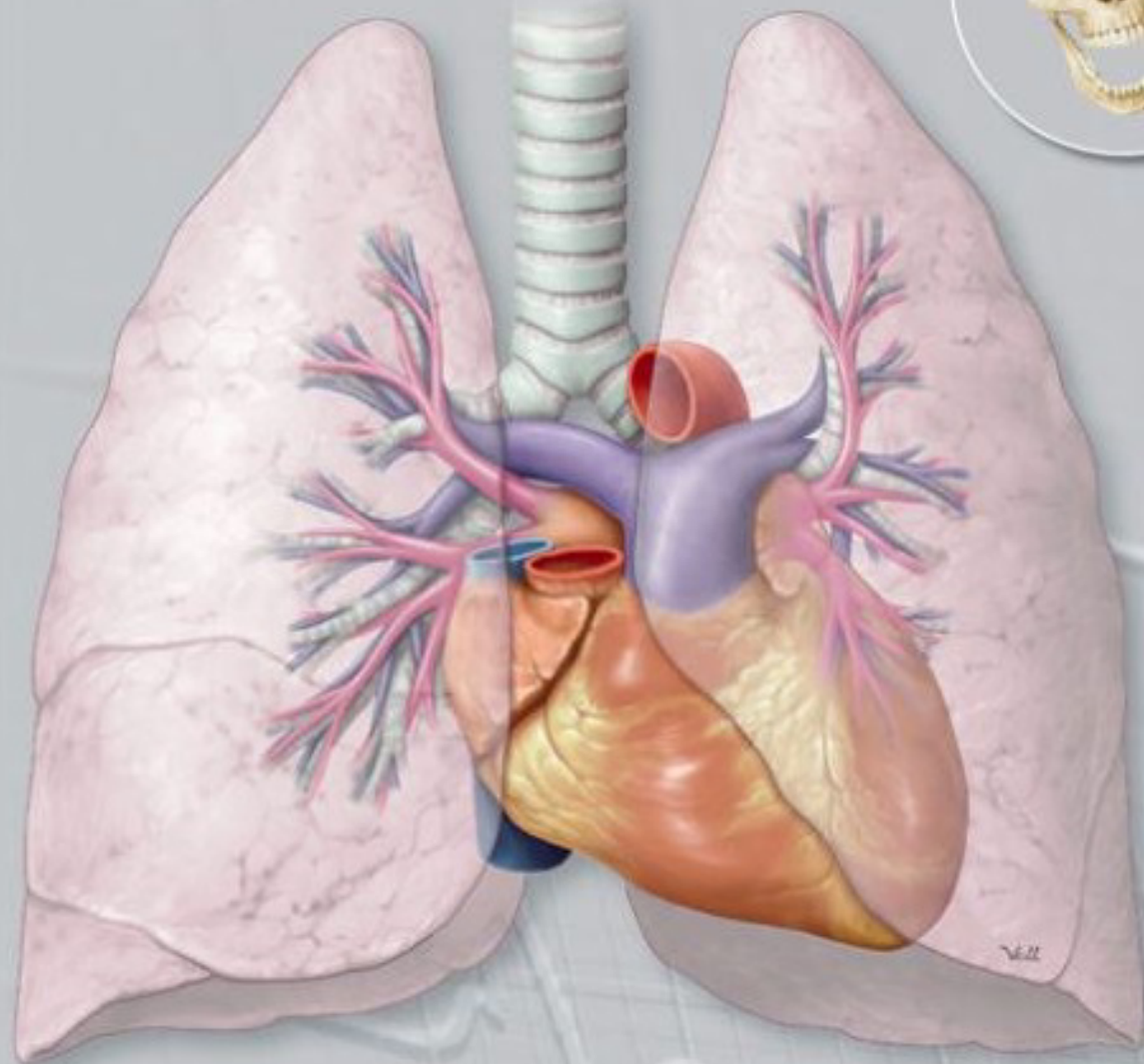


Education

Principles of Medical Physiology

Second Edition

Sabyasachi Sircar



Thieme

Principles of Medical Physiology

2nd edition

Dedicated to the memory of my parents

Who were the best teachers I have ever known

Principles of Medical Physiology

2nd edition

Sabyasachi Sircar, MD

Professor and Head of Department of Physiology

All India Institute of Medical Sciences

Jodhpur, Rajasthan

India

811 Illustrations

116 Tables

19 Text Boxes

Thieme Medical and Scientific Publishers Private Limited

A-12, Second Floor, Sector - 2

Noida, Uttar Pradesh - 201 301, India

Email: customerservice@thieme.in

www.thieme.in

Thieme

Delhi • Stuttgart • New York • Rio

Thieme Publishers Delhi
A - 12, Second Floor, Sector - 2
Noida, Uttar Pradesh - 201 301, India

Publishing Manager: Dr. Sonu Singh
Development Editor: Dr. Gaurav Singh
Head Editorial Services: Rachna Sinha
Assistant Manager - Editorial Production: Kumar Kunal
Sales and Marketing Director: Arun Kumar Majji
Chief Executive Officer: Ajit Kohli

Copyright © 2014 by Thieme Medical and Scientific
Publishers Private Limited.

1st Edition Published in 2007.

Principles of Medical Physiology /
Sabyasachi Sircar, 2nd ed.

Important note: Medical knowledge is ever changing. As new research and clinical experience broaden our knowledge, changes in treatment and drug therapy may be required. The authors and editors of the material herein have consulted sources believed to be reliable in their efforts to provide information that is complete and in accord with the standards accepted at the time of publication. However, in view of the possibility of human error by the authors, editors, or publishers of the work herein or changes in medical knowledge, neither the authors, editors, nor publishers, nor any other party who has been involved in the preparation of this work, warrants that the information contained herein is in every respect accurate or complete, and they are not responsible for any errors or omissions or for the results obtained from use of such information. Readers are encouraged to confirm the information contained herein with other sources. For example, readers are advised to check the product information sheet included in the package of each drug they plan to administer to be certain that the information contained in this publication is accurate and that changes have not been made in the recommended dose or in the contraindications for administration. This recommendation is of particular importance in connection with new or infrequently used drugs. Some of the product names, patents, and registered designs referred to in this book are in fact registered trademarks or proprietary names even though specific reference to this fact is not always made in the text. Therefore, the appearance of a name without designation as proprietary is not to be construed as a representation by the publishers that it is in the public domain.

Published by Thieme Publishers Delhi
A - 12, Second Floor, Sector - 2, Noida - 201 301,
Uttar Pradesh, India, +911204556600
Email: customerservice@thieme.in
www.thieme.in
Printer: Saurabh Printers Pvt Ltd

5 4 3 2 1

ISBN 978-93-82076-53-7
eISBN: 978-93-82076-64-3

This book, including all parts thereof, is legally protected by copyright. Any use, exploitation, or commercialization outside the narrow limits set by copyright legislation without the publisher's consent is illegal and liable to prosecution. This applies in particular to photostat reproduction, copying, mimeographing or duplication of any kind, translating, preparation of microfilms, and electronic data processing and storage.

Preface to the Second Edition

The second edition of Principles of Medical Physiology remains committed to the quartet of utility, lucidity, brevity, and contextual relevance of the text. The changes in this edition broadly fall in four categories: *First*, the errors that crept into the first edition are weeded out. *Second*, at the end of each chapter, revision questions are added. The questions, which somewhat define the “learning objectives” of the chapter, are mostly conceptual though some require only factual recall. Indeed, a quick look at these questions before going through the chapter should provide the necessary focus for appreciating the text. *Third*, several chapters have been substantially modified. Notable among them are the chapters on Membrane Excitation & Action Potential, Electrocardiography, Cardiac Output, Hemodynam-

ics, and Respiratory mechanics. Chapters 16 (Characteristics of Muscle Contraction) and 17 (Muscle Elasticity) of the previous edition are merged into a single chapter (Characteristics of Skeletal Muscle Contraction) while Chapter 38 (Circulatory Pathway & Hemodynamics) is split into two. *Finally*, if the first edition of the book was aimed at providing a text that is “easy to understand”, the second edition is aimed at making the text “easy to revise” by highlighting important parts of the text so that the rigorous structure of the text comes alive, making it even more suitable for reading at bedtime for relaxation!

I welcome, indeed beseech, the feedback of readers which may be emailed to me at pmp.sircar@gmail.com

Sabyasachi Sircar
January, 2014

Preface to the first Edition

The real voyage of discovery consists not in seeking new landscapes but in having new eyes.

Marcel Proust

About a decade ago, I set out to write a book of human physiology that undergraduate medical students could read at bedtime for relaxation, and the result is now in your hands! The four principal considerations that have shaped this book are utility, lucidity, brevity, and contextual relevance.

Every sentence in this text has been carefully weighed for its utility before it was granted admission to this book. The utility could be its clinical relevance, conceptual elegance, or its importance in examinations. For deciding on the utility, I have used the simple test of time: Surely, any physiological fact that I fail to recall myself despite teaching physiology for more than two decades need not be taught to the student either. Admittedly though, I have erred and included in this book a lot more than what I can recall effortlessly.

Lucidity has been attempted through numerous full-color illustrations and flow charts (which I have drawn myself), tables, analogies, cartoons, anecdotes, and caveats, and by restricting the text to the basics.

Brevity has been achieved, not by compromising on the elucidation of important facts and concepts but through a rigorous economy of words, elimination of introductory passages to chapters, avoiding repetitions, and substituting wherever appropriate a well-labeled picture in place of “a thousand words.”

Ensuring contextual relevance has been the toughest of all and is also by far the most distinctive feature of this book. I have spent months planning as to what goes in where and mulling over the rubrics. Sundry facts that could not be woven into a flowing narrative have been omitted unless deemed especially important. Some topics do not fit nicely

anywhere but are too small to be presented as a separate chapter. Others seem to be appropriate at too many places. There are still others that can be presented as a separate chapter only at the cost of extensive repetition. Hence, topics like exercise physiology or neonatal physiology are not presented as separate chapters but are discussed at several places throughout the book, wherever relevant.

Conspicuous by their paucity or absence in this book are the references to the work of scientists, and the normal range of laboratory values. I do not discount the importance of these but leave them for students to find out from reference books. I have witnessed with dismay heated arguments among teachers on what should be considered the “normal” range of physiological parameters, often sidetracking vital conceptual questions, and would not like to join the fray! Rather, I would like to impress on the readers that there can be no global range of normal biological values and that in all such considerations, the context is important, as was strikingly brought out by the comment of the copy-editor of this book, “Are these data real?” referring to the 60 kg weight of a “reference man” (Table 78.3) that is well below western norms. Hence all “normal” values mentioned in this book should be taken as more illustrative than authoritative as they are drawn from myriad sources and altered slightly to make them easy to remember.

I hope this book ushers in happier times for students interested in understanding medical physiology and doing well in competitive examinations.

Sabyasachi Sircar
August, 2007

Photo Acknowledgments

The following figures are reproduced/redrawn/adapted, with permission, from other sources.

From Thieme Publications

Figs. 4.2, 8.5 Reproduced from Kuehnel W. **Color Atlas of Cytology, Histology, and Microscopic Anatomy**. Stuttgart-New York: Thieme ; 2003.

Fig. 92.6 Reproduced from Passarge E. **Color Atlas of Genetics**. Stuttgart-New York: Thieme; 2007.

Figs. 28.1, 28.3, 29.1, 29.2, 32.3, 32.5, 32.6 Reproduced from Theml N, Diem H, Haferlach T. **Color Atlas of Hematology**. Stuttgart-New York: Thieme; 2004.

Fig. 30.2 Reproduced from Burmester G-R, Pezzutto A. **Color Atlas of Immunology**. Stuttgart-New York: Thieme; 2006.

Figs. 103.20, 105.2 Reproduced from Rohkamm R. **Color Atlas of Neurology**. Stuttgart-New York: Thieme; 2006.

Figs. 56.3, 56.4, 56.5, 76.4, 77.4, 84.4, 99.4 Reproduced from Riede U-N, Werner M. **Color Atlas of Pathology**. Stuttgart-New York: Thieme; 2005.

Figs. 8.10, 49.19 Reproduced from Despopoulos A, Silbernagl. **Color Atlas of Physiology**. Stuttgart-New York: Thieme; 2006

Figs. 45.7, 107.1, 107.3, 108.2, 120.2, 102.5 Reproduced from Mumenthaler M, Mattle H, Taub E. **Fundamentals of Neurology**. Stuttgart-New York: Thieme; 2006.

Figs. 76.1, 76.3, 84.5 Reproduced from Eastman G, Wald C. Crossin J. **Getting Started in Radiology**. 2006, Thieme Stuttgart.

Figs. 1.17, 8.2, 9.1, 9.7, 37.6, 37.8, 37.9, 38.5, 96.1. Reproduced from Schuenke M, et al. **Thieme Atlas of Anatomy. General Anatomy and Musculoskeletal System**. Stuttgart-New York : Thieme; 2006.

Figs. 33.1, 35.1, 37.2, 44.1, 46 .1, 47.4, 48.2 , 49. 1, 57.1, 70.3, 71.1, 71.3, 72.1, 83.1, 84.2, 84.6, 85.1, 87.1, 89.6, 90.4, 91.1, 91.2, 91.4, 91.9, 93.2, 101.1

Reproduced from Schuenke M, et al. **Thieme Atlas of Anatomy. Neck and Internal Organs**. Stuttgart -New York: Thieme; 2006.

From Other Publications

Fig. 10.6A Redrawn, with permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc., from Curtis HJ, Cole KS: Membrane resting and action potentials from the squid giant axon . J Cell Comp Physiol, 1942,19:139.

Fig. 11.4 Redrawn from Hodgkin AL, Katz B: The electrical activity of the giant axon of the squid. J Physiol (Lond), 1949, 109:245 .

Fig. 11.7 Redrawn from Mortimer JT: Motor prosthesis. In Brooks VB, editor: Handbook of Physiology. ed 2, Washington, 1981, Motor control, American Physiological Society.

Fig. 12.8A Redrawn from Ackerman MJ, Clapham DE: Ion channels: Basic science and clinical disease. N Eng J Med 1997; 336:1575.

Fig. 13.4 Redrawn from Rushton WAH: A theory of the effects of fiber size in medullated nerve. J Physiol (Lond), 1951,115:116.

Fig. 16.1 Redrawn from Jewell BR, Wilkie DR: The mechanical properties of relaxing muscle. J Physiol (Lond), 1960, 152:30.

Fig. 16.3B Redrawn from Gordon AM, Huxley AF, Julian FJ: The variation in isometric tension with sarcomere length in vertebrate muscle fibers. J Physiol (Lond), 1960, 152:30.

Fig. 29.3, A, B. Redrawn, with kind permission of Springer Science and Business Media, from Faucett DW: A textbook of histology, ed 12, New York , 1994, Chapman & Hall.

Fig. 38.10 Redrawn from O'Brien ET, O'Malley K: British Medical Journal, 1979, 2(6197), 1048.

Fig. 42.6 Redrawn from O'Brien ET, O'Malley K: BMJ, 1979, 2(6198), 1126

Fig. 45.5 *Courtesy* Dr. M Abid Geelani, Professor of Cardiothoracic Surgery, GB Pant Hospital, Maulana Azad Medical College. New Delhi.

Fig. 49.9 Redrawn from Rahn H, Otis AB, Chadwick LE, Fenn WO :The pressure-volume diagram of the thorax and lung. Am J Physiol, 1946,146:170.

Fig. 49.15 Adapted from Sherman TF: A simple analog of lung mechanics. Am J Physiol, 1993, 265:S33.

Fig. 52.3 Redrawn from Sircar SS: The oxygen-carrying flask: A representational transform of oxygen dissociation curve of hemoglobin. Am J Physiol, 1994, 266:S34.

Figs. 54.1, 58.3, 80.8, 80.9, 80.10, 82.5, 84.2, 92.5 Adapted from Ganong WF, Review of Medical Physiology. Boston. McGraw Hill; 2005.

Fig. 55.2 *Courtesy* Lt. Col. (Dr.) DJ Chakrabarty, Army Medical Corps.

Fig. 77.2 Redrawn from WHO/FCH/CAH/ Integrated management of childhood illness.

Sections

Section 1	General Physiology
Section 2	Nerve and Muscle
Section 3	Blood and Immune System
Section 4	Cardiovascular System
Section 5	Respiratory System
Section 6	Renal System
Section 7	Gastrointestinal System
Section 8	Nutrition and Metabolism
Section 9	Endocrine System
Section 10	Reproductive System
Section 11	Central Nervous System
Section 12	Special Senses

Table of Contents

Preface	vii	Section 3 Blood and Immune System	
Section 1 General Physiology			
1. Principles of Physics in Physiology	1	21. Body Fluids and Blood	132
2. Principles of Physical Chemistry in Physiology	18	22. Red Blood Cells	136
3. Principles of Control Systems in Physiology	24	23. Hemoglobin	142
4. The Cell	28	24. Hematinic Factors	151
5. Cell Membrane	35	25. Blood Grouping and Transfusion	158
6. Cell Cycle	44	26. Blood Platelets and Hemostasis	163
7. Applied Genetics	49	27. Hemostatic Balance	170
Section 2 Nerve and Muscle		28. Granulocytes	176
8. Functional Anatomy of Nerve and Muscle	54	29. Agranulocytes and Lymphoid Organs	182
9. Degeneration and Regeneration of Nerve and Muscle	63	30. Immunity, Tolerance, and Hypersensitivity	187
10. Resting Membrane Potential	69	31. Immune Mechanisms	193
11. Membrane Excitation and Action Potential	76	32. Hemopoiesis	200
12. Electrophysiology of Ion Channels	81	Section 4 Cardiovascular System	
13. Conduction of Nerve Impulses	86	33. Cardiac Excitation and the Electrocardiogram	205
14. Neuromuscular Transmission	93	34. Abnormalities of Cardiac Excitation	217
15. Mechanism of Striated Muscle Contraction	97	35. Cardiac Cycle	224
16. Characteristics of Skeletal Muscle Contraction	101	36. Cardiac Output	231
17. Electromyography and Electroneurography	111	37. Circulatory Pathway	245
18. Muscle Mechanics	117	38. Hemodynamics	252
19. Smooth Muscles	122	39. Capillary Circulation and Lymphatic Circulation	262
20. Cardiac Muscle	127	40. Chemical Control of the Cardiovascular System	268
		41. Neural Control of the Cardiovascular System	273
		42. Blood Pressure Regulation	280
		43. Circulatory Shock	286
		44. Coronary Circulation	291

45. Cerebral Circulation	295	Section 7 Gastrointestinal System	
46. Pulmonary and Pleural Circulation	303	70. Events in the Mouth and Esophagus	437
47. Cutaneous, Muscle, and Splanchnic Circulation	307	71. Events in the Stomach	442
Section 5 Respiratory System		72. Events in the Duodenum	448
48. Functional Anatomy of the Respiratory System	313	73. Events in the Small Intestine	454
49. Mechanics of Pulmonary Ventilation	318	74. Events in the Colon	458
50. Measurement of Pulmonary Ventilation	332	75. Gastrointestinal Hormones	462
51. Alveolar Ventilation, Perfusion, and Gas Exchange	340	76. Gastrointestinal Disorders	466
52. Transport of Gases	347	Section 8 Nutrition and Metabolism	
53. Control of Respiratory Rhythm	352	77. Dietary Nutrients and Fibers	473
54. Control of Pulmonary Ventilation	358	78. Nutritional Assessment and Dietary Planning	481
55. High- and Low-Pressure Breathing	364	79. Metabolic Pathways	485
56. Pulmonary Function Tests and Respiratory Disorders	368	80. Metabolic States and the Liver	494
Section 6 Renal System		Section 9 Endocrine System	
57. Functional Anatomy of the Kidney	376	81. Mechanism of Hormonal Action	504
58. Glomerular Filtration and Tubular Reabsorption	382	82. Hypothalamic and Pituitary Hormones	510
59. Renal Handling of Sodium	387	83. Thyroid Hormones	517
60. Renal Regulation of Urine Volume and Osmolarity	392	84. Calcitropic Hormones	525
61. Body Fluid and Electrolyte Balance	396	85. Adrenocortical Hormones	534
62. Renal Regulation of Acid–Base Balance	402	86. Adrenomedullary Hormones	544
63. Renal Regulation of Potassium Balance	408	87. Pancreatic Hormones	547
64. Renal Handling of Miscellaneous Substances	411	Section 10 Reproductive System	
65. Hormones Acting on the Kidney	414	88. Testicular and Ovarian Hormones	554
66. Quantitation of Renal Functions	418	89. Puberty and Gametogenesis	560
67. Urine Analysis and Renal Function Tests	423	90. Menstrual Cycle	566
68. Renal Syndromes	427	91. Sperm Transport and Fertilization	570
69. Urinary Bladder and Micturition	431	92. Sexual Differentiation of the Fetus	580
		93. Pregnancy	585
		94. Parturition and Lactation	591
		Section 11 Central Nervous System	
		95. Anatomy of the Central Nervous System	595

96. Spinal Cord and Brain Stem	606	110. Regulation of Food Intake	727
97. Cerebellum	615	111. Memory and Learning	731
98. Diencephalon	622	112. Language and Speech	740
99. Basal Ganglia	626		
100. Cerebral Cortex	632	Section 12 Special Senses	
101. Autonomic Nervous System	642	113. Functional Anatomy of the Eye	744
102. Synaptic Mechanisms and Neurotransmitters	648	114. Visual Optics	749
103. Sensory Mechanisms	660	115. The Retina and Phototransduction	760
104. Regulation of Muscle Length and Tone	677	116. Visual Pathways and Image Processing	766
105. Motor Planning, Programming, and Execution	685	117. Oculomotor Mechanisms	776
106. Thought, Emotion, and Conation	699	118. Auditory Mechanisms	783
107. Electroencephalogram and Epilepsies	706	119. Vestibular Mechanisms	794
108. Sleep	715	120. Olfactory and Gustatory Mechanisms	799
109. Regulation of Body Temperature	720	Index	803

1

Principles of Physics in Physiology

Several areas of physiology cannot be appreciated without an elementary knowledge of physics. If only to emphasize the point, selected principles of physics that are important in physiology are outlined below. Also discussed are mathematical principles that are inseparable from physics.

Mathematics

■ Vectors

Vectors are quantities that have a **magnitude** as well as a **direction**. A vector is represented with an arrow that indicates its direction. The length of the arrow gives a measure of the magnitude of the vector. A vector can be resolved into **components**. Conversely, two or more vectors can be combined into a **resultant** vector. Physiological applications of vectors are found in biomechanics, electrocardiography, motor control, and vestibular mechanisms, among others. The basic concepts of vector resolution are explained with the following examples.

Suppose a person travels 150 km from point O to point P (**Fig. 1.1A**). In doing so, he travels 138 km to the south, 60 km to the east, and 140 km to the southeast. The distance he travels in a given direction can be calculated simply by dropping a perpendicular from point P onto the axis representing that direction. This basic method can be applied to any vector quantity. For example, if the person travels from point O toward point P with a velocity of 150 km/h, then he moves south with a velocity of 138 km/h, toward east with a velocity of 60 km/h, and toward southeast with a velocity of 140 km/h. Conversely, if it is known at what velocity he is moving toward south and east, the actual velocity and direction of his movement can be calculated graphically.

A vector does not have any component perpendicular to it. This is self-evident; a person traveling due south does not move toward east or west. The calculation of the **mean electrical axis of the heart** from the recorded voltages in the electrocardiogram (ECG) leads (see **Fig. 34.11**) is based on the same principle. It explains why an ECG wave that has the highest amplitude in lead I would have the lowest amplitude in lead augmented vector foot (aVF) which is perpendicular to lead I. Vector principles also explain why the **effect of gravity on**

blood circulation gets nullified when the body is supine. In the supine position, the direction of gravity is perpendicular to the direction of blood flow in most of the large vessels.

A special case of the vector principle is illustrated in **Fig. 1.1B**. Suppose a person is standing at point O situated at the center of an equilateral triangle formed by points A, B, and C. When the person moves toward point P with a velocity of (say) 70 km/h, he is moving toward point A at 51 km/h, toward point B at 16 km/h, and toward point C at -67 km/h. (The minus sign denotes that he is actually moving away from point C.) On adding the three velocities, we get zero. This principle forms the basis of the zero potential obtained by interconnecting the vertices of **Einthoven triangle of electrocardiography**.

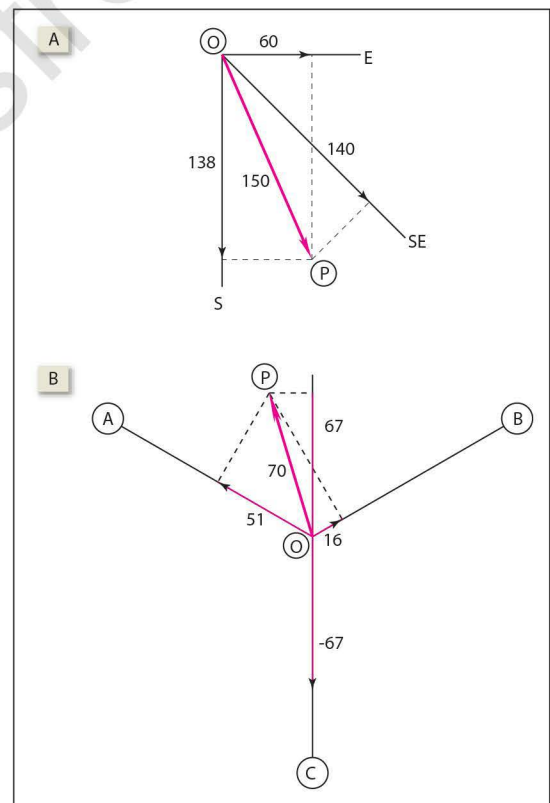


Fig. 1.1 Vector principles. [A] Resolution of a vector into its components. [B] Vector principles underlying Einthoven hypothesis of electrocardiography.

Calculus

There are two broad branches of calculus: differential and integral. Both find applications in several areas of physiology.

Differential calculus makes it possible to calculate the slope of a curve at any point. The slope of a line segment is expressed mathematically as $\delta y/\delta x$, where δy is the small vertical distance through which a point on the segment moves up as it shifts a small distance δx to the right. In the case of a straight line, this ratio remains the same anywhere along the line but in the case of a curved line, this ratio is different for every small segment of the line (Fig. 1.2A). Clearly, a precise description of the slopes of various parts of a curve requires that we break up the curve into infinitesimally small segments. The slope of each segment, which is small enough to be called a *point*, is denoted by the expression dy/dx .

Integral calculus makes it possible to calculate the area under the curve in a Cartesian plane. The area under the curve can be calculated by breaking up the area into rectangular blocks. However, if the curve is irregular, the blocks are not perfect rectangles and the result obtained is inaccurate. Greater accuracy is possible only if the width of blocks is infinitesimally small (dx) so that each block becomes a perfect rectangle. The area under the curve (A) is then given by:

$$A = \int y \, dx$$

Or stated simply, A is the sum of areas of an infinite number of rectangles with an infinitesimally small width (dx) so that the area of each rectangle is given by $y \times dx$. It should be readily apparent from Fig. 1.2B that $\int y \, dx$ is *not the same* as $\int x \, dy$.

The graphic calculation of the work done by the lung and heart, the pressure-time index of the heart and mean blood pressure, and the amount of dye flowing in the indicator dilution method of blood flow estimation are based on the principles of integral calculus.

Logarithms

The logarithm of a number is the exponent to which 10 must be raised so as to equal the number. Expressed mathematically, if $10^a = N$, then $\log N = a$. For example, $\log 1,000 = 3$ and $\log 0.001 = -3$. It may be noted here that 1,000 and 0.001 are reciprocal numbers. In other words, if $\log N = a$, then $\log 1/N = -a$. Logarithms appear in the Nernst equation (p 19). In accordance with the logarithmic property

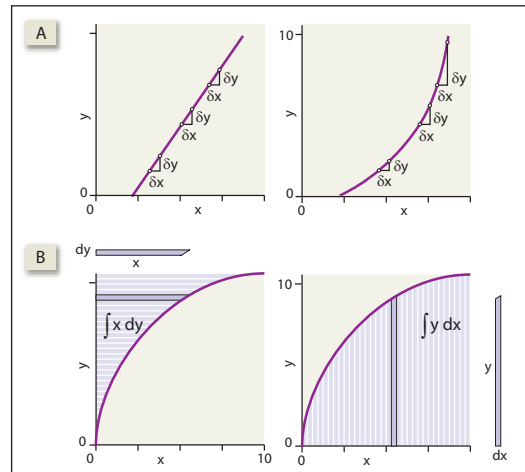


Fig. 1.2 [A] (Left) The slope ($\delta y/\delta x$) is the same in all segments of a straight line. (Right) For the same δx , the δy varies in different segments of the curve. Hence the slope ($\delta y/\delta x$) is different in all segments. [B] Areas defined by the terms $\int x \, dy$ (left) and $\int y \, dx$ (right).

just explained, the Nernst equation can be written in either way as follows:

$$E_m = \frac{61}{z} \log \frac{[C_o]}{[C_i]}$$

or

$$E_m = \frac{-61}{z} \log \frac{[C_i]}{[C_o]}$$

An important application of logarithms is the logarithmic scale which is a convenient scale for describing certain types of data. Consider the sound pressure from various sources (Table 1.1). The inadequacy of a linear scale in representing these data graphically is obvious from Fig. 1.3. On the other hand, if the data are converted into their logarithms and then plotted, the advantage is instantly apparent. The logarithmic scale of sound intensity is called the decibel scale. Similarly, the logarithmic scale of acidity is the pH scale.

Table 1.1 Sound pressures at auditory threshold and in different situations

	Sound pressure in dyne/cm ²	Sound pressure as a multiple of threshold	Log of the multiple
Threshold	0.002	1	0
Conversation	0.02	10	1
Factory noise	2	1,000	3
Discomfort	200	100,000	5

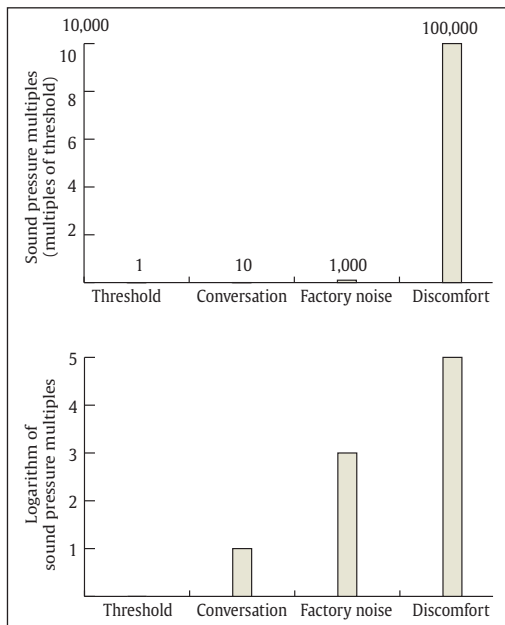


Fig. 1.3 Sound pressures expressed as multiples of hearing threshold are depicted graphically on a linear scale (*above*) and a logarithmic scale (*below*).

Mechanics

■ Laws of motion

The **first law of motion** states that a body continues to be stationary or to move in a straight line with uniform velocity until it is acted upon by an external force. The law helps in defining force itself (see below). The law is also known as the **law of inertia**. Because of inertia, a stationary body cannot start moving on its own and a moving body cannot stop on its own. The concept of inertia helps in understanding the effect of *coup and counter-coup injuries* of the brain and the working of the *semicircular canals* of the vestibular apparatus.

The **second law of motion** tells us that the *acceleration or retardation* of a body is *directly proportional to the force applied* on it and *inversely proportional to the mass* of the body. The product of the mass and acceleration of a body gives the force acting on it.

The **third law of motion** states that every action has an equal and opposite reaction. The third law has been put to use in *ballistocardiography*. It also helps in understanding *muscle action* (see **Fig. 49.3**).

■ Force

Force is *anything that changes or tends to change the uniform motion of a body in straight line*. The SI

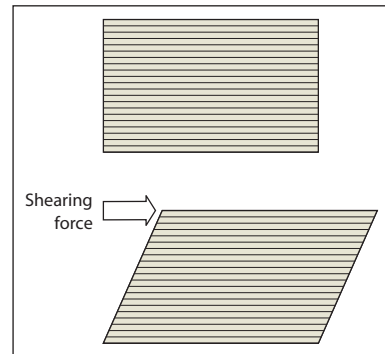


Fig. 1.4 Shearing force.

unit of force is the Newton (N). A force of 1 N, when applied to a 1 kg mass placed on a horizontal frictionless surface, produces an acceleration of 1 m/s^2 .

A **shearing force** is a force that is directed tangentially to the surface of a body (**Fig. 1.4**). Under the effect of a shearing force, the constituent laminae of a body move through different distances. An understanding of the nature of shearing force is important in the context of *otolith organs*. Shearing force is also *exerted by blood flow on the capillary endothelium*.

Gravitational force, as applicable on earth, is a special kind of force that has two remarkable features: (1) it is *always directed toward the center of the earth* and, thereby, defines the vertical. (2) It is *directly proportional to the mass of the body*. The acceleration produced by gravitational force is called **acceleration due to gravity**. A heavier body is pulled with a greater gravitational force (in accordance with Newton law of gravitation). However, for a given force, a heavier body has less acceleration (in accordance with Newton second law of motion). Therefore the acceleration due to gravity does not vary with the mass of a body and remains absolutely constant at 9.8 m/s^2 . This acceleration due to gravity, denoted by g , gives us *the feel of weight*.

G forces are *accelerative forces other than gravity* that may act on the body in specific situations. The accelerative force is called G_x when it is directed *anteroposteriorly*, G_y when it is directed *laterally*, and G_z when it is directed *superoinferiorly* in reference to the body. Together, the G_x , G_y , and G_z forces are called **G forces**.¹ The organs for sensing G forces are the *otolith organs*.

The weight felt by a subject under the effect of G_z is called the **apparent g**. Changes in G_z have important effects on cardiac output, blood pressure, and

¹G conventionally denotes the universal gravitational constant. In certain special contexts, as in aerospace medicine, G denotes a variable whose value is expressed in multiples of g , for example, 2, -1, or 0 g . The term G-force is a misnomer because a change in G-force is almost always due to acceleratory forces other than gravitational force.

pulmonary ventilation that are described on p 257. **High and low Gz** are experienced when there is acceleration or deceleration along the long axis of the body. Very high Gz occurs *when a space rocket takes off*. More commonly, such changes in Gz are experienced during *looping of airplanes* and during *parachute jumping* when the parachute is suddenly opened out after a period of free fall: It is called the **opening shock** load. A slight feel of high and low Gz is experienced in an elevator when it starts or stops.

Zero Gz is perceived by astronauts in orbiting satellites where the gravitational force is precisely counterbalanced by the centrifugal force generated by the orbiting satellite. On earth, zero Gz is experienced during free fall under the effect of gravity. Stated mathematically

$$Gz = g - a$$

where

Gz = apparent acceleration due to gravity

a = actual acceleration in the direction of gravity

g = 9.8 m/s²

When a = g, Gz becomes 0. In other words, *a body falling under the effect of gravity is weightless* (see Fig. 38.15).

■ Pressure

Pressure is the *force exerted per unit area*. The pressure of a column of fluid is called **hydrostatic pressure**. It depends only on its *vertical height*; neither the width of the column nor its inclination makes any difference to the pressure exerted (Fig. 1.5A). This concept is of importance in understanding the measurement of **jugular venous pressure** and the **direct manometry** of blood pressure. The dependence of fluid pressure on the height of the fluid column also explains why the *venous pressure is higher in the dependent parts of the body* and why the *atmospheric pressure decreases at high altitudes*. Its knowledge helps in estimating the *pressure at different depths* of the sea, which is important in deep-sea diving.

The *SI unit of pressure* is N/m², which is also called **Pascal** (Pa). The atmospheric pressure at sea level is taken as 101.29 kilopascals (kPa). This pressure is also known as **1 atmosphere absolute (ATA)**. The value of 1 atmospheric pressure, when rounded off to 100 kPa, is called **1 bar**. However, fluid pressure is often expressed simply in terms of the *vertical height of a fluid column* because the pressure of a fluid column is given by **hρg** where "h" is the height of the fluid column and "ρ" is the density of the fluid. The fluid of reference is usually

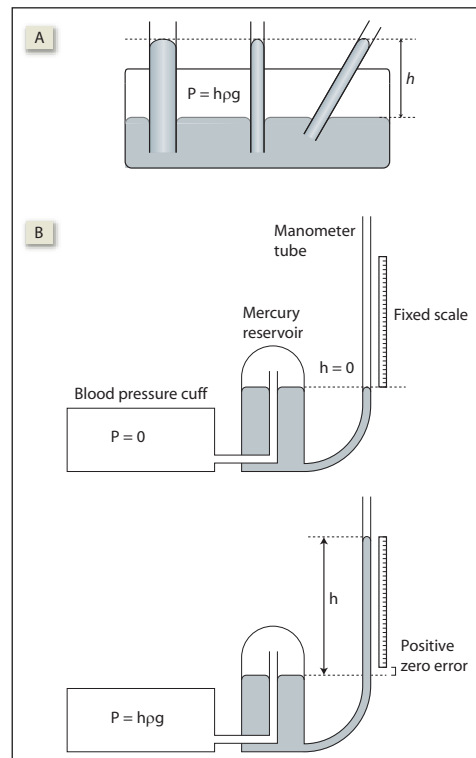


Fig. 1.5 [A] Hydrostatic pressure depends only on the vertical height of a fluid column. It remains unaffected by the diameter and inclination of the tube. [B] A mercury manometer connected to a rubber cuff is used in sphygmomanometry. (Above) The cuff pressure is zero and the mercury levels in the reservoir and manometer tube are equal. (Below) As the cuff pressure increases, the mercury level in the tube rises while the mercury level in the reservoir falls. The difference *h* in the mercury levels gives the cuff pressure. Due to the fall in mercury level in the reservoir, the fixed scale gives a false-high reading due to positive zero error. The error can be reduced somewhat by scaling down the calibration. It can be verified that in a sphygmomanometer, the 1 cm calibration is actually a little less than 1 cm.

mercury but can be water, saline, or even blood, whichever is convenient. The pressure exerted by **1 mm of mercury column** is called **1 Torr**. Nearly all fluid pressures inside the body are a few millimeters of mercury above or below the atmospheric pressure, which makes the Torr a convenient unit for expressing physiological pressures. Since mercury is 13.6 times denser than water, **1 Torr equals 13.6 mm of water pressure**.

Even atmospheric pressure is commonly expressed in terms of the height of the mercury column that would counterbalance it. The height can

be calculated, given that the density of mercury is $13.6 \times 10^3 \text{ kg/m}^3$:

$$\begin{aligned} \text{Atmospheric pressure} &= h \times \rho \times g \\ 101.29 \times 10^3 &= h \times (13.6 \times 10^3) \times 9.8 \\ \therefore & \quad \quad \quad h = 0.76 \text{ m} \\ & \quad \quad \quad = 760 \text{ mm} \end{aligned}$$

Hence, atmospheric pressure is also expressed as 760 mm of mercury, or 760 Torr.

The *mercury manometer* (Fig. 1.5B) is commonly used for measuring pressure. One limb of the manometer is connected to the system whose pressure is to be measured. The other limb of the manometer is left open to the atmosphere. The difference in the mercury column in the two limbs indicates the pressure of the system in excess of the atmospheric pressure.

Levers

A lever is a rigid bar that is acted on by forces that tend to rotate the bar about its pivot or fulcrum. Levers are of three types depending on the relationship between the fulcrum, load (weight), and effort (force) applied. If the fulcrum (F) is central, it is a *class-I lever*. If the load (L) is central, it is a *class-II lever*. If the effort (E) is central, it is called a *class-III lever*.

The *mechanical advantage* of a lever is the ratio of the load force to the effort force. The ratio is proportional to the ratio of effort arm to load arm (Fig. 1.6). The greater the mechanical advantage, the less will be the force required to move a load.

$$\text{Mechanical advantage} = \frac{\text{Load}}{\text{Effort}}$$

The *velocity ratio* is the ratio of the distance moved by the effort to that moved by the load. The greater the velocity ratio, the greater is the distance through which an effort has to move for lifting a load.

$$\text{Velocity ratio} = \frac{\text{Distance moved by effort}}{\text{Distance moved by load}}$$

In a frictionless lever, *mechanical advantage equals velocity ratio*. Therefore, in all three classes, what is gained in excursion of the load is lost in the effective force acting on the load and vice versa. *Class-II levers* always produce force gains, and *class-III levers* always produce *excursion gains*. Class-I levers can produce either of the two. The principles of levers are essential to the understanding of the *biomechanics* of the musculoskeletal system (Chapter 18).

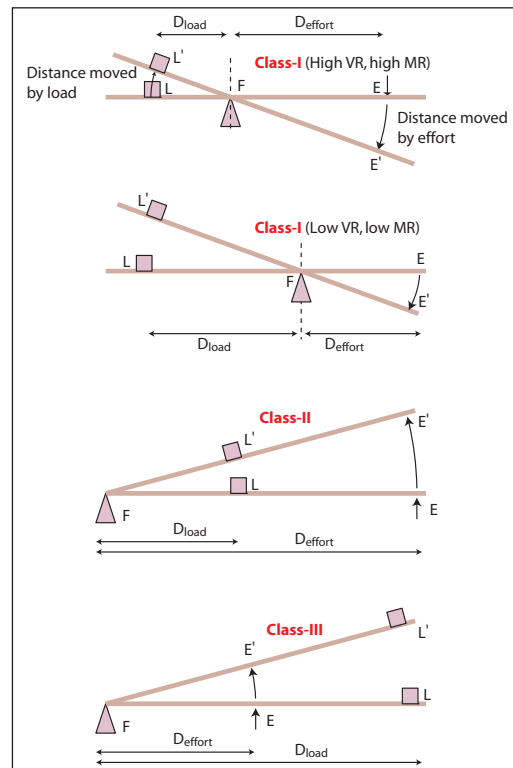


Fig. 1.6 The physics of a lever (explanation in text).

Work and energy

Work is done when the point of application of a force (F) moves through a certain distance (D). Stated mathematically,

$$\text{Work done (W)} = F \times D$$

The SI unit of work is the *Joule*. One joule of work is done when a force of 1 N moves a body through 1 m.

Energy is the capacity for doing work and the unit of energy is the same as that of work, that is, *Joule*. The energy associated with motion is called *kinetic energy (KE)* and is given by the formula:

$$KE = \frac{1}{2}mv^2$$

where *m* is the mass of the body and *v* is its velocity. The amount of kinetic energy that a body can gain by falling under the effect of gravity gives a measure of its *potential energy (PE)* and is given by the formula:

$$PE = mgh$$

where *m* is the mass of the body, *g* is acceleration due to gravity, and *h* is the height through which the body can fall.

When work is done *by a body* by expending its own energy, the work done is said to be **positive**. For example, a body that falls through a certain height expends its potential energy and therefore the work done *by the body* is positive. Conversely, a body that is lifted against gravity gains in potential energy. Therefore, work is done *on the body*, and the work done is **negative**.

When a fluid is compressed by application of pressure P , the fluid gains in energy, which is at least partly in the form of heat. The work done on the fluid is therefore negative and is given by:

$$\text{Work done (W)} = -\int P dV$$

where dV is the small decrement in volume through which it has been compressed. Conversely, for expanding against an incumbent pressure, the fluid has to expend energy and therefore it loses energy. The work done *by the fluid* is therefore positive and is given by:

$$\text{Work done (W)} = +\int P dV$$

These concepts help us understand why the work done by the lungs or the heart is given by the area enclosed in its pressure–volume loop.

In the case of the lungs, *the work done on the lungs is negative during inspiration* because the

lungs themselves do not expend any energy; rather, they are made to inflate by the inspiratory muscles. During inspiration, the lungs gain in elastic recoil energy. This recoil energy stored in the lungs is expended during expiration as the lungs deflate. Therefore, *the work done by the lungs during expiration is positive*. The work done on and by the lungs during inspiration and expiration are shown graphically in **Fig. 1.7A**.

When the work done during both inspiration and expiration are added, there is a *net small amount of negative work done* that is represented by the area enclosed within the **pressure–volume loop** (called the **hysteresis loop**). The negative work signifies that at the end of one breathing cycle, work has been done *on the lungs* by the respiratory muscles which have expended their energy. The lungs have not spent energy of their own; rather, they have gained some energy. The *energy gained* is the heat energy that has been generated by the frictional (viscous) forces inside the lungs.

In the same way, the area inside the **ventricular pressure–volume loop** of the cardiac cycle gives the positive work done by the ventricle in overcoming viscous resistance of blood flow (**Fig. 1.7B**).

■ Buoyancy

When a body is immersed in a fluid, *the fluid displaced pushes the body up so that the body loses weight*. The loss in weight is called **buoyancy** and it

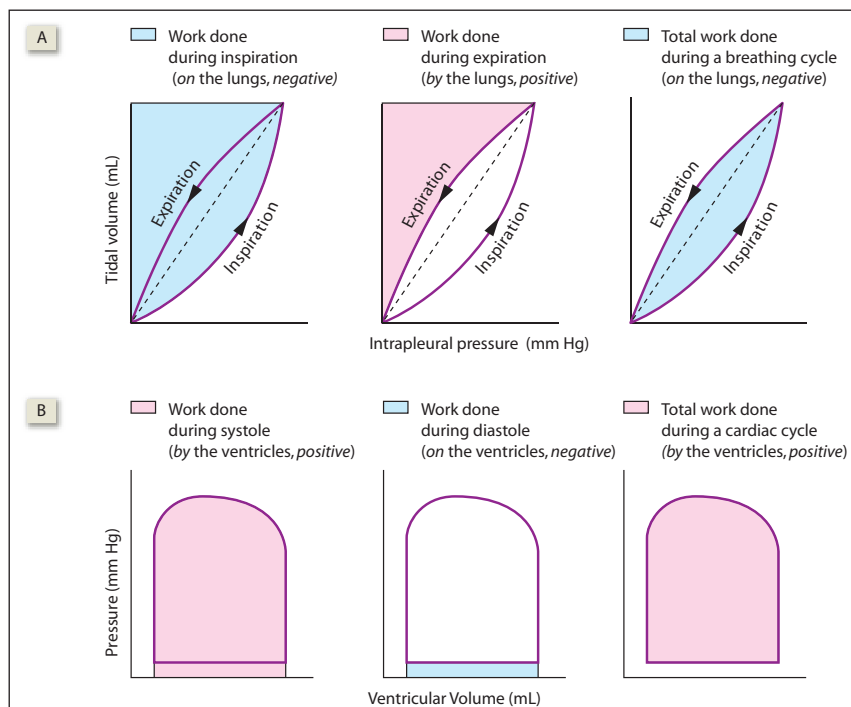


Fig. 1.7 [A] Work done on the lung during breathing. [B] Work done by the ventricle during a cardiac cycle.

is equal to the weight of the fluid displaced. Providing buoyancy to the brain is an important function of the *cerebrospinal fluid*.

■ Surface tension

Surface tension is a property of liquids due to which a *liquid surface behaves like a stretched membrane*. Because of surface tension, a drop of fluid *minimizes its surface area*. It therefore *assumes a spherical shape* because the surface to volume ratio is minimum for a sphere. The fact that a bubble does not collapse *indicates* that the air inside it provides a distending pressure that exactly balances the collapsing pressure of the fluid shell.

The **law of Laplace** states that tension (T) in the wall of a cylinder is equal to the product of the transmural pressure (P) and the radius (R). The formula has several variations (**Table 1.2**) depending on the geometry of the surface (spherical or cylindrical) and its composition (liquid drop, liquid bubble, or air bubble) (**Fig. 1.8**). Applications of Laplace law in physiology assume $P = 2T/R$ as an approximation, which is the formula for a spherical air bubble in air.

The law of Laplace helps explain several important physiological phenomena. (1) It explains how the *thin-walled capillaries* are able to withstand an internal pressure as high as 25 mm Hg, which is the normal capillary hydrostatic pressure. This is possible because capillaries have a very small radius R.

Table 1.2 Formulas defining Laplace law

	Spherical	Cylindrical
Liquid drop in air	$2T/R$	T/R
Air bubble in liquid	$4T/R$	$2T/R$
Air bubble in air	$2T/R$	T/R

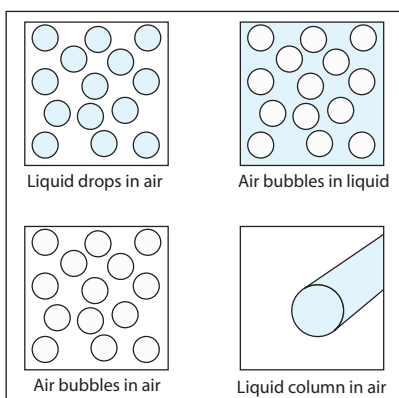


Fig. 1.8 The Laplace law is applicable to any of the above (see **Table 1.2**).

Even though the wall tension is low, the small value of R in the denominator makes a very high P possible. (2) The law of Laplace *puts the dilated heart at a disadvantage*. When the radius R increases, the wall tension T must go up proportionately if the ventricular pressure P is to be maintained. (3) In accordance with the Laplace law, the lower the *functional residual capacity* of the lungs, the more difficult it is to inflate it. (4) Because of Laplace law, the greater the gastric filling, the lower is the pressure that causes *gastric emptying*. This is obviously beneficial to the process of digestion. (5) Laplace law explains why a reduction in detrusor muscle tension prevents a rise of *intravesical pressure* even as the urinary bladder fills up to greater volume.

■ Viscosity

Viscosity is *fluid friction*. When fluid moves along a tube, it does so in the form of multiple layers or laminae that slip on one another, moving at different velocities due to friction between the layers. The lamina at the middle of the blood vessel moves fastest while the one closest to the vessel wall does not move at all. Thus, there is a velocity gradient of laminae, which is called the *shear rate*. A fluid has **1 poise** of viscosity if there is a frictional force of 1 dyne/cm² between its layers when flowing at a shear rate of 1 cm/s. A **Newtonian fluid** is one in which the *viscosity is independent of the shear rate*. Blood is not a Newtonian fluid.

Just as friction affects the velocity of a body, viscosity affects the velocity of fluid flow. In a long narrow tube of uniform radius, the relation of the flow rate (Q) with the pressure gradient ($P_A - P_B$), that is, the pressure difference at the two ends of the tube, fluid viscosity (η), tube radius (R), and tube length (L) is given by the **Poiseuille-Hagen formula**:

$$Q = (P_A - P_B) \times \frac{\pi}{8} \times \frac{1}{\eta} \times \frac{R^4}{L}$$

This law is important in the understanding of hemodynamics. Since resistance varies inversely with the fourth power of the radius, even *small changes in the vessel diameter cause large variations in blood flow through it*, enabling effective regulation of blood flow through vascular beds. The physiological relevance of the Poiseuille-Hagen formula is best brought out by the differences in the blood flow in the renal cortex and the renal medulla (see p 377), where blood viscosity, capillary length, and capillary diameter all have a role.

At high velocities, the flow of fluid becomes turbulent and does not remain streamlined. The

probability of turbulence is related to the diameter of the vessel and the viscosity of the blood. This probability is expressed by the **Reynolds number**:

$$Re = \frac{\rho DV}{\eta}$$

where Re is the Reynolds number, ρ is the density of the fluid, D is the diameter of the tube (in cm), V is the velocity of the flow (in cm/s), and η is the viscosity of the fluid (in poise). The higher the value of Re, the greater the probability of turbulence. When $Re < 2,000$, flow is usually not turbulent whereas if $Re > 3,000$, turbulence is almost always present. Turbulence of blood is responsible for *cardiac murmurs* and the *Korotkov sounds*.

The Poiseuille-Hagen formula can be rewritten as:

$$(P_A - P_B) = Q \times \frac{8}{\pi} \times \eta \times \frac{L}{R^4}$$

Thus when the flow rate, fluid viscosity, and tube radius remain constant, the pressure drop ($P_A - P_B$) along a tube is directly proportional to the length of the tube and inversely proportional to the tube radius, as explained in **Fig. 1.9**.

■ Bernoulli principle

When a constant amount (Q) of fluid flows through a tube, the total fluid energy, that is, the sum of its **kinetic energy**, **pressure energy**, and **potential energy**, remains constant (**Fig. 1.9**). This is known as **Bernoulli principle** and it helps in understanding several important hemodynamic principles. It explains, for example, why the fluid pressure is low in a blood vessel at places where its radius is less² (**Fig. 1.9**). It also explains the suction effect of venous blood flow on the thoracic duct terminating in the vein (see **Fig. 39.7**).

■ Heat

Heat always flows from a higher temperature to a lower temperature. When the temperatures of two bodies are equal, no heat transfer occurs between them. (1) In **conduction**, transfer of heat occurs through a medium whose molecules cannot move about freely. Thus, when we hold an iron rod in fire, the iron atoms vibrate and transfer the heat from the fire to the hand through the rod. However, the

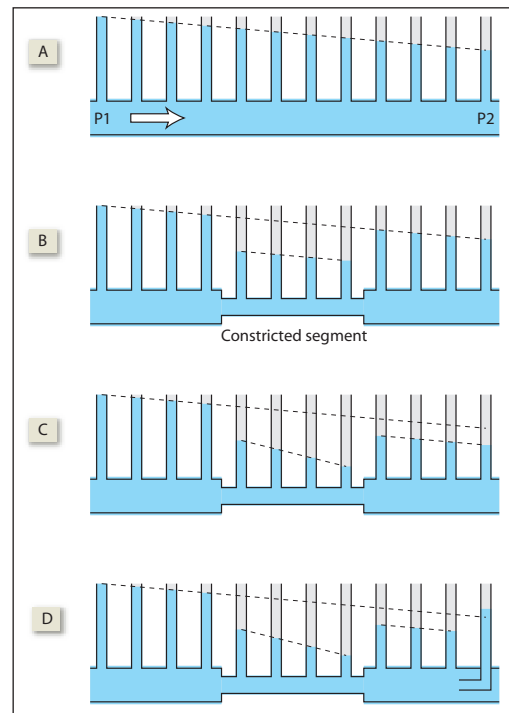


Fig. 1.9 A horizontal tube is fitted with uniformly spaced manometers for measuring the pressure drop. [A] The pressure falls uniformly with distance in accordance with the Poiseuille-Hagen formula. [B] The pressure falls in the constricted segment in accordance with Bernoulli principle. Beyond the constricted segment, the pressure rises again. The depiction is not entirely correct: the correct pressures are depicted immediately below. [C] The pressure in the constricted segment is low in accordance with Bernoulli principle and the pressure drop along the segment is steeper in accordance with the Poiseuille-Hagen formula. (D) The fluid rises higher in the last manometer because it captures a part of the fluid's kinetic energy and converts it into pressure energy in accordance with Bernoulli principle.

iron atoms do not move from their fixed places. (2) In **convection**, the molecules of the medium transfer heat by actually moving about, carrying the heat with them. **Evaporative cooling** is an example: The water molecules vaporize and move away, carrying the heat with them. The **caloric stimulation test** for testing vestibular functions is based on the convective currents set up inside the vestibular apparatus. (3) In **radiation**, no medium is required. It is through radiation that heat from the sun travels through space and reaches us.

²The fact that both Poiseuille-Hagen formula and Bernoulli principle give the effect of tube diameter on fluid pressure need not be cause for confusion. The two formulas deal with different aspects of fluid energy. The Poiseuille-Hagen formula deals with the conversion of fluid pressure into heat and the resultant loss of fluid pressure as the fluid flows through a long tube of uniform diameter. It does not deal with the other forms of fluid energy like kinetic energy or potential energy. On the other hand, Bernoulli principle deals with the interconversion of the three different forms of fluid energy and not with its conversion into heat energy.

Heat transfer occurs through all possible modes. If a medium is present and the molecules are free to move about, heat will travel through convection. If the molecules cannot move about, heat will travel through conduction. Regardless of whether a medium is present or not, some heat will always travel through radiation.

A knowledge of the modes of heat transfer is important to the understanding of *thermoregulation* in the body and explains our behavioral responses to changes in environmental temperature. For example, it is through convection that the circulating blood maintains a fairly uniform temperature throughout the body. The body is cooled through the *evaporation* of sweat (convection). We seek shade because it cuts off the *radiation heat* from the sun although we still feel the *convective heat* of hot air or the *conductive heat* of the ground we lie upon. We wear dark, coarse-textured clothes in winter because *dark surfaces absorb radiated heat*. We wear light-colored clothes in summer because they *light surfaces reflect back heat radiation*. Thick woolen clothes that trap air in them reduce conductive and convective losses of body heat in winter.

■ Light

Light bends as it passes from air into a denser medium or when it emerges out of the medium into air. This bending is known as refraction and the extent to which it bends while passing in and out of the medium is given by the *refractive index* (μ) of the medium. The refractive index of water is 1.33. The refractive indices of the various compartments of the eye are shown in **Fig. 114.1**.

Refraction of light underlies the formation of an image by the lens. A knowledge of how lenses work helps in understanding the refraction in the eye. The basic formula for lenses is:

$$\frac{1}{f} = \frac{1}{v} - \frac{1}{u}$$

where u is the distance of the object from the lens, v is the distance of the image from the lens, and f is the focal length of the lens. The distances u , v , and f may be positive or negative, depending on whether the measurements are made along the direction of the light or against it (**Fig. 1.10**). The focus is the point where parallel rays passing through the lens would converge (in the case of a convex lens) or appear to diverge (in the case of a concave lens). The *focal length of a convex lens is positive while the focal length of a concave lens is negative*.

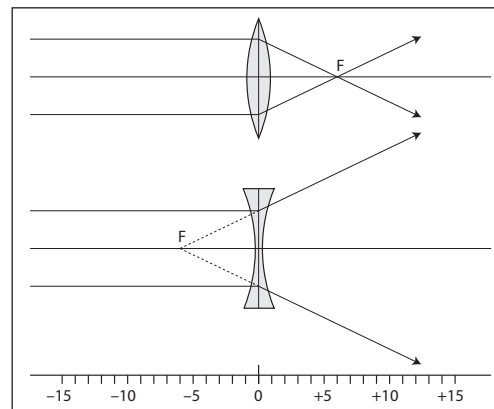


Fig. 1.10 Ray diagrams showing the focal point of a convex lens (above) and a concave lens (below). In the examples illustrated here, the focal length of the convex lens is +6 cm and the focal length of the concave lens is -6 cm.

The **power of a lens** is expressed in *diopter (D)* and is the *reciprocal of its focal length in meters*. Thus the power of a convex lens with focal length of +25 cm is +4 diopter, whereas the power of a concave lens of -50 cm is -2 diopter. The power of the intraocular lens can be increased by contracting some of the intraocular muscles. This phenomenon is called *accommodation*. When two coaxial lenses are placed together, their *dioptric powers are added algebraically*. This forms the basis of *correction of refractory errors by using external lenses*.

The physics of image formation is best understood by integrating it with the physiology of sensory perception, because whatever we see around us is essentially the *sensory projection* of images formed on our retina. Consider a point object (O) and its point image I formed by a convex lens L (**Fig. 1.11**). A point image is formed when all the rays diverging from a point object are made to converge at a single point using a convex lens. In **Fig. 1.11A**, the image I_r (r for real) is formed in front of the eye. After converging at I_r , the rays diverge again and therefore the point image (the point of convergence of rays) behaves like a point object (the point of divergence of rays). Hence the image is called a **real image**. The diverging rays from I_r converge on the retina after passing through the convex lens of the eye. Past sensory experience of the observer's brain tells it that the rays forming the retinal image are originating from point I_r and, therefore, the brain "sees" an image at point I_r . This "calculated guess" made by the brain is known as *sensory projection*. In making the projection, the brain takes into consideration, among other factors, the extent of accommodation required by the intraocular muscles.

In **Fig. 1.11B**, the convex lens L bends the divergent rays from the object but not enough to

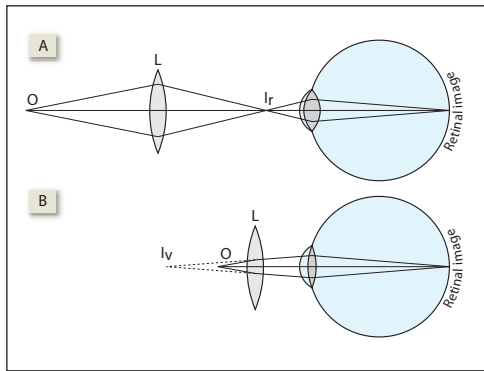


Fig. 1.11 The optics of image formation of point objects.

make them converge in front of the eye. It is the intraocular lens that finally converges the rays on to a point on the retina. As in the previous case, the brain projects the retinal image to the point where the divergent rays incident on the eye seem to be originating from. Hence, the brain “sees” an image at point I_v . Unlike I_r , no rays are actually diverging from I_v ; they only appear to do so due to sensory projection. Hence, I_v is called a **virtual image** (virtual = amounting to). Stated simply, an image is real if it is formed by the convergence of rays. *The image is virtual if it is visible at a site where there is no convergence of rays.* Going by this definition, the image formed on the retina of the eye is a real image.

The images formed when the object is located at different distances from the lens are shown in Fig. 1.12. These ray diagrams can be easily

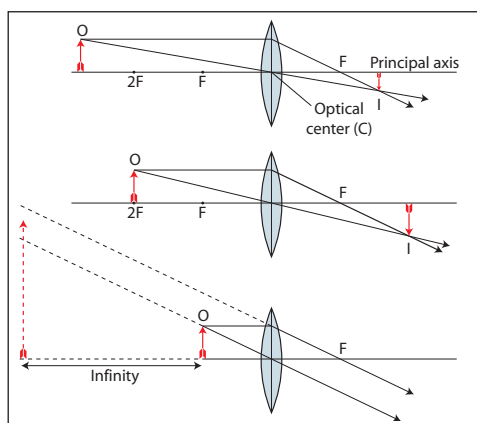


Fig. 1.12 The optics of image formation of objects placed greater than two focal lengths away (above), at twice the focal length (middle), and at focal length (below).

constructed by remembering that (1) rays parallel to the principal axis converge at the focus after passing through the lens and (2) the rays passing through the optical center (3) of the lens do not deviate.

How big does an object appear to the observer? The answer is that the nearer an object is to the eye, the bigger is the retinal image and the bigger it appears to the observer. However, how close we can bring an object to our eyes is limited by the power of our accommodation. *Short-sighted people (myopes) see bigger images* because they can bring the object nearer to the eye and still see it clearly. Another question arises here. Given two objects of different sizes placed at different distances from the eye, which one will appear bigger? The answer lies in the *visual angle* subtended by the two objects: *The object that subtends a greater visual angle looks bigger* (Fig. 1.13). The visual angle forms the basis of the testing of **visual acuity**. The visual angle subtended by an object is given by the formula:

$$\text{Visual angle} = \tan^{-1} \frac{\text{Height of the object}}{\text{Distance of the object from the eye}}$$

Thus, if an 8.75 cm high object³ is kept at a distance of 60 m (i.e., 6,000 cm), the visual angle subtended by it will be:

$$\begin{aligned} &= \tan^{-1} [8.75/6,000] \\ &= \tan^{-1} [0.001458] \\ &= 0.0835 \text{ degree} \\ &= 0.0835 \times 60 \text{ minutes} \\ &= 5.01 \text{ minutes} \end{aligned}$$

The retina is normally situated at the focus of the intraocular lens (see Fig. 114.4) and therefore *when there is no accommodation, only objects at infinity are clearly visible*, for example, the stars in a night sky. Anything nearer appears hazy. This can be verified by putting atropine (a cycloplegic) on the eye, which abolishes the power of accommodation. For the same reason, we see the maximally magnified images without straining our eyes when *the object is kept at the focus of a magnifying glass* and its virtual image is formed at infinity.⁴ Due to sensory projection, parallel rays falling on the eye result in a virtual image that is infinitely large and is formed at infinity. How big does an infinitely large image located at infinity appear to the eye? As already explained, the answer depends on the visual angle subtended at the eye by the virtual image.

³In the Snellen chart, the topmost letter is 8.75 cm high, which is readable by a normal person from a distance of 60 m (see Fig. 114.12).

⁴This point is stressed here because several textbooks of high school physics mention that when an object is kept at the focus of a convex lens, the rays become parallel and therefore no image is formed. That is incorrect.

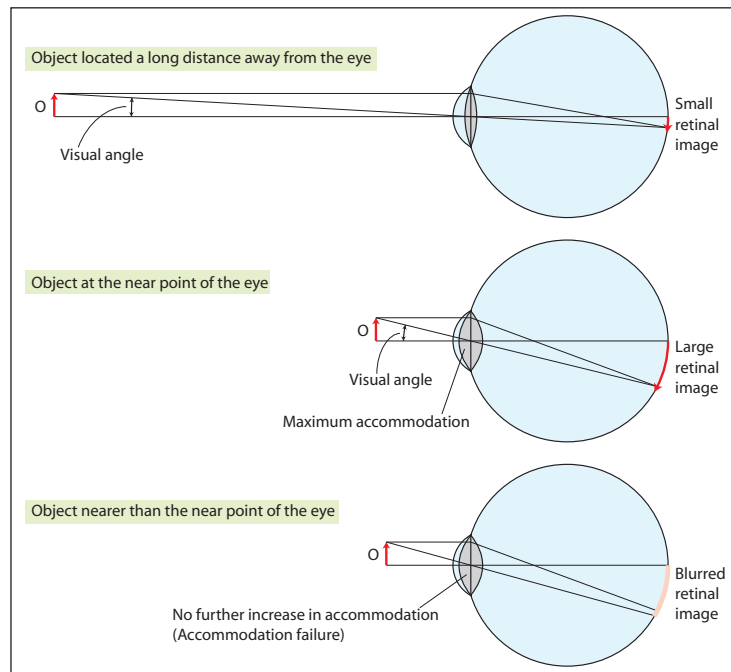


Fig. 1.13 The visual angle increases as the object is brought nearer to the eye. Note that the greater the visual angle, the larger is the retinal image. When the object comes nearer than the near point, the retinal image becomes blurred.

Because the eyes do not have to be strained (accommodated) for seeing a virtual image at infinity, all optical instruments, including microscopes, are so designed that the *image formed is virtual and is situated at infinity*. A commonly observed virtual image at infinity is the *red eye* well known to photographers: The pupil appears red in a photograph that has been shot in dim light using flashlight (**Fig. 1.14**). Because the pupils are dilated in dim light, the *red choroid layer* beneath the retina is illuminated by the flashlight and gets photographed. The illuminated choroid, which is situated at the focus of the intraocular lens, sends out



Fig. 1.14 Red eye (Bruckner) reflex in a photograph shot in the dark using flashlight. The principle of direct ophthalmoscopy is similar.

parallel rays from the pupil. These parallel rays are captured by the camera even from a distance of several meters. Image formation when the object is at focus is also of importance in understanding direct ophthalmoscopy and retinoscopy.

Fig. 1.12 also shows that the nearer an object is to the focal point F , the greater the distance between the image and the lens. In the eye, the distance between the lens and the retina is fixed at 17 mm and the maximum power of the lens that can be attained after full accommodation is 69 diopters which is equivalent to a focal length of 14.5 mm. Substituting these values in the lens formula, we get $u = 10$ mm, which is as close the object can get to our eyes without blurring the retinal image and is called the *near point*.

■ Waves and sound

A wave propagates through a medium when the particles of the medium oscillate rhythmically about their mean position. The wave is called *transverse or longitudinal* depending on whether the plane of oscillation of the particles is perpendicular or parallel to the direction of wave motion. A *transverse wave* has alternate crests and troughs (**Fig. 1.15**). The ocean waves or the ripples in a pool of water or the fluttering of a flag in the wind are examples of transverse waves. Transverse waves

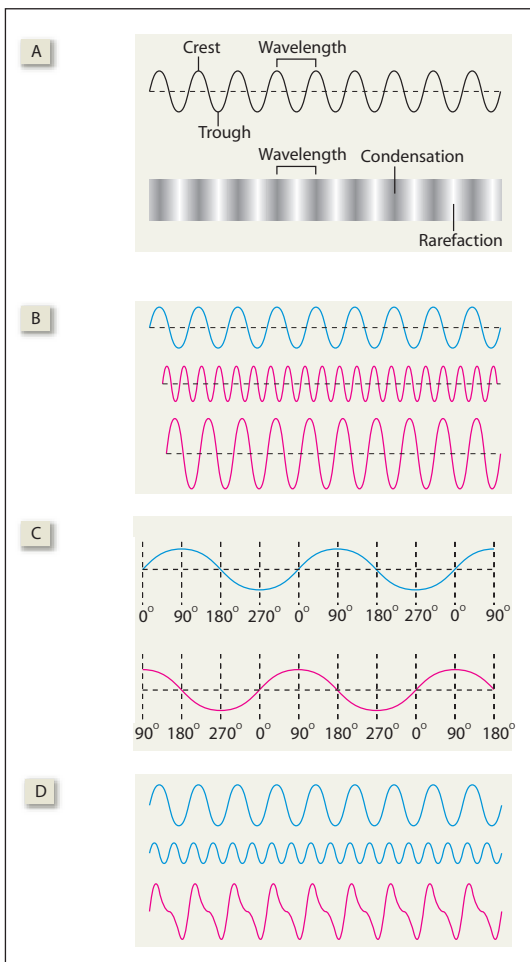


Fig. 1.15 Characteristics of wave motion. [A] Transverse and longitudinal waves. [B] Waves with different frequency and amplitude. [C] Waves with a phase difference. [D] Wave summation. Note that the wave in red color is formed by the summation of the two waves shown in blue, one of which has a lower frequency (fundamental wave) and the other has double the frequency (harmonic wave). The shape of the summated wave is different from the component waves, and gives the wave a characteristic sound quality or timbre.

are also set up on the basilar membrane of the cochlea (see **Fig. 118.6**). A **longitudinal wave** has alternate zones of condensation and rarefaction. Sound waves in air are longitudinal waves.

The **frequency of a wave** is the number of waves produced in 1 second and is expressed in **Hertz (Hz)**. It is inversely related to its wavelength. Two waves with the same frequency can have a **phase difference** or **phase lag** (**Fig. 1.15C**). For example, if at any point of time, one wave is at its crest (i.e., at 90 degrees) and the other is at its trough (i.e., at 270 degrees), the two waves are 180 degrees out

of phase. The term **phase difference** or **phase lag** has been used in this book in the context of fluctuations in blood gases in Cheyne–Stokes breathing and in the ionic basis of cardiac pacemaker potentials. Two waves with different frequencies and amplitudes can summate to produce a **complex wave**. Such complex waves form the basis of the quality or timbre of sound waves (**Fig. 1.15D**). The characteristics of a sound wave, namely, its pitch, loudness, and timbre, are discussed below.

Pitch The pitch (or tone) of the sound is what is perceived by the human ear as higher or lower musical notes: It is directly related to its frequency. The greater the frequency, the higher is the tone. **Pitch discrimination** by the human ear is best in the range of 1,000 to 3,000 Hz.

Intensity and loudness A sound wave is essentially a traveling pressure wave and therefore the amplitude of a sound wave is expressed as the maximum pressure variation in it. The **intensity** of a sound wave is the amount of energy (in joule) transported by the sound wave in unit time (1 s) per unit area (1 m²), across a surface perpendicular to the direction of propagation. The intensity of a sound wave is directly proportional to the square of its pressure amplitude. The pressure amplitude of the faintest sound that can be heard is about 3×10^{-5} Pa and the corresponding intensity is 10^{-12} J/s/m². A more convenient unit for expressing sound intensity is the **decibel (dB)**.

$$\text{Number of dB} = 10 \log \frac{\text{Intensity of the sound}}{\text{Intensity of barely audible sound}}$$

The term **loudness** refers to the listener's subjective perception of the magnitude of a sound sensation. Loudness increases with intensity, and the relation between the two is empirical, which is given by the **Weber–Fechner law** or any of its modifications (see **Fig. 103.17**).

Since the decibel scale is a logarithmic scale, every 10 dB represents a 10-fold increase in sound intensity. The **threshold of hearing** (i.e., the intensity of the faintest audible sound) is **zero decibel**. A sound is 10 times louder than threshold at 10 dB, 100 times louder at 20 dB, a million times louder at 60 dB, and 10 billion times louder at 100 dB. The usual loudness of speech is 65 dB. A 100 dB sound damages the peripheral auditory apparatus while 120 dB causes pain, and 150 dB causes permanent damage to the hearing apparatus.

Timbre or quality Most musical sounds are made of complex waves formed by the summation of a high-amplitude wave (fundamental frequency) with several other waves of smaller amplitudes whose

frequencies are multiples of the fundamental frequency (harmonics or overtones). The **fundamental frequency** determines the pitch of the sound while the **overtones** give the sound its characteristic timbre or quality. Variations in timbre enable us to distinguish tones of the same pitch produced by different musical instruments.

A mathematical theorem called the **Fourier transform** makes it possible to express any wave with a given frequency and amplitude as the sum of a fundamental frequency and its numerous harmonics. Fourier transform has tremendous application in engineering. A well-known example is the musical synthesizer in which the tone of any musical instrument can be synthesized by mixing several multiples of a fundamental frequency in appropriate proportions of their amplitudes. Fourier transform has been applied to biological signals too, for example, for *data reduction in electroencephalogram*. Finally, it is interesting to note that the *basilar membrane of the cochlea works out the Fourier transform of the sound waves*: The sound wave breaks up into its component frequencies on the basilar membrane and each component registers its maximum amplitude at a different site of the basilar membrane. This information is conveyed separately to the brain which is then able to “synthesize” the quality of the sound.

Electricity

Electrical charge Atoms are made of protons, neutrons, and electrons. A proton has a unit positive charge, an electron has a unit negative charge, and neutrons do not have any charge on them. A body containing equal numbers of protons and electrons is electrostatically neutral. When the number of electrons in a body exceeds the number of protons in it, the body develops a **negative charge**. Conversely, it develops a **positive charge** when the number of electrons is less than the number of protons in it. The SI unit of charge is the **Coulomb** which equals the charge of 6.242×10^{18} electrons.

Current The *flow of charged particles* is called an electric current. The SI unit of current is the **Ampere**. A flow of 1 C of charge every second is 1 A current. Electricity is the flow of negatively charged electrons from a negative to positive potential. Conventionally, however, electricity is assumed to be the flow of positively charged particles from a positive to negative potential.

Potential The potential of a body is the work done when a charge of +1 C approaches the body from infinity under the effect of its attractive or repulsive force. If the work done is 1 J, the potential of

the body is said to be 1 V, which is the SI unit of potential. Two charged bodies have **1 V potential difference** if 1 J of work is done in moving a charge of +1 C from one body to the other.

Resistance The resistance of a conductor determines how much current it allows to pass through it when a potential difference is applied across it: for a given potential difference across a conductor, the current flowing through the conductor is inversely proportional to its resistance. This is known as the **Ohm law**. If the current is 1 A when the potential difference is 1 V, the resistance is said to be **1 Ω** , which is **the SI unit of resistance**. Conductance is the reciprocal of resistance and its SI unit is Mho.

$$\text{Voltage} = \text{Current} \times \text{Resistance}$$

An important derivation of Ohm law is that when resistances ($R_1, R_2, R_3 \dots$) are **connected in series** (Fig. 1.16A), the total resistance (R) is given by:

$$R = R_1 + R_2 + R_3 \dots$$

When R_1, R_2 , and R_3 are **connected in parallel** (Fig. 1.16B), the total resistance is given by the formula:

$$1/R = 1/R_1 + 1/R_2 + 1/R_3 \dots$$

The resistance of a conductor varies directly with its **length** and inversely with its **cross-sectional area**. This principle is important in the understanding of nerve conduction and explains why *nerve conduction velocity increases with axon diameter*.

Capacitance The potential of a body is directly proportional to the amount of charge it contains. The greater the amount of charge a body contains,

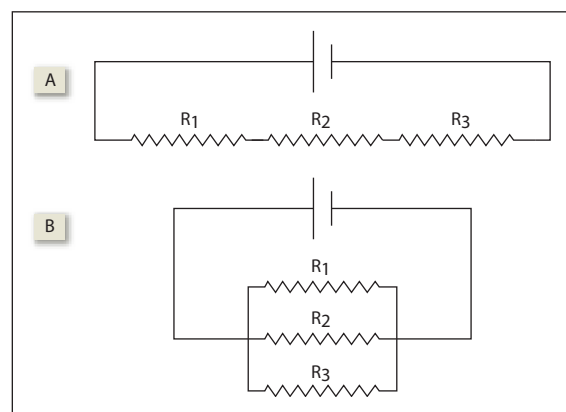


Fig. 1.16 Resistances in series [A] and parallel [B].

Box 1.1 What is Zero Potential?

Although earth is a stable reference for all electric potentials, it is impractical to express all potentials with reference to ground potential. Consider Fig. 55.2 showing the Warila pass. Its altitude is shown as 17,300 feet. The reference point here is the sea level. A more stable reference would be the center of the earth but that is unnecessary and inconvenient. Similarly, it is more convenient to measure the heights of the local peaks in reference to the average altitude of the surrounding terrain. Thus, all bioelectric potentials are expressed in reference to the potential of adjacent areas of the body, or to the average potential of the body.

the greater is its potential. However, the extent to which the potential rises when a given amount of charge is imparted to it will differ for different bodies. *A body that can hold a large amount of charge without much rise in its potential is said to have a high capacitance.* Stated mathematically,

$$\text{Capacitance} = \frac{\text{Charge}}{\text{Potential}}$$

Larger bodies have higher capacitance and, therefore, the *earth has almost an infinite capacitance.* Thus the potential of earth (ground potential) remains unchanged no matter how much charge enters it. This stable **ground potential** is assumed to be zero and serves as a *convenient reference for all electrical measurements* (Box 1.1). For the same reason, any charged body, when grounded, loses all its charge to the earth and its potential becomes zero.

A device that can store a large amount of charge is called a capacitor. *A capacitor can hold a lot of charge and yet its potential remains low.* On the other hand, a body with a low capacitance develops a high potential when it is charged. At high potentials, charge tends to “leak out” and hence a body with low capacitance cannot hold much charge. A capacitor can be built by sandwiching a thin sheet of dielectric (nonconducting material) between two parallel metal plates. Such a capacitor is called a **parallel plate capacitor**. When a potential difference is applied to the plates, a large amount of charge is stored in the plates.

The *nerve membrane behaves like a dielectric*, and when both its surfaces are layered with ions of opposite charges, it acts like a parallel plate capacitor (see Fig. 10.1). **Membrane capacitance** is an important determinant of membrane excitability and conduction velocity of nerve membranes.

■ Electricity and fluid dynamics

The concepts of electricity are relevant to fluid dynamics too and help in understanding several important hemodynamic phenomena. Conversely, simple analogies of fluid mechanics help in understanding certain electrical phenomena like capacitance. These analogies are explained below.

Fluid pressure is analogous to **electric potential** and fluid usually flows from high pressure to low pressure just as electric current flows from high to low potential. The amount of fluid flowing in unit time is analogous to **electric current**. Fluid resistance is quite similar to **electrical resistance**. The relationship between resistance, pressure drop, and flow rate in the blood vessels is similar to the relationship between the resistance, potential difference, and current:

$$\text{Resistance} = \text{Pressure drop/flow rate}$$

A similar formula is employed for defining blood flow resistance in vascular beds in terms of flow rate and perfusion pressure. Even the formulas for electrical resistances in series and parallel are valid for fluid dynamics and have important implications in hemodynamics. For example, most capillary beds are disposed in parallel to the heart while the *portal beds are in series* (Fig. 1.17). The physics of parallel circuits explains the phenomena of **circulatory “steal”** like coronary steal or subclavian steal: The electrical analogy is that when multiple resistors are connected in parallel, more current flows through resistors with lower resistance and vice versa. The electrical analogy also helps in understanding why the *blood pressure falls when the resistance in any of the capillary beds decreases.*

The **electrical capacitance** of a body can be compared with the fluid capacity of a container. A container with a large base is similar to a capacitor because it can store a large amount of fluid (analogous to charge) without its hydrostatic pressure (analogous to potential) increasing too much. On the other hand, the pressure in a narrow tube rises quickly when filled with fluid. For the same reason, *leakage of charge is higher when the capacity is low* (Fig. 1.18A).

Another similarity between electricity and fluid dynamics pertains to the measurement of potential. Consider two containers with fluid filled to different levels. For measuring their average potential, the two containers are connected to a graduated measuring tube as shown in Fig. 1.18B. The height of the fluid column in the graduated

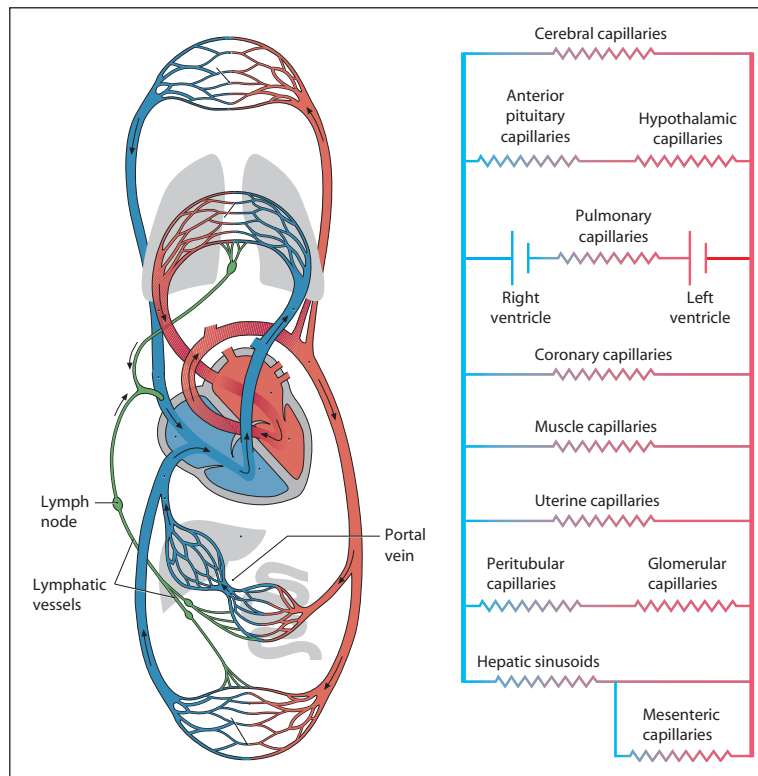


Fig. 1.17 Circulatory beds in series and parallel with the heart. Note that the capillary beds that are in series constitute a portal system.

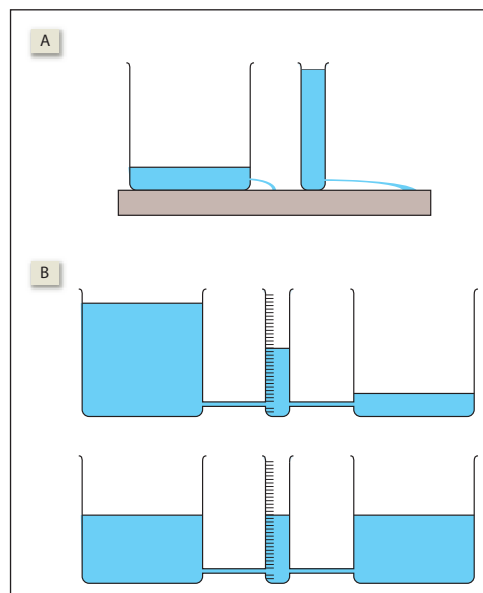


Fig. 1.18 Fluid analogy of electricity. [A] Fluid pressure and capacity are analogous to electric potential and capacitance. Shown here are two containers containing the same volume of fluid and with both having a small hole near their base. The container on the left has a broader base and therefore the hydrostatic pressure is low and leakage is much less. The container on the right has a narrow base. Hence the hydrostatic pressure is high and leakage is more. [B] A graduated tube interconnects two fluid-filled containers and is equidistant from both. The fluid pressure in the tube gives the mean of the fluid pressures in the two containers. However, the pressures in the two containers soon equalize unless the connecting limbs have high resistance to fluid flow.

tube gives the average fluid pressure in the two containers. The problem with this method is that since the two containers are interconnected, their pressures equalize very soon. The problem can be tackled by increasing the resistance of the connecting limbs. This analogy helps in understanding why it is important in electrocardiography to interconnect the three limb electrodes *only through a high (5,000 Ω) resistor* for obtaining zero potential (see Fig. 34.4).

■ Electric field of dipoles

Electric field refers to the area of influence of an electric charge (or charges). The electric field of a single charged particle is depicted in Fig. 1.19A. Each vector in the diagram represents *the direction and force with which a positive charge of 1 C moves (or tends to move) under the effect of the electric field*. If an electric field is applied to a medium containing mobile charged particles, the charged particles start moving along

the lines of the electric field, resulting in the flow of an electric current. This happens in body fluid where an electric field is set up by the heart.

A **dipole** is a pair of opposite charges lying close together. The electric field set up by a dipole is more complex. Each dipole results in the flow of current between the opposite charges along multiple pathways as shown in Fig. 1.19B. Although the conventional current flows from the positive to the negative charge, to a recording electrode placed at a distance, the current appears to flow along a vector directed from the negative to the positive charge.

Multiple dipoles located along a line result in a large resultant current. Each dipole generates a large field around it as shown in Figs. 1.19C and 1.19D. Only currents that travel straight along the axis of the dipole add up and become strong enough to be picked up by a distant recording electrode. A knowledge of dipoles and their fields is essential to the understanding of *electrocardiography and electroencephalography*.

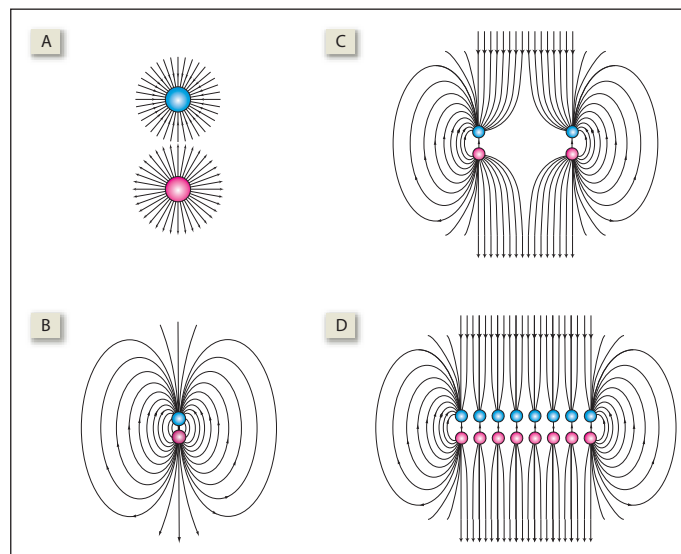


Fig. 1.19 [A] The electric field of a positive charge (pink) and a negative charge (blue). Each line represents the direction of the repulsive or attractive force exerted on a charge of +1 C. [B] The lines of force around a dipole. [C] The lines of force when two dipoles are placed side by side. [D] When several dipoles are lined up, the current flowing along the long axis of the dipoles summate and is recordable by a distant electrode.

REVISION QUESTIONS

- Gravity affects the blood pressure when the subject is standing but not when he is supine. Why?
- When the vertices of the Einthoven triangle is connected together, the nodal point is at zero potential. Which branch of mathematics explains it?
- What is the difference between gravitational force and g-force? What are G_x , G_y and G_z ? What is zero G_z and when is it experienced?
- In a sphygmomanometer, the 1 cm calibration is actually a little less than 1 cm. Why?

5. The total work done on the lung during the breathing cycle is said to be negative. Why?
6. The lower the functional residual capacity of the lungs, the more difficult it is to inflate it. This observation is in accordance with which law of physics?
7. Is blood a Newtonian fluid? Give reasons.
8. Even small changes in vessel diameter cause large variations in blood flow through it. This observation is in accordance with which law of physics?
9. What are the factors that (1) affect the flow rate of a fluid, and (2) tend to make the flow of a fluid turbulent?
10. Circulation of blood maintains a fairly uniform temperature throughout the body? Which mode of heat transfer is at work in this case?
11. During near-vision, short-sighted people (myopes) see things bigger and better. Give reasons.
12. If the eye is photographed in dim light using flashlight, the pupil appears red. Why?
13. If the distance between the intraocular lens and retina is 17 mm and the power of the intraocular lens after maximum accommodation is 50 diopter, what will be the near point of the eye?
14. Which physical characteristic of sound determines its (1) pitch, (2) intensity or loudness, and (3) quality or timbre?
15. The basilar membrane of the cochlea performs a Fourier transform on the sound waves. What is Fourier transform?
16. How does the diameter of a conductor affect its resistance?
17. What is zero potential?
18. What is the product of fluid resistance and the flow rate of the fluid?
19. *True or false:* System capillary beds are disposed in parallel while portal capillary beds are disposed in series.
20. What is circulatory steal?
21. *True or false:* if two capacitors contain the same amount of charge, the one with lower capacity will have lower potential. Give reasons.
22. In electrocardiography, the zero potential for reference is obtained by interconnecting the three limb electrodes through a high resistor. Why is a high resistor used?

2

Principles of Physical Chemistry in Physiology

Gases

Gas laws

Universal gas law The volume of a gas is directly proportional to the absolute temperature (T) of the gas (*Charles' law*) and inversely proportional to the pressure (P) applied upon it (*Boyle's law*). The volume is also directly proportional to the number of moles of gas (n). Stated mathematically,

$$V (\text{volume}) = n \times \frac{T}{P}$$

or

$$PV = nRT$$

where R is the proportionality constant and is called the *universal gas constant*.

The variability in gas volume with temperature and pressure makes it imperative to express all gas volumes at standard temperature and pressure (STP). The *standard temperature* is 0°C (273 Kelvin) and the *standard pressure* is 1 atmosphere (760 mm Hg.)

In physiology, gas volumes are often expressed at *body temperature (37°C)* and *1 atmosphere pressure*, or *BTP* for short.

Dalton law of partial pressure When two or more gases are mixed, the total pressure exerted by the mixture is equal to the sum of the pressures exerted by the constituent gases when occupying the same volume.

Suppose 500 mL of O₂ and 500 mL of N₂ at 0°C and 760 mm Hg are mixed together to make a 1 mL gaseous mixture (**Fig. 2.1**). What will be the pressure exerted by the mixture? Since the final volume of the gaseous mixture is 1 L, both the gases will come to occupy 1 L each

(and not 500 mL each). In accordance with Boyle law, the pressure of each gas will drop to 380 mm Hg. Finally, in accordance with Dalton law of partial pressure, the pressure of the mixture will be 380 + 380 = 760 mm Hg.

In the above example, the pressure of 380 mm Hg will be called the *partial pressure* of each of the constituent gases. It is calculated by the formula:

$$\begin{aligned} \text{Partial pressure of gas (380 mm Hg)} &= \text{Total pressure (760 mm Hg)} \times \frac{\text{Volume of the gas at STP (500 mL)}}{\text{Volume of the mixture at STP (1 L)}} \\ &= \text{Total pressure} \times \text{Concentration of a gas in a gaseous mixture} \end{aligned}$$

Taking another example: the concentration of O₂ in atmospheric air = 21%. Therefore, its partial pressure in atmospheric air will be 760 × 21% = 160 mm Hg.

The partial pressure of a gas dissolved in a liquid is equal to the partial pressure of the gas in a gaseous mixture that is in equilibrium with the liquid.

Aqueous tension When water vapor saturates a gas or a gas mixture, it exerts an additional pressure of 47 mm Hg at 37°C. This pressure is called the *saturated vapor pressure* or the *aqueous tension*. When air is saturated with water vapor, the concentration of all the constituent gases decreases. However, the percentage composition of a gas mixture is always expressed assuming that there is no moisture in it. For example, when we say that O₂ constitutes 21% of atmospheric air, we actually mean 21% of "dry" atmospheric air. If the air is saturated with

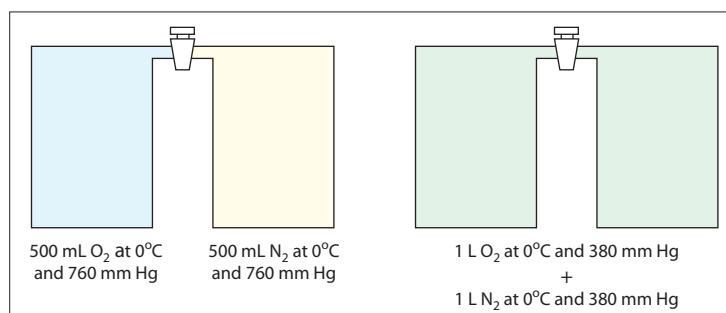


Fig. 2.1 An example of Dalton law of partial pressures.

water vapor, the saturated vapor pressure has to be deducted from the atmospheric pressure before calculating the partial pressure of constituent gases individually. For example, the partial pressure of O_2 (P_{O_2}) in humidified atmospheric air (at $37^\circ C$) is $(760 - 47) \times 21\% = 150$ mm Hg.

In physiology, it is a common practice to record the volume of gases *saturated with water vapor at body temperature and at atmospheric pressure*. Volumes thus recorded are denoted as *BTPS* ("S" for saturated).

Solutions

■ Diffusion of solutes

Simple diffusion If there are two solutions of different solute concentrations separated by a membrane, the *solutes diffuse from higher to lower concentration*. This is because the molecules in solution, due to their constant random motion, spread out to fill all of the available volume. The greater the concentration of a substance in solution, the greater is its tendency to spread out. Hence, although solutes in both the solutions are diffusing farther apart, there is a net flux of solutes from higher to lower concentration.

The diffusion of a solute from one region to another is directly proportional to the **cross-sectional area** across which the diffusion takes place and the **concentration gradient**, which is the difference in concentration of the diffusing substance divided by the thickness of the partition (*Fick law of diffusion*). Thus,

$$J = -DA \frac{\Delta c}{\Delta x}$$

where J is the net rate of diffusion, D is the diffusion coefficient, A is the area of the partition that separates the two solutions, and $\Delta c/\Delta x$ is the concentration gradient, that is, the rate of change of concentration with distance. The minus sign is a sign convention that indicates the direction of diffusion. When solutes diffuse from higher to lower concentration, $\Delta c/\Delta x$ is negative and therefore multiplying by $-DA$ gives a positive value of J .

Nernst potential When two ionic solutions A and B of different concentrations (C_A and C_B) are separated by a permeable membrane, the ions tend to diffuse along their concentration gradient. Since ions are charged particles, their diffusion can be stopped by an appropriate electrical potential (E) applied across the membrane. The magnitude and polarity of the potential that must be applied to

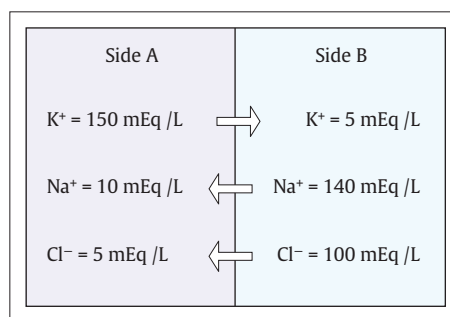


Fig. 2.2 Ionic distribution in two compartments for illustrating the principles of Nernst potential.

side A of the membrane for stopping the diffusion of ions (E_A) are given by the *Nernst equation*:

$$E_A = \frac{61}{z} \log \frac{[C_B]}{[C_A]}$$

where z is the valence of the ion. This equation can be also written as:

$$E_A = \frac{-61}{z} \log \frac{[C_A]}{[C_B]}$$

In the example illustrated in **Fig. 2.2**, the application of Nernst equation to the concentration of K^+ on side A (150 mEq/L) and side B (5 mEq/L) gives a value of -90 mV. This potential is called the *Nernst potential* for K^+ (E_K):

$$E_K = \frac{-61}{+1} \log \frac{150}{5}$$

It has two implications. **(1)** It means that -90 mV applied to side A will prevent any outward diffusion of the K^+ . **(2)** It also means that if the K^+ concentrations are equal on both sides of the membrane, a potential difference of -90 mV applied to the membrane will produce the same rate of diffusion as an ionic concentration ratio of 30 (i.e., $150 \div 5$).

Similarly, the diffusion of Na^+ will be prevented by a potential of $+70$ mV (E_{Na}) applied to side A and the diffusion of Cl^- (E_{Cl}) will be prevented by applying a potential of -80 mV to side A:

$$E_{Na} = \frac{-61}{+1} \log \frac{10}{140}$$

$$E_{Cl} = \frac{-61}{-1} \log \frac{5}{100}$$

Thus, every ion distributed asymmetrically across the membrane has its own Nernst potential (E) that will prevent its diffusion. In other

words, the *Nernst potential of an ion* gives the *electrical equivalent of its diffusion energy*.

Gibbs-Donnan equilibrium If two compartments are separated by a semipermeable membrane and one of the compartments contains diffusible ions, the ions will soon distribute themselves equally across the two compartments. If, however, *one of the compartments contains impermeable ions* (Fig. 2.3), the typical equilibrium (with equal numbers of ions on both sides of the membrane) as observed in simple diffusion will not occur. Instead, a different type of equilibrium called the *Gibbs-Donnan equilibrium* will occur in which two conditions must be satisfied: (1) both the compartments must be electroneutral and (2) the product of *diffusible ions* (anions and cations) must be equal in both compartments. The solution is shown in Table 2.1.

At equilibrium, compartment A will have a larger concentration of solutes (18) than compartment B (12). As a result, water will move from B to A due to osmosis and, thereby, decrease the solute concentration in compartment A. However, since the impermeable anion continues to be present in compartment A, the Gibbs-Donnan equilibrium is reestablished. As a result, compartment A would continue to have a higher solute concentration, no matter how much osmosis occurs!

A living cell contains *impermeable protein anions*. This imposes the Gibbs-Donnan equilibrium, so that

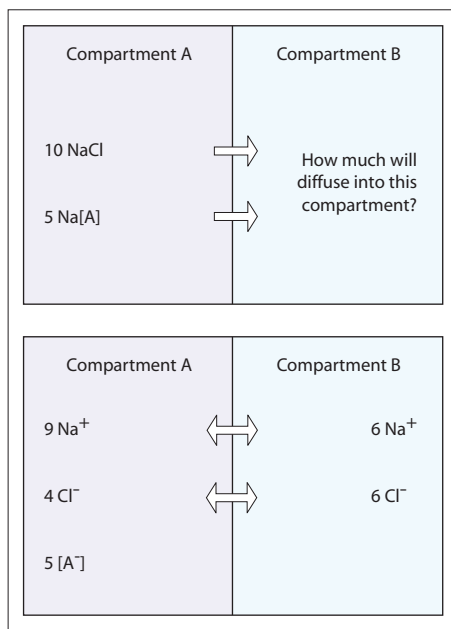


Fig. 2.3 Ionic distribution in two compartments for illustrating the principles of Gibbs-Donnan equilibrium. [A⁻] denotes an impermeable anion.

Table 2.1 Ionic distribution at Gibbs-Donnan equilibrium

	Inside	Outside
Total +ve charge	9 (Na ⁺)	6 (Na ⁺)
Total -ve charge	9 (4 Cl ⁻ + 5A ⁻)	6 (Cl ⁻)
Product of diffusible ions [Na ⁺] × [Cl ⁻]	36 (9 Na ⁺ × 4 Cl ⁻)	36 (6 Na ⁺ × 6 Cl ⁻)
Total number of ions	18 (9 + 4 + 5)	12 (6 + 6)

there are more ions inside the cell than outside it. As a result, water continuously moves into the cell by osmosis, which will inevitably rupture the cell. *To survive, a cell must continuously pump out excess ions*. It does so with the help of the adenosine triphosphate (ATP)-driven Na⁺-K⁺ pump. *When a cell dies, the pump stops and the cell swells up*, the microscopic appearance of which is known to pathologists as the *cloudy swelling*.

■ Diffusion of solvent

Osmosis is the movement of solvent across a semipermeable membrane that separates two solutions of different solute concentrations and restricts the movement of solutes across it. For osmosis to occur, it is not necessary for the membrane to be totally impermeable to the solute. A solute with limited permeability also produces osmosis, though to a lesser extent. The solute permeability is given by the *reflection coefficient*, which varies between 0 (freely permeable) and 1 (totally impermeable). The reflection coefficient is the *probability of a solute molecule reflecting back from the membrane instead of passing through it*. The reflection coefficient of a solute is not an absolute constant but varies with the type of membrane. It will be zero for all solutes if the membrane has very large pores in it. Osmosis is proportional to the reflection coefficient as well as to the concentration gradient. *Osmosis does not occur when the reflection coefficient is zero*, that is, when the membrane is freely permeable to the solute.

Osmotic pressure The osmosis from a dilute to a concentrated solution can be prevented by increasing the hydrostatic pressure of the concentrated solution. The *hydrostatic pressure necessary to prevent the osmosis* is called the osmotic pressure of the solution (Fig. 2.4).

Osmotic pressure, like freezing-point depression and boiling-point elevation, is a *colligative property*, that is, it depends on the number rather than the size or type of particles

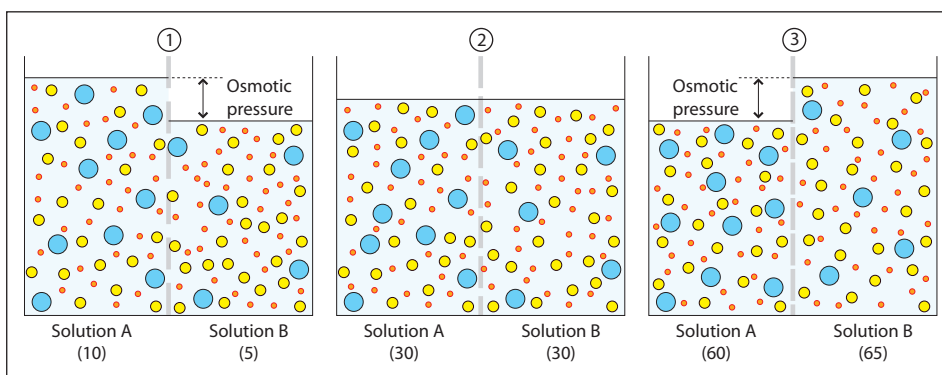


Fig. 2.4 A membrane separates two solutions, A and B, containing ions of three different sizes. Solution A contains 60 particles: 10 large, 20 medium-sized, and 30 small. Solution B contains 65 particles: 5 large, 25 medium-sized, and 35 small. The tonicity of the solutions depends on the membrane pore size. The figures below each solution indicate the number of osmotically active particles in it. (1) The pores allow only the medium-sized and small particles to pass through, making solution A hypertonic. (2) The pores allow only the small particles to pass through, making the two solutions isotonic. (3) The pores are impermeable to particles of all sizes, making solution B hypertonic.

in a solution. It is given by the **Van't Hoff relationship**:

$$\pi = RT \times \phi nC$$

where π is the osmotic pressure, R is the universal gas constant, T is the absolute temperature, ϕ is the osmotic coefficient, n is the number of ions yielded by the dissociation of a solute molecule, and C is the molar concentration (moles/L) of the solute.

The Van't Hoff equation shows that at a constant temperature, osmotic pressure is proportional to the number of particles in solution per unit volume of solution (given by $n \times C$ in the formula). The **osmotic coefficient** (ϕ) in the formula is not the same as the reflection coefficient. The osmotic coefficient does not depend on membrane characteristics: It is constant for a solute molecule and is assumed to be 1.0 in all approximate calculations.

Osmoles, osmolarity, and osmolality One mole¹ of osmotically active particles is called one **osmole** (Osm). Thus, a molar solution of glucose contains 1 osmole, a molar solution of NaCl contains 2 osmoles (1 mole of Na⁺ and 1 mole of Cl⁻), while a molar solution of CaCl₂ contains 3 osmoles (1 mole of Ca²⁺ and 2 moles of Cl⁻). The osmolar concentration of a solution in Osm/L is called **osmolarity**. When expressed in Osm/kg of solution, it is called **osmolality**.

Osmolarity is affected by temperature, which changes the volume of solution. Also, dissolution of solutes is associated with a slight rise in the

volume of the solution. This increase is different for different solutes. Hence, when 1 mole of glucose is dissolved in 1 L of water, the osmolarity will be slightly less than 1 Osm/L. However, **osmolality is unaffected by temperature** or the increased volume of the solution that accompanies dissolution.

The measurement of osmolarity (nC) is based on the principle that the freezing point of a solution is depressed in proportion to the number of osmoles present in it. It is given by the formula:

$$nC = \Delta T_f \div 1.86$$

where ΔT_f is the reduction in the freezing temperature. The freezing point of normal human plasma averages -0.54°C , which corresponds to an osmolal concentration in plasma of 290 mOsm/L. This is equivalent to an osmotic pressure against pure water of 7.3 atm.

Two solutions having identical osmolarity are called **isoosmolar**. They exert the same osmotic pressure and are therefore also called **isoosmotic**. If one of the two has greater osmolarity, it is said to be **hyperosmolar or hyperosmotic** in comparison to the other, which is called **hypoosmolar or hypoosmotic**. Except immediately after a sudden change in composition, all fluid compartments of the body are in osmotic equilibrium.

Tonicity When a membrane with unknown characteristics separates two solutions with different osmolarity, it does not necessarily mean that there

¹A "mole" is just a number. Just as a dozen equals 12 and a score equals 20, a mole equals 6.023×10^{23} . The weight of 1 mole of atoms or molecules of any element or compound is exactly equal to the atomic or molecular weight of the element or compound expressed in grams. For example, the molecular weight of NaCl is 58.5. Therefore, 6.023×10^{23} molecules (i.e., 1 mole) of NaCl weighs exactly 58.5 grams.

will be osmosis from the hypoosmolar to the hyperosmolar solution. This is made clear by the example illustrated in **Fig. 2.4** which explains why the magnitude and direction of osmosis cannot be predicted without knowing the pore size of the membrane. In real-life situations, the complexity of biological membranes makes it impossible to accurately predict whether a solute particle will pass through it. Moreover, body fluids contain innumerable types of solutes, which is another reason why their tonicity cannot be predicted. Therefore, *the only way of estimating the tonicity is to determine it experimentally.*

In clinical parlance, the word tonicity always refers to the tonicity of a solution *with respect to an erythrocyte*. In other words, it is the erythrocytic cell membrane across which the tonicity is tested. If the erythrocyte shrinks in a solution by losing water through osmosis, the solution is **hypertonic**. If the erythrocyte swells up in a solution by gaining water through osmosis, the solution is **hypotonic**. If the erythrocyte neither shrinks nor swells in the solution, the solution is called **isotonic**.

The fluid used in clinics for intravenous transfusion is both isotonic and isoosmolar to the plasma. A **0.9% saline solution** is most often used for the purpose. The 0.9% NaCl solution is roughly isoosmolar (308 mOsm/L) to the body fluids (290 mOsm/L) as can be calculated easily.² A **5% glucose solution** is also isotonic initially when infused intravenously, but as the glucose is metabolized, the solution gradually becomes hypotonic. An **isoosmolar solution of urea will not be isotonic** since urea rapidly diffuses into the erythrocytes.

Oncotic pressure Colloid osmotic pressure is also called oncotic pressure. The following example explains the necessity of distinguishing osmotic and oncotic pressures. If a hypertonic electrolyte solution (solution-A) is added to an isotonic colloid solution (solution-B), the mixture (solution-C) will have a higher osmotic pressure but a lower oncotic pressure than solution-B.

■ Buffers

The pH of a solution is maintained at near constant levels with the help of buffers (**Fig. 2.5**). Most body buffers comprise weak acids [HA] with their conjugate bases [A⁻]. By the law of mass action:

$$k = \frac{[H^+][A^-]}{[HA]}$$

²Molecular weight of NaCl = 58.5.

∴ 58.5 g of NaCl contain 1 mole of NaCl molecules.

∴ A solution containing 58.5 g NaCl will contain 2 osmoles (1 osmole of Na⁺ and 1 osmole of Cl⁻).

∴ A solution containing 9 g of NaCl contains (2 × 9) / 58.5 moles = 0.308 osmoles (or 308 milliosmoles).

∴ The osmolarity of a 9 g/L (0.9 g%) solution = 308 mOsm/L.

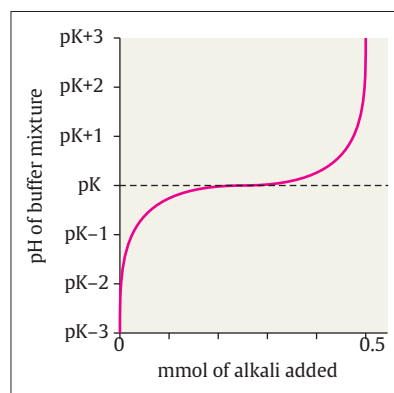


Fig. 2.5 The titration curve of a buffer mixture with alkali. The buffering power is maximum at pK.

where k is the dissociation constant.

$$[H^+] = k \times \frac{[HA]}{[A^-]}$$

$$-\log [H^+] = -\log k - \log \frac{[HA]}{[A^-]}$$

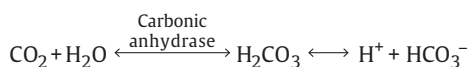
$$pH = pK + \log \frac{[A^-]}{[HA]}$$

The above equation is known as the **Henderson-Hasselbalch equation**.

The **buffering power** of the system is the number of moles of acid or base that must be added to 1 mole of the buffer to change its pH by 1 unit. The maximum buffering power of any buffer is 0.575. The buffering power of a buffer is maximal when the pH of the solution equals the pK. From the **Henderson-Hasselbalch equation**, it is seen that when $pH = pK$, $[A^-] = [HA]$. Hence for maximum buffering power, *a buffer solution should contain a weak acid and its conjugate base in equimolar proportion*. The buffering power also depends on the concentration of the buffer.

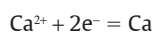
■ Chemical equilibrium

All chemical reactions are *potentially reversible* but they tend to *proceed unidirectionally* if one of the reactants or products is removed from the field of reaction. Consider the reaction shown below:



The reaction proceeds unidirectionally if *at least one of the products is quickly removed*, for example, if the H^+ is buffered. On the other hand, if the CO_2 escapes from the reaction mixture, the reaction will proceed in the backward direction.

Another important example is the *ionization of calcium in acidic solution*, which is explained by the following reaction:



Acids are electron acceptors. When the acid mops up electrons, the above reaction proceeds backward, causing ionization of calcium.

■ Electrophoresis

Electrophoresis is the migration of charged solutes or particles in a liquid medium under the influence of an electrical field (**Fig. 2.6**). Cations migrate toward the cathode and anions migrate toward the anode. An **ampholyte** (earlier called *zwitterion*) takes on a positive charge in an acidic solution and a negative charge in a basic solution. Depending on its pH, therefore, an ampholyte will migrate toward the anode or cathode. At a particular pH called the **isoelectric pH**, an ampholyte will be neutral and therefore will not move in an electric field. **Proteins** contain ionizable amino (NH_2) and carboxyl ($COOH$) groups and, therefore, they behave as ampholytes in solutions. **Nucleic acids** too are ampholytes.

The **rate of electrophoretic migration** depends on (1) the net electrical charge of the molecule, (2) the size and shape of the molecule, (3) electrical field

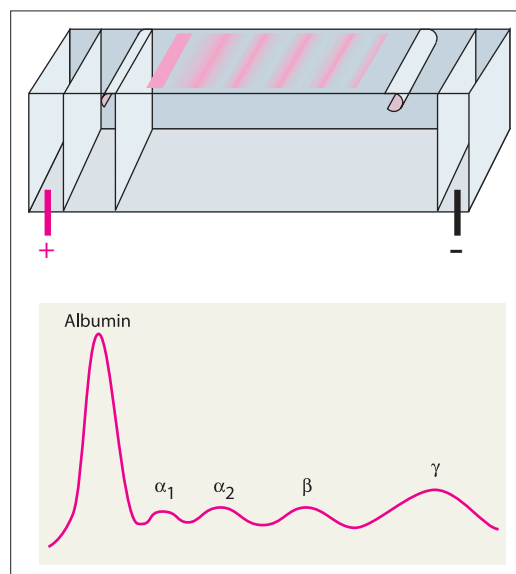


Fig. 2.6 (Above) Electrophoresis apparatus. (Below) Major components of plasma proteins separated out through electrophoresis.

strength, (4) properties of the supporting medium, and (5) temperature of operation. Electrophoretic mobility is defined as the rate of migration (cm/s) per unit field strength (V/cm). Since electrophoretic mobility depends, among other things, on the size and shape of the molecule, electrophoresis is used for the *separation of different proteins in a solution*.

REVISION QUESTIONS

1. In physiology, gas volumes are expressed in BTP. What is BTP?
2. If the air is saturated with water vapor, the saturated vapor pressure has to be deducted from the atmospheric pressure before calculating the partial pressure of constituent gases. Why?
3. If a permeable membrane separates two NaCl solutions having Na^+ concentrations of 100 and 10 mEq/L, calculate the potential that will prevent the diffusion of Na^+ across the membrane. Calculate also the potential that will be produced if the membrane is permeable only to Na^+ ions.
4. When a cell dies, it swells. Why?
5. Osmosis does not occur when the reflection coefficient is zero. What is reflection coefficient?
6. If two solutions are isosmotic, will they be isotonic too? Give reasons.
7. What is meant by an "isoelectric pH"?

3

Principles of Control Systems in Physiology

To control means to change something in a desired manner. A **control system** can therefore be defined as “a system with a goal.” The factors that have to be controlled to achieve the goal are called **variables**. The goal of a control system may be to increase or decrease a variable. The goal may also be to maintain the variable within a narrow range, in which case it is a **regulatory control**. Thus, a regulatory control can be defined as a special type of control meant for maintaining the constancy of a variable. On the other hand, a **servo control** sets the variable at any predetermined level that is not necessarily within a narrow range.

The above concepts can be understood easily by considering the example of a fan regulator. The regulator controls (increases or decreases) the speed of the fan (a variable), which in turn controls the amount of evaporation (another variable) of our sweat. The final result is the regulation of our body temperature (a third variable) that is maintained within physiological limits. The regulation is achieved through the servo control of the fan. Similarly, there is a control system in our body that controls the cardiac output (increases or decreases it) in order to regulate (maintain constant) tissue oxygenation. Other physiological examples are the regulation of blood pressure, regulation of blood glucose, regulation of body temperature, regulation or control of posture, control of heart rate, and control of breathing.

Homeostasis A special type of regulation is *homeostasis*, which refers to the regulation of the internal environment or *milieu intérieur* of the body. The importance of homeostasis becomes apparent when we consider the environmental adversities faced by a unicellular organism like the amoeba with the cozy environment of any human cell. The human cell enjoys a steady supply of oxygen at 100 mm Hg partial pressure and glucose at a concentration of 100 mg/dL. Its metabolites are promptly washed away from its immediate vicinity. It basks in the cozy comfort of 37°C and is secure in the protective cover of an elaborate defense system. All this is made possible by the cooperation of all the cells that constitute a multicellular organism in which each cell plays its part and, together, ensure the maintenance of a steady internal environment. For example, when the body temperature falls it is sensed by a group of cells in the brain that signals to the muscles to step up their activity so that the

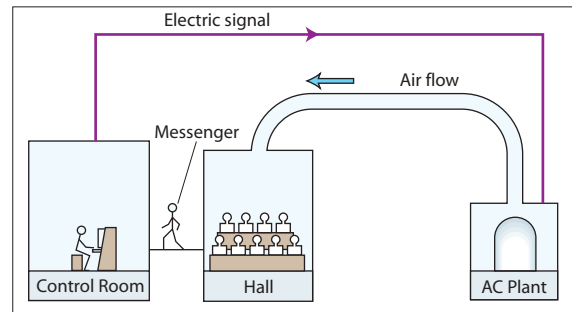


Fig. 3.1 A manually controlled temperature-regulatory system in a large hall can explain several principles of a control system.

temperature can be restored. Likewise, *each parameter that is held at a constant level by the body is the result of a unique homeostatic mechanism.*

Regulatory mechanisms

Several broad principles of regulatory system can be appreciated by considering the example of an air-conditioned hall (**Fig. 3.1**). The temperature inside the air-conditioned hall is regulated so that it remains constant (say, at 25°C). Whenever the temperature of the hall rises, the message is signaled to the control room. In a few minutes, an extra blast of cold air cools the hall. Conversely, when the hall temperature falls below 25°C, the cooling power of the air conditioner is reduced. All the essential elements of a regulatory system are present in the above example. There is a sensor or receptor that senses the variable (temperature inside the hall). There is an input that signals the information to the control center and there is an output (electrical power supply to the air-conditioning plant) that acts on the motor or the effector (the pump delivering the cold blast) that restores the variable to its original level.

Closed-loop (guided) control The control system described above is an example of a closed-loop control. The loop begins with the sensing of any change in the variable and ends with the correction of the same variable. In other words, *a closed-loop control system begins and ends with the controlled variable.* Such a system can be a crude one, manned by an untrained janitor who would only know when to increase or decrease the cooling power but not by

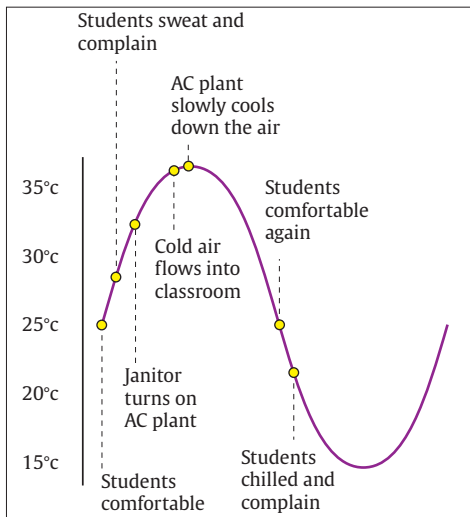


Fig. 3.2 Genesis of temperature oscillations in an air-conditioned hall. The temperature regulatory system is shown in Fig. 3.1.

how much. Hence, it is likely that each time the temperature drifts from the set point, it gets over-corrected resulting in temperature oscillations (Fig. 3.2). Despite its simplicity and its obvious propensity for oscillations, this system guarantees through repeated corrections that the hall temperature will hover around the desired temperature. It is called *guided control* and *requires virtually no knowledge of the factors that affect the variable*.

Open-loop (parametric, anticipatory) control Not all control systems have closed loops. Consider the same example of an air-conditioned hall. A brilliant engineer decides that he will not let the hall temperature drift at all so that there is no need whatsoever of correcting the temperature. So he derives a complex formula for calculating the amount of heat lost from the hall each minute and the amount of body heat produced by the students present in the hall. From this formula, he calculates the exact amount of cold air that is to be delivered to the hall each minute so that the hall temperature does not change. The engineer realizes that his formula needs to incorporate the number of students present in the hall (and therefore the amount of heat produced in the hall). He therefore installs at the door a sensor that counts the number of students entering the lecture hall. This type of control system is an open-loop control system. Here, *the controlled variable (hall temperature) is not the same as the feedback (number of students) provided by the sensor*. The control system tries to cool the hall “in anticipation” even before it starts warming up. The system

is *extremely prompt but not accurate* since it does not have any temperature sensors to judge if the hall temperature has been properly corrected. The open-loop control system is also called a *parametric control* system since it is *heavily dependent on parameters* and calculations based on them. Parametric control might appear to be only a theoretical proposition that is impracticable. That is not true. The *vestibuloocular reflex* (p 777) is an example of parametric control in which the vestibular apparatus (sensor) detects jerky head movements and reflexly corrects ocular gaze (controlled variable).

Normally, a parametric control system does not depend on information related to the controlled variable. However, if the same information is provided to it, parametric control becomes more efficient and is called *parametric feedforward control*. However, it still will never match the accuracy of a guided control system. This can be explained by extending the example of the air-conditioned hall as follows. Suppose at the end of a lecture, the students present in the hall complain to the engineer that the temperature of the hall was warmer than it should have been. The engineer realizes that he needs to revise his temperature-correcting formula. He therefore goes back to the hall and does a detailed survey of the hall to find out the reasons for the failure of his formula. He realizes that the projector used by the professor was generating a lot of heat and therefore incorporates this factor too in his new revised formula which turns out to be better and the students are satisfied. It is important to realize that the feedback regarding the controlled variable was not a continuous feedback and, therefore, is essentially different from guided control. In parametric feedforward control, *the control system learns and becomes wiser with time* and the control formula becomes more and more accurate.

What happens if the engineer is unable to detect the cause of failure to maintain the hall temperature at the set point? After all his efforts to detect the cause have failed, he will probably increase the cooling power a little above that suggested by his formula. He does so without understanding why his formula has failed, which could be an undetected leak in the air-conditioning duct. Such a system is called *parametric feedback control*.

Combined control Finally, there will always be certain factors, known or unknown, that are so unpredictable and random that no adjustments in the formula are possible for them. Such disturbances are called *noise*, which is an unavoidable component of any control system. Parametric control systems have no way