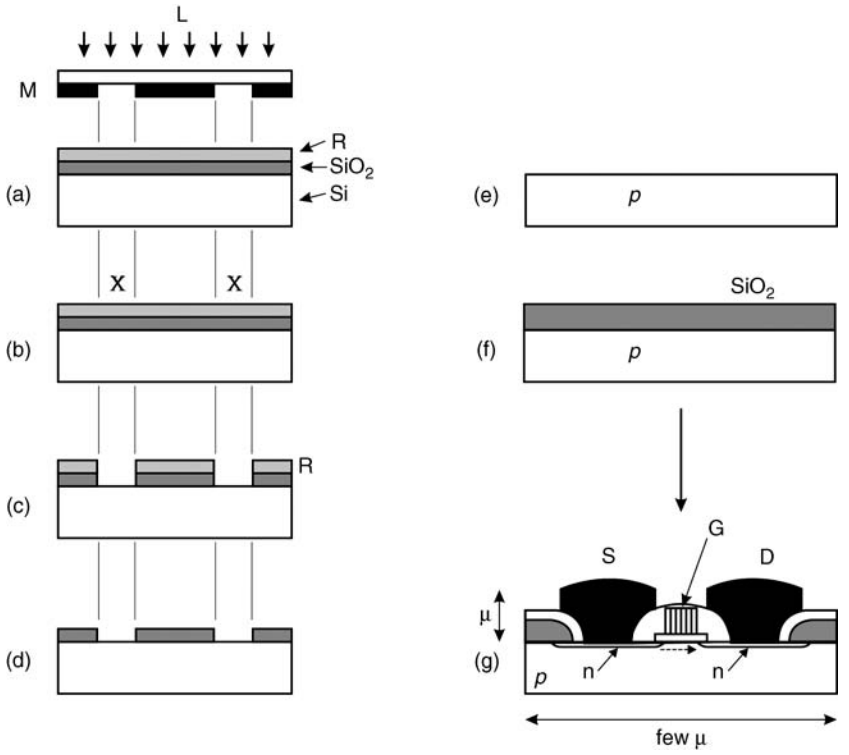

FIGURE 9.5
(a) to (d): Lithography to cut windows in SiO_2; (e) to (g): NMOS transistor fabrication.

When illuminated by collimated UV light L through a close mask M, regions X are exposed, but the others are unaffected, as shown in Figure 9.5(b). In the associated developing process, the exposed resist whose nature has been altered and rendered soluble by the exposure is dissolved away in the developing solution. This leaves the parts of R to their prime function of protecting the silicon oxide underneath in the subsequent etching process to remove the oxide, shown completed in Figure 9.5(c). Finally, the remaining resist is dissolved away to leave the oxide pattern in Figure 9.5(d). Thus the pattern of this simple mask in which two windows are cut is transferred to the surface below. Negative resists work in the opposite way — exposed areas are hardened against subsequent dissolution in the developing process.

Lithography is used at various stages in the fabrication of ICs. For a given patterning operation, the mask consists of the basic mask geometry for one die replicated to contain the required number of units. Masks are made by optical demagnification of much larger and precisely computer-drawn versions. The minimum spacing of features in the pattern on the mask is about 1 μm; diffraction of the waves of the patterning radiation, between 0.2 and 0.5 μm for UV light, and subsequent edge effects in the processing, limit the attainable scale of fineness. When augmented or replaced by electron or

ion-beam lithography, which is the trend as smaller, more closely spaced components are sought, different resists and procedures are needed. Processing is carried out in vacuum in order that the beams can be appropriately deflected and the waste products pumped away. The micromechanical manipulation to move masks and maintain registry between successive patternings with different masks is extremely sophisticated.

Etching is also increasingly carried out by "dry" processes, either by plasma etching in which surfaces are eroded by chemical reaction with the active species of a specific plasma or by ion-beam etching (also known as ion-beam milling), in which surfaces are sputtered away in a controlled way. In both cases, vacuum pumping is needed to operate the technique and deal with the resulting products. Another important part of the fabrication process is the means of doping the silicon to induce the required type of impurity semiconduction, either *p*- or *n*- type, in its subsurface. This may be accomplished either by high-temperature diffusion of species into the silicon or by their implantation with an ion beam, both of which require vacuum. Various conducting or insulating overlayers required prior to their patterning are grown by the CVD or PVD techniques that were described in Section 9.3. In some cases the conditions of growth are chosen so that the layers grow epitaxially, which means that the deposit grows in an orderly way as a monocrystalline structure continuing in registry with the crystallographic structure of the substrate underneath. Otherwise, deposits are polycrystalline.

The techniques described above are deployed at various stages in the fabrication process. As an example of an actual device, Figure 9.5(e), (f), and (g) demonstrate the starting and finishing points in making a silicon-gate NMOS transistor on a substrate of *p*-doped silicon. The acronym signifies *n*-type semiconduction in a metal–oxide–semiconductor structure. A thick insulating thermal oxide layer (f) is first produced by heating the wafer in oxygen or steam. Such layers isolate the device to be grown from its neighbors and provide barriers to diffusion/ion implantation in later stages. Nine processing steps (not shown, but of the types described) lead to the transistor structure of (g). Metallic connections S and D make contact with the underlying *n*-regions that form the source and drain of the device, and G is the gate that controls the flow of current between them. The potential applied to G determines the amount of conduction in the channel between the source and drain so that the device can serve as a voltage-controlled resistor or as a switch, if driven between conducting and nonconducting states by an appropriate binary signal. Contact to the thin-gate oxide layer is by deposited polycrystalline silicon (polysilicon) G that conducts sufficiently because of *n*-type impurities in it from the doping procedure. It extends backwards "into the page," so to speak, to be adequately separated from the S and D contacts.

The relative scales of the representation in (g) are considerably distorted for clarity. In the vertical, the height of the active areas is a micron or so compared with a wafer thickness of about 500; (1000 μm = 1 mm). Feature sizes and spacings laterally mean that the extent w of a single device is a

few microns. To give a feel for orders of magnitude, a w value of 5 μ implies 200 spaces per mm, 4×10^4 units per mm^2, and therefore 10^6 in a 5-mm square — conservative figures in the light of current trends. Halving the characteristic width enables the packing density to be increased by a factor four, though achieving such aims is not simple and not just from a processing point of view. Bringing components into closer proximity increases problems of interference between their functions. Eventually, of course, limits are reached where quantum-size effects start to dominate device behavior, but current aims to manufacture at fractional submicron levels involve advances in processing technique rather than new types of physical behavior.

As mentioned earlier, the ICs are not produced singly but in batches on the surface of large, round monocrystalline silicon wafers, and the wafers themselves are processed through the fabrication facility in batches of 25–200. Once processing is complete, the wafer is cut into the dice that carry the individual ICs prior to their testing, the making of connections, and packaging. Because processing is multistaged and complex, yields are inevitably less than 100%. Currently, ICs are being produced on both 150-mm (6-in) and 200-mm (8-in) wafers. Wafers of size 300 mm are projected. The number of dice increases rapidly with wafer size. On a 150-mm wafer, just less than 200 dice, 10-mm square, can be accommodated, but this rises to about 320 on a 200 mm wafer. Thus, to use larger wafers with an improved yield of dice that carry circuits of reduced feature size and hence a higher number of circuits per chip, is the driving economic incentive. The industrial plants in which ICs are made — wafer fabrication facilities as they are called — are extraordinarily sophisticated. Their capital cost is several hundred million dollars, and getting high yields is critical to the economics of operation. There is scrupulous attention to matters such as process monitoring and control, and the precision, cleanliness, and efficiency of operations. Vacuum processing in the necessary ultraclean conditions is discussed by O'Hanlon (2003). Microelectronics is arguably the finest achievement of modern science, technology, and manufacturing.

9.5 Flat-Panel Displays

For more than a hundred years, vacuum technology has played a vital role in the evolution and manufacture of the cathode ray tube, enabling the passage of an electron beam from source to screen in a variety of display devices such as oscilloscopes, radar tubes and, at a very high level of technical sophistication, the modern color TV tube. In essence, within an evacuated glass envelope, electrons released thermionically from a hot filament are accelerated through a potential difference of 15 kV or more and focused to form an electron beam whose impact on a thin phosphor coating generates light, visible through the envelope, with a brightness that depends on the

beam current. By deflecting the electron beam to move across the phosphor, in a scanning motion in the case of TV, and onto pixels (picture elements) that contain phosphors to create light in the primary colors red, blue, and green, the information encoded in the incoming signal is presented visually to the viewer.

The TV tube and other CRTs are necessarily rather bulky because the electron gun assembly has to be set some distance behind the screen in order that the beam can be scanned across the whole of its area, and they tend to be rather heavy. In recent decades, there has been an increasing demand for slimmer displays with a much smaller depth dimension, particularly for portable laptop computers, and this has been met most successfully by liquid-crystal display (LCD) technology. The active-matrix liquid-crystal display, described in the following text, has come to occupy a commanding position in this application, and now, also as the display screen for desktop computers with diagonal screen size of up to 18 in. There are, however, other applications for which "flat-panel displays" or FPDs as they have come to be known, would be highly desirable, most notably for domestic "picture on the wall" TV, with larger screens up to 30 in. and more. Because of the huge size of this potential market and the profits to be made, a number of different schemes for FPDs are being very actively pursued, of which the active-matrix LCD is one. Plasma and field-emission displays are two other technologies that are also highly developed. One of the principal problems of securing a substantial market for these products is to bring down the costs of manufacture to be competitive with the conventional, long-established and mature CRT TV technology. It is likely that eventually it will be superseded.

As in the case of the CRT, vacuum plays a significant part in all of the new technologies — either in their manufacture or operation, or both. Articles on the basic physics of all these new types of display were published in the June 1997 issue of *Physics World.* The text of Luther and Inglis (1999) provides a good introduction to the electronic and video engineering aspects of the subject, and the website www.atip.org/fpd is informative about current developments. Because they are examples of existing and developing technologies, the principles of the active-matrix LCD, plasma, and field emission flat-panel displays are described in the following text.

The structure of a representative ***active-matrix LCD*** is shown schematically in Figure 9.6(a). Liquid-crystal material LC is contained between two thin sheets of glass G1 and G2, with polarizing sheets P1 and P2 outside them, crossed at 90°. A backlight B illuminates the whole area of the display with white light and the view is from above. On the inner surface of G2, a two-dimensional grid formed by a fine metallic matrix M (called the "black matrix"), covers the whole screen area to define the array of pixels and subpixels and to prevent crosstalk between them. Each oblong pixel area consists of three subpixels of red, blue, and green color filters r, b, and g, as in the view of G2 in Figure 9.6(b). A thin, transparent conducting layer of indium tin oxide (ITO) covers the whole area to act as a reference electrode. On the inner surface of sheet G1 are patterned thin film transistor (TFT)

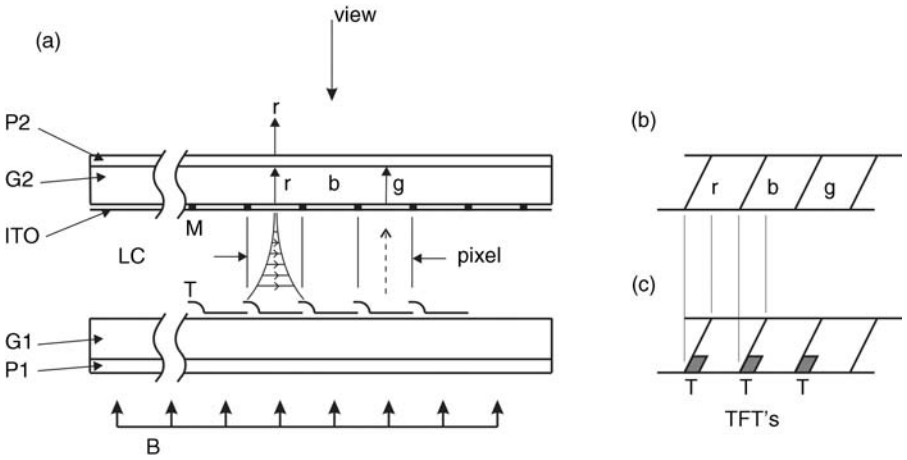

FIGURE 9.6
(a) Active-matrix LCD, (b), (c) detail.

structures in an array that matches the array of subpixels on G2, as suggested in Figure 9.6(c). The TFT occupies only a small fraction of the area of each subpixel; the rest is a transparent conducting area connected to the TFT drain electrode, which sets its potential. The TFTs are individually addressed by means of orthogonal "row and column" conducting lines (not shown), so that a signal applied to a TFT gate determines its drain potential, which sets up an electric field across the LC material of the subpixel. This array of TFTs constitutes the "active matrix."

The LC material has long polar molecules that affect the polarization of light passing through it. The surfaces adjacent to G1 and G2 with which the LC makes contact are microgrooved, which causes the long molecules of the LC to align with them, and these grooves are aligned parallel with the directions of polarization in the adjacent polarizers. Thus, the LC molecules form a twisted spiral between G1 and G2, as suggested in pixel "r" of Figure 9.6(a). Optically, this means that, although the polarizers P1 and P2 are crossed, the presence of the LC spiral causes the polarization of the light upcoming through P1 to be rotated through 90° and transmitted through P2. Having passed through a red filter, it makes a red contribution to the pixel color. Thus, in the unexcited state, the LC transmits all the light incident through G1. However, when a signal causes a finite potential on a subpixel electrode, the electric field set up across the LC tends to orientate the molecules in its own direction and out of the spiral. Depending on the signal level, the polarizing capability of the LC is partly or wholly annulled. In the latter case, the light passes through with unchanged polarization and is blocked completely at polarizer P2, as suggested at subpixel "g". This is the basic mechanism by which individual pixels are activated to give the correct color mix and brightness, essentially by subtraction from a white background field. In an LCD display with 14 in. diagonal, typical for a laptop screen, in

a 285-mm horizontal dimension there are 1024 pixels, and therefore a pixel size of 0.28 mm, consistently with the resolving power of the eye at typical user distance. The details of signal processing and frame rates, etc. are discussed in the text of Luther and Inglis (1999).

Vacuum technology is involved extensively in the manufacture of this type of display, using a number of the processes described in Sections 9.3 and 9.4. For example, the black metallic matrix M and ITO layer on G2 are deposited by sputtering in argon and reactive sputtering in an argon–oxygen mixture, respectively. On G1, the laying down of the TFTs is by plasma-enhanced CVD, with associated patterning by lithography and etching. Perrin (1999) describes the processes in detail. To successfully make a display of 1024 × 768 = 786, 432 pixels on glass sheets G1 and G2 with dimensions 285 × 216 mm is a formidable technical challenge. As in the case with chip manufacture (Section 9.4) processing yields are critical and an important factor in determining cost and therefore profitability.

Plasma displays exploit the ultraviolet light emission from minute, highly localized electrical discharges in low-pressure gas to excite visible light from adjacent color phosphors. The light-emitting process involved is rather similar to that in the tubular fluorescent lamp. Figure 9.7(a) illustrates the basic arrangement of the alternating current matrix (ACM) structure, so called because each color subpixel is addressed by an appropriately modulated alternating square wave voltage signal. The rear glass plate G1 carries a set of column electrodes E in the form of conducting strips that run across the length of the plate, and this surface is covered by a continuous dielectric layer of MgO. On top of this are barriers B — small walls symmetrically located with respect to the underlying electrodes — that also run the length of the plate. Red, blue, and green phosphor coatings cover the barrier walls and part of the dielectric as shown, forming long troughs, each associated with an electrode E. On the front glass plate G2, there are similar but orthogonal rows of electrodes F, also covered with dielectric. Thus each subpixel r, b, or g of the pixel (r + b + g) can be addressed and a voltage difference presented across it. The space between the plates contains a mixture of neon with a few percent xenon at a pressure of the order of 1 mbar. This is arranged by initial evacuation to high vacuum and then backfilling with the gas. The outer edges of the two plates are sealed together with a low-temperature glass, through which the electrical connections E and F protrude. Small, regularly spaced insulating posts, not visible to a viewer, keep the plates parallel against the force of atmospheric pressure.

The discharge that is initiated by a voltage signal applied across a subpixel is confined locally as shown in Figure 9.7(a). The insulating surfaces over the electrodes play a key role in the character and control of the discharge. Xenon atoms excited by the discharge deexcite with the emission of ultraviolet radiation. This in turn causes excitation of the phosphor material and its de-excitation by giving out blue light. The modulated square wave that is applied to each subpixel alternates at a frequency of typically 70 kHz between various levels that switch the discharge on, sustain it, or turn it off.

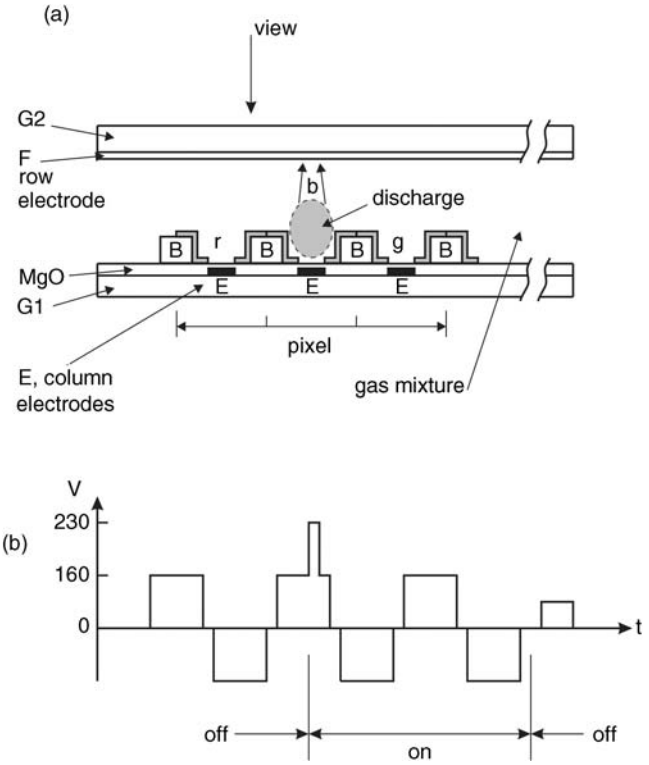

FIGURE 9.7
(a) Plasma flat-panel structure, (b) subpixel driving signal waveform.

An example of the waveform is shown in Figure 9.7(b). In the first interval shown, the signals at + and –160 V are too small to trigger the discharge and the cell addressed is OFF. When the level is increased to 230 V, the discharge is triggered and sustained ON by the subsequent 160 V excursions. To turn it off again, a lower voltage pulse is applied that is insufficient to maintain the discharge but necessary to take away the charge established on the dielectric in the previous excursion, leaving it in a sustained OFF condition. The rather complex physics of the processes involved is discussed by Deschamps and Doyeux (1997).

The brightness of an activated subpixel is determined by the discharge and is fixed. Its perceived brightness, therefore, is controlled electronically by varying the duration of the ON time in the interval between successive refreshments of the image as a whole, and exploiting the viewer's persistence of vision and the eye's integrated response to discrete excitations. Thus, for example, a brightness of half maximum is achieved with excitation for half the interval. The perceived light from the pixel is the sum of the subpixel contributions to give a single color of the correct brightness. The driving electronics refreshes the image at a frequency sufficient to avoid "flicker,"

typically 24 times per second. Because these are of the order of one million pixels in a display screen, much signal processing is involved.

The fabrication technology for plasma displays is simpler than that for LCDs, in which the provision of a TFT is necessary at each subpixel address. Large screens with 42-in. diagonal and more are commercially available, though expensive. Including their associated power sources and driver electronics, they are only about 10 cm in depth and so can literally be "pictures on a wall." Traditional CRTs at this screen size would be impossibly bulky. Plasma displays are a feature of many public information displays in airports, museums, etc., and usually enhance their surroundings. Smaller displays for domestic use are available but still expensive.

The working principle of the ***field emission display*** (FED) resembles that of the original CRT in that electrons are accelerated in vacuum from source to a phosphor-coated anode screen. However, in the FED, each color subpixel has a dedicated group of micron-sized field emission tips for its electron source so that emitters and screen can be very close together, resulting in a thin device. Field emission is the process in which electrons are extracted from a surface by quantum mechanical tunneling in the presence of a very strong electric field. It is therefore a cold-cathode device with no power requirement for electron emission, in contrast with the CRT. Shah (1997) discusses FEDs in detail.

A representative structure is shown in Figure 9.8(a). Front-glass-plate G2, about 1 mm thick, has a thin conducting transparent ITO layer on its inner side, on which are deposited a black matrix and then an array of pixels each comprising subpixels with red, blue, and green phosphors. The manufacture of this plate involves techniques that are close to those of existing CRT technology. The rear-plate G1 has cathode electrode strips K and control-gate electrode strips CG that run orthogonally for addressing purposes. The K strips carry FE tips, typically made of molybdenum and about a micron high with a very small radius of curvature at their point. They are spaced apart by a few microns in groups whose size is appropriate to the excitation of the area of a subpixel. Only four as 2 × 2 are shown in the figure for clarity, but in practice there are more.

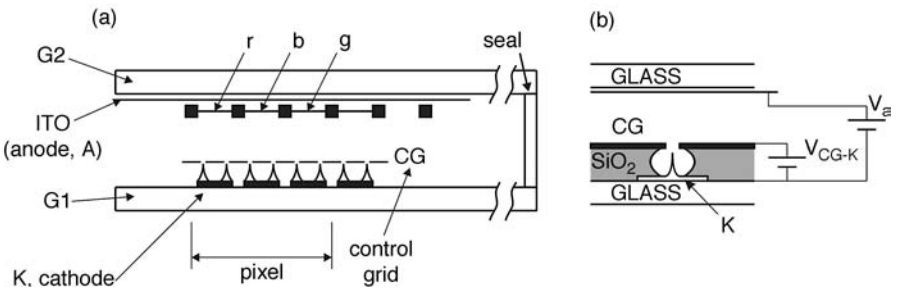

FIGURE 9.8
(a) Field emission display, (b) detail.

The geometry of one FE tip and its control-gate electrode are shown in Figure 9.8(b). They are named Spindt tips after the scientist who developed their fabrication process. Each control grid strip contains holes that match the individual FE tips. The ability to make the whole complex structure, which involves lithography and etching, is an outcome of techniques developed for microelectronic fabrication, discussed in Section 9.4. The whole assembly, sealed at the edges through which the K and CG electrodes emerge, has to operate at 10^{-7} mbar and contains distributed getters to maintain this pressure, which is necessary in order that the FE tips are not degraded by contamination, and to prevent ionization and sputtering. Insulating spacers that are not visible to the viewer prevent geometrical distortion due to atmospheric pressure. The cathode-to-anode distance is a little more than a millimeter.

In operation, the anode plate A is held at high potential with respect to K and an addressed subpixel gate is presented with a signal voltage V_{CG} that causes electron emission from its FE tips. The electrons are accelerated to the phosphor and produce light of the designated color following impact. Traditional CRT phosphors that are efficient with 10-keV excitation are much less so at the lower voltage levels best suited to thin FE displays. In order to have low anode voltages, ideally less than 1 kV so as to prevent electrical breakdown across the insulating spacers and retain the close proximity of source and screen, which promotes good picture definition, new "low-voltage" phosphors are being developed.

Achieving large screen sizes requires the fabrication of FE tips over a large area using very sophisticated techniques. The field is very active and many variants of the concept are being investigated, including alternative tip materials such as diamond and diode rather than triode structures, which would considerably simplify the fabrication of the emissive backplate. There seems to be little doubt of the eventual demise of the CRT, but which technology will replace it in the TV application remains an open question.

9.6 Mirrors of the Very Large Telescope (VLT)

Located in the Atacama Desert in northern Chile on Paranal Mountain at an altitude of 2635 m, the Very Large Telescope consists of four linked telescopes, each with a main mirror 8.2 m in diameter, supplemented by three auxiliary 2-m telescopes. The total light-collecting area of the four main units is 200 m^2, equivalent to an instrument of 16 m diameter, and in conjunction with the auxiliary units, the whole configuration can be used in an interferometric mode with very high angular resolution. The useable wavelength range extends beyond the visible into the infrared. The Paranal Observatory is operated by the European Southern Observatory (ESO) whose

headquarters are in Garching, Germany, to which it is connected by a permanent satellite link for communication and data transmission. A comprehensive description of ESO's purpose and methods can be found on its web site www.eso.org.

Each of the four unit telescopes consists of the main 8.2-m mirror weighing 24 tons and two others in the Cassegrain configuration. This optical cell is supported in a sophisticated mechanical structure that directs the instrument and allows celestial objects to be tracked in their diurnal motion. The correct mirror shape is preserved as its orientation is changed by a distribution of computer-controlled supports acting on the rear that compensate for the slight weight-induced distortions that would otherwise occur. Each enclosed assembly is about 20 m high, equivalent to a 5-story building, and weighs 430 tons, of which about 200 move.

It is interesting to consider the manufacture and maintenance of the main 8.2-m mirrors. Even in the good atmospheric conditions that exist at the VLT, its mirror surfaces are subject to wear by airborne dust particles, and their remetallization is necessary periodically at approximately 2-year intervals. For this reason the laboratory, vacuum-coating plant, and other maintenance facilities after being built, commissioned, and first used in Germany, were dismantled, shipped, and reassembled at a location some 2 km away from the VLT and 300 m below it in altitude.

The 8.2-m mirrors have a focal length of 108.83 m and were cast from Zerodur glass blanks 175 mm thick, which were rotated while molten to a natural meniscus shape that was retained after cooling and solidification, and then machined to their exact desired shape by lengthy grinding and test procedures. The central hole, a necessary feature for Cassegrain operation, is 1 m in diameter. The large vacuum vessel in which they are aluminized was designed by engineers of Linde AG in Germany and BOC Coating Technology in the U.K. It is 9.4 m in diameter, with a central maximum height of about 1.7 m and a volume of approximately 120 m^3. It consists of upper and lower halves, horizontally divided, that are shaped and constructed to withstand the huge atmospheric force following evacuation, and sealed with two concentric "O" rings. It is shown in a highly simplified schematic form in Figure 9.9 and fully described by Ettlinger et al. (1999). The lower half, which contains the mirror mounted on special supports B, rests on an air-cushion vehicle V whose eight inflatable cushions A enable it to be raised into sealing contact with the upper half, which is at a fixed height above the floor, or lowered so that it can be moved away on the vehicle to other services and the loading/unloading position.

The fixed upper half contains the aluminum coating sources M and shutters S, gauging, other sensors, and connections P to primary pumps. For secondary pumping, eight directly attached cryogenic pumps C are supplemented by a peripheral copper loop W that can be chilled with liquid nitrogen to provide very large pumping speed for water vapor. The base pressure of the system is 2×10^{-6} mbar. The coating sources are aluminum bars, 99.995% pure, aligned radially and bonded to copper plates that are water

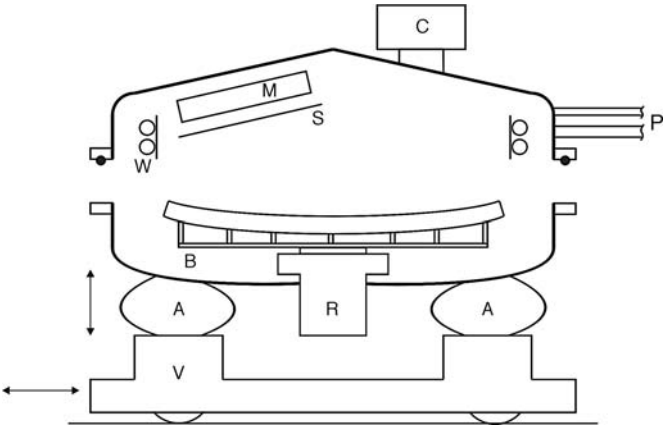

FIGURE 9.9
Highly simplified representation of VLT mirror-coating plant.

cooled. Permanent magnets located behind these plates provide magnetic fields that enable magnetron sputtering of the aluminum, as described in Section 9.3. Uniformity of coating is of course critical, and this is achieved in a radial direction by careful design, characterization, and location of the source, and in a circumferential direction by slowly rotating the mirror at an accurately constant prescribed speed on its mounting. In order to do this, the support system B is itself supported on a Ferrofluidic rotating seal R (see Section 10.5), which can be driven externally. The low base pressure is necessary to minimize the incorporation of residual gas species into the deposited film, particularly those derived from ionization of water vapor in the plasma discharge when the source is working.

Once a mirror is loaded and the chamber sealed, several stages are involved in getting to base pressure. All operations are under full computer control, with inputs from a variety of sensors. Initially, the full speed of the primary pumps is restricted by the use of a choking aperture to allow a "soft start" that prevents turbulence in the chamber and thus avoids the possibility of generating flow-borne particulates that would degrade the mirror surface if they reached it. Pumping for about 2 h in this mode reduces the pressure to 10 mbar, beyond which the pumps work at their full unrestricted speed without risk of turbulence, rapidly reducing the pressure to 2.5×10^{-2} mbar, at which point they are valved off and the cryopumps take over the pumping and, together with the Meissner traps, bring the chamber vacuum to its base pressure in a further few hours. The mirror is given a preliminary cleaning by rotating it under a glow discharge in dry air at 10^{-2} mbar. Just prior to the coating process, and with the mirror rotating at about five revolutions per hour, argon is backfilled into the system to the discharge pressure and the plasma ignited with the shutters closed to clean the aluminum targets. Once this is accomplished, the shutters are opened so that coating proceeds, with an integral number of revolutions to deposit a highly pure and uniform

sputtered aluminum coating approximately 80 nm in mean thickness, with less than 5% deviation. Venting the system up to atmospheric pressure (770 mbar at the prevailing altitude) is carried out very slowly through filters and with prewarmed air so that the initial cooling due to expansion as air initially enters the vacuum, expands, and cools, does not cause frost formation on the new mirror.

9.7 Synchrotron Radiation Sources

A particle of mass m and charge q moving with velocity v in a direction perpendicular to a uniform magnetic field B experiences a force qvB perpendicular to both v and B, causing it to move in the arc of a circle of radius R. The value of R, determined by equating qvB to the product of m and the centripetal acceleration v^2/R, is $R = mv/qB$. This relationship, which we will refer to as the "orbit equation" is of prime importance for machines in which charged particles, guided by magnetic fields, move in circular orbits at high energies. Because the acceleration of electrical charge causes the emission of electromagnetic radiation, there will be radiation from these orbits associated with the continuous inward acceleration.

Synchrotrons are large modern machines for accelerating electrons, protons, or their antiparticles to very high energies. The synchrotron principle may be described with reference to Figure 9.10(a), which is a schematic plan view of a ring-shaped vacuum system in which particles of charge q may circulate at radius R. Over the area indicated by dotted lines, external electromagnets create a vertically directed magnetic field B that can be increased with time in a controlled way. A radio frequency voltage source, whose frequency may also be varied, provides an electric field of the correct phase to accelerate particles circumferentially as they pass through a resonant cavity. Particles are injected tangentially into the system at a velocity and magnetic field B such that the orbit equation is satisfied, and then accelerated to higher and higher energies by applying an RF field of increasing frequency while increasing the value of the magnetic field commensurately so as to maintain them in the same circular orbit. Such a scheme allows for the relativistic mass increase of the particles with energy, which limited the attainment of earlier types of accelerating machines.

Radiation from these orbits has come to be known as synchrotron radiation. In electron synchrotrons, it is particularly strong and the principal mechanism by which the circulating beam loses energy. The electrons travel with speeds very close to that of light, and the radiated photons are emitted tangentially in a forward direction in a very narrow cone — for 1 GeV electrons, the semi-angle of this cone is only about 0.2°, so the radiation is highly collimated. From the whole circulating beam, therefore, there is a pattern of emission that resembles the spray of water that is thrown off the

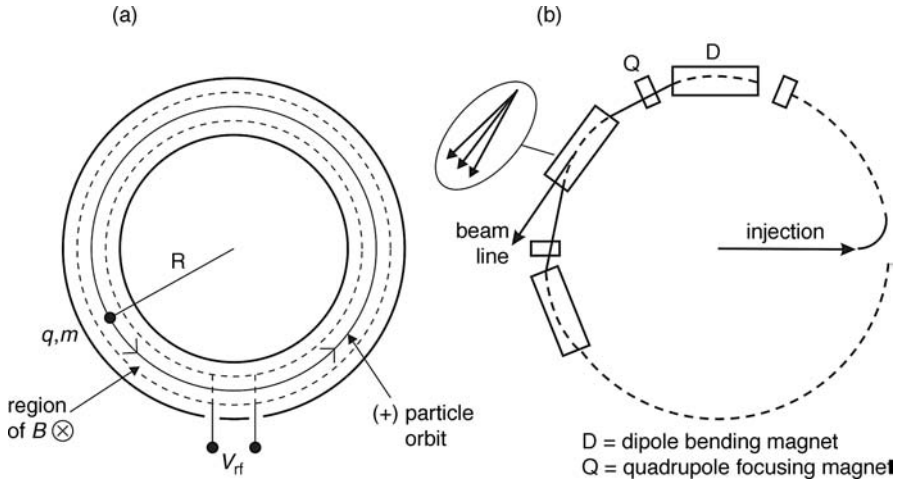

FIGURE 9.10
(a) Principle of synchrotron, (b) practical arrangement with curved and straight sections.

rim of a wet wheel rotating at high speed. It occupies a thin disc of space that extends radially outwards in the plane of the circulating charges. The emitted radiation is intense and spans a broad spectral range, typically from millimeter infrared wavelengths through the visible and UV to soft and hard X-rays, corresponding to photon energies less than 1 eV up to more than 10 keV. Furthermore, it is highly polarized in the plane of the orbit. Typical synchrotron sources are brighter than traditional X-ray tubes by a factor of order 10^5 or more.

In recent decades, a number of electron synchrotrons have been built specifically to provide radiation for scientific investigations on the structure of materials. They have beam energies in the range 1 to 10 GeV and require vacua of about 10^{-10} mbar so that the high-energy beam, once established, can circulate continuously to provide a steady supply of well-characterized radiations for a number of hours. They are described as electron storage rings. The lifetime of the beam is determined by the loss of electrons due to their scattering by residual gas molecules. The RF voltage maintains the energy of those remaining against the radiation losses. It is in the nature of the processes involved that the current is bunched, consisting of very short pulses of less than a nanosecond duration and this is therefore also true of the output radiation.

While Figure 9.10(a) illustrates the principle, Figure 9.10(b), though highly simplistic, is more representative of matters in practice. Storage rings are not simply circular but are made up of alternate curved and straight sections. In the curved sections, dipole magnets cause beam-bending and therefore synchrotron emission as described in the preceding text. The limited arc length means that the radiation is in the form of a thin fan shape. The straight sections accommodate the primary beam injection arrangement and

accelerating cavities and allow the insertion of special devices that enhance performance. Also located externally at these sections are magnetic quadrupole lenses that, allied with the focusing role that the dipole magnets have for electron trajectories close to the central orbit, are part of the beam focusing and control system. The focusing role of the dipole magnet arises by making their pole faces slightly divergent rather than parallel, so that small changes in the vertical field are created in a direction transverse to the beam. Because the vacuum tube has to pass between the poles of magnets, its cross section is relatively small and typically elliptical with vertical minor axis of a few centimeters, and the horizontal major axis rather more.

A fraction of the fan-shaped radiation from a particular dipole magnet may be selected by apertures and extracted into a tangentially directed beam line, while the rest impinges on the vacuum walls. There are, therefore, a number of junctions made with the main vacuum ring to allow this. Inside a beam line are housed apertures, mirrors, monochromators, and other optical elements appropriate to the wavelength components of the selected radiation. Beam lines are dedicated to specific purposes, infrared, ultraviolet, and various x-ray investigative techniques, diffraction, microscopy, and grazing incidence diffraction, for example. Woodruff and Delchar (1986) describe particular techniques that are valuable in surface science, and Field (2001) discusses further applications over a wider range.

The Synchrotron Radiation Source (SRS) at Daresbury, U.K. (www.srs.ac.uk) is an example of a machine offering world-class facilities to a large community of researchers in diverse fields. The main storage ring is 96 m in circumference and contains 16 bending magnets, each 2 m long. Inside it a current in the range of 150 to 300 mA circulates at an energy of 2 GeV. The injection system consists of a 12 MeV linear accelerator that feeds pulses of electrons into an intermediate booster synchrotron, accelerates them up to 600 MeV, and injects them into the main ring. When sufficient current has been built up at this energy, it is accelerated up to 2 GeV. Useful currents are maintained for 20 h or more, and 36 experimental stations are supplied with electromagnetic radiations of various wavelengths. Some of these are obtained by the insertion of "wigglers" and "undulators" in the straight sections to produce enhanced effects. In these devices the electron beam is made to pass through magnetic field configurations that, in the former case, locally force it to be highly curved with the enhancement of radiation brightness at very short wavelengths. In the latter case, a periodic field variation leads to interference and the emergence of a number of very bright, almost monochromatic, beams.

The vacuum system of the SRS is rather complex, and aspects of its design and early performance were described by Trickett (1987). Reid (1997) has given a brief account of some aspects of its operation more recently. It is an all-metal UHV system in which scrupulous attention is paid to the avoidance of hydrocarbon contamination, which would have serious effects in beam scattering and the degradation of surfaces. The main ring is divided into sections that can be isolated by gate valves, and the beam lines can be

similarly isolated. The machine stays at UHV for long periods during which many users simultaneously carry out experiments on a nonstop shift basis at its various stations. The pumps for ultrahigh vacuum are distributed ion pumps at the bending sections that utilize the field of the bending magnets in an extended pumping cell parallel and adjacent to the beam path. In the straight sections there are large 400 l s^{-1} ion pumps on top of which are mounted titanium sublimation pumps. Sections are brought to atmosphere rarely, but when this is necessary, re-establishment of UHV starts with oil-free rough pumping by diaphragm pumps and turbomolecular pumps with magnetic bearings. Items to be inserted into the system are rigorously cleaned, subjected to prior vacuum degassing in special vacuum furnaces, and also glow-discharge cleaned.

Pressure is sensed at places distributed throughout the systems by BAG gauges that have to be carefully located so that their function is not impaired by the various fields and the electron beam. Quadrupole residual gas analyzers are also strategically located to monitor the system and give warning of undesirable changes such as the opening of small leaks. The pressure around the ring is not uniform because there is some distance between pumps, and the pipe conductance is limited by its cross section and the presence of apertures. A pressure profile as a function of distance therefore shows minima nearest to pumps and maxima remote from them. The profiles with and without the beam are considerably different because when the beam is on it, it produces a large gas load due to photon stimulated desorption of molecules off the wall in the vicinity of the bending regions where the high photon flux of the radiation impinges. This dynamic effect is the main gas load in electron storage rings. Fortunately, it progressively diminishes with accumulating photon dose to acceptable levels, an effect referred to as "beam cleaning." This can frequently be exploited in restoring UHV conditions when such occasions arise, rather than the lengthier procedure of sector bakeout. The modeling of the photon-induced gas load is therefore a crucially important aspect in the design of the vacuum systems of an electron storage ring. The vacuum in the beam lines has of course to be compatible with that in the ring, except in the case of beam lines that can be isolated from it by physical windows that are transparent to the radiation, as is beryllium for high energy x-rays. The vacuum conditions are then slightly less stringent.

The pressure measured in the straight sections of the ring in operation can be related to the mean pressure around it, which determines the beam lifetime. The state of this very complex machine is monitored by numerous pressure, temperature, and other sensors together with status indicators for valves, pumps, etc., and associated interlocks. The strategy for the control of the machine by computer and its implementation was first discussed by Reid (1982). Recent developments in this area are described on the Web site.

The SRS was the first large high-current synchrotron built as a dedicated x-ray light source, and its value to the scientific community has paved the way for the design and building of Diamond, a third-generation synchrotron

source, and currently U.K.'s largest scientific project, whose first beam lines are scheduled to come into operation in 2006. This machine will be 560 m in circumference, with beam energy 3 GeV, current 300 mA, and sophisticated facilities at many user stations. Details may be found at the web site www.diamond.ac.uk

9.8 Controlled Thermonuclear Fusion

In the long term, on a scale of the order of 100 years, nuclear fusion arguably offers the only sustainable means for the generation of electric power in a quantity sufficient to satisfy the world's needs. Solar, wind, and wave power can make valuable but limited contributions to the energy budget, and fossil fuels, reserves of which are finite, become increasingly scarce and costly, with oil needed for purposes other than energy generation. In addition, their burning creates soundly based environmental concerns. Electricity generation by nuclear fission, once thought to be the solution to the perceived long-term problem, is no longer so regarded for reasons partly political and partly technical, particularly those to do with long-term storage of its radioactive waste products.

Although the technological challenges of controlled thermonuclear fusion and capital cost of research facilities are enormous, and many challenges remain, the fact that the ingredients that fuel the reaction are abundant in nature and easily acquired, that there are no long-lived radioactive waste products, and that considerable progress has already been made, are powerful incentives for the furtherance of the work to its long-term goal. The underlying principles of the subject and its current state are described in a recent book by Harms et al. (2000), and an essay by Pert (2002) reviews past work and future prospects. Research in this subject has been under way since the 1950s, and from nationally financed activities initially, often with military connections, it has evolved to one of substantial open international collaboration. Designs and planning for ITER, the International Thermonuclear Experimental Reactor, are well advanced.

Of a number of thermonuclear reactions between light nuclei that yield products of less total mass and the consequent release of energy equivalent to the mass difference, attention is concentrated on that between the hydrogen isotopes deuterium D and tritium T to produce an alpha particle and a neutron, in which 17.6 MeV of energy is released as kinetic energy of the products:

$$ {}_1\mathrm{D}^2 + {}_1\mathrm{T}^3 \rightarrow {}_2\mathrm{He}^4\ (3.5\ \mathrm{MeV}) + {}_0\mathrm{n}^1(14.1\ \mathrm{MeV}) $$

For this reaction to occur, the positively charged D and T nuclei must approach each other with sufficient speed to overcome the strong repulsive Coulomb force. For a gaseous deuterium–tritium mixture, this requires that

the temperature be of the order of 10^8 K, so hot that the mixture is completely ionized to form an electrically neutral plasma of intimately mixed positive ions and electrons. Heating and confining the plasma are problematic. In order to have a sustained reaction in an ignited plasma, such that the energy released is greater than that lost from it by radiation, a criterion due to Lawson has to be satisfied, namely, that the "fusion triple product" of ion density N, energy containment time t_E, and temperature T exceed a specified value. The value of t_E is determined by the rate of energy loss from the plasma by radiation. At temperature 10^8 K, the product $N\,t_E$ has to be greater than about 2×10^{20} m^{-3} s for the net release of energy. For a t_E value of 1 second, the N value is that associated with gas at 10^{-2} mbar (see Table 3.1). Since the early 1980s, experimental progress towards this goal has been quite rapid, with increases in the realizable values of N, t_E, and T. Values of the triple product have increased by many orders of magnitude and are now only a factor of about 10 below that for ignition. The ITER machine previously mentioned is intended to take investigations into the realm of ignited plasmas. Ultimately, the generation of electricity will be achieved in plant that surrounds the burning plasma with a lithium blanket. While the energetic alpha nuclei generated will interact strongly with the plasma to oppose cooling losses by radiation and maintain its temperature, the neutrons will escape to be absorbed in the blanket. Here they have two functions: by further nuclear reactions to produce more tritium that is sent back to feed the reaction, and, from the heat generated as they are slowed down, to raise steam for the electrical power plant. Helium nuclei that would otherwise accumulate in the plasma from the spent D/T fuel constitute "ash" and have to be removed.

Containing and controlling the hot plasma against instabilities remains one of the central problems of the activity, and since the earliest days, strong magnetic fields in various special configurations have been investigated for this purpose. The considerable progress of recent decades has been due to the adoption of the tokamak configuration in which the plasma is contained in a toroidal (essentially doughnut-shaped) vacuum vessel, subject to a toroidal magnetic field that is augmented by other stabilizing components, such that the charged particles tend to spiral around field lines. The diagrams to be found in texts need to be consulted to appreciate how the field affects particle trajectories. The plasma is created by pulsed transformer action — electric fields induced by rapidly changing magnetic fields initiated by a large current pulse through an external, primary winding cause ionization of a low-pressure D/T gas mixture and its ohmic heating. The plasma pulse lasts for seconds or more, in which interval the diagnostic measurements are made under sophisticated computer control. In this interval, higher temperatures may be achieved by additional RF heating and, in some cases, the injection of beams of energetic neutrals tangentially into the plasma. Such beams are created by partial neutralization of energetic ion beams directed towards the plasma, with the deflection of the non-neutralized fraction electromagnetically into a dump.

To illustrate the scale and operation of these large fusion experiments, we will consider the vacuum system of the JET tokamak. The acronym signifies Joint European Torus, reflecting its funding by the European Communities, now transferred to the European Fusion Development Agreement, EFDA. First operated in the early 1980s and, like related experiments in the U.S., Japan, and elsewhere, continuously developed in advancing programs of work to discover the complex physics of high temperature plasmas and their confinement, a full description of the project will be found at www.jet.efda.org. The vacuum systems are described by Duesing (1987).

The JET toroidal vessel is a very large ultrahigh vacuum system with major radius of about 3 m and D-shaped torus cross section of height 4.2 m and minor radius 2.6 m as indicated in Figure 9.11. The need for ultrahigh vacuum conditions, commensurate leak tightness, and special attention to wall conditioning in order to minimize outgassing products and control their identity is discussed below. The torus, constructed from eight octants welded together, has an all-metal double-walled structure with numerous access ports of various sizes to allow for diagnostic probes, robotic manipulation, additional heating, and other services. The independently pumped interspace between the relatively thin inner and outer walls, appropriately braced for strength, facilitates leak detection and allows the circulation of hot helium gas for bakeout. The surface area of the inner wall is about 1000 m^2 and the torus volume 189 m^3. Two chambers C, one of which is indicated, have horizontal entry to the torus via 1.2-m diameter ports and at their base carry 400-mm inlet turbomolecular pumps T, with matching all-metal gate valves V, backed by primary pumps. The total effective pumping speed at the torus is 6000 l s^{-1}. Prior to this secondary pumping, the rough pumping from atmosphere to 0.1 mbar, using several pumping stations each with capacity 2000 m^3 h^{-1}, takes less than 2 h. The torus wall is degassed by heating it at up to 500°C by blowing hot helium through the interspace, and in operation the wall is held at 300°C. In addition, surface conditioning by extensive glow discharge cleaning is carried out, after which partial pressures of H_2O, CO, CH_4, and C_2H_4 are less than or of order 10^{-9} mbar, with a hydrogen partial pressure about 10^{-7} mbar. Higher hydrogen partial pressures are tolerable because the torus is backfilled with a very pure mixture of its D and T isotopes to a pressure of about 10^{-1} mbar prior to the creation of the plasma. The pulsed operation of a tokamak is a relatively violent event. The large toroidal currents induced, 5 MA or more, impose large mechanical forces of magnetic origin on the structure, and occasionally if plasma reaches the wall inadvertently, where it is quenched, a small perforation may result. An ingenious scheme of sectioning the interspace to detect such leaks in the inner wall is described by Orchard and Scales (1999).

One of the most critical aspects of operation is to minimize the presence of impurities in the background gas and their entry into the plasma by wall outgassing and by the surface interactions that occur when the hot plasma

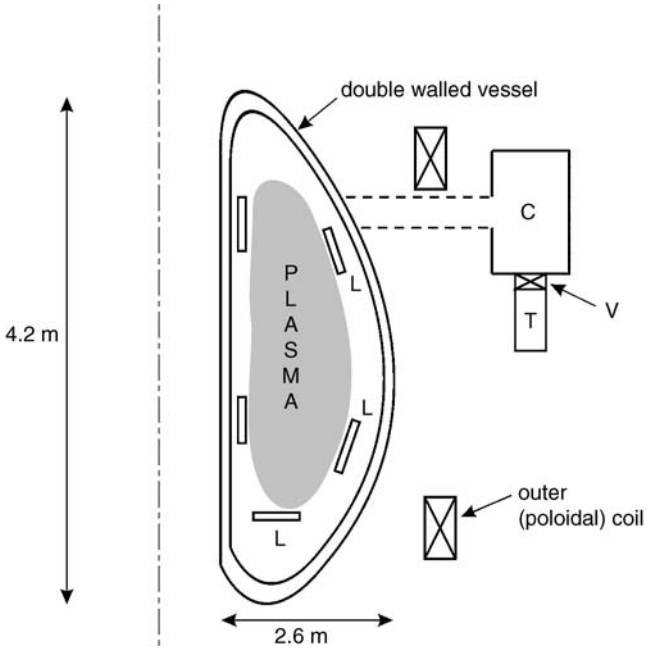

FIGURE 9.11
Schematic diagram of torus cross section.

touches the plates L that serve as limiters to define its shape and keep it away from the inner vessel wall. This is because energy loss from the plasma, as well as due to a small amount of actual particle loss, is principally by radiation, predominantly Bremsstrahlung (braking radiation) due to the inevitable accelerations and decelerations of electrons in the plasma as they move in the electric fields of positive ions. This radiation scales with the atomic number as Z^2. Low Z materials are therefore necessary for the limiter surfaces that define the "first-wall" with which the plasma comes into direct contact. Graphite ($Z = 6$) and Be ($Z = 3$) are used, the latter where the plasma flux is especially high. As noted by Dylla (1998), in an article on the evolution of large machines for particle physics and fusion experiments, it was advances and collaborations in solving first-wall problems that have enabled machines such as JET, TFTR in the U.S., and JT60U in Japan to achieve plasma parameters sufficiently near to the Lawson criterion minimum to justify the next phase of development of the significantly larger and more expensive ITER machine.

9.9 Gravitational Wave Detection

Optical astronomy, our window on the universe for some hundreds of years, has been augmented in the last half century by others that use different parts of the electromagnetic spectrum, such as infrared, x-ray, and radio waves, with a consequent dramatic increase of knowledge and understanding of the universe and its origins. Observatories that detect gravitational rather than electromagnetic waves are expected to become operational in the early years of this millennium, and to extend our knowledge in new and fundamental ways.

The emission of gravitational waves from accelerating mass was predicted early in the last century by Einstein in his General Theory of Relativity. They may be visualized as distortions of spacetime that travel away from their source with the speed of light. They have not yet been directly detected, but there is strong indirect experimental evidence for their existence. It comes from observations of the orbital frequency of the Hulse–Taylor binary pulsar PSR1913+16, which consists of two neutron stars orbiting each other under mutual gravitational attraction. As they slowly spiral inwards, falling towards their common center of mass, the gradual increase of orbital frequency, detected by periodic Doppler effect shifts in the frequency of the radio wave from the pulsar, is exactly as predicted by the theory of gravitational wave emission. Other sources of gravity waves are catastrophic events, such as supernovae explosions and the collapse of stars to black holes. There is also a general low-level stochastic background that is a legacy of such events throughout the universe. A general introduction to the subject will be found in the article by Barish and Weiss (1999). The character, sources, and detection of gravitational radiation are extensively discussed in a review by Ju et al. (2000).

Gravitational waves have yet to be directly detected because the size of the effects to be measured is exceedingly small. It is estimated that a strong gravitational wave signal originating in our own galaxy would create a space strain of only 10^{-21}. In a distance of 1 km, this corresponds to a change of 10^{-18} m, smaller by a factor of about 1000 than the size of a nuclear particle such as a proton! Making instruments sufficiently sensitive to detect them has been the preoccupation of a small community of scientists for some time. In the pioneering work of Weber in the 1960s on resonant mass detectors, a suspended heavy mass was designed to vibrate resonantly in a detectable way when excited by the passage of a gravitational wave, responding rather like a tuning fork at just one frequency near 1 kHz from an excitation with a broad range of frequency components. Developments of this sort continue, but in the last few decades attention has become increasingly focused on

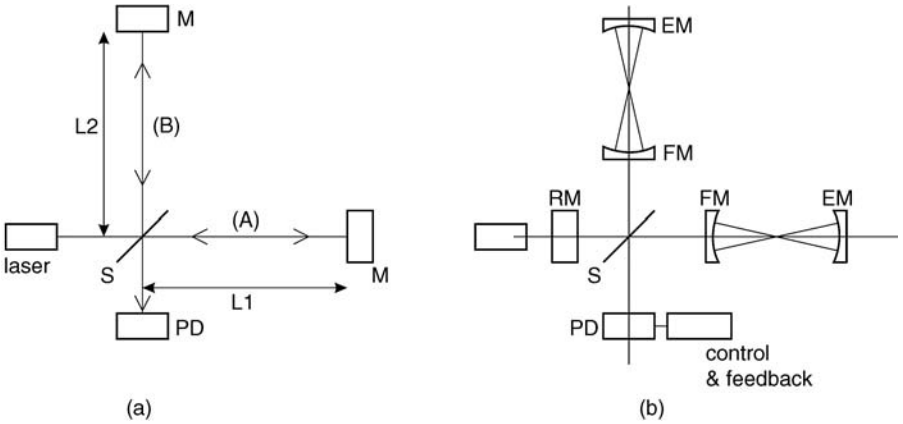

FIGURE 9.12
(a) Interferometer principle, (b) with refinements.

laser interferometer detectors based on the principle of the Michelson interferometer. In contrast with resonating masses, these will be broadband instruments, able to detect disturbances over a range of frequencies, from a few Hz to several kHz.

The basic principle of the measurement is illustrated in the diagram of Figure 9.12(a). A beam of laser light incident at 45° on a beam splitter S produces two beams of equal intensity, one transmitted (A), the other reflected (B), which travel in orthogonal directions to mirrors M which are part of freely suspended but otherwise mechanically isolated heavy masses, at equal distances L of the order of 1 km from S. The suspended masses are free to move laterally in the direction of the light beams. The mirrors are orientated to return the beams back along their incoming paths, so that part of returning beam B is transmitted through S to combine with the reflected part of returning beam A at the photodetector PD. If the optical paths L are equal then, because of the 180° phase change of beam A reflecting off S, the beams at the photodetector interfere destructively and there is a "dark fringe" — no light output. The passage of a gravitational wave, by first slightly lengthening one arm and shortening the other, and then *vice versa*, makes the times of travel in the interferometer arms unequal, creating a phase change between the recombining beams and hence a light output at the detector. Such is the basic principle. Creating a corresponding instrument in the context of such weak signals is a formidable task. Numerous sources of noise exist that mask the signal. For example, the instrument must be isolated from mechanical disturbances and seismic noise of the ground that supports it, the laser frequency must be stable and pure, and the random, thermally induced fluctuations in the refractive index of the gas through which the light beams pass must be reduced to acceptable levels. Theory indicates that a vacuum of about 10^{-9} mbar is necessary to achieve this. Housing the instrument in vacuum also reduces both light loss by scattering

and the molecular bombardments of the suspended masses that contribute to their thermal motion. The vacuum-related issues of their design were first discussed in a paper by Bennett (1987) on the proposal for a British long-baseline gravitational wave detector, activity that has fortunately been continued by combination with German work in the GEO collaboration listed below.

Worldwide, six interferometers are being built. They are:

LIGO (U.S.) The Laser Interferometer Gravity-Wave Observatory, consisting of two identical instruments with 4 km arms located at Hanford, Washington State, and Livingston, Louisiana, over 3000 km apart, (www.ligo.caltech.edu)

VIRGO (near Pisa, Italy), a French–Italian collaboration, 3 km arms (www.ego-gw.it)

GEO600 (Germany), a German–British collaboration, 600 m arms (www.geo600.uni-hannover.de)

TAMA300 (Japan), 300 m arms (www.tamago.mtk.ano.ac.jp)

AIGO, Perth, W. Australia, 80 m arms (www.gravity.pd.uwa.edu.au)

The design of these instruments involves refinements of the basic concept of Figure 9.12(a) in various ways, some of which are shown in Figure 9.12(b). A substantial increase in sensitivity is achieved by incorporating partially reflecting mirrors FM in each beam to create Fabry–Perot optical cavities with the end mirrors EM. The beams thus pass to and fro many times, effectively increasing the interferometer length. In addition, the insertion of a partially reflecting mirror RM allows beam energy that would otherwise be returned to the laser source and wasted to be recycled, thus building up the energy in the optical fields. Because of the increased intensity, the statistics of photon counting at the detector is improved and hence the sensitivity. The control and sensing systems are sophisticated. A null detection technique is employed in which the instrument is servo-controlled to lock on to a dark fringe by means of minor adjustments made to the mirror positions by actuators. When there is a disturbance, such as the passage of a gravity wave, the error signals that drive the actuators to maintain the null give a measure of its magnitude. The mirrors used are of exceptional state-of-the-art quality, massive, and of about 30-cm aperture in order to limit diffraction broadening of the beams. The design of the vibration-isolating supports is particularly crucial, involving cascaded stages of isolation to give progressive attenuation of input disturbances originating at the ground.

The vacuum systems, essentially L-shaped tubes, may be illustrated with reference to that of the 4-km Hanford LIGO instrument, whose vacuums, like those of TAMA and GEO, are operational at the time of writing (2003). Each arm consists of sections of stainless steel corrugated bellows of diameter 1.2 m welded together and 4 km long, in two 2-km sections to allow intermediate stations at 2 km for testing and development purposes. The 1.2-m

cross-sectional diameter allows the multiple passage of the laser beams, which individually are about 10 cm in girth. The total volume in each arm is 4500 m^3 and the internal gassing area 1.6×10^4 m^2. Gate valves isolate the tube sections from the corner station and the intermediate and end stations in order that the vacuum chambers at these locations that contain the optical elements can be serviced without venting the whole instrument to atmospheric pressure. At various stages along the tube lengths, cryopumps provide the secondary pumping. Bakeout lasts for 3–4 weeks and is accomplished by wrapping a thick layer of fiberglass insulation around the 2-km section of the tube and passing about 1800 A of current lengthwise through the vacuum wall using large magnet power supplies. This ohmic heating raises the temperature to 150°C or more. After bakeout, the beam tube pressure achieved is the required 10^{-9} mbar, and the outgassing rate, determined by isolating the system from its pumps and measuring the pressure rise Δp in a time Δt, ($V\,\Delta p = Q_T\,\Delta t$), was shown to be less than 10^{-13} mbar l s^{-1} per cm^2, dominantly of hydrogen.

As these instruments come into operation, they will be progressively developed to lower their "noise floor" and increase their signal-to-noise performance. The correlation of signals between different detectors, already part of the planning for the two LIGO instruments, will be an essential aspect of validating data globally. Extraterrestrially, and rather further into the future, NASA and ESA have an ambitious joint project proposal for a Laser Interferometer Space Antenna (LISA), a three-arm interferometer consisting of spacecraft in the fixed formation of an equilateral triangle of side 5 million km. The vacuum comes free, but the rest is formidably expensive.

9.10 Particle Physics: the LEP and LHC

The nature of matter has been a primary concern of philosophers since the earliest times, and modern science has brought our knowledge of the subject to a highly refined, though still dynamic, and far from final, state. The quest to identify and understand the elementary particles of which matter is composed and the forces that act between them is a continuing search involving the interplay of theory and experiment. The theory is at a very high level and the experimentation extremely sophisticated and expensive. Undergraduate texts on "modern physics" give introductory accounts of these matters, and the Oxford Dictionary of Science has an extended and informative entry on "elementary particles." The reader will find the book of Zee (2000) a good general resource on all these matters.

Quite recently, in the 1960s, there were thought to be four fundamentally distinct forces in nature, namely, the gravitational force, the electromagnetic force between charged particles, and the strong and weak nuclear forces, the former involved in binding nuclear particles together in the nucleus, the

latter in the process of radioactive decay. But experiments at CERN and elsewhere, mentioned in the following text, have reduced this number to three by establishing that the electromagnetic and weak nuclear forces can be regarded as different aspects of just one electro-weak force. Gravity is by many orders of magnitude the weakest and remains outside any unifying scheme, but the electro-weak and the strong nuclear forces are very elegantly incorporated into the current "standard model" of particles and their interactions. In this quite complex scheme, matter particles are made up of two basic types of elementary particle, quarks and leptons, and the forces between them are transmitted by "messenger particles," which in the nature of their role have only a transient existence. In this scheme of things, particles made up from quarks that interact by the strong force are called hadrons. A missing link in the model relates to how particles have mass. It is thought to come about by interaction with the Higgs field, though the Higgs boson that mediates it has yet to be discovered.

The experimental search for these transient particles has three aspects: the acceleration of stable particles to high energies to collide with others, the detection of new particles that may be produced as a result of the collision, and the data-processing systems needed to analyze the large amount of information from the detectors. We will deal with only the first. The possibility of creating new particles in interparticle collisions arises because of the equivalence of mass and energy, as proposed by Einstein in his famous equation $E = m \times c^2$. A particle of mass m_0 at rest has rest energy $E_0 = m_0 \times c^2$. It is 0.51 MeV for the electron and 940 MeV = 0.94 GeV for the proton, for example. According to relativistic mechanics, mass increases with velocity and the total energy $E_T = m \times c^2 = m_0 \times c^2 + E_K$, where E_K is the kinetic energy. The relativistic momentum is $p = m \times v$ and total energy and momentum are conserved in collisions. In machines in which an accelerated beam is made to collide with the nuclei of a stationary target, the conservation of momentum requires that an appreciable fraction of the available energy is consumed in giving momentum to the products. However, in colliding beam machines, two beams of the same energy and momentum but traveling in opposite directions are made to collide almost head-on so that virtually all the energy available, i.e., about twice the beam energy, is available for the creation of new particles. The Large Electron Positron Collider at CERN in Geneva (the LEP), operational from 1989 to 2000, was such a machine. It was built to study in detail the W_+, W_-, and Z_0 particles, the messenger particles of the weak force, which had been discovered in 1983 in proton–antiproton collisions investigated in the super proton synchrotron (SPS) at CERN. These particles have masses W = 80 GeV and Z_0 = 91 GeV, in contrast with the massless photon that mediates the electromagnetic force.

In the LEP, beams consisting of bunches of electrons and positrons circulated in opposite directions at high energy in a vacuum system 27 km in circumference, crossing at four collision stations where detectors were located to study their interactions. The particles were injected into it at lower energy from smaller machines. The principle of the synchrotron accelerating

mechanism and the concept of the storage ring were introduced in Section 9.7. The vacuum pipe was a storage ring for the two beams, consisting of alternating curved and straight sections with external dipole magnets to guide them, magnetic quadrupole fields for focusing, electric fields to steer them, and radio frequency cavities to accelerate and maintain their energy against losses. The four interaction regions were just 7 m long. In the early experiments, the beam energy was 50 GeV, giving a collision energy of 100 GeV. The machine was housed in an underground tunnel straddling the border between France and Switzerland. Its large size was dictated by the need to limit energy losses by synchrotron radiation in the bending arcs, the principal loss mechanism, to an acceptable level by having the largest practicable radius of curvature, which was about 3 km. The bending arcs occupied more than three quarters of the 27 km circumference.

Creating the vacuum of about 10^{-12} mbar that is necessary for beam lifetimes of many hours duration was achieved in the straight sections, which were made of stainless steel, by sputter ion-pumps augmented by titanium sublimation pumps. In the beam-bending sections, where photon-induced gas loads were generated by synchrotron radiation striking the pipe walls, the beam pipe was made up of extruded aluminum sections 12 m long and pumped by a continuous strip of nonevaporable getter (NEG, Section 6.4.5), with some supplemental localized sputter-ion-pumping for inert gases and methane. The beam pipe cross section is shown in Figure 9.13, the shape reflecting its location between the poles of dipole magnets.

The aluminum chambers incorporated beam and pumping regions, together with channels for water cooling during operation. That part in which the beams circulate was elliptical in section with height 70 mm and width 120 mm approximately, connected by a continuous pumping slot to the region in which the NEG pumping strip was supported by insulating ceramic pillars. The strip was a constantan ribbon 0.2 mm thick and 30 mm

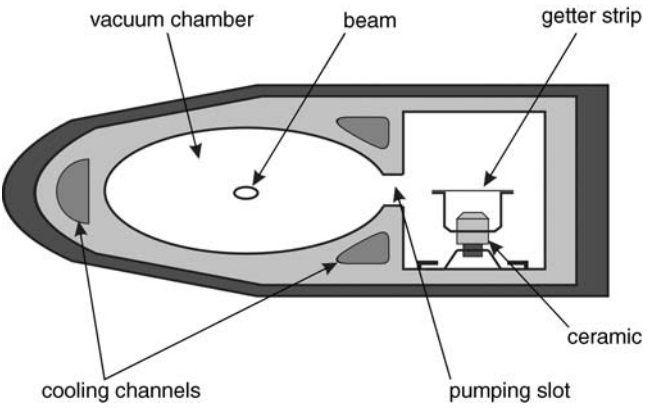

FIGURE 9.13
LEP beam pipe cross section.

wide, coated with the pumping alloy; electrical feedthroughs at the end of each section allowed the periodic activation of the NEG. In the nature of NEG operation, its continuous activation was unnecessary and the magnetic field generated would have interfered with the beam optics. Initial bakeout at 150° was carried out by circulating superheated water at high pressure through the channels. Sector gate valves split the system into manageable sections. Many metal joining gaskets, stainless steel bellows to allow for thermal expansion and fine positioning, monitoring gauges, and other sensors were distributed around the system that was managed under computer control. An important element in the building strategy of the machine was to fully test all sections in the laboratory prior to their installation below ground. The CERN Web site www.cern.ch gives information about the machine's design and operation.

The machine operated very successfully throughout its 11-year life. In the first 5 years, at about 100 GeV collision energy, the very short-lived Z_0 (91 GeV) particle was extensively studied by its decay properties. Among the huge total number of collision events with their various outcomes, more than 17 million Z decay events were identified and confirmation given to the scheme of the standard model. In 1995, the machine was upgraded to higher energy by the installation of superconducting rf cavities that enabled the collision energy to be raised to 160 GeV, under which conditions W+ and W– pairs could be created. The W particles, mass 80 GeV, were produced in large quantities, and the work on their properties brought the understanding of the weak force to a very satisfactory state, confirming the character of the electro-weak force in the standard model.

In the latter stages of operation, the continuing search for the Higgs particle, which had given some tantalizing hints of its presence in the data, caused the machine energy to be raised first to 103 and finally to 104.5 GeV per beam, a record and a very considerable enhancement of the machine performance over its initial design values. At this stage, 288 superconducting cavities had been installed and the search was centered on a Higgs particle at 114 GeV. Despite extensions of the work in the last stages, however, the information gained was not conclusive, and the decision was taken, not without controversy, to shut down the LEP so that the planned schedule for the building of the new device to occupy the tunnel, the large hadron collider (the LHC), should not be delayed.

The LHC, which has been in planning since the mid-1980s, is currently building and due to become operational in the spring of 2007 in the old LEP tunnel. Designed to accelerate countercirculating proton beams to 7000 GeV = 7 TeV energy with current 540 mA and collide them, its purpose is to probe further into the nature of matter, find the Higgs boson, search for so-called supersymmetric particles, and throw light on the matter–antimatter imbalance that exists in the universe. The 7 TeV beams will be accelerated to this energy from proton beams injected at 450 GeV from subsidiary machines. The beam system geometry is closely similar to that of the LEP, with crossing of the two beams at four interaction stations, as shown in Figure 9.14(a), but

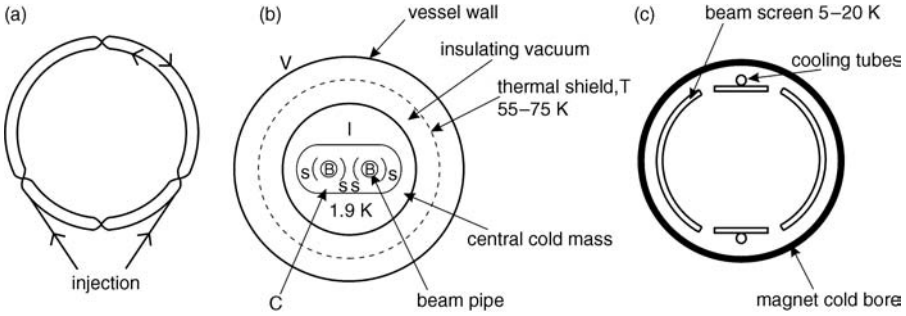

FIGURE 9.14
(a) LHC intersecting beam paths, (b) simplified cross section, (c) 1.9 K cold bore.

the vacuum system that houses the beams (Gröbner, 1996) is entirely different, as will be described in the following text. To guide proton beams of this energy around the circular path of the machine requires magnetic field strengths of 8.4 T (84,000 Gauss), and to achieve this, superconducting magnets operated at 1.9 K are necessary. Therefore, cryogenic engineering is involved, that can also be exploited in the pumping. The central cold mass that is held at 1.9 K by the circulation of superfluid helium is shown in a simplified schematic cross section in Figure 9.14(b), which is abstracted from full-color diagrams on the CERN Web site that are to scale and give informative detail. It is appropriate to recall that CERN invented the Web as a means of communication; the detailed description of the LHC available there should be consulted. The central mass consists of an iron yoke I surrounding a nonmagnetic collar C that contains the two beam pipes B. These are 50 mm in bore and 194 mm apart. Superconducting coils S are located in pairs as shown and are individually 14 m long. Heat exchange with liquid helium in a superfluid state at 1.9 K makes thermal conduction very efficient.

The vacuum vessel V encloses this central assembly, providing an outer vacuum of about 10^{-6} mbar for thermal insulation to minimize heat input to the central mass. This is assisted by the presence of a thermal shield T held at 55–75 K and a radiation reflecting surface on the central mass. The outer vacuum will be created initially by turbopumps but maintained by cryogenic pumping onto the central mass once the running conditions are reached. The arcs and cryogenic zones in which the beams are bent occupy more than 24 of the machine's 27-km circumference. The remaining straight sections that contain the beam accelerating and control devices, and the detection regions where the beams emerge from their separate pipes to cross, are at room temperature. It is estimated that the initial cool down of the massive cryogenic assembly will consume 12 million l of liquid nitrogen and that the total amount of liquid helium in use in the cryogenic systems will be 0.7 million l.

The vacuum in the beam pipes has to be the equivalent of 10^{-9} mbar or less to allow the beams to circulate with useful currents for times of the order of a day. The 1.9 K cryogenic surfaces provide ample pumping speed, but

they also have to absorb the energy associated with the passage of the beam, and although this amounts to less than 1 W per meter, it would present a substantial load on the cooling system that is to maintain the 1.9 K temperature. This load arises from various sources:

1. Synchrotron radiation (much smaller for protons than electrons)
2. Dissipation of the image currents induced in the walls that enclose the beams
3. The impact of photoelectrons and secondaries that are accelerated back and forth in the strong electric fields set up by the passage of the proton bunches of the beams
4. The high kinetic energy particles that are lost from them by scattering

The elegant solution to this problem is to provide a perforated beam screen to fit closely inside the 1.9 K cold bores as shown for one beam in Figure 9.14(c). The screens are made of thin 1-mm stainless steel coated on the inside with copper and held at 5–20 K by gaseous helium passing through attached cooling tubes. These surfaces prevent the energy generated from reaching the 1.9 K surfaces of the cold bore, while the molecules deriving from scattered beam particles and those desorbed by synchrotron radiation eventually migrate safely via the perforations to the 1.9 K surfaces where they are pumped by condensation.

An aspect of the pumping analysis that is distinctly different from the usual methods is that for the cold-beam tubes it must be carried out in terms of molecular number densities n in the first instance rather than room temperature pumping speeds, pressures, throughputs, and volumetric flow rates. Vacuum conditions are not measured by the usual gauges but are inferred from gas density measurements based on residual gas ionization in suitably biased electrode configurations.

Like the LEP before it, the LHC is an extraordinarily complex machine that stands at the forefront of modern science, technology, manufacturing, and project management. It is international in the scope of both its human and material resources and promises large advances in our understanding.

10

Vacuum Components, Construction, and Leak Detection

10.1 Introduction

In approaching a vacuum system the user needs to be aware of how its various parts — the chamber with its access port and door, pumps, valves, pipes, gauges, etc. — are connected in a way that seals the vacuum in the system from the atmosphere outside. Also, and especially if contemplating the design of items to be used in vacuum (such as a sample holder, for example), the user should be aware of the limitations imposed on the choice of materials that can be used.

In most cases, experimental actions inside a vacuum system are made possible by a range of flange-mounted feedthroughs, many of them available from manufacturers. Fitted to suitably located ports, these feedthroughs allow services to be taken through the vacuum wall. These services include mechanical motion of various kinds (linear, rotary, and their combinations), the piping in and out of fluids (cooling water, for example), and electrical feedthroughs, of which there is a great variety. Thus, there are electrical feedthroughs for signals at low voltage (coaxial and multipin), for thermocouple voltages, for high currents of many amperes at low voltage (to produce resistive heating, for example), and, conversely, for high voltages (kV) at low current as might be employed to set electrode potentials. Windows made of glass or sometimes of other materials allow the entry and exit of laser or other light beams in some applications. But, more generally, windows (or "viewports," as they are usually called) are important and arguably essential because they allow visual inspection of devices and monitoring of processes occuring in the vacuum chamber.

Devices of the type mentioned above and others of a specialized nature are available from many manufacturers, whose Web sites and catalogs should be consulted for the full range of possibilities. In the U.K., Caburn and Thermovac, for example, offer a very wide range of such components. In some applications, however, investigators may have to design and make (or have made in the workshop) items to their own specifications; sometimes

a commercially available item may be too expensive. In these cases designs reported in the literature may be consulted, particularly those in the journals *Vacuum* and the *Journal of Vacuum Science and Technology.* The structured index of the latter is particularly helpful in this regard.

A large body of knowledge and expertise about sealing methods, materials, and fabrication techniques has been built up over many years. Like developments in pumping and measurement, it has evolved and continues to evolve in response to technical requirements created in the various applications of the subject. This knowledge is thoroughly surveyed in Roth's book of 1990. O'Hanlon's text (2003) is an indispensable guide to modern developments and good practice, which may be consulted for detailed information or for further reference on all topics. The purpose of this short chapter is to outline some basic topics about materials, components and construction, and leak testing.

10.2 Materials

A number of requirements have to be met by any construction material planned to be used in vacuum practice, either as part of the vacuum wall or in the internal fixtures. A primary concern is that the amount of gas released under vacuum conditions must be as small as possible. Thus, materials with a porous structure that can harbor gas, or others with flawed structures due to cracking on a microscopic scale, or that are otherwise permeable to gas, are not permissible. Materials that can be used must have adequately small outgassing rates after cleaning treatments, must be stable over the range of temperatures encountered, and must exert an insignificant vapor pressure. Mechanically, they should have adequate strength and be easy to fabricate into the shapes required. Techniques must be available for joining them to other parts of the same material and to other vacuum-compatible materials.

For the vacuum enclosing wall, austenitic stainless steel, aluminium, and glass are good materials in all the above respects, though in the case of the metals the welding process, in which parts to be joined together are locally melted at the join to fuse and form a continuous structure, is more problematic for aluminium. Nevertheless, considerable expertise exists and aluminium systems are commercially available. With the obvious exception of the TV tube, glass, which was at one time very widely employed as a vacuum enclosure in experimental physics, is now used very little. This is primarily because most laboratory-sized vacuum systems are nowadays bigger than they were hitherto, being usually equipped with numerous devices to allow the various fed-through facilities referred to in the introduction. These are too heavy for glass to support and in many cases need to be detached from the system for servicing. Thus, stainless steel is the most widely used

material in the construction of vacuum systems. Stainless steel with low carbon content and nickel and chromium content of around 18 and 8% respectively, and members of the "300" series (304, 316, and 321, in particular) are strong, tough, and versatile, with good machining and welding properties, and good corrosion resistance and outgassing characteristics on account of the stable chromium oxide layer that forms on their surfaces. They are also nonmagnetic, though occasionally their joins may acquire some magnetism as a result of welding. The use of 'L' grades helps to minimize this if it is a potential problem.

Because of these favorable characteristics, stainless steel is also the preferred material for the construction of a substantial number of parts of many of the feedthrough-mounted devices previously mentioned, and for internal fixtures. An aluminium alloy such as Dural, with its excellent outgassing and machining properties, is also frequently very suitable for internal constructions, as it is available from suppliers in useful channel and right-angle sections that are often helpful in the design of fixtures, as well as in the form of solid rods. Oxygen-free high conductivity copper (OFHC copper), with its high electrical and thermal conductivity, allied with good mechanical and outgassing properties, is a natural choice for electrodes and for applications in which efficient heat transfer is important. Nickel is also a good vacuum-compatible metal, though magnetic, which might prohibit its use in some applications. The refractory metals tantalum and molybdenum have good outgassing properties but are very difficult to machine. However, when supplied as a sheet, tantalum in particular can be cut and easily bent into useful shapes to form, for example, radiation shields. A spot-welding facility is useful in fashioning such pieces.

A widely used insulating material for in-vacuum constructions is alumina, which is supplied as rods, tubes, or flat slabs. Using a sawing tool such as a diamond-toothed cutting wheel, it is easily possible to fabricate from these a range of pieces in various useful shapes. For example, insulating washers, rings, and collars may be made from alumina tube so that conducting rods, wires, or screws can be enclosed and electrically isolated. Machinable glass ceramics are available for making insulating pieces with non-simple shapes. Borosilicate glass (Pyrex) is an excellent vacuum-compatible insulator but items made of it need to be used with appropriate care because of its fragility.

Glass-to-metal and ceramic-to-metal seals are necessary features in many components. Adhesion between the two and similar values of thermal expansion coefficients are critically important. For glasses, which are very weak when subject to mechanically or thermally induced stresses that are not purely compressive, good matching of thermal expansion coefficients between glass and metal is possible with certain combinations. Sometimes, for a glass and metal that are not adequately matched, it is possible to make a graded seal in which selected glasses of slightly differing expansion coefficients are fused together in a sequence that effectively distributes the mismatch into smaller, acceptable values. Generally, glasses bond to the oxidized metal rather than with the metal itself. Ceramics are tougher than glass, and in ceramic-to-metal

seals alumina, typically, is premetallized using sophisticated metallurgical techniques, with a tightly bonded very thin layer of titanium or molybdenum and then a layer of nickel before brazing to the bulk metal.

Plastic materials generally desorb large amounts of gas and absorbed water vapor and are not recommended for use in vacuum. However, there are exceptions such as PTFE (Teflon®), which is an insulator, can be baked, has a low outgassing rate, and most notably, a very low friction coefficient, making it useful for bearings. Also PEEK®, a high temperature polymer, can be machined to useful shapes, and kapton is an insulating covering for conducting wires.

10.3 Demountable Connections

The design and construction of modern vacuum systems is considerably assisted by the existence of standardized patterns and dimensions for interconnecting flanges and their sealing mechanisms. This means, for example, that a gate valve from manufacturer A may be easily connected to a pump from manufacturer B or C. In addition, based on such flange designs, a whole range of components is available. Thus, there are basic pipeline components, isolation valves, and other devices with entrances and exits that terminate in such flanges, as well as accessories that require just one flange, such as vacuum gauges. Tees, crosses, elbows, and reducers that connect one flange size to the next (the fittings shown schematically in Figure 10.1) are particularly valuable, available from a number of manufacturers in various sizes and flange types, and competitively priced. Based on these demountable joints, therefore, substantial subsystems can be constructed according to a modular scheme. They are quickly assembled and easily adapted and modified.

There are three widely used connection systems of the type described above: the ISO–KF and ISO–K connections, which are used in applications down to about 10^{-7} mbar, and the ConFlat® CF flange used in UHV practice. ConFlat is a registered design of Varian, Inc. (Palo Alto, CA). The ISO (International Standards Organisation) designation indicates that the designs are internationally recognized standards. The KF (kleinflansch, small flange) is

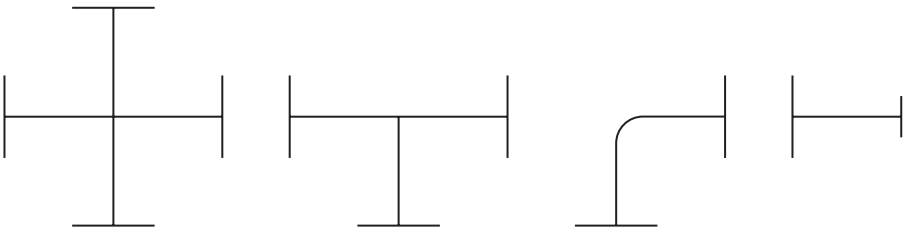

FIGURE 10.1
Various forms of connections: cross, tee, elbow, and reducer.

an original design of the Leybold Company. It is available in sizes DN10, 16, 25, 40, and 50 mm, where the DN number signifies the nominal inner bore diameter of the pipe size with which it is associated. Thus, a KF flange designation 25 means a flange that is nominally matched to a 25 mm (1 in.) bore pipe. The value of the designating number is given by some manufacturers as a similar (and frequently identical) NW number rather than DN. As shown in Figure 10.2(a), the connection consists of two symmetrical flanges, a centering ring that holds in place an O-ring gasket, and a clamping ring whose tightening squashes down the gasket to a thickness determined by the centering ring to make the seal between the flange faces. The O-rings are made of an elastomer, either neoprene or viton. The latter is less permeable and can be used up to about 150°C. In addition to their use in the tee configurations, etc., previously mentioned, flanges are supplied as blanks, or with short sections of pipe already attached, to terminate lines and allow for the attachment of extensions, respectively. They are made of either stainless steel or aluminium (the latter are less expensive but more prone to damage by accidental misuse). Manufacturers catalogs may be consulted for details of construction and dimensions, and it is important to study the dimensions carefully when designing parts for construction that will be attached to them by welding or brazing. Brazing is the process in which a molten filler alloy, of melting temperature less than the metals to be joined, penetrates the gap between them by capillary action so as to form a strong metallic bond on cooling.

For larger bores, the ISO–K system allows connections with bores from 63 up to 1000 mm. DN values 100, 160, and 200 correspond to nominal bores of 4, 6, and 8 in. The sealing strategy is similar to that with centered O-rings, but the flanges are pulled together with an appropriate number of claw clamps distributed around the periphery, as shown in Figure 10.2(b). In other patterns, bolts are used for this purpose. In circumstances where a nonstandard size of connection is necessary, it is possible to have a groove machined in a flange into which an O-ring fits, with bolts providing the closure force to make the seal. O-rings are available in many diameters and cross sections, and Roth (1990), as well as other texts, gives practical details of groove shape and dimensioning.

The seal in the CF flange system used in UHV applications is created by the compression and consequent flow of an OFHC copper gasket between knife edges, as shown in Figure 10.3. Tightening the bolts to bring the stainless steel flanges closer together from initial knife-edge contact on the gasket causes the copper, whose movement at its outer edge is restricted, to flow axially and radially so as to form compressive seals on the the opposed sloping faces. Being all metal and because of its design, the flange can be heated for baking puposes up to 450°C. ConFlat® flanges are available in sizes CF 10, 16, 40, 63, 100, 160, 200, and 250, and larger, where the number again signifies the nominal bore of the maximum size of pipe to which it may be attached.

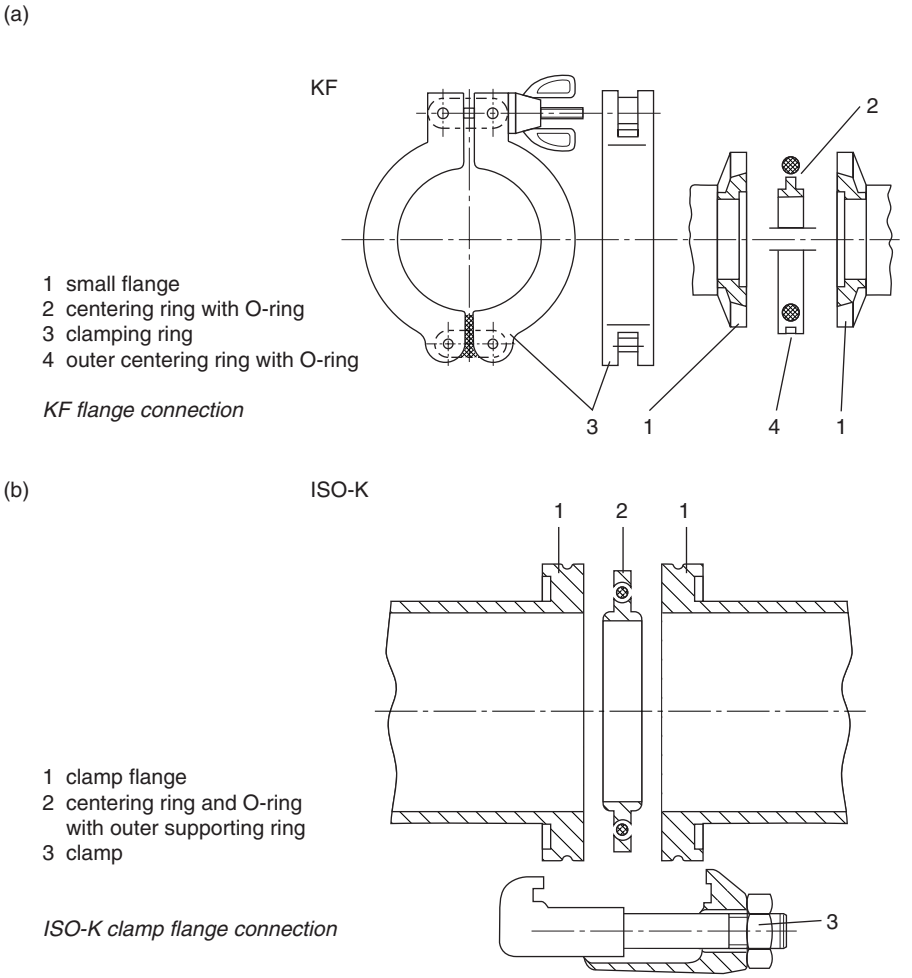

FIGURE 10.2
(a) KF flange, (b) ISO–K flange. (Used with permission of Leybold, Ltd.)

Though very widely used, CF seals are not the only type of UHV seal and for flanges larger than the CF 250 (and for smaller ones also), compressed wire seals of gold, silver, aluminum, or indium are possible, the last named with some restriction on operating temperature imposed by its low melting temperature of 156°C. The Helicoflex® seal, in which an aluminium, copper, or silver outer sheath that is supported on a looped helical inconel spring is compressed between plane flange faces, is an example of a commercially available large seal. Like wire seals, it can be formed into shapes other than circular. These alternatives are discussed by O'Hanlon (2003).

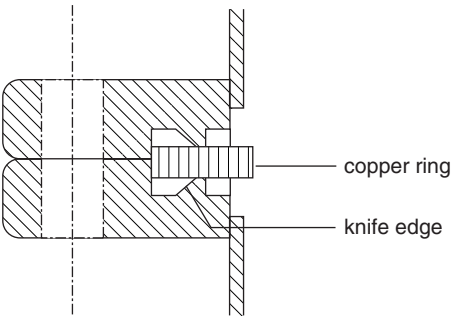

FIGURE 10.3
Cross section of a CF flange. (From Chambers, A., Fitch, R.K., and Halliday, B.S., *Basic Vacuum Technology*, 2nd ed., Institute of Physics Publishing, Bristol, 1998. With permission.)

10.4 Valves

Valves used in vacuum practice fall into two broad categories. Their function is either to isolate one region from another, allowing free flow of gas when open and completely blocking it when closed, or to allow controlled flow from one region to another, usually at a relatively small rate, including zero at complete closure. Among the latter type are manually controlled variable fine-leak valves for the admission of gas into operational systems, simple air inlet valves for venting to the atmosphere in systems that are being shut down or cycled, and sophisticated devices for the regulation of gas flows under automatic control in industrial processing, such as chemical vapor deposition, which depend on the precise control of displacement in a variable conductance.

Isolation valves should have a conductance that is zero when closed and high when open. When closed, adequate leak tightness has to be sustained against a pressure differential of up to an atmosphere, and sometimes more. Inlet and outlet flange connections are of the types described earlier, depending on the region of vacuum in which they are used, so that, for example, an ultra high vacuum system will have ordinary low-vacuum valves in its roughing lines. The sizes of valves depend on the loads being handled, with flange connections to a matching pipeline. They usually have an in-line or right-angle configuration. The latter arrangement facilitates provision of the force necessary to make the seal. Valves may be located in line in pumping networks to define different pathways such as roughing and backing lines or, for example, in the tail of a T junction to isolate a component at vacuum during venting. Between a work chamber attached to a high-vacuum pumping station, isolation is usually by a gate valve (Figure 10.6), whose conductance in the open state is the maximum possible. Gate valves are used to isolate sections of beam lines of various kinds, as discussed in Chapter 9.

The construction of valves requires vacuum-sealed access to their interior for the actuating mechanism, whose movement opens the valve and must close it with sufficient force to create a good seal. Though this may be achieved in a number of ways, for example, by the rotation or translation of a round rod in an O-ring, it is nowadays frequently accomplished with a bellows seal.

Figure 10.4 (a) shows a simple and effective diaphragm-sealed in-line valve suitable for work at low vacuum. The flexing diaphragm may be made of either neoprene or viton, and actuation is by a screw-driven plunger. Another relatively inexpensive type of valve for the same region (not illustrated) is the ball valve, in which a sphere containing a diametral hole is rotated in a spherical housing through 90° between open and closed states. Figures 10.4(b) and (c) show the principles of right-angle valves with actuation through a sliding seal and bellows respectively, using elastomer O-rings to make the main isolating seal.

Figure 10.5 shows a frequently used type of bellows-sealed UHV all-metal right-angle valve, in which a knife edge is forced into an OFHC gasket to make the seal. These have a long life of many operations between gasket renewal, if treated carefully and not tightened beyond the recommended torque on closure. They are, of course, bakeable. If such a valve is at the junction between a baked and an unbaked zone, it is important that the unbaked side be kept at low vacuum during heating so as to avoid serious oxidation of part of the copper gasket. Black copper oxide, if formed in quantity, is sufficiently hard, abrasive, and loosely bound as to easily cause damage to the sealing surfaces in subsequent use.

Figure 10.6 shows the essential structure of a typical gate valve. A carriage that moves down the side sleeve enables the sealing plate, which carries an O-ring, to be forced against the sealing face to make the main seal, and is transported away from it completely to give the high-conductance opening. A number of ingenious mechanisms have been invented for this purpose. Gate valves with a Viton O-ring are bakeable up to 150°C. All-metal UHV gate valves are considerably more expensive.

10.5 Bellows and Motion Feedthroughs

Stainless steel bellows are used extensively in vacuum devices and are made in two basic forms, convoluted and edge-welded, as shown in Figure 10.7(a) and Figure 10.7(b), respectively. They are available as individual components in a range of sizes, lengths, and flange terminations, or unterminated, in which state they can be brazed or welded into connecting parts. In short lengths, the convoluted type of bellows, which are formed under hydrostatic pressure from a thin-walled pipe, can be compressed and extended and formed into a simple bend. In longer lengths of 50 cm or more, with bores

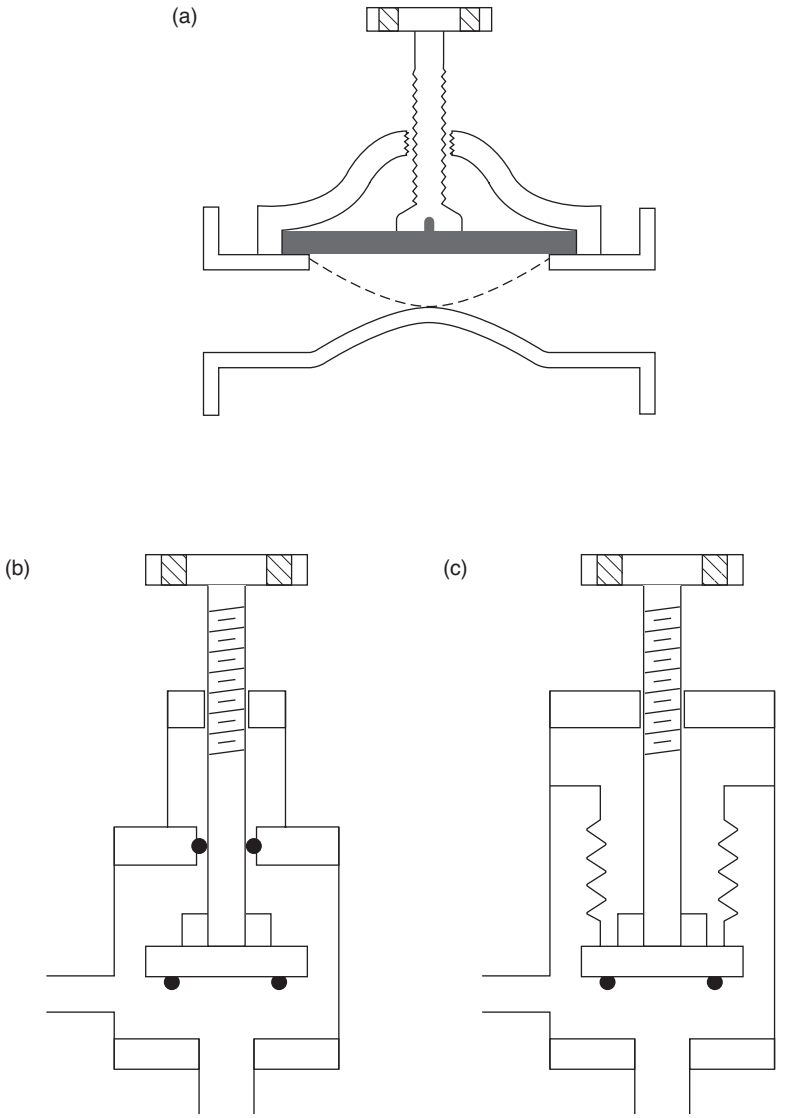

FIGURE 10.4
(a) Diaphragm valve, (b), and (c) elastomer-sealed right-angle valves.

of 10 to 50 mm typically, they are more pliant, and provided they are attached at one end to a rotatable flange, are useful as flexible connecting lines. (Neither pattern of bellows will sustain a simple twist about their axes.) The edge-welded pattern of bellows, which are more expensive, are size-for-size more pliant than the convoluted type and, in addition, can take bends in a shallow S-shape at small lengths so that attached flanges can be displaced transversely in a way that keeps the flange faces in parallel planes. They can

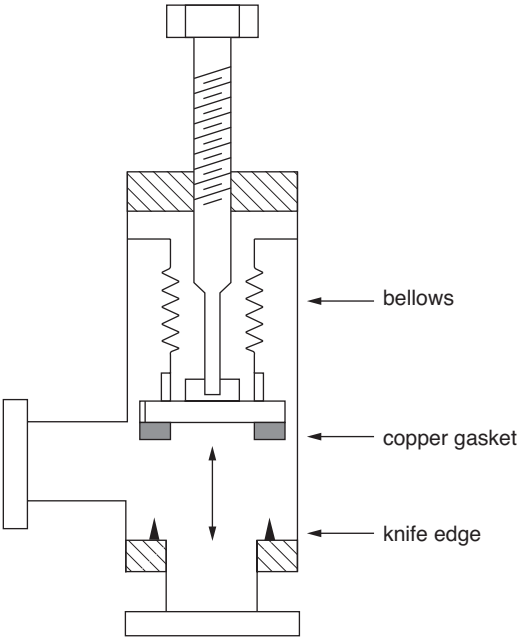

FIGURE 10.5
Metal-sealed right-angle valve.

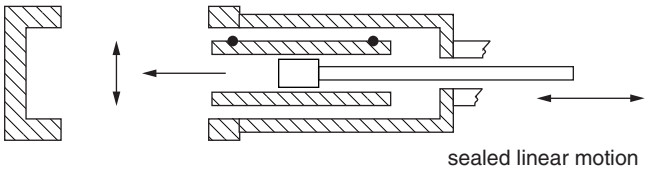

FIGURE 10.6
Representative gate valve structure.

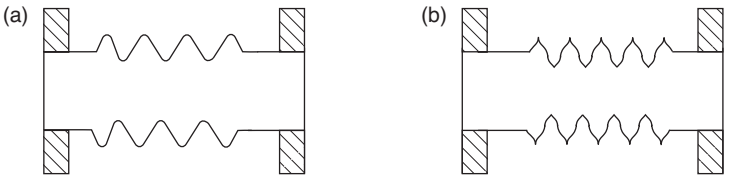

FIGURE 10.7
Bellows sections (a) convoluted, (b) edge-welded.

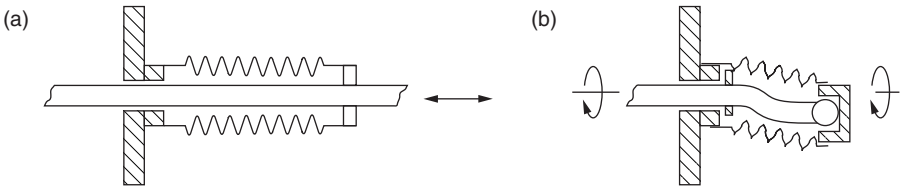

FIGURE 10.8
Bellows arrangements allowing (a) linear and (b) rotational motion.

be extended and compressed to a much greater degree than the convoluted type and in long lengths have the floppiness similar to a "slinky spring."

How bellows may be exploited to feed simple translation and rotational motion through the vacuum wall is illustrated in Figure 10.8(a) and Figure 10.8(b), respectively. Both functions are also realizable relatively cheaply by a mechanism (not illustrated) in which a rod of circular section may slide in an O-ring seal. This technique is restricted to non-UHV conditions, and one of the undesirable features of translation in such arrangements is that surface-adsorbed gas is carried into the interior when the movement is inward. The extent to which the use of bellows allows sophisticated manipulative mechanisms may be judged from the commercially available devices that are illustrated on Web sites.

It is also possible to transmit mechanical displacement through the vacuum wall magnetically, exploiting the coupling between internal and external permanent magnets, the internal one "following" its external partner, provided that limiting forces and torques are not exceeded. Thus, Figure 10.9 illustrates the principle of a long-stroke translation and rotation mechanism that can be useful in sample manipulation and exchange. A widely used rotary feedthrough, capable of very high rotation rates, is the Ferrofluidic® seal shown in Figure 10.10. The seal is made by a fluid of very low vapor pressure containing a suspension of ferromagnetic particles, which is

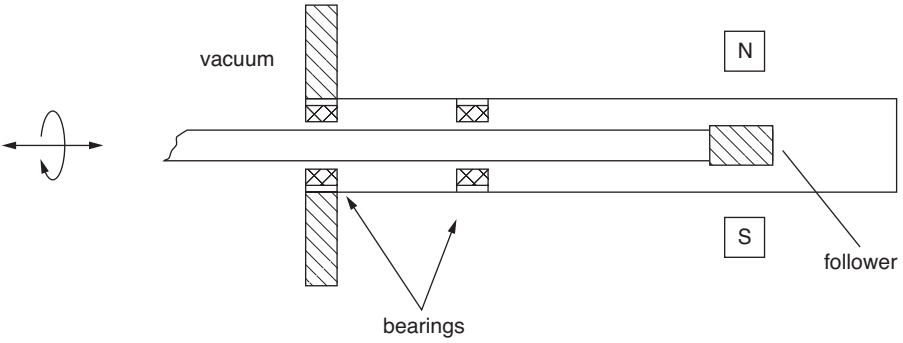

FIGURE 10.9
Long-stroke translation with rotation.

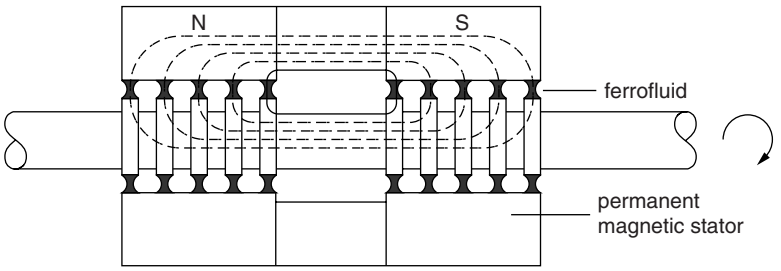

FIGURE 10.10
Ferrofluidic® seal. (From Chambers, A., Fitch, R.K., and Halliday, B.S., *Basic Vacuum Technology,* 2nd ed., Institute of Physics Publishing, Bristol, 1998. With permission.)

induced to reside in a sequence of sealing rings between the stationary and rotating parts by the permanent magnetic field of a shaped stator.

A recent development in providing rotary motion and others derivable from it by rack-and-pinion-type devices, which avoids the need to convey mechanical motion through the vacuum wall, is to use internal stepper motors directly. Such motors have been specially developed for use in vacuum by AML Microelectronics (Arundel, U.K.) and others. Suitably supported in vacuum, the only feedthroughs required are those to supply the control current. Once degassed, they are UHV-compatible, bakeable to 200°C, and usable at 10^{-10} mbar.

10.6 In-House Design

The investigator faced with the design of items for construction in-house may consult a number of texts, particularly those of Dennis and Hepple (1968) and Green (1968), for guidance on good practice and established principles. Innovative design ideas covering a wide range of applications in laboratory-size vacuum systems are regularly reported in the "shop notes" of the *Journal of Vacuum Science and Technology.*

A very important and frequently given piece of advice, repeated here, is to avoid creating trapped volumes from which air, initially isolated at atmospheric pressure, can be pumped away only slowly under vacuum. This might happen, for instance, with a screw thread held in a tapped blind hole, i.e., a hole that terminates inside a material rather than going through it. As shown in Figure 10.11(a), the problem thus created is easily avoided by filing a small flat on the thread or by drilling a hole down through its center, a procedure called "vacuum relieving." For the same reason, it is important in welding a pipe into a flange, for example, to avoid trapped volumes by welding at A rather than B in Figure 10.11(b). If for some reason extra strength were required of the join, then welding at location B could be made at just

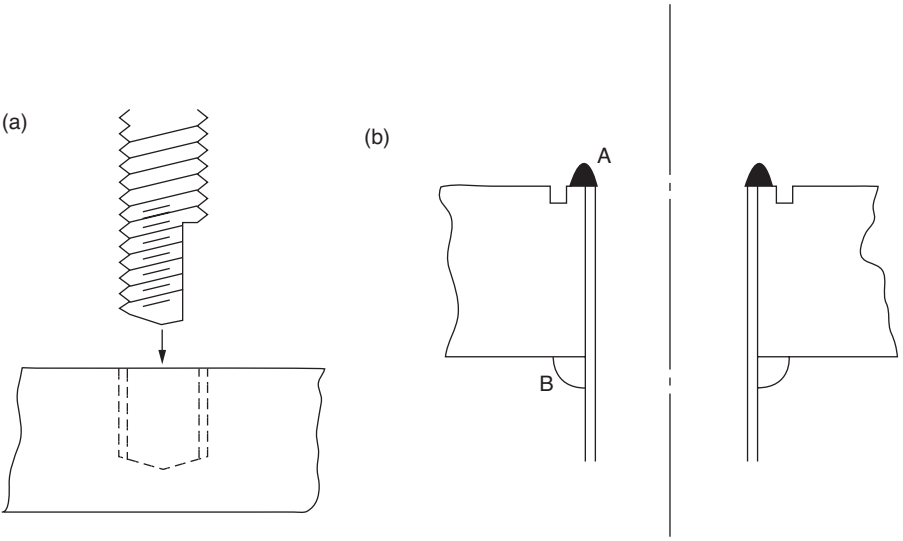

FIGURE 10.11
(a) Blind hole, vacuum relieved, (b) good and bad welding practices.

a few points distributed around the periphery. A complete weld around both B and A would be bad practice because a hole developing in A would produce a leak that was not detectable.

The principles that underly the geometric design of items for installation in vacuum are in many cases closely related to those of classical "kinematic" design, in which the degrees of mechanical freedom necessary in an assembly of parts can be adjusted independently of each other and reproducibly, with dedicated and simply made securing arrangements that use the minimum force to achieve a particular configuration. In vacuum practice they are entirely consistent with the avoidance of trapped volumes. Thus, in the arrangement of Figure 10.12, a milled rectangular slot in one of two plates serves as a guide to clamp them to a rod of circular section, with minimal pressure from a securing piece with a clamping screw. There are two degrees of freedom: the position along the straight circular rod and the orientation around it, so that an item attached to the free end can be put in its desired location.

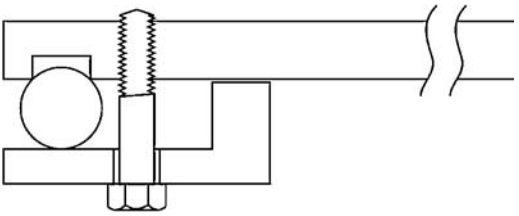

FIGURE 10.12
Clamping arrangement.

10.7 Cleaning

Parts supplied by manufacturers are normally received in a thoroughly cleaned state, well packaged, and ready for insertion into vacuum. Parts that are fabricated in-house need to be cleaned before use. Halliday (1998) gives an account of general cleaning procedures and Venables (2000) gives valuable guidance on procedures for UHV work, sample preparation in particular. As will be appreciated from Chapter 5, in leak-free systems that are working properly, the sole gas load for the pumps is from surfaces, and cleanliness of internal surfaces prior to pumping is therefore of paramount importance for rapid pump-down and optimum vacuum. Regardless of the ultimate level of vacuum sought, it is sensible to clean articles as thoroughly as possible. These matters are critical in UHV applications but are also highly important in good high-vacuum practice. A reliable set of procedures for parts that have been washed free of machining fluids or for items such as stainless steel nuts and bolts purchased from ordinary (i.e., non-vacuum) suppliers is to (1) clean in an ultrasonic tank in water containing a small concentration of proprietary detergent such as RBS 25®, (2) rinse with tap water, (3) clean in an ultrasonic tank in deionized water, (4) dry in a clean hot air blast, and (5) wrap in aluminum foil and store in a dry location such as a hot cabinet prior to installation.

In the sequence above, parts should not be handled with bare hands but with clean latex surgical gloves or clean tweezers. Technical grade tissues that will not release their fibers in use, when wetted with isopropyl alcohol can be used (once) to remove light fingerprint grease from parts inadvertently touched or otherwise suspect, but steps 3 to 5 should be undertaken subsequently for UHV work. It is advisable in assembling parts to work on a clean tray with a rim to retain small items — nuts and washers, typically — that are easily dropped accidentally, and it is very important to ensure that tools and spanners that have been thoroughly cleaned remain so during use. A small spanner dropped on the floor is a spanner lost to the operation in hand, and retrieving it, potentially now soiled, would likely also cause soiling of a glove. In the absence of a spare or an assistant, a frequent state of affairs, an interruption of the original operation is therefore necessary to restore clean conditions. These comments apply to work in small-scale laboratory operations. More extensive procedures are used in large systems, and in any operations that involve the use of chemical cleaning, strict safety precautions must be observed.

10.8 Leaks and Their Detection

In working with vacuum systems, the user will from time to time encounter leaks that cause performance to be below expectation. Their origins and

counter-measures will be considered in due course. It is appropriate first to consider their characteristics and specification. In a typical dynamic (i.e., pumped) system in which the pump and gauges are working properly, a higher pressure than is normal signifies a gas load that is too large and, therefore, a leak of some kind. Basically, there are just two types of leaks, described as real or virtual. A leak is said to be real when there is a route by which atmospheric air outside the system can find its way inside, perhaps through a connection whose seal has not been fully made or through a crack that develops in a weld, for example. Rather more rarely, a route for a real leak may be due to permeation; as noted elsewhere, atmospheric helium can permeate a viton seal at a slow rate, though it is usually insignificant. Virtual leaks, on the other hand, occur when the gas originates inside the system. Air trapped in an internal volume, as mentioned in Section 10.6, is an example and water vapor from imperfectly dried components is another. Newly assembled systems may take longer to pump down than expected because of large initial surface outgassing rates that will diminish with cycling and regular use.

Real and virtual leaks may be distiguished in principle, and a real leak identified, if it is possible to isolate a volume from its pump and monitor the ensuing pressure on a gauge. The various outcomes after the isolation at time t_1 are shown in pressure-time plots of Figure 10.13. In practice, logarithmic scales for both axes are advisable to plot the data.

Figure 10.13(a) represents a virtual leak because the pressure reaches an eventual maximum in which either the supply of gas that is trapped is exhausted, resulting in a uniform pressure throughout or, in the case of evaporating vapor, gaseous and surface phases come into equilibrium. Such behavior is also observed, but on a small scale, when high or ultrahigh vacuum chambers are isolated from their pumps. Figure 10.13(b) shows a linear rise associated with the influx of atmospheric air at a constant rate. (Usually in small systems even serious leaks allow the attainment of pressures of a few mbar, so the leak is driven by a pressure difference that is essentially constant.) Figure 10.13(c) shows the general case, which is a

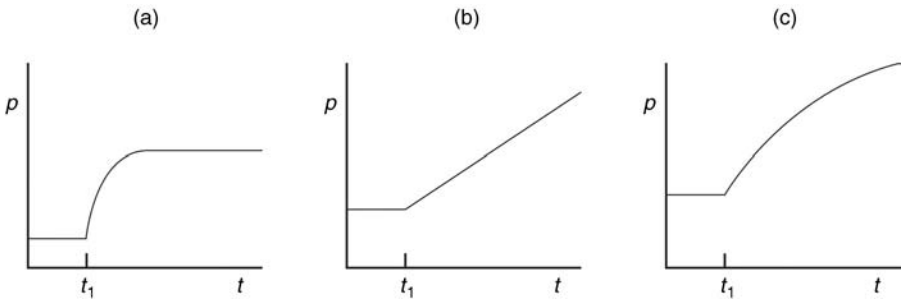

FIGURE 10.13
Pressure vs. time plots for (a) a virtual leak, (b) a real leak, (c) a combination of the two.

combination of the two possibilities. In practice, it would be important to continue to test for long periods of time if possible to establish an eventually linearly rising section that signifies a real leak. As Bigelow (1994) has discussed, carrying out this procedure as a test for the presence of a leak can be problematic and does not always give unambiguous information. The range spanned by the gauge from the starting condition may be insufficient, especially in a high vacuum context where the outgassing rates are relatively large.

If the volume V of a vacuum chamber and values of (dp/dt) are known, a leak rate may be quantified by its throughput Q_L as $Q_L = V(dp/dt)$. A leak rate may thus be expressed in mbar l s^{-1} or Pa m^3 s^{-1}, or the information may be converted to other units such as the sccm or standard cc per second, as discussed in Section 2.2. Thus 1 mbar l s^{-1} = (1000/1013) = 0.987 std cc per second. Mass flow rates could be determined from Q_L values using Equation 5.4.

It is instructive to consider how the magnitude of Q_L compared with other gas loads affects the vacuum achievable in a typical case. Suppose that a high vacuum chamber of volume 50 l has internal surfaces of area 5000 cm^2, which outgas at a rate $q_G = 2 \times 10^{-8}$ mbar l s^{-1} per cm^2, and that the pumping speed is $S = 100$ l s^{-1}. The load due to outgassing is therefore 10^{-4} mbar l s^{-1}, and with no leak the pressure achieved is $p = 10^{-4}/100 = 10^{-6}$ mbar. Table 10.1 shows how leaks Q_L of various magnitudes would increase the total gas load Q_T and modify the pressure. Some leaks may be tolerable and not worth eliminating.

In this context, leak 1 would clearly be insignificant, and probably leak 2 also. Leak 3 might be regarded as acceptable depending on the purpose at hand, but leak 4 would be unacceptable and require corrective action. The rate of pressure rise (after isolation) with this leak would be $(dp/dt) = Q_L/V = 10^{-4}/50 = 2 \times 10^{-6}$ mbar per second, with an equal contribution initially from the gassing rate, and therefore 4×10^{-6} mbar per second in total.

A leak of any given magnitude would, of course, be less significant in a chamber of larger volume. Thus, in the example above, suppose that the dimensions were doubled so that the volume increased eightfold to 400 l and surface areas correspondingly quadrupled to roughly 20,000 cm^2. The gassing rate would be 4×10^{-4} mbar l s^{-1} and with an installed pump of four

TABLE 10.1

Magnitudes of Leaks Q_L and Effects on Total Gas Load (Q_T) and Pressure

Leak Number	Q_L mbar l s^{-1}	Q_T mbar l s^{-1}	$p = Q_T/S$ mbar
	0	1.0×10^{-4}	1.0×10^{-6}
1	10^{-8}	$10^{-4}(1 + 10^{-4})$	1.0001×10^{-6}
2	10^{-6}	$10^{-4}(1 + 10^{-2})$	1.01×10^{-6}
3	10^{-5}	$10^{-4}(1 + 10^{-1})$	1.1×10^{-6}
4	10^{-4}	2×10^{-4}	2×10^{-6}

times the speed, 400 l s^{-1}, the same leak-free pressure of 10^{-6} mbar would be achieved, but a 10^{-4} mbar l s^{-1} leak would now raise this base pressure by only 0.25×10^{-6} mbar to 1.25×10^{-6} mbar, compared with 2×10^{-6} mbar in the smaller system.

Returning to the chamber of 50 l volume, leaks 1, 2, and even 3 might be tolerable at high vacuum, but matters would be very different if the chamber were part of an ultrahigh vacuum system. Thus, with the same $S = 100$ l s^{-1} and an outgassing load reduced by four orders of magnitude to 10^{-8} mbar l s^{-1} by baking, the base pressure would be $p = 10^{-8}/100 = 10^{-10}$ mbar. Thus, even a leak of 10^{-8} mbar l s^{-1} would double the pressure, allowing atmospheric constituents, particularly oxygen and water vapor, into the system and would be unacceptable. Thus, UHV systems with volumes and speeds of this order ought to be leak-tight at a level 10^{-10} mbar l s^{-1} if a leak giving a 1% rise in base pressure is tolerable, though there will be situations in which it is not. At the other end of the pressure scale, in the low-vacuum region, a large leak that held the system pressure at 0.5 mbar, say, and prevented a primary pump with a speed of 2 l s^{-1} from reaching the crossover condition to secondary pumping, would have a magnitude of 1 mbar l s^{-1}. A range of leak rates is, therefore, encountered in practice, with levels of tolerance that vary with the regime of vacuum.

The need for very high degrees of leak tightness such as 10^{-10} mbar l s^{-1} is not restricted just to UHV practice. Many industrial processes carried out at higher pressures, particularly in the fabrication of semiconductor devices, make similar demands on equipment to avoid contamination of products. Also, static vacuum devices in which there is no pumping need to be very effectively sealed. Thus, a leak of 10^{-12} mbar l s^{-1} into a device of volume 10 cm^3 (0.01 l) sealed off at 10^{-8} mbar, say, gives a rate of pressure rise $Q_L/V = 10^{-12}/10^{-2} = 10^{-10}$ mbar s^{-1}. If the functioning of this device is degraded by a pressure rise to 10^{-4} mbar, then this would occur in a time $10^{-4}/10^{-10} = 10^6$ seconds, which is less than two weeks, and clearly be unacceptable. Sealing requirements for such devices are extremely stringent.

Turning to the subject of the origins and location of leaks in vacuum systems, one may in some cases, depending on the pattern of use, distinguish the situation in which a high or ultrahigh vacuum system has been running well for some time and then apparently develops a leak, from that in which such a system has recently been at atmospheric pressure for some purpose (to reload sources, for example), and will not pump down as it should. The former situation may be more problematic because in the latter case recent actions and their possible implications in causing a leak can be reviewed. Newly installed feedthroughs should, in any case, have been individually leak tested on a leak detector (see below) prior to installation, thus removing one source of uncertainty. Thinking hard about what might have changed since the previous problem-free operation is usually helpful. The importance of keeping a log book of significant actions, especially with equipment that has more than one user, cannot be overemphasized.

As mentioned above, some leaks may be sufficiently gross to prevent the cross-over condition for switching from primary to secondary pumping being reached. A long primary pump-down time is a sure sign of a problem. Careless placement of a Conflat gasket so that it is not properly located can mean that the application of the usual bolt-down forces deforms the copper in unintended, nonsealing ways. This situation is most likely to arise when the flange faces and gasket are in a vertical plane. A good test of correct location before tightening of bolts is to rotate the flange of the free item back and forth slightly about the fixed flange of the receiving port, when anything mechanically wrong becomes immediately apparent. A gasket may be correctly located but inadequate closure force applied. Manufacturers' recommendations should be followed and a feeler gauge may be employed to measure flange face separations near closure. Great force is not required to make a proper seal, though many of us apply it before learning so. The even tightening of the flange bolts progressing circumferentially, perhaps using a torque wrench for larger flanges, is advised.

A more subtle and elusive cause of slow primary pumping may be a virtual leak arising from the saturation with prolonged use of the sorbing material in a foreline trap. Once recognized, the problem is rapidly solved by replacing the sorbent, and taking care to service the trap at regular intervals. It is sensible to know the system time constant (V/S) and thence the time per decade $2.3 \times (V/S)$ of a primary pumping arrangement, so that serious deviation from expected behavior is quickly obvious. An adequate estimate of the system volume can be easily made and divided by the rated speed of the primary pump. The pumping speed of oil-sealed primary pumps is not necessarily an unchanging quantity and may deteriorate if the pumps are not regularly serviced and their oil changed.

In networks of piping and components that contain isolation valves and gauges, it may be possible to hunt down a leak by a logical sequence of valve openings and closings while observing the effect on indicated pressures. Thus, if a suspiciously high value falls in a pumped location when a particular valve is closed, it is a good indication that a leak exists in the region beyond the valve. Occasionally, if the results of such testing defy logic, the fault may be in the valve itself, a leak in its bellows, for example. Locating a leak that restricts pressures to the low vacuum region can sometimes be simply accomplished by looking for changes in a Pirani gauge reading that occur when an externally directed flow of isopropyl alcohol from a squeeze bottle covers a leak, briefly blocking it, and then passing through to vaporize internally. The response of this gauge is quite strongly dependent on the identity of the gas, as noted in Chapter 7, and this is the basis of the method. Other organic solvents such as diethyl ether are particularly effective but nowadays their use is likely to be limited by Health and Safety Regulations. Extensive use of organic solvents in this way is discouraged because they get absorbed in the elastomer and can subsequently permeate into the system.

In high- and ultrahigh vacuum systems where a leak is suspected, perhaps because of longer than usual times for system pressures to fall with secondary

pumping, or because of an unexpected sudden increase in pressure from a previous stable and acceptable condition, a pressure vs. time plot following isolation, of the type previously described, may assist in confirming the presence of a leak. Relatively simple techniques exploiting an installed thermionic ionization gauge may be helpful in locating it, if more sophisticated means are not immediately at hand. Thus, a change of its reading may be brought about by a leak being temporarily interrupted with isopropyl alcohol, although a surer method is to explore surfaces with a fine flow of helium gas through a probe and throttled from a high-pressure storage cylinder, working from the uppermost regions downwards because the helium will rise. The change in pressure reading caused is most easily discerned with the older type of analog instrumental display and depends on three factors. First, the air being conducted by the leak is replaced by helium which, being more mobile (molecular flow conductances are proportional to $1/\sqrt{M}$), has a higher throughput, so the pressure should rise. Second, if the pumping speed for helium is less than for air, as it would be with turbomolecular pumping, the pressure will rise on account of this also; if it is greater, the two effects will oppose each other. Third, as we have seen in Chapter 7, the gauge itself will be less sensitive to helium than air, so this will tend to reduce the magnitude of the change in the indicated pressure. It is unlikely, however, that these factors combine to annul the value of the technique, and helium is a good probe gas on account of its mobility, inertness, lack of toxicity, and low concentration in the normal atmosphere. As noted by Delchar (1993), leaks as small as 10^{-7} mbar l s^{-1} can be detected using this ion gauge technique.

In systems that are equipped with a residual gas analyzer, the presence of an oxygen signal at mass 32 amu and a nitrogen fragment signal at 14 amu can be a good indicator of the existence of an air leak. And, of course, the quadrupole instrument itself may serve as an inbuilt leak detector by tuning to mass 4 and looking for signal changes that can be associated with the manipulation of an external helium probe. Many instruments have a software facility, a "leak mode," that assists in this activity. However, the purpose-built mass spectrometer helium leak detector (MSLD), frequently portable nowadays, is an almost indispensable piece of laboratory equipment for the testing of newly made and repaired items, as well as for testing whole systems. The various designs, operational methods, and applications of MSLDs comprise a very large subject, whose importance extends beyond the laboratory to industrial processing and manufacturing. Semiconductor fabrication lines and the manufacture of aerosol cans and refrigerator elements are just a few examples. Hablanian's text (1997) surveys these applications besides discussing methods.

Mass spectrometer leak detectors usually use a 180° magnetic sector analyzer tuned to mass 4 as the mass filter on account of its simplicity, sensitivity, and general robustness. This helium sensor, which has to operate in a vacuum of 10^{-4} mbar or less, is incorporated in a pumping configuration such that with a system or a component attached to the inlet port for testing, any helium entering via a leak flows with other gases through the MSLD and is

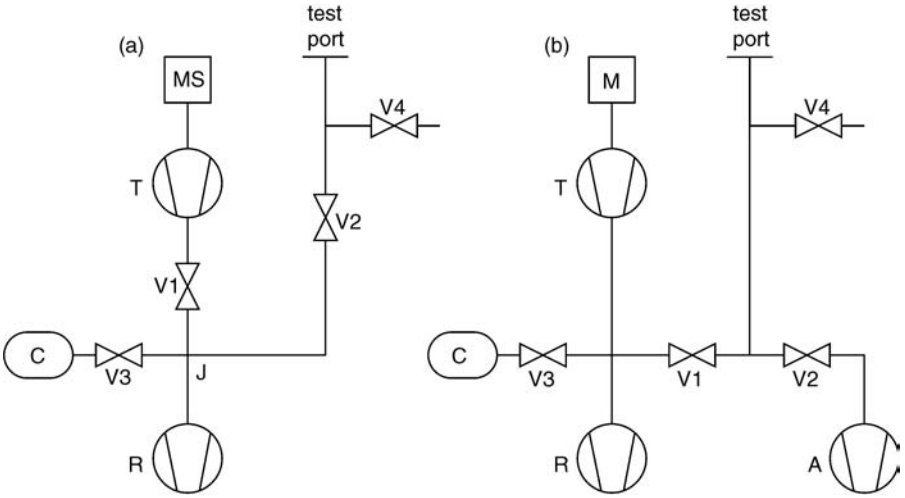

FIGURE 10.14
Contra-flow leak detector (a) portable configuration, (b) static arrangement with auxiliary pump.

detected. Of the two principal configurations, direct-flow or counter-flow, only the latter, which has recently come into more prominent use for reasons that will be given, is described.

The essential configuration of the counter-flow instrument is shown in Figure 10.14(a), which is the portable version. The mass spectrometer head MS is held at high vacuum by the turbo pump T, which is connected to a backing/roughing pump R via valve V1. The test port to which the item under test is attached is connected to R via valve V2, and also to the outlet of the turbo when V1 is open. The principle of the counter-flow technique is that when a gas mixture containing helium is flowing at the junction J, at the turbo-backing pressure of about 0.1 mbar, some of the helium will diffuse through the pump in a direction opposite to the normal gas flow to reach MS and be detected. This is because the turbo pump's compression ratio K for helium is relatively small, of the order of 100, so that a situation is quickly established by back-diffusion that for a partial pressure p_{He} at J, there is an associated pressure p_{He}/K at the MS. The back-flow of the other gases, for which K values are many orders of magnitude greater, is negligible and they flow wholly to R in the normal way.

The detection technique is as follows. With the instrument operational and V1 open and V2 closed, the item for testing is attached to the test port. Valve V1 is then closed to briefly isolate the turbo and V2 opened to pump it down. When the turbo backing pressure is reached, V1 is again opened. Any helium entering the test item as the helium probe is applied reaches J, and some of it MS, so that the leak is detected and its location known. Valve V4 is for venting and V3 to allow the system to be checked when necessary with a calibrated helium leak, C. The various actions of the sequence are carried

out automatically, with a meter display of the detected level of a helium signal above any background level. This can arise from previous helium retention in rotary pump oils and its subsequent release, a problem discussed by Chew (1999). The presence of a leak is also announced sonically by an audible rising tone; a particularly valuable feature when the meter is outside the probe operator's field of vision, as sometimes happens with leaks in awkward locations. The response of the instrument to a step change of helium input through a leak is not immediate, but rises with time constant (V/S), where V is the evacuated volume and S the pumping speed. A signal that is 95% of the final steady value is reached after three time constants and reaches essentially 100% after five.

The calibrated helium leaks against which system performance can be checked depend either on its permeation through a quartz membrane or slow passage through a fine capillary from volumes sealed at atmospheric pressure. Devices of the former type cover the Q_L range 10^{-10} to 10^{-7} and the latter 10^{-8} to 10^{-4} mbar l s^{-1}. At its most sensitive, the MSLD can detect leak rates of about 10^{-11} mbar l s^{-1}. Among the advantages of the counterflow system are that that the system under test has to be pumped down to only 0.1 mbar, and that the sensor head MS is always maintained at high vacuum. Also, the turbo-pumping mechanism acts as a buffer that prevents any pressure surges in the test line reaching the MS. The conventional direct-flow type of MSLD required the protection of a liquid nitrogen trap in the path to the MS head in order to trap vapors, and the pressure had to be brought to a value better than 10^{-4} mbar by primary and secondary pumping, which took longer, prior to testing.

In the nonportable version of the counter-flow detector shown in Figure 10.14(b), the principle is the same but an auxiliary pump A does the preliminary pumping. This is necessary if the volume to be pumped is large and also facilitates operations if the equipment is in continuous use. If a large volume has to be tested with the portable system, an auxiliary pump can be employed to lower its pressure initially. If the large volume is a system with its own pumping, it should be possible to attach it to the leak detector in a way that allows the system pumps to be exploited in conjunction with that of the MSLD.

In conclusion, there are a number of informative sources describing practical methods of leak hunting. They include Harris (2005), who gives a useful general survey and Delchar (1993), who emphasizes UHV concerns and also techniques appropriate in glass systems. Hablanian, as previously mentioned (1997), deals particularly with high-vacuum practice and so does O'Hanlon (2003). Bigelow's 1994 text is aimed at electron microscope users but contains much sound advice and instruction that is transferable to other laboratory situations.

Answers to Exercises and Problems

2.1 (a) 300 torr, (b) 40 kPa

2.2 2.45 MPa, 24.5 bar

2.3 2.45 bar

2.4 10^5 mbar l

2.5 Argon — 10 mbar, methane — 30 mbar

2.6 Total pressure — 33 mbar.
Partial pressures: oxygen and argon — 11 mbar, nitrogen and carbon monoxide — 5.5 mbar

2.7 (a) 2000, (b) 3983, (c) 7983 mbar

2.8 0.147 mol helium and 0.326 mol oxygen. Total pressure — 892 mbar.
Partial pressures: helium — 277, oxygen — 615 mbar

2.9 2.38 cm^3, (a) 235 mbar, (b) 560 mbar

2.10 57.8 l, 1.35 mbar, greater.

3.1 (a) 1766 (b) 588 (c) 394 m s^{-1}

3.2 (a) 658 (b) 241 m s^{-1}

3.3 (a) 0.64 mm (b) 0.21 mm

3.4 7.4×10^5

3.5 5%

3.6 Bigger; no, it is of similar order even though helium is one of the smallest molecules, 1.73 m

3.7 Molecular

3.8 2.9/6.02 = 0.48 mol

3.9 (Equation 3.26.) 5.5×10^5 s = 1 week approx. Mass flow rate, 3.6×10^{-9} kg m^{-2} s^{-1}, mass per m^2, 2×10^3 kg

3.10 2.48 kg m^{-2} s^{-1}, 15 cm (6 in.) per minute

3.11 Reasoning as Section 3.9, transitional.

4.1 1 min

4.2 6×10^{17}, 2.4 Pa = 0.024 mbar

4.3 $J = 2.9 \times 10^{22}\ p\ \mathrm{m^{-2}\ s^{-1}}$, p in Pa. (take care to distinguish between a measure and the unit of measure)

4.5 (a) 3.47 h, (b) 23 h, (c) 52 min

4.6 10^{13}, 0.01 ML

4.7 From equation 3.23, J is proportional to $1/\sqrt{M}$. Therefore ratio is $\sqrt{11} = 3.3$

4.8 (a) 0.1 μs (b) 5.4×10^{6} s, about 2 months

4.9 Number of spherical voids in a 1 cm cube = $(2 \times 10^{6})^3 = 8 \times 10^{18}$. Each has area $\pi D^2 = 25\pi \times 10^{-18}$ m^2, giving a total area 628 m^2! 25 cm^3 of gas.

4.10 Flux from unit area per second into solid angle $2\pi \sin\theta d\theta$ at θ is $dI = I(0)\cos\theta \times 2\pi\sin\theta\ d\theta$. Integrating from 0 to θ gives the flux into a cone of semi-angle θ, say, I_θ, to be $I_\theta = [I(0)\pi/2][1 - \cos 2\theta]$. Hence the total flux is $I_T = I_{90} = \pi\, I(0) = J \times A$ with J given by equation 3.22.

4.11 From problem 4.10 $I_\theta/I_{90} = (1/2)[1 - \cos 2\theta]$, from which $I_{30}/I_{90} = 0.25$, $I_{45}/I_{90} = 0.5$ and $I_{80}/I_{90} = 0.97$

4.12 The change in the number of molecules in the enclosure in a small time-interval dt is $dN = -AJdt = -An\,\bar{v}\,dt/4$ by equation 3.13. dN/V gives dn the change in number density, from which $dn/n = dp/p = -A\,\bar{v}\,dt/4V$. This integrates to give the result stated, an exponential fall with time constant $4V/A\,\bar{v}$, with value 226 s = 3.7 min for the conditions specified. A mean free path of 1 mm or more for hydrogen will occur for pressures less than about 0.1 mbar, as can be deduced from Table 3.1, bearing in mind that molecules of hydrogen are slightly smaller than those of nitrogen.

4.13 Use Equation 3.12 to get dE_v, the energy brought by molecules with speeds between v and $v + dv$. Then substitute for dn_v and integrate. The integral below is useful.

$$\int_0^\infty x^{2n+1} \exp\left(-\beta^2 x^2\right) dx = n!/2\beta^{2n+2}$$

4.14 4.2×10^{-12} Pa m^3 s

5.1 0.2 mbar, 2×10^{-3} mbar

5.2 3×10^{-3} mbar l s^{-1}

5.3 1 mbar l s^{-1} = 0.1 Pa m^3 s^{-1}, 1 l s^{-1} = 3.6m^3 h^{-1}

5.4 0.01 kg s^{-1}

5.5 Q = 0.2 mbar l s^{-1}, flow rate 2 l s^{-1}

5.6 6×10^{-3} mbar l s^{-1}, 0.6 l s^{-1}

5.7 Recall electrical analogies

5.8 (a) 91 l s^{-1}, (b) 80 l s^{-1}, (c) 50 l s^{-1}

5.9 Continuum, laminar by Equations 5.1 and 5.13, respectively

5.10 Choked flow, speed = 0.2 l s^{-1} in each case, throughputs: (a) 2 mbar l s^{-1}, (b) 0.2 mbar l s^{-1}

5.11 90 l s^{-1}

5.12 441 l s^{-1}

5.13 (a) 1870 l s^{-1}, (b) 399 l s^{-1}

5.14 ×8

5.15 19%, 397 l s^{-1}. Compare with result of 5.12 — Equation 5.44 estimates high, as discussed.

5.16 (a) 17 l s^{-1}, (b) 24 l s^{-1}

5.17 83 m^3 h^{-1} (23 l s^{-1} !), a large pump

5.18 74 s, time per decade 37 s.

5.19 0.2 s

5.20 2×10^{-6} mbar

5.21 Speed cannot be greater than conductance at entrance — molecular flow.

5.22 $p_U/p_D = 1.2$, 41.7 l s^{-1}

5.23 232 l s^{-1}, 58 l s^{-1}, 14.5 l s^{-1}

5.24 320 l s^{-1}, $Q = 6.4 \times 10^{-6}$ mbar l s^{-1}, 1.07×10^{-9} mbar l s^{-1} per cm^2, $V\Delta p = Q\Delta t$ gives (a) 4.6 s, (b) 47s, or perhaps a little longer if there is re-adsorption. $V/S = 0.1$ s

5.25 As text

5.26 A/R = 4, B/R = 3, graph gives 0.244, only slightly less than for straight tube

5.27 The flow equation is $V(dp/dt) = Q_{0C} = 20A \times p_0$. Integrating from zero pressure up to 0.53 p_0 gives $t = [0.53/20]V/A$. For $V = 100$ l and $A = 0.04$ cm^2, $t = 66$ s

5.28 0.18, from Equation 5.48

8.1 Pipe conductance 3.25 $\bar{p}$ l s^{-1}. Decade times from equation 3.53 are: 41.4 s, 41.4 s, 44 s, 53 s, and 230 s. Total time 410 s or approximately 7 min

8.2 1.85×10^{-6} mbar, 3.5×10^{-5} mbar

8.5 0.015 mbar l s^{-1}, 150 l s^{-1}, 1/0.837 = 1.19.

Bibliography

A User's Guide to Vacuum Technology, 3rd edition, O'Hanlon JF, 2003, John Wiley & Sons, New York.

An arguably indispensable text for the active vacuum user. Comprehensive in scope, with good discussion of underlying principles, descriptions of equipment, operational matters, and contemporary applications. It includes many useful data tabulations and appendices. Particularly useful in residual gas analysis.

High-Vacuum Technology: A Practical Guide, 2nd edition, Hablanian MH, 1997, Marcel Dekker, New York.

One of the finest books on its subject, with broad in-depth coverage. Informatively and engagingly written by an acknowledged expert.

Theory and Practice of Vacuum Technology, Wutz M, Adam H, and Walcher W (Eds.), 1989 (Braunschweig: Vieweg, marketed in the U.K. and U.S.A. by John Wiley & Sons).

Translated into English by W Steckelmacher, this is a comprehensive survey of the subject, thoroughgoing in its treatment, and a valuable resource. A distinctly valuable feature is the inclusion of worked examples throughout.

Vacuum Technology, 3rd edition, Roth A, 1990, Elsevier Science, Amsterdam.

Comprehensive coverage of the subject with substantial early chapters, comprising about a third of the book, which deal succintly with the physics and physical chemistry of the underlying phenomena.

Vacuum Physics and Techniques, Delchar TA, 1993, Chapman & Hall, London.

A text in the publisher's series *Physics and its Applications*. A very good text for undergraduate physicists and others coming to the subject with a background in physics or chemistry. Contains useful information on glass UHV systems.

Modern Vacuum Practice, Harris NS, 2005, 3rd edition, available direct from nigel.harris1@virgin.net.

There is probably no better general and accessible introduction to the subject from the practical point of view. The treatment is largely nonmathematical and the author's experience in teaching the principles and practice of the subject to those directly involved in using or maintaining vacuum equipment is evident throughout. A good introduction to the subject for readers from many backgrounds and a useful work of reference on practical matters.

Vacuum Methods in Electron Microscopy, Bigelow WC, 1994, Portland Press, London.

A very broad range of vacuum techniques is employed in electron microscopy, so that despite the apparently specialist audience indicated in the title, to whom it is probably indispensable, this is a text with a wider appeal. The underlying principles of the subject and their application are clearly explained in a thorough and accessible way, and there is much sound advice on practical matters throughout.

Foundations of Vacuum Science and Technology, Lafferty JM (Ed.), 1998, John Wiley & Sons, New York.

This completely new and updated version of a standard work, Dushman's *Scientific Foundations of Vacuum Technique*, is a valuable resource with the scope and authority as its predecessor.

Handbook of Vacuum Science and Technology, Hoffman DM, Singh B, and Thomas JH III, 1998, Academic Press, San Diego.

Surveys the subject broadly with a number of valuable contributions from distinguished practitioners.

The Physical Basis of Ultrahigh Vacuum, Redhead PA, Hobson JP, and Kornelsen EV, 1968, Chapman & Hall, London, reprinted in the American Vacuum Society's series of classic texts, 1993, AVS, Woodbury, OH.

Invaluable for its description of fundamental principles and their application, this was a work of seminal importance, and is still a prime resource with enduring relevance.

Ultrahigh Vacuum Practice, Weston GF, 1985, Butterworths, London.

A text surveying the principles and practice of attaining and measuring ultrahigh vacuum.

Capture Pumping Technology: an Introduction, Welch KM, 1991, Pergamon, Oxford.

Emphasizing practical aspects of its subject matter and presenting the underlying fundamentals in a very accessible way, this is a most informative and valuable book.

Total and Partial Pressure Measurement in Vacuum Systems, Leck JH, 1989, Blackie, Glasgow.

A valuable resource that discusses its subject in depth.

Basic Vacuum Technology, Chambers A, Fitch RK, and Halliday BS, 2nd edition, 1998 Institute of Physics Publishing, Bristol.

This is an introductory text based on short course teaching for newcomers to the subject. It contains appendices that deal briefly with the maintenance of various types of equipment and the techniques for measuring flow.

In the English language the *Journal of Vacuum Science and Technology*, which is the official journal of the American Vacuum Society, published by the American Institute of Physics, and the journal *Vacuum*, published by Elsevier Science, are primary resources. They report new developments in the subject and the proceedings of conferences that are held on the wide variety of topics embraced by the technology.

A special issue of the *Journal of Vacuum Science and Technology* (*J. Vac. Sci. Technol*. A 21, Number 5, Supplement, 2003) celebrates 50 years (1953–2003) of Science, Technology, and the American Vacuum Society, and contains many useful articles by distinguished authors about the progress and applications of vacuum technology.

References

Akimichi H, Arai T, Takeuchi K, Tuzi Y, and Arakawa I, 1997, *J. Vac. Sci. Technol.*, A15, 753.

Alpert D, Matland CG, and McCoubrey AC, 1951, *Rev. Sci. Instrum.*, 22, 370.

Andrade EN da C, 1960, *Advances in Vacuum Science and Technology, Proceedings of the 1st International Congress on Vacuum Techniques*, E Thomas, Ed., 14, Pergamon Press, New York.

Arnold PC and Borichevsky SC, 1994, *J. Vac. Sci. Technol.*, A12, 568–573.

Atkins PW, 1998, *Physical Chemistry,* 6th ed., Oxford University Press, Oxford, U.K.

Barish BC and Weiss R, 1999, *Phys. Today,* 52(10), 44–49.

Bayard RT and Alpert D, 1950, *Rev. Sci. Inst.*, 21, 571.

Beams JW, Spitzer DM, Jr., and Wade JP, Jr., 1962, *Rev. Sci. Inst.*, 33, 151.

Beams JW, Young JL, and Moore JW, 1946, *J. Appl. Phys.*, 17, 886.

Becker W, 1958, *Vak.–Tech.*, 7, 149.

Beeck U and Reich G, 1972, *J. Vac. Sci. Technol.*, 9, 126.

Bennett JRJ, 1987, *J. Vac. Sci. Technol.*, A5, 2363–2366.

Benvenuti C and Hauer M, 1980, *Proceedings of the 8th Int. Vac. Congress* (*Cannes*) Vol. 2, p. 199, Société Francaise du Vide, Paris.

Berman A, 1996, *Vacuum,* 47, 327.

Bigelow WC, 1994, *Vacuum Methods in Electron Microscopy,* Portland Press, London.

Butler BP, Music V, and Redgrave FJ, 1999, *Vacuum,* 53, 163–166.

Carter G and Colligon JS, 1968, *Ion Bombardment of Solids*, Heinemann, London.

Chapman S and Cowling TG, 1991, *The Mathematical Theory of Non-Uniform Gases*, 3rd ed., Cambridge University Press, Cambridge, U.K.

Chew AD, 1999, *Vacuum,* 53, 243–246.

Clausing P, 1932, *Ann. Phys.–Leipzig,* 12, 961 and English translation *J. Vac. Sci. Technol.*, 8, 636–646, 1971, courtesy Veeco Instruments Inc.

Cole RJ, 1976, *Rarefied Gas Dynamics,* 10(1), 261–272.

Comsa G, 1994, *Surface Science,* 299/300, 77–91.

Comsa G, Fremery JK, Lindenau B, Messer G, and Rohl P, 1980, *J. Vac. Sci. Technol.*, 17, 642.

Davis DH, 1960, *J. Appl. Phys.*, 31, 1169–1176.

Dawson PH, Ed., 1976, *Quadrupole Mass Spectrometry and its Applications*, Elsevier, New York.

Dayton BB, 1957, *Vacuum Symposium Transactions*, (1956), pp. 5–11, Pergamon Press, London.

Dayton BB, 1962, *Transactions of the 8th National Vacuum Symposium,* 1961, Vol. 1, p. 42, Pergamon Press, New York.

Dayton BB, 1998, in *Foundations of Vacuum Science and Technology,* J Lafferty, Ed., John Wiley & Sons, New York.

De Boer JH, 1953, *The Dynamical Character of Adsorption,* Clarendon Press, Oxford, U.K.

De Boer JH, 1969, in *Molecular Processes on Solid Surfaces,* Drauglis E, Gretz RJ, and Jaffee RI, Eds., McGraw-Hill, New York.
Delchar TA, 1993, *Vacuum Physics and Techniques*, Chapman and Hall, London.
Dennis WTM and Hepple TA, 1968, *Vacuum System Design*, Chapman and Hall, London.
Deschamps J and Doyeux H, 1997, *Phys. World*, 10, 39–43.
Drinkwine MJ and Lichtman D, 1995, *Partial Pressure Analysers and Analysis,* AVS Monograph Series, AIP Press, Woodbury, NY.
Duesing G, 1987, *Vacuum,* 37, 309–315.
Durakiewicz T and Halas S, 1995, *Vacuum,* 46 101.
Dushman S, 1962, *Scientific Foundations of Vacuum Technique*, 2nd ed., Lafferty JM, Ed., John Wiley & Sons, New York.
Dylla HF, 1998, in *Handbook of Vacuum Science and Technology,* Hoffmann DM, Singh B, and Thomas JH, Eds., Academic Press, San Diego, CA.
Edwards JG and Gilles PW, 1966, *J. Chemical Phys.*, 44, 4426–4430.
Ellefson RE, 1998, in *Foundations of Vacuum Science and Technology,* J Lafferty, Ed., John Wiley & Sons, New York.
Elsey RJ, 1975, *Vacuum,* 25, 299–306 and 347–352.
Ettlinger E, et al., 1999, *Reports on Science and Technology.*
Fay JA, 1994, *Introduction to Fluid Mechanics,* MIT Press, Cambridge, MA.
Feather R, 1961, *An Introduction to the Physics of Vibrations and Waves*, Edinburgh University Press, Edinburgh.
Ferrario B, 1998, in *Foundations of Vacuum Science and Technology,* J Lafferty, Ed., John Wiley & Sons, New York.
Field D, 2001, *Contemp. Phys.,* 42(5), 275–283.
Fremerey JK, 1982, *Vacuum,* 32, 685.
Fremerey JK, 1999, *Vacuum,* 53, 197–201.
Frenkel J, 1924, *Z. Physik,* 26, 117.
Glang R, 1970, in *Handbook of Thin Film Technology,* Maissel LI and Glang R, McGraw-Hill, New York.
Green GW, 1968, *The Design and Construction of Small Vacuum Systems,* Chapman and Hall, London.
Gröbner O, 1996, *Vacuum,* 47, 591–595.
Hablanian M, 1984, *J. Vac. Sci. Technol.,* A2(2), 118–126.
Hablanian M, 1997, *High Vacuum Technology,* 2nd ed., Marcel Dekker, New York.
Haken H and Wolf HC, 1996, *The Physics of Atoms and Quanta,* Springer, Berlin.
Halliday BS, 1998, in *Basic Vacuum Technology,* Chambers A, Fitch RK, and Halliday BS, IOP, Bristol, U.K.
Harms AA, Schoepf KF, Miley GH, and Kingdon DR, 2000, *Principles of Fusion Energy,* World Scientific.
Harris NS, 2005, *Modern Vacuum Practice*, 3rd ed. contact author at nigel.harris1@virgin.net. Also available from BOC Edwards; call 01.293.528844 (U.K.) or 1-800-848-9800 (U.S.).
Helmer JC and Hayward WH, 1966, *Rev. Sci. Instrum.,* 37, 1652.
Henning H, 1998, in *Foundations of Vacuum Science and Technology,* J Lafferty, Ed., John Wiley & Sons, New York.
Hinkle LD, 1998, in *Handbook of Vacuum Science and Technology,* Hoffmann DM, Singh B and Thomas JH, Ed., Academic Press, San Diego, CA.
Hinkle LD and Uttaro FL, 1996, *Vacuum,* 47, 523.

Hobson JP 1961, *Transactions of the 8th National Vacuum Symposium,* Vol. 1, p. 26, Pergamon Press, New York.
Hobson JP and Redhead PA, 1958, *Can. J. Phys.*, 36, 271.
Hobson JP and Salzman DB, 2000, *J. Vac. Sci. Technol.*, A18(4), 1758–1765.
Holkeboer DH, Jones DW, Pagano F, and Santeler DJ, 1993, *Vacuum Technology and Space Simulation*, American Vacuum Society Classics, AIP, New York.
Hudson JB, 1992, *Surface Science, an Introduction,* Butterworth–Heinemann, Boston, MA.
Hudson JB, 1998, in *Foundations of Vacuum Science and Technology,* Lafferty JM, Ed., John Wiley & Sons, New York.
James AP, 1987, *Vacuum,* 37, 677.
Jaeger RC, 2002, *Introduction to Microelectronic Fabrication,* 2nd ed., Addison-Wesley Reading, MA.
Jeans JH, 1982, *An Introduction to The Kinetic Theory of Gases,* Cambridge University Press, Cambridge, U.K.
Jitschin W, Migwi JK, and Grosse G, 1990, *Vacuum,* 40, 293.
Ju L, Blair DG, and Zhao C, 2000, *Rep. Prog. Phys.,* 63, 1317–1427.
Kaminsky M, 1965, *Atomic and Ionic Impact Phenomena on Metal Surfaces,* Springer, Berlin.
Kaye GWC and Laby TH, 1995, *Physical and Chemical Constants,* 16th ed., John Wiley & Sons, New York.
Kempe's Engineers Year-Book 2001, CMP Information Ltd., Tonbridge, U.K.
Kendall BRF and Drubetsky E, 1997, *J. Vac. Sci. Technol.,* A15, 740.
Kennard EH, 1938, *Kinetic Theory of Gases,* McGraw-Hill, New York.
Knudsen M, 1909, *Ann. Phys. Lpz.,* 28, 75–130.
Knudsen M, 1934, *Kinetic Theory of Gases*, Methuen, London.
Kruger CH and Shapiro AH, 1961, *Transactions of the 7th National Vacuum Symposium (1960)*, Pergamon, New York, pp. 6–12.
Lafferty JM, Ed., 1998, *Foundations of Vacuum Science and Technology,* John Wiley & Sons, New York.
Leck JH, 1970, *Vacuum,* 20, 369.
Leck JH, 1989, *Total and Partial Pressure Measurement in Vacuum Systems,* Blackie, Glasgow and London.
Levenson LL, Milleron N, and Davis DH, 1963, *Le Vide*, 103, 42.
Li M and Dylla HF, 1993, *J. Vac. Sci. Technol.*, A11, 1702–1707; 1994, *J. Vac. Sci. Technol.*, A12, 1772–1777; 1995, *J. Vac. Sci. Technol.*, A13, 1872–1878.
Lindenau B, 1988, *Vacuum,* 38, 893.
Lindenau BE and Fremerey JK, 1991, *J. Vac. Sci. Technol.*, A9, 2737–2743.
Livesey RG, 1998, in *Foundations of Vacuum Science and Technology,* Lafferty J, Ed., John Wiley & Sons, New York.
Loeb LB, 1961, *The Kinetic Theory of Gases,* Dover, New York.
Luther AC and Inglis AF, 1999, *Video Engineering*, McGraw-Hill, New York.
Madey TE, 1984, *J. Vac. Sci. Technol.*, A2(2), 110–117.
Maissel LI and Francombe MH, 1991, *An Introduction to Thin Films*, Gordon and Breach, New York.
McCash EM, 2001, *Surface Chemistry,* Oxford University Press, Oxford, U.K.
McCulloh KE, Wood SD, and Tilford CR, 1985, *J. Vac. Sci. Technol.*, A3, 1738.
Morgan DV and Board K, 1994, *An Introduction to Semiconductor Microtechnology,* John Wiley & Sons, Chichester, U.K.
Mossharrafa M, 1970, *Ind. Res. Dev.,* March, p. 24.

Nash PJ, 1987, *Vacuum,* 37, 643.
Nash PJ and Thompson TJ, 1983, *J. Vac. Sci. Technol.,* A1, 172.
Oatley CW, 1957, *Br. J. Appl. Phys.,* 8, 15.
O'Hanlon JF, 1994, *J. Vac. Sci. Technol.,* A12, 921–926.
O'Hanlon JF, 2003, *A User's Guide to Vacuum Technology,* 3rd ed., John Wiley & Sons, New York.
Orchard JC and Scales SC, 1999, *Vacuum,* 53, 357–361.
Peacock RN, Peacock NT, and Hauschulz DS, 1991, *J. Vac. Sci. Technol.,* A9, 1987.
Peacock RN, 1998, in *Foundations of Vacuum Science and Technology,* Lafferty J, Ed., John Wiley & Sons, New York.
Pendlebury JM, 1985, *Kinetic Theory,* IOP, Bristol, U.K.
Perrin J, 1999, *Vacuum Solutions,* May/June 25–28, IOP, Bristol, U.K.
Pert GJ, 2002, *Contemp. Phys.,* 43, 303–306.
Pirani M, 1906, *Verh. der Deutsch Phys. Ges.,* 8, 686.
Poulter KF, 1977, *J. Phys. E. Sci. Instrum.,* 10, 112–125.
Poulter KF, 1981, *Vide-Couches Minces,* 36, 521.
Poulter KF, Rodgers M-J, Nash PJ, Thompson TJ, and Perkin MP, 1983, *Vacuum,* 33(6), 311–316.
Power BD, 1966, *High Vacuum Pumping Equipment*, Chapman and Hall, London.
Powers MJ, 1998, in *Handbook of Vacuum Science and Technology,* Hoffmann DM, Singh B, and Thomas JH, Eds., Academic Press, San Diego, CA.
Prescott J, 1961, *Applied Elasticity,* Dover, New York.
Present RD, 1958, *Kinetic Theory of Gases*, McGraw-Hill, New York.
Redhead PA, 1960, *Rev. Sci. Instrum.,* 31, 343.
Redhead PA, 1984, *J. Vac. Sci. Technol.,* A2(2), 132–138.
Redhead PA, 1987, *J. Vac. Sci. Technol.,* A5, 3215.
Redhead PA, 1993, *Vacuum,* 44, 559–564.
Redhead PA, 1998, in *Foundations of Vacuum Science and Technology,* Lafferty J, Ed., John Wiley & Sons, New York.
Redhead PA, 1999, *Vacuum,* 53, 137–149.
Redhead PA, Hobson JP, and Kornelsen EV, 1968, *The Physical Basis of Ultra-high Vacuum,* reprinted 1993 in AVS Classics Series, AIP Press, Woodbury, NY.
Reid RJ, 1982, *J. Vac. Sci. Technol.,* 20, 1156–1158.
Reid RJ, 1997, *Vacuum Solutions*, 1.1, IOP, Bristol, U.K.
Reif F, 1965, *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill, New York.
Roth A, 1990, *Vacuum Technology,* 3rd ed., Elsevier Science, Amsterdam.
Santeler DJ, 1986, *J. Vac. Sci. Technol.,* A4, 348–352 and 338–343.
Santeler DJ and Boeckmann MD, 1987, *J. Vac. Sci. Technol.,* A5, 2493–2496.
Sears FW, 1965, *An Introduction to Thermodynamics, Kinetic Theory of Gases, and Statistical Mechanics*, Addison-Wesley, Reading, MA.
Sears FW and Sallinger GL, 1986, *Thermodynamics, Kinetic Theory, and Statistical Mechanics*, Addison-Wesley, Reading, MA.
Shah SI, 1995, *Handbook of Thin Film Process Technology,* IOP, Bristol, U.K.
Shah I, 1997, *Phys. World,* 10, 45–47.
Shin YH, Lee KJ, and Chung Jhung KH, 1996, *Vacuum,* 47, 679.
Singleton JH, 1998, in *Handbook of Vacuum Science and Technology,* Hoffmann DM, Singh B, and Thomas JH, Eds., Academic Press, San Diego, CA.
Singleton JH, 2001, *J. Vac. Sci. Technol.,* A19, 1712–1719.
Somorjai GA, 1994, *Introduction to Surface Chemistry and Catalysis,* John Wiley & Sons, New York.

Steckelmacher W, 1966, *Vacuum,* 16, 561–583.
Steckelmacher W, 1986, *Rep. Prog. Phys.,* 49, 1083–1107.
Steckelmacher W, 1987, *Vacuum,* 37, 651.
Stuart RV, 1983, *Vacuum Technology, Thin Films, and Sputtering*, Academic Press, New York.
Su Y, Sawada T, Takemito J, and Haga S, 1996, *Vacuum,* 47, 815–818.
Sullivan JJ, 1985, *J. Vac. Sci. Technol.,* A3, 1721.
Tate JT and Smith PT, 1932, *Phys. Rev.,* 39, 270.
Temperley HNV, 1961, *Properties of Matter,* University Tutorial Press, London.
Tilford CR, Dittman S and McCulloh KE, 1988, *J. Vac. Sci. Technol.,* A6, 2853.
Tilford CR, 1990,
Tilford CR, 1992, *Physical Methods of Chemistry,* 2nd ed., BW Rossiter and Baetzold, Eds., John Wiley & Sons, pp. 101–73.
Tilford CR, Filippelli AR, and Abbott PJ, 1995, *J. Vac. Sci. Technol.,* A13, 485.
Trickett BA, 1987, *Vacuum,* 37, 747–755.
Troup AP and Dennis NTM, 1991, *J. Vac. Sci. Technol.,* A9, 2048–2052.
Van Atta CM, 1965, *Vacuum Science and Engineering,* McGraw-Hill, New York.
Venables JA, 2000, *Introduction to Surface and Thin Film Processes,* Cambridge University Press, Cambridge, U.K.
Walton AJ, 1983, *The Three Phases of Matter,* 2nd ed., Clarendon Press, Oxford, U.K.
Watanabe F, 1993, *J. Vac. Sci. Technol.,* A11, 1620.
Watanabe F, 1996, *Vacuum,* 47, 567.
Welch KM, 1991, *Capture Pumping Technology, an Introduction*, Pergamon Press, Oxford, U.K.
Wehner GK and Anderson GS, 1970, in *Handbook of Thin Film Technology,* Maissel LI and Glang R, McGraw-Hill, New York.
Weston G, 1980, *Vacuum,* 30, 49.
Westwood WD, 2003, *Sputter Deposition*, American Vacuum Society Education Committee.
Woodruff DP and Delchar TA, 1986, *Modern Techniques of Surface Science*, Cambridge University Press, Cambridge, U.K.
Weston G, 1985, *Ultrahigh Vacuum Practice,* Butterworth, London.
Wutz M, Adam H, and Walcher W, 1989, *Theory and Practice of Vacuum Technology* trans. W Steckelmacher, Vieweg, Braunschweig, Germany.
Zee A, 2000, *Fearful Symmetry: the search for beauty in modern physics*, Princeton University Press, Princeton, NJ.
Zemansky MW and Dittman RH, 1997, *Heat and Thermodynamics,* 7th ed., McGraw-Hill, New York.

List of Symbols

The range of the physical quantities needed means that unavoidably some share the same symbol, but the context will distinguish the intended meaning.

A	area
B	magnetic flux density
$c(x,t)$	concentration as a function of place and time
C	conductance
C_L	conductance of a long pipe
C_A	conductance of an aperture
C_O	conductance of a circular aperture
d	diameter of a molecule
D	diameter of a pipe or circular opening, diffusion coefficient
e	electronic charge
f	function, frequency
F	force
h	height, Planck's constant
H	speed factor, enthalpy
H_A	molar heat of adsorption
Ho	Ho coefficient
i, I	current
i_r	residual current
$I+$	ion current
$I-$	electron current
I	molecular flux per second from unit area into unit solid angle at specified angle
j	diffusion or permeation flux per unit area per second
J	flux of molecules per unit area per second
J_C	condensation flux (molecules per unit area per second)
J_E	evaporation flux (molecules per unit area per second)

k	Boltzmann's constant, general constant of proportionality
K	gauge constant, thermal conductivity, compression ratio
K_0	compression ratio for zero flow
K_p	pressure ratio
Kn	Knudsen number
l	length
L	pipe length
ML	monolayer
m	mass of a molecule
M	molar mass (gram molecular weight)
M_r	relative mass number, but taken as M in Section 7.7 to simplify notation
n	neutron
n	number density of molecules
n_a	number of molecules adsorbed per unit area
n_M	number of moles
N	number of molecules
N_A	Avogadro's number
N_0	number of molecules in a monolayer, taken to be 10^{15} per cm^2
$p+$	proton
p	pressure
p_i	partial pressure of i th gas, inlet pressure
p_e	equilibrium vapor pressure
p_u	ultimate pressure
p_U	upstream pressure
p_D	downstream pressure
$\bar{p}$	mean pressure
p_o	outlet pressure
p_0	atmospheric pressure, initial pressure
P	pressure offset (spinning rotor gauge)
q	binding energy of adsorbed molecule expressed in eV
q_G	specific gassing rate
q_1	specific gassing rate after 1 hour of pumping
Q	throughput
Q_G	throughput due to outgassing
Q_L	throughput due to leaks (leak rate)
Q_P	throughput due to process-created gas

Q_T	throughput corresponding to total gas load
Q_{0C}	throughput in choked flow
$\dot{Q}$	rate of loss of heat
r	radial coordinate
r_0	equilibrium distance of adsorbed molecule from a surface
R	radius, relative sensitivity
R_0	universal gas constant
RH	relative humidity
Re	Reynolds' number
s	sticking coefficient
s_r	speed ratio U/v_a
S	pumping speed (volumetric flow rate), entropy
S^*	pumping speed at a pump inlet
S_D	volumetric flow rate of free air displacement
S_{AC}	pumping speed of choked aperture
t	time, dimensionless ratio of elapsed time of outgassing normalized to one hour, thickness
T	time constant
T	absolute temperature
T_C	critical temperature
u	average drift velocity of a molecule, velocity of bulk gas
U	imposed velocity, potential
v	velocity of a molecule
v_α	most probable velocity of a molecule
$\bar{v}$	average velocity of a molecule
$\overline{v^2}$	mean square velocity of molecules
V	volume of a chamber, volume of gas, potential
$\dot{V}$	volumetric flow rate
w	width
W	mass of gas, probability of pumping
$\dot{W}$	mass flow rate
x_i	mole fraction
x, y, z	spatial coordinates
Z	atomic number
α	transmission probability
α_E	energy accommodation coefficient

φ	potential, angle
γ	ratio of principal specific heats of a gas, c_p/c_v
η	viscosity coefficient
k	thermal conductivity
λ	mean free path
θ	angle measured from surface normal, fractional coverage (n_a/N_0)
ρ	gas density
σ	ionization cross section, tangential momentum accommodation coefficient
τ	time constant
τ_0	reciprocal of atomic attempt frequency (kT/h)
τ_a	stay time
ω	angular velocity
$\dot{\omega}$	angular deceleration/acceleration

Index

Absolute temperature, 11
Activation energy, 57
Active-matrix LCD, 273
Adsorbed phase, 61
Adsorption, 49, 55
Adsorption stay time, 58
Aperture, 88, 97
Argon instability, 179
Atmosphere, constituents of, 17
Atomic mass unit, 12
Atomic mass, relative, 12
Average molecular speed, 26, 29
Avogadro's number, 12

Back-migration, 156
Back-streaming, 155
Baffles, 155
Bakeout, 66, 69
Base pressure, 109
Bayard-Alpert gauge, BAG, 209
Bellows, 306
Bent beam gauge, 214
Bernoulli's equation, 85
BET isotherm, 61
Binding energy, molecule to surface, 55
Boltzmann factor, 57
Boltzmann's constant, 26
Bourdon gauge, 190
Boyle's law, 11, 218
Bremsstrahlung, 288

Calibrated helium leak, 319
Capacitance diaphragm gauge, 190
Capsule gauge, 190
Capture pumps, 160, 119
Ceramics, 301
Channel electron multiplier, 229
Chemical vapour deposition, CVD, 264
Chemisorption, 55
Choked flow, 88
Claw pump, 132
Cleaning, 312
Cold cap, 156
Cold cathode gauge, 215
Collision frequency, 38
Collision ratio, surface/volume, 42
Compressibility, effects of, 86
Compression ratio, 121, 128
Condensation, 39
Conductance, 80
Conflat flange, 302
Continuum flow, 82
Continuum state of a gas, 42
Contra-flow leak detector, 317
Cosine law, 53
Critical backing pressure, 155
Critical foreline pressure, 155
Critical temperature, 19
Crossover, 120, 159, 251
Cryopump, 162
Cryosorption pump, 160
CVD, chemical vapour deposition, 264, 267

Dalton's law, 16
Decade time, pumping, 110
Degassing, 7, 212, 258
Desorption, 55
Desorption energy, 56
Deuterium/tritium reaction, 285
Diaphragm gauge, 190
Diaphragm pump, 130
Differential manometer, 187, 194
Differential mode, diaphragm gauge, 194
Differential pumping, 113, 244
Diffuse scattering, 53
Diffusion, 66
Diffusion coefficient, 66
Diffusion pump, 153
Diode sputter ion pump, 176
Direct pressure gauge, 184
Display, mass spectra, 232
Dissociation products, 233
Drying, 7, 237
Dushman formula, 95
Dynamic expansion, 220

Effusion, 53
Elastomer seals, 303
Electrical feedthroughs, 299
Electron stimulated desorption, ESD, 212
Electron synchrotron, 281
Energy accommodation coefficient, 52
Entrance and exit effects, 102
Entry length, 92
Equilibrium vapour pressure, 17
Evaporation, 39
Evaporation rate, 40
Exit loss, 104
Extractor gauge, 213
Extreme high vacuum, 7

Faraday cup detector, 229
Fick's laws, 66
Field emission display, 277
Filaments, ionization gauge, 212, 213
Fine leak valve, 305
First wall, 288
Flanges, 302
Flat panel displays, 272
Flow regimes, 77
Fluid feedthroughs, 299
Foreline trap, 125
Fractional coverage, adsorbed layer, 60
Fragmentation, 222
Freeze drying, 239
Fully developed flow, 92
Fusion triple product, 286
Fusion, thermonuclear, 285

Gas ballast, 124
Gas laws, 11
Gas load, 108
Gas mixtures, 16
Gas sources, 108
Gaseous state, 25
Gases and vapours, 21
Gate valve, 306
Gauge calibration, 218
Gauge factor, 210
Gauges, vacuum, 183
Gettering, 171, 176
Gifford-McMahon cycle, 164
Glow discharge cleaning, 73
Gravitational waves, 289
Guericke, von, 4

Heat conduction, 45
Heat of adsorption, 56
Higgs boson, 293
High vacuum, 7
Ho coefficient, 120
Horror vacui, 2
Hot cathode gauge, 205

Ideal gas, 11, 26
Impact rate, 26, 30, 38
Indirect pressure gauge, 183
Integrated circuit, 269
Inter-molecular separation in a gas, 25
Inverted magnetron gauge, 216
Ion beam assisted deposition, IBAD, 267
Ion beam sputtering, 266
Ion burial, 178
Ion collector, 208
Ion source box, 222
Ionization gauge, 205
Ionization-cross section, 206
Isentropic flow, 86
ISO-K flanges, 302
Isolation valve, 305
Isotherms, 20
Isotope effects, 233

JET tokamak, 287

KF flanges, 302
Kinetic pressure, 86
Kinetic theory of gases, 26
Knudsen cell, 54
Knudsen cosine law, 53
Knudsen number, 41

Laminar flow, 83
Langmuir isotherm, 61
Large Electron Positron Collider, LEP, 292
Large Hadron Collider, LHC, 295
Laser interferometer, 290
Leak detection, 312
Leakage rate, 314
Leaks, 312
Leaks and pumping, 314
LEP, 292
LHC, 295
Liquid crystal active matrix display, 273
Liquid state, 25
Liter, 6
Lithography, 269
Logarithmic display, gas analysis, 235
Long pipe, molecular flow, 95

Low vacuum, 7

m/e ratio, 222
Mach number, 87
Magdeburg hemispheres, 4
Magnetic sector analyser, 230
Magnetron sputtering, 266
Mass flow controller, MFC, 268
Mass flow rate, 79
Mass spectrum, 222
Materials, 300
Mathieu equations, 227
Maxwell-Boltzmann distribution, 27
McLeod gauge, 187
Mean free path, 26, 34, 38
Mean pressure, 94
Medium vacuum, 7
Metal seals, 303
Michelson interferometer, 290
Microelectronic fabrication, 269
Millibar, 6
Millibar-liter, 13
Millibar-liter per second, 79
MOCVD, 269
Modulator gauge, 211
Molar mass, 12
Mole, 12
Molecular beaming, 104
Molecular diameter, 25, 34
Molecular drag, 70
Molecular drag pump, 138
Molecular flow, 77, 95
Molecular flow conductance, aperture, 97
Molecular flow conductance, long pipe, 95
Molecular flow conductance, short pipe, 100
Molecular flux, 40
Molecular impingement rate, 26, 30
Molecular number density, 1, 26, 38
Molecular state of a gas, 42
Molecular weight, 12
Molecule, 12
Momentum accommodation, 203
Momentum transfer pumps, 119
Monolayer formation time, 60
Monolayer, ML, 59
Monte-Carlo methods, 102
Motion feedthroughs, 306
Multi-layer adsorption, 61
Multiply-charged ions, 233

NEG pump, 173
NMOS transistor, 271
Non-evaporable getter pump, 173
Nude gauge, 212
Number density, of molecules, 1, 38

Oil-sealed rotary pump, 122
Orifice flow technique, 220
Outgassing, 62
Outgassing rate, 63

Parallel connection, 81
Partial pressure, 16
Particle flow rate, 79
Particle physics, 292
Pascal, 1, 5
Penning gauge, 215
Permeation, 68
Photolithography, 269
Photoresist, 269
Physical adsorption, 55
Pipe conductance, molecular, 95, 100
Pipe conductance, viscous, 94
Pirani gauge, 198
Pitot tube, 86
Pixel, 273
Plasma, 286
Plasma display, 275
Poiseuille formula, 93, 240
Positive displacement pumps, 122
Pressure, 1
Pressure ratio, 121
Primary pump, 120
Principle of detailed balance, 51
Process gas, 108
Pump-down time, 110
Pumping equation, 109
Pumping process, 108
Pumping speed, 80
Pumping time, 110
PVD, physical vapour deposition, 264

Quadrupole mass filter, 224

Ranges of vacuum, 6
Rate of adsorption, 60
Rate of desorption, 60
Real gas, 13
Real leak, 313
Regeneration, 170
Relative gauge sensitivity, 210
Relative humidity, 17
Relative molecular mass, 12
Residual current, 208

Residual gas, 221
Residual gas analyser, RGA, 222
Resist, 269
Resolution, 10% valley, 223
Resolving power, 223
Reynolds' number, 83
Right-angle valves, 306
Roots pump, 126
Rotary vane pump, 122

Saturated vapour pressure, 19
Scattering, molecules from surfaces, 50
Screw pump, 133
Scroll pump, 135
Seals, 302
Secondary pump, 120
Semiconductor fabrication, 269
Series connection, 81
Series expansion, 218
Short ducts, 89, 100
Silicon 'chip', 269
Silicon oxide, 270
Sojourn time, 58
Solid angle, 30
Sorption pump, 160, 257
Space-charge effects, 177, 208, 215
Specific outgassing rate, 63
Specular reflection, 50
Speed factor, 120
Spinning rotor gauge, 201
Sputter-ion pump, 176
Sputtering, 72
Sputtering yield, 72
Staging ratio, 152, 247
Stagnation point, 86
Standard atmosphere, 5
Standard cc per minute(SCCM), 80
Standard conditions, 2
Standard liter per minute(SLM), 80
Standard model, 293
Static pressure, 86
Stepper motor, 310
Sticking coefficient, 60
Sub-pixel, 273
Superconducting magnets, 296
Surface analysis, 74
Surface scattering, 50
Survival probability, 36
Synchrotron radiation, 281
Synchrotron radiation source, 281

Tangential momentum accommodation, 53, 203
TFT, thin film transistor, 273
Thermal agitation, 25, 39
Thermal conduction, 45, 195
Thermal transpiration, 46
Thermally activated process, 40, 58, 66
Thermionic ionization gauge, 205
Thermonuclear fusion, 285
Thermocouple gauge, 200
Thin film coatings, 264
Throughput, 78
Time constant, 110
Time-of-flight analyser, 230
Titanium sublimation pump, 171
Tokamak, 287
Torr, 5
Torricelli's experiment, 3
Torricellian vacuum, 3
Total pressure, 16
Transferable reference gauge, 195, 205
Transitional flow, 77
Transmission probability, 102
Trap, foreline, 125
Trapped volume, 310
Triode sputter-ion pump, 179
Turbo-drag pump, 153
Turbomolecular pump, 143, 253
Turbulent flow, 83

U-tube manometer, 187
UHV, 7
Ultimate pressure, 109, 122
Ultra-high vacuum, 7, 66, 253, 257
Units, 5

Vacustat, 189
Vacuum, 1
Vacuum coating, 264
Vacuum deposition, 264
Vacuum drying, 41, 237
Vacuum gauges, 183
Vacuum insulated panels, 261
Vacuum molding, 7
Vacuum pumps, 119
Vacuum tube, 4
Vacuum, units of, 5
Valves, 305
Van der Waals equation, 14
Van der Waals forces, 15
Vapor, 17
Vapor jet pump, 153
Vapor pressure, 17
Velocity profile, 83
Venting, 92, 252, 255

Virial equation of state, 14
Virtual leak, 313
Viscosity, 43
Viscous flow, 82
VLT, Very Large Telescope, 278
Volumetric flow rate, 79

Water vapor in the atmosphere, 17
Water vapor pump, 171
Water vapor, outgassing of, 69
Water vapor, pumping of, 237
Water, vapor pressure, 22
Windows, 299

X-ray effect, 209
X-ray limit, 209
XHV, 7
XHV gauges, 213

Zeolites, 160