

Principles of optics

*Electromagnetic theory of propagation,
interference and diffraction of light*

MAX BORN

MA, Dr Phil, FRS

Nobel Laureate

Formerly Professor at the Universities of Göttingen and Edinburgh

and

EMIL WOLF

PhD, DSc

Wilson Professor of Optical Physics, University of Rochester, NY

with contributions by

A. B. BHATIA, P. C. CLEMMOW, D. GABOR, A. R. STOKES,
A. M. TAYLOR, P. A. WAYMAN AND W. L. WILCOCK

SEVENTH (EXPANDED) EDITION



CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa
<http://www.cambridge.org>

Seventh (expanded) edition © Margaret Farley-Born and Emil Wolf 1999

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 1959 by Pergamon Press Ltd, London
Sixth edition 1980
Reprinted (with corrections) 1983, 1984, 1986, 1987, 1989, 1991, 1993
Reissued by Cambridge University Press 1997
Seventh (expanded) edition 1999
Reprinted with corrections, 2002
Reprinted 2003

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data

Born, Max
Principles of Optics—7th ed.
I. Optics
I. Title II. Wolf Emil
535 QC351 80-41470

ISBN 0 521 642221 hardback

Contents

Historical introduction	xxv
I Basic properties of the electromagnetic field	1
1.1 The electromagnetic field	1
1.1.1 Maxwell's equations	1
1.1.2 Material equations	2
1.1.3 Boundary conditions at a surface of discontinuity	4
1.1.4 The energy law of the electromagnetic field	7
1.2 The wave equation and the velocity of light	11
1.3 Scalar waves	14
1.3.1 Plane waves	15
1.3.2 Spherical waves	16
1.3.3 Harmonic waves. The phase velocity	16
1.3.4 Wave packets. The group velocity	19
1.4 Vector waves	24
1.4.1 The general electromagnetic plane wave	24
1.4.2 The harmonic electromagnetic plane wave	25
(a) Elliptic polarization	25
(b) Linear and circular polarization	29
(c) Characterization of the state of polarization by Stokes parameters	31
1.4.3 Harmonic vector waves of arbitrary form	33
1.5 Reflection and refraction of a plane wave	38
1.5.1 The laws of reflection and refraction	38
1.5.2 Fresnel formulae	40
1.5.3 The reflectivity and transmissivity; polarization on reflection and refraction	43
1.5.4 Total reflection	49
1.6 Wave propagation in a stratified medium. Theory of dielectric films	54
1.6.1 The basic differential equations	55
1.6.2 The characteristic matrix of a stratified medium	58
(a) A homogeneous dielectric film	61
(b) A stratified medium as a pile of thin homogeneous films	62
1.6.3 The reflection and transmission coefficients	63
1.6.4 A homogeneous dielectric film	64
1.6.5 Periodically stratified media	70
II Electromagnetic potentials and polarization	75
2.1 The electrodynamic potentials in the vacuum	76

2.1.1	The vector and scalar potentials	76
2.1.2	Retarded potentials	78
2.2	Polarization and magnetization	80
2.2.1	The potentials in terms of polarization and magnetization	80
2.2.2	Hertz vectors	84
2.2.3	The field of a linear electric dipole	85
2.3	The Lorentz–Lorenz formula and elementary dispersion theory	89
2.3.1	The dielectric and magnetic susceptibilities	89
2.3.2	The effective field	90
2.3.3	The mean polarizability: the Lorentz–Lorenz formula	92
2.3.4	Elementary theory of dispersion	95
2.4	Propagation of electromagnetic waves treated by integral equations	103
2.4.1	The basic integral equation	104
2.4.2	The Ewald–Oseen extinction theorem and a rigorous derivation of the Lorentz–Lorenz formula	105
2.4.3	Refraction and reflection of a plane wave, treated with the help of the Ewald–Oseen extinction theorem	110
III	Foundations of geometrical optics	116
3.1	Approximation for very short wavelengths	116
3.1.1	Derivation of the eikonal equation	117
3.1.2	The light rays and the intensity law of geometrical optics	120
3.1.3	Propagation of the amplitude vectors	125
3.1.4	Generalizations and the limits of validity of geometrical optics	127
3.2	General properties of rays	129
3.2.1	The differential equation of light rays	129
3.2.2	The laws of refraction and reflection	132
3.2.3	Ray congruences and their focal properties	134
3.3	Other basic theorems of geometrical optics	135
3.3.1	Lagrange’s integral invariant	135
3.3.2	The principle of Fermat	136
3.3.3	The theorem of Malus and Dupin and some related theorems	139
IV	Geometrical theory of optical imaging	142
4.1	The characteristic functions of Hamilton	142
4.1.1	The point characteristic	142
4.1.2	The mixed characteristic	144
4.1.3	The angle characteristic	146
4.1.4	Approximate form of the angle characteristic of a refracting surface of revolution	147
4.1.5	Approximate form of the angle characteristic of a reflecting surface of revolution	151
4.2	Perfect imaging	152
4.2.1	General theorems	153
4.2.2	Maxwell’s ‘fish-eye’	157
4.2.3	Stigmatic imaging of surfaces	159
4.3	Projective transformation (collineation) with axial symmetry	160
4.3.1	General formulae	161
4.3.2	The telescopic case	164
4.3.3	Classification of projective transformations	165
4.3.4	Combination of projective transformations	166
4.4	Gaussian optics	167
4.4.1	Refracting surface of revolution	167

4.4.2	Reflecting surface of revolution	170
4.4.3	The thick lens	171
4.4.4	The thin lens	174
4.4.5	The general centred system	175
4.5	Stigmatic imaging with wide-angle pencils	178
4.5.1	The sine condition	179
4.5.2	The Herschel condition	180
4.6	Astigmatic pencils of rays	181
4.6.1	Focal properties of a thin pencil	181
4.6.2	Refraction of a thin pencil	182
4.7	Chromatic aberration. Dispersion by a prism	186
4.7.1	Chromatic aberration	186
4.7.2	Dispersion by a prism	190
4.8	Radiometry and apertures	193
4.8.1	Basic concepts of radiometry	194
4.8.2	Stops and pupils	199
4.8.3	Brightness and illumination of images	201
4.9	Ray tracing	204
4.9.1	Oblique meridional rays	204
4.9.2	Paraxial rays	207
4.9.3	Skew rays	208
4.10	Design of aspheric surfaces	211
4.10.1	Attainment of axial stigmatism	211
4.10.2	Attainment of aplanatism	214
4.11	Image-reconstruction from projections (computerized tomography)	217
4.11.1	Introduction	217
4.11.2	Beam propagation in an absorbing medium	218
4.11.3	Ray integrals and projections	219
4.11.4	The N -dimensional Radon transform	221
4.11.5	Reconstruction of cross-sections and the projection-slice theorem of computerized tomography	223
V	Geometrical theory of aberrations	228
5.1	Wave and ray aberrations; the aberration function	229
5.2	The perturbation eikonal of Schwarzschild	233
5.3	The primary (Seidel) aberrations	236
(a)	Spherical aberration ($B \neq 0$)	238
(b)	Coma ($F \neq 0$)	238
(c)	Astigmatism ($C \neq 0$) and curvature of field ($D \neq 0$)	240
(d)	Distortion ($E \neq 0$)	243
5.4	Addition theorem for the primary aberrations	244
5.5	The primary aberration coefficients of a general centred lens system	246
5.5.1	The Seidel formulae in terms of two paraxial rays	246
5.5.2	The Seidel formulae in terms of one paraxial ray	251
5.5.3	Petzval's theorem	253
5.6	Example: The primary aberrations of a thin lens	254
5.7	The chromatic aberration of a general centred lens system	257
VI	Image-forming instruments	261
6.1	The eye	261
6.2	The camera	263
6.3	The refracting telescope	267
6.4	The reflecting telescope	274

6.5	Instruments of illumination	279
6.6	The microscope	281
VII	Elements of the theory of interference and interferometers	286
7.1	Introduction	286
7.2	Interference of two monochromatic waves	287
7.3	Two-beam interference: division of wave-front	290
7.3.1	Young's experiment	290
7.3.2	Fresnel's mirrors and similar arrangements	292
7.3.3	Fringes with quasi-monochromatic and white light	295
7.3.4	Use of slit sources; visibility of fringes	296
7.3.5	Application to the measurement of optical path difference: the Rayleigh interferometer	299
7.3.6	Application to the measurement of angular dimensions of sources: the Michelson stellar interferometer	302
7.4	Standing waves	308
7.5	Two-beam interference: division of amplitude	313
7.5.1	Fringes with a plane-parallel plate	313
7.5.2	Fringes with thin films; the Fizeau interferometer	318
7.5.3	Localization of fringes	325
7.5.4	The Michelson interferometer	334
7.5.5	The Twyman–Green and related interferometers	336
7.5.6	Fringes with two identical plates: the Jamin interferometer and interference microscopes	341
7.5.7	The Mach–Zehnder interferometer; the Bates wave-front shearing interferometer	348
7.5.8	The coherence length; the application of two-beam interference to the study of the fine structure of spectral lines	352
7.6	Multiple-beam interference	359
7.6.1	Multiple-beam fringes with a plane-parallel plate	360
7.6.2	The Fabry–Perot interferometer	366
7.6.3	The application of the Fabry–Perot interferometer to the study of the fine structure of spectral lines	370
7.6.4	The application of the Fabry–Perot interferometer to the comparison of wavelengths	377
7.6.5	The Lummer–Gehrcke interferometer	380
7.6.6	Interference filters	386
7.6.7	Multiple-beam fringes with thin films	391
7.6.8	Multiple-beam fringes with two plane-parallel plates	401
	(a) Fringes with monochromatic and quasi-monochromatic light	401
	(b) Fringes of superposition	405
7.7	The comparison of wavelengths with the standard metre	409
VIII	Elements of the theory of diffraction	412
8.1	Introduction	412
8.2	The Huygens–Fresnel principle	413
8.3	Kirchhoff's diffraction theory	417
8.3.1	The integral theorem of Kirchhoff	417
8.3.2	Kirchhoff's diffraction theory	421
8.3.3	Fraunhofer and Fresnel diffraction	425
8.4	Transition to a scalar theory	430
8.4.1	The image field due to a monochromatic oscillator	431
8.4.2	The total image field	434

8.5	Fraunhofer diffraction at apertures of various forms	436
8.5.1	The rectangular aperture and the slit	436
8.5.2	The circular aperture	439
8.5.3	Other forms of aperture	443
8.6	Fraunhofer diffraction in optical instruments	446
8.6.1	Diffraction gratings	446
	(a) The principle of the diffraction grating	446
	(b) Types of grating	453
	(c) Grating spectrographs	458
8.6.2	Resolving power of image-forming systems	461
8.6.3	Image formation in the microscope	465
	(a) Incoherent illumination	465
	(b) Coherent illumination – Abbe's theory	467
	(c) Coherent illumination – Zernike's phase contrast method of observation	472
8.7	Fresnel diffraction at a straight edge	476
8.7.1	The diffraction integral	476
8.7.2	Fresnel's integrals	478
8.7.3	Fresnel diffraction at a straight edge	481
8.8	The three-dimensional light distribution near focus	484
8.8.1	Evaluation of the diffraction integral in terms of Lommel functions	484
8.8.2	The distribution of intensity	489
	(a) Intensity in the geometrical focal plane	490
	(b) Intensity along the axis	491
	(c) Intensity along the boundary of the geometrical shadow	491
8.8.3	The integrated intensity	492
8.8.4	The phase behaviour	494
8.9	The boundary diffraction wave	499
8.10	Gabor's method of imaging by reconstructed wave-fronts (holography)	504
8.10.1	Producing the positive hologram	504
8.10.2	The reconstruction	506
8.11	The Rayleigh–Sommerfeld diffraction integrals	512
8.11.1	The Rayleigh diffraction integrals	512
8.11.2	The Rayleigh–Sommerfeld diffraction integrals	514
IX	The diffraction theory of aberrations	517
9.1	The diffraction integral in the presence of aberrations	518
9.1.1	The diffraction integral	518
9.1.2.	The displacement theorem. Change of reference sphere	520
9.1.3.	A relation between the intensity and the average deformation of wave-fronts	522
9.2	Expansion of the aberration function	523
9.2.1	The circle polynomials of Zernike	523
9.2.2	Expansion of the aberration function	525
9.3	Tolerance conditions for primary aberrations	527
9.4	The diffraction pattern associated with a single aberration	532
9.4.1	Primary spherical aberration	536
9.4.2	Primary coma	538
9.4.3	Primary astigmatism	539
9.5	Imaging of extended objects	543
9.5.1	Coherent illumination	543
9.5.2	Incoherent illumination	547

X	Interference and diffraction with partially coherent light	554
10.1	Introduction	554
10.2	A complex representation of real polychromatic fields	557
10.3	The correlation functions of light beams	562
10.3.1	Interference of two partially coherent beams. The mutual coherence function and the complex degree of coherence	562
10.3.2	Spectral representation of mutual coherence	566
10.4	Interference and diffraction with quasi-monochromatic light	569
10.4.1	Interference with quasi-monochromatic light. The mutual intensity	569
10.4.2	Calculation of mutual intensity and degree of coherence for light from an extended incoherent quasi-monochromatic source	572
	(a) The van Cittert–Zernike theorem	572
	(b) Hopkins' formula	577
10.4.3	An example	578
10.4.4	Propagation of mutual intensity	580
10.5	Interference with broad-band light and the spectral degree of coherence. Correlation-induced spectral changes	585
10.6	Some applications	590
10.6.1	The degree of coherence in the image of an extended incoherent quasi-monochromatic source	590
10.6.2	The influence of the condenser on resolution in a microscope	595
	(a) Critical illumination	595
	(b) Köhler's illumination	598
10.6.3	Imaging with partially coherent quasi-monochromatic illumination	599
	(a) Transmission of mutual intensity through an optical system	599
	(b) Images of transilluminated objects	602
10.7	Some theorems relating to mutual coherence	606
10.7.1	Calculation of mutual coherence for light from an incoherent source	606
10.7.2	Propagation of mutual coherence	609
10.8	Rigorous theory of partial coherence	610
10.8.1	Wave equations for mutual coherence	610
10.8.2	Rigorous formulation of the propagation law for mutual coherence	612
10.8.3	The coherence time and the effective spectral width	615
10.9	Polarization properties of quasi-monochromatic light	619
10.9.1	The coherency matrix of a quasi-monochromatic plane wave	619
	(a) Completely unpolarized light (natural light)	624
	(b) Complete polarized light	624
10.9.2	Some equivalent representations. The degree of polarization of a light wave	626
10.9.3	The Stokes parameters of a quasi-monochromatic plane wave	630
XI	Rigorous diffraction theory	633
11.1	Introduction	633
11.2	Boundary conditions and surface currents	635
11.3	Diffraction by a plane screen: electromagnetic form of Babinet's principle	636
11.4	Two-dimensional diffraction by a plane screen	638
11.4.1	The scalar nature of two-dimensional electromagnetic fields	638
11.4.2	An angular spectrum of plane waves	639
11.4.3	Formulation in terms of dual integral equations	642
11.5	Two-dimensional diffraction of a plane wave by a half-plane	643
11.5.1	Solution of the dual integral equations for E -polarization	643
11.5.2	Expression of the solution in terms of Fresnel integrals	645
11.5.3	The nature of the solution	648

11.5.4	The solution for H -polarization	652
11.5.5	Some numerical calculations	653
11.5.6	Comparison with approximate theory and with experimental results	656
11.6	Three-dimensional diffraction of a plane wave by a half-plane	657
11.7	Diffraction of a field due to a localized source by a half-plane	659
11.7.1	A line-current parallel to the diffracting edge	659
11.7.2	A dipole	664
11.8	Other problems	667
11.8.1	Two parallel half-planes	667
11.8.2	An infinite stack of parallel, staggered half-planes	669
11.8.3	A strip	670
11.8.4	Further problems	671
11.9	Uniqueness of solution	672
XII	Diffraction of light by ultrasonic waves	674
12.1	Qualitative description of the phenomenon and summary of theories based on Maxwell's differential equations	674
12.1.1	Qualitative description of the phenomenon	674
12.1.2	Summary of theories based on Maxwell's equations	677
12.2	Diffraction of light by ultrasonic waves as treated by the integral equation method	680
12.2.1	Integral equation for E-polarization	682
12.2.2	The trial solution of the integral equation	682
12.2.3	Expressions for the amplitudes of the light waves in the diffracted and reflected spectra	686
12.2.4	Solution of the equations by a method of successive approximations	686
12.2.5	Expressions for the intensities of the first and second order lines for some special cases	689
12.2.6	Some qualitative results	691
12.2.7	The Raman–Nath approximation	693
XIII	Scattering from inhomogeneous media	695
13.1	Elements of the scalar theory of scattering	695
13.1.1	Derivation of the basic integral equation	695
13.1.2	The first-order Born approximation	699
13.1.3	Scattering from periodic potentials	703
13.1.4	Multiple scattering	708
13.2	Principles of diffraction tomography for reconstruction of the scattering potential	710
13.2.1	Angular spectrum representation of the scattered field	711
13.2.2	The basic theorem of diffraction tomography	713
13.3	The optical cross-section theorem	716
13.4	A reciprocity relation	724
13.5	The Rytov series	726
13.6	Scattering of electromagnetic waves	729
13.6.1	The integro-differential equations of electromagnetic scattering theory	729
13.6.2	The far field	730
13.6.3	The optical cross-section theorem for scattering of electromagnetic waves	732
XIV	Optics of metals	735
14.1	Wave propagation in a conductor	735

14.2	Refraction and reflection at a metal surface	739
14.3	Elementary electron theory of the optical constants of metals	749
14.4	Wave propagation in a stratified conducting medium. Theory of metallic films	752
14.4.1	An absorbing film on a transparent substrate	752
14.4.2	A transparent film on an absorbing substrate	758
14.5	Diffraction by a conducting sphere; theory of Mie	759
14.5.1	Mathematical solution of the problem	760
	(a) Representation of the field in terms of Debye's potentials	760
	(b) Series expansions for the field components	765
	(c) Summary of formulae relating to the associated Legendre functions and to the cylindrical functions	772
14.5.2	Some consequences of Mie's formulae	774
	(a) The partial waves	774
	(b) Limiting cases	775
	(c) Intensity and polarization of the scattered light	780
14.5.3	Total scattering and extinction	784
	(a) Some general considerations	784
	(b) Computational results	785
XV	Optics of crystals	790
15.1	The dielectric tensor of an anisotropic medium	790
15.2	The structure of a monochromatic plane wave in an anisotropic medium	792
15.2.1	The phase velocity and the ray velocity	792
15.2.2	Fresnel's formulae for the propagation of light in crystals	795
15.2.3	Geometrical constructions for determining the velocities of propagation and the directions of vibration	799
	(a) The ellipsoid of wave normals	799
	(b) The ray ellipsoid	802
	(c) The normal surface and the ray surface	803
15.3	Optical properties of uniaxial and biaxial crystals	805
15.3.1	The optical classification of crystals	805
15.3.2	Light propagation in uniaxial crystals	806
15.3.3	Light propagation in biaxial crystals	808
15.3.4	Refraction in crystals	811
	(a) Double refraction	811
	(b) Conical refraction	813
15.4	Measurements in crystal optics	818
15.4.1	The Nicol prism	818
15.4.2	Compensators	820
	(a) The quarter-wave plate	820
	(b) Babinet's compensator	821
	(c) Soleil's compensator	823
	(d) Berek's compensator	823
15.4.3	Interference with crystal plates	823
15.4.4	Interference figures from uniaxial crystal plates	829
15.4.5	Interference figures from biaxial crystal plates	831
15.4.6	Location of optic axes and determination of the principal refractive indices of a crystalline medium	833
15.5	Stress birefringence and form birefringence	834
15.5.1	Stress birefringence	834
15.5.2	Form birefringence	837

15.6	Absorbing crystals	840
15.6.1	Light propagation in an absorbing anisotropic medium	840
15.6.2	Interference figures from absorbing crystal plates	846
	(a) Uniaxial crystals	847
	(b) Biaxial crystals	848
15.6.3	Dichroic polarizers	849
<i>Appendices</i>		853
I	The Calculus of variations	853
1	Euler's equations as necessary conditions for an extremum	853
2	Hilbert's independence integral and the Hamilton–Jacobi equation	855
3	The field of extremals	856
4	Determination of all extremals from the solution of the Hamilton–Jacobi equation	858
5	Hamilton's canonical equations	860
6	The special case when the independent variable does not appear explicitly in the integrand	861
7	Discontinuities	862
8	Weierstrass' and Legendre's conditions (sufficiency conditions for an extremum)	864
9	Minimum of the variational integral when one end point is constrained to a surface	866
10	Jacobi's criterion for a minimum	867
11	Example I: Optics	868
12	Example II: Mechanics of material points	870
II	Light optics, electron optics and wave mechanics	873
1	The Hamiltonian analogy in elementary form	873
2	The Hamiltonian analogy in variational form	876
3	Wave mechanics of free electrons	879
4	The application of optical principles to electron optics	881
III	Asymptotic approximations to integrals	883
1	The method of steepest descent	883
2	The method of stationary phase	888
3	Double integrals	890
IV	The Dirac delta function	892
V	A mathematical lemma used in the rigorous derivation of the Lorentz–Lorenz formula (§2.4.2)	898
VI	Propagation of discontinuities in an electromagnetic field (§3.1.1)	901
1	Relations connecting discontinuous changes in field vectors	901
2	The field on a moving discontinuity surface	903
VII	The circle polynomials of Zernike (§9.2.1)	905
1	Some general considerations	905
2	Explicit expressions for the radial polynomials $R_n^{\pm m}(\rho)$	907
VIII	Proof of the inequality $\mu_{12}(\nu) \leq 1$ for the spectral degree of coherence (§10.5)	911
IX	Proof of a reciprocity inequality (§10.8.3)	912
X	Evaluation of two integrals (§12.2.2)	914
XI	Energy conservation in scalar wavefields (§13.3)	918
XII	Proof of Jones' lemma (§13.3)	921
	Author index	925
	Subject index	936

I

Basic properties of the electromagnetic field

1.1 The electromagnetic field

1.1.1 Maxwell's equations

THE state of excitation which is established in space by the presence of electric charges is said to constitute an *electromagnetic field*. It is represented by two vectors, \mathbf{E} and \mathbf{B} , called the *electric vector* and the *magnetic induction* respectively.*

To describe the effect of the field on material objects, it is necessary to introduce a second set of vectors, viz. *the electric current density* \mathbf{j} , *the electric displacement* \mathbf{D} , and *the magnetic vector* \mathbf{H} .

The space and time derivatives of the five vectors are related by *Maxwell's equations*, which hold at every point in whose neighbourhood the physical properties of the medium are continuous:†

$$\text{curl } \mathbf{H} - \frac{1}{c} \dot{\mathbf{D}} = \frac{4\pi}{c} \mathbf{j}, \quad (1)$$

$$\text{curl } \mathbf{E} + \frac{1}{c} \dot{\mathbf{B}} = 0, \quad (2)$$

the dot denoting differentiation with respect to time.

* In elementary considerations \mathbf{E} and \mathbf{H} are, for historical reasons, usually regarded as the basic field vectors, and \mathbf{D} and \mathbf{B} as describing the influence of matter. In general theory, however, the present interpretation is compulsory for reasons connected with the electrodynamics of moving media.

The four Maxwell equations (1)–(4) can be divided into two sets of equations, one consisting of two homogeneous equations (right-hand side zero), containing \mathbf{E} and \mathbf{B} , the other of two nonhomogeneous equations (right-hand side different from zero), containing \mathbf{D} and \mathbf{H} . If a coordinate transformation of space and time (relativistic Lorentz transformation) is carried out, the equations of each group transform together, the equations remaining unaltered in form if \mathbf{j}/c and ρ are transformed as a four-vector, and each of the pairs \mathbf{E} , \mathbf{B} and \mathbf{D} , \mathbf{H} as a six-vector (antisymmetric tensor of the second order). Since the nonhomogeneous set contains charges and currents (which represent the influence of matter), one has to attribute the corresponding pair (\mathbf{D} , \mathbf{H}) to the influence of matter. It is, however, customary to refer to \mathbf{H} and not to \mathbf{B} as the *magnetic field vector*; we shall conform to this terminology when there is no risk of confusion.

† The so-called Gaussian system of units is used here, i.e. the electrical quantities (\mathbf{E} , \mathbf{D} , \mathbf{j} and ρ) are measured in electrostatic units, and the magnetic quantities (\mathbf{H} and \mathbf{B}) in electromagnetic units. The constant c in (1) and (2) relates the unit of charge in the two systems; it is the velocity of light in the vacuum and is approximately equal to 3×10^{10} cm/s. (A more accurate value is given in §1.2.)

They are supplemented by two scalar relations:

$$\operatorname{div} \mathbf{D} = 4\pi\rho, \quad (3)$$

$$\operatorname{div} \mathbf{B} = 0. \quad (4)$$

Eq. (3) may be regarded as a defining equation for the electric charge density ρ and (4) may be said to imply that no free magnetic poles exist.

From (1) it follows (since $\operatorname{div} \operatorname{curl} \equiv 0$) that

$$\operatorname{div} \mathbf{j} = -\frac{1}{4\pi} \operatorname{div} \dot{\mathbf{D}},$$

or, using (3),

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \mathbf{j} = 0. \quad (5)$$

By analogy with a similar relation encountered in hydrodynamics, (5) is called the *equation of continuity*. It expresses the fact that the charge is conserved in the neighbourhood of any point. For if one integrates (5) over any region of space, one obtains, with the help of Gauss' theorem,

$$\frac{d}{dt} \int \rho \, dV + \int \mathbf{j} \cdot \mathbf{n} \, dS = 0, \quad (6)$$

the second integral being taken over the surface bounding the region and the first throughout the volume, \mathbf{n} denoting the unit outward normal. This equation implies that the total charge

$$e = \int \rho \, dV \quad (7)$$

contained within the domain can only increase on account of the flow of electric current

$$\mathcal{I} = \int \mathbf{j} \cdot \mathbf{n} \, dS. \quad (8)$$

If all the field quantities are independent of time, and if, moreover, there are no currents ($\mathbf{j} = 0$), the field is said to be *static*. If all the field quantities are time independent, but currents are present ($\mathbf{j} \neq 0$), one speaks of a *stationary field*. In optical fields the field vectors are very rapidly varying functions of time, but the sources of the field are usually such that, when averages over any macroscopic time interval are considered rather than the instantaneous values, the properties of the field are found to be independent of the instant of time at which the average is taken. The word *stationary* is often used in a wider sense to describe a field of this type. An example is a field constituted by the steady flux of radiation (say from a distant star) through an optical system.

1.1.2 Material equations

The Maxwell equations (1)–(4) connect the five basic quantities \mathbf{E} , \mathbf{H} , \mathbf{B} , \mathbf{D} and \mathbf{j} . To allow a unique determination of the field vectors from a given distribution of currents

and charges, these equations must be supplemented by relations which describe the behaviour of substances under the influence of the field. These relations are known as *material equations** (or *constitutive relations*). In general they are rather complicated; but if the field is time-harmonic (see §1.4.3), and if the bodies are at rest, or in very slow motion relative to each other, and if the material is *isotropic* (i.e. when its physical properties at each point are independent of direction), they take usually the relatively simple form†

$$\mathbf{j} = \sigma \mathbf{E}, \quad (9)$$

$$\mathbf{D} = \varepsilon \mathbf{E}, \quad (10)$$

$$\mathbf{B} = \mu \mathbf{H}. \quad (11)$$

Here σ is called the *specific conductivity*, ε is known as the *dielectric constant* (or *permittivity*) and μ is called the *magnetic permeability*.

Eq. (9) is the differential form of Ohm's law. Substances for which $\sigma \neq 0$ (or more precisely is not negligibly small; the precise meaning of this cannot, however, be discussed here) are called *conductors*. Metals are very good conductors, but there are other classes of good conducting materials such as ionic solutions in liquids and also in solids. In metals the conductivity decreases with increasing temperature. However, in other classes of materials, known as *semiconductors* (e.g. germanium), conductivity increases with temperature over a wide range.

Substances for which σ is negligibly small are called *insulators* or *dielectrics*. Their electric and magnetic properties are then completely determined by ε and μ . For most substances the magnetic permeability μ is practically unity. If this is not the case, i.e. if μ differs appreciably from unity, we say that the substance is *magnetic*. In particular, if $\mu > 1$, the substance is said to be *paramagnetic* (e.g. platinum, oxygen, nitrogen dioxide), while if $\mu < 1$ it is said to be *diamagnetic* (e.g. bismuth, copper, hydrogen, water).

If the fields are exceptionally strong, such as are obtained, for example, by focusing light that is generated by a laser, the right-hand sides of the material equations may have to be supplemented by terms involving components of the field vectors in powers higher than the first.‡

In many cases the quantities σ , ε and μ will be independent of the field strengths; in other cases, however, the behaviour of the material cannot be described in such a simple way. Thus, for example, in a gas of free ions the current, which is determined

* There is an alternative way of describing the behaviour of matter. Instead of the quantities $\varepsilon = D/E$, $\mu = B/H$ one considers the differences $\mathbf{D} - \mathbf{E}$ and $\mathbf{B} - \mathbf{H}$; these have a simpler physical significance and will be discussed in Chapter II.

† The more general relations, applicable also to moving bodies, are studied in the theory of relativity. We shall only need the following result from the more general theory: that in the case of moving charges there is, in addition to the conduction current $\sigma \mathbf{E}$, a convection current $\rho \mathbf{v}$, where \mathbf{v} is the velocity of the moving charges and ρ the charge density (cf. p. 9).

‡ Nonlinear relationship between the displacement vector \mathbf{D} and the electric field \mathbf{E} was first demonstrated in this way by P. A. Franken, A. E. Hill, C. W. Peters and G. Weinrich, *Phys. Rev. Lett.*, **7** (1961), 118.

For systematic treatments of nonlinear effects see N. Bloembergen, *Nonlinear Optics* (New York, W. A. Benjamin, Inc., 1965), P. N. Butcher and D. Cotter, *The Elements of Nonlinear Optics* (Cambridge, Cambridge University Press, 1990) or R. W. Boyd, *Nonlinear Optics* (Boston, Academic Press, 1992).

by the mean speed of the ions, depends, at any moment, not on the instantaneous value of \mathbf{E} , but on all its previous values. Again, in so-called *ferromagnetic* substances (substances which are very highly magnetic, e.g. iron, cobalt and nickel) the value of the magnetic induction \mathbf{B} is determined by the past history of the field \mathbf{H} rather than by its instantaneous value. The substance is then said to exhibit *hysteresis*. A similar history-dependence will be found for the electric displacement in certain dielectric materials. Fortunately hysteretic effects are rarely significant for the high-frequency field encountered in optics.

In the main part of this book we shall study the propagation in substances which light can penetrate without appreciable weakening (e.g. air, glass). Such substances are said to be *transparent* and must be electrical nonconductors ($\sigma = 0$), since conduction implies the evolution of Joule heat (see §1.1.4) and therefore loss of electromagnetic energy. Optical properties of conducting media will be discussed in Chapter XIV.

1.1.3 Boundary conditions at a surface of discontinuity

Maxwell's equations were only stated for regions of space throughout which the physical properties of the medium (characterized by ϵ and μ) are continuous. In optics one often deals with situations in which the properties change abruptly across one or more surfaces. The vectors \mathbf{E} , \mathbf{H} , \mathbf{B} and \mathbf{D} may then be expected also to become discontinuous, while ρ and \mathbf{j} will degenerate into corresponding surface quantities. We shall derive relations describing the transition across such a discontinuity surface.

Let us replace the sharp discontinuity surface T by a thin transition layer within which ϵ and μ vary rapidly but continuously from their values near T on one side to their value near T on the other. Within this layer we construct a small near-cylinder, bounded by a stockade of normals to T ; roofed and floored by small areas δA_1 and δA_2 on each side of T , at constant distance from it, measured along their common normal (Fig. 1.1). Since \mathbf{B} and its derivatives may be assumed to be continuous throughout this cylinder, we may apply Gauss' theorem to the integral of $\text{div } \mathbf{B}$ taken throughout the volume of the cylinder and obtain, from (4),

$$\int \text{div } \mathbf{B} dV = \int \mathbf{B} \cdot \mathbf{n} dS = 0; \quad (12)$$

the second integral is taken over the surface of the cylinder, and \mathbf{n} is the unit outward normal.

Since the areas δA_1 and δA_2 are assumed to be small, \mathbf{B} may be considered to have constants values $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ on δA_1 and δA_2 , and (12) may then be replaced by

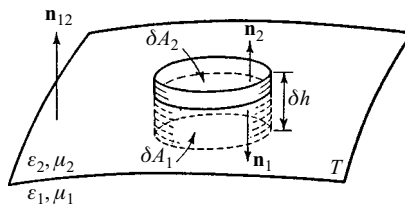


Fig. 1.1 Derivation of boundary conditions for the normal components of \mathbf{B} and \mathbf{D} .

$$\mathbf{B}^{(1)} \cdot \mathbf{n}_1 \delta A_1 + \mathbf{B}^{(2)} \cdot \mathbf{n}_2 \delta A_2 + \text{contribution from walls} = 0. \quad (13)$$

If the height δh of the cylinder decreases towards zero, the transition layer shrinks into the surface and the contribution from the walls of the cylinder tends to zero, provided that there is no surface flux of magnetic induction. Such flux never occurs, and consequently in the limit,

$$(\mathbf{B}^{(1)} \cdot \mathbf{n}_1 + \mathbf{B}^{(2)} \cdot \mathbf{n}_2) \delta A = 0, \quad (14)$$

δA being the area in which the cylinder intersects T . If \mathbf{n}_{12} is the unit normal pointing from the first into the second medium, then $\mathbf{n}_1 = -\mathbf{n}_{12}$, $\mathbf{n}_2 = \mathbf{n}_{12}$ and (14) gives

$$\mathbf{n}_{12} \cdot (\mathbf{B}^{(2)} - \mathbf{B}^{(1)}) = 0, \quad (15)$$

i.e. *the normal component of the magnetic induction is continuous across the surface of discontinuity.*

The electric displacement \mathbf{D} may be treated in a similar way, but there will be an additional term if charges are present. In place of (12) we now have from (3)

$$\int \text{div } \mathbf{D} \, dV = \int \mathbf{D} \cdot \mathbf{n} \, dS = 4\pi \int \rho \, dV. \quad (16)$$

As the areas δA_1 and δA_2 shrink together, the total charge remains finite, so that the volume density becomes infinite. Instead of the volume charge density ρ the concept of *surface charge density* $\hat{\rho}$ must then be used. It is defined by*

$$\lim_{\delta h \rightarrow 0} \int \rho \, dV = \int \hat{\rho} \, dA. \quad (17)$$

We shall also need later the concept of *surface current density* $\hat{\mathbf{j}}$, defined in a similar way:

$$\lim_{\delta h \rightarrow 0} \int \mathbf{j} \, dV = \int \hat{\mathbf{j}} \, dA. \quad (18)$$

If the area δA and the height δh are taken sufficiently small, (16) gives

$$\mathbf{D}^{(1)} \cdot \mathbf{n}_1 \delta A_1 + \mathbf{D}^{(2)} \cdot \mathbf{n}_2 \delta A_2 + \text{contribution from walls} = 4\pi \hat{\rho} \delta A.$$

The contribution from the walls tends to zero with δh , and we therefore obtain in the limit as $\delta h \rightarrow 0$,

$$\mathbf{n}_{12} \cdot (\mathbf{D}^{(2)} - \mathbf{D}^{(1)}) = 4\pi \hat{\rho}, \quad (19)$$

i.e. *in the presence of a layer of surface charge density $\hat{\rho}$ on the surface, the normal component of the electric displacement changes abruptly across the surface, by an amount equal to $4\pi \hat{\rho}$.*

* For later purposes we note a representation of the surface charge density and the surface current density in terms of the Dirac delta function (see Appendix IV). If the equation of the surface of discontinuity is $F(x, y, z) = 0$, then

$$\rho = \hat{\rho} |\text{grad } F| \delta(F), \quad (17a)$$

$$\mathbf{j} = \hat{\mathbf{j}} |\text{grad } F| \delta(F). \quad (18a)$$

These relations can immediately be verified by substituting into (17) and (18) and using the relation $dF = |\text{grad } F| dh$ and the sifting property of the delta function.

Next, we examine the behaviour of the tangential components. Let us replace the sharp discontinuity surface by a continuous transition layer. We also replace the cylinder of Fig. 1.1 by a 'rectangular' area with sides parallel and perpendicular to T (Fig. 1.2).

Let \mathbf{b} be the unit vector perpendicular to the plane of the rectangle. Then it follows from (2) and from Stokes' theorem that

$$\int \text{curl } \mathbf{E} \cdot \mathbf{b} \, dS = \int \mathbf{E} \cdot d\mathbf{r} = -\frac{1}{c} \int \dot{\mathbf{B}} \cdot \mathbf{b} \, dS, \quad (20)$$

the first and third integrals being taken throughout the area of the rectangle, and the second along its boundary. If the lengths P_1Q_1 ($= \delta s_1$), and P_2Q_2 ($= \delta s_2$) are small, \mathbf{E} may be replaced by constant values $\mathbf{E}^{(1)}$ and $\mathbf{E}^{(2)}$ along each of these segments. Similarly $\dot{\mathbf{B}}$ may be replaced by a constant value. Eq. (20) then gives

$$\mathbf{E}^{(1)} \cdot \mathbf{t}_1 \delta s_1 + \mathbf{E}^{(2)} \cdot \mathbf{t}_2 \delta s_2 + \text{contribution from ends} = -\frac{1}{c} \dot{\mathbf{B}} \cdot \mathbf{b} \delta s \delta h, \quad (21)$$

where δs is the line element in which the rectangle intersects the surface. If now the height of the rectangle is gradually decreased, the contribution from the ends P_1P_2 and Q_1Q_2 will tend to zero, provided that \mathbf{E} does not in the limit acquire sufficiently sharp singularities; this possibility will be excluded. Assuming also that $\dot{\mathbf{B}}$ remains finite, we obtain in the limit as $\delta h \rightarrow 0$,

$$(\mathbf{E}^{(1)} \cdot \mathbf{t}_1 + \mathbf{E}^{(2)} \cdot \mathbf{t}_2) \delta s = 0. \quad (22)$$

If \mathbf{t} is the unit tangent along the surface, then (see Fig. 1.2) $\mathbf{t}_1 = -\mathbf{t} = -\mathbf{b} \times \mathbf{n}_{12}$, $\mathbf{t}_2 = \mathbf{t} = \mathbf{b} \times \mathbf{n}_{12}$, and (22) gives

$$\mathbf{b} \cdot [\mathbf{n}_{12} \times (\mathbf{E}^{(2)} - \mathbf{E}^{(1)})] = 0.$$

Since the orientation of the rectangle and consequently that of the unit vector \mathbf{b} is arbitrary, it follows that

$$\mathbf{n}_{12} \times (\mathbf{E}^{(2)} - \mathbf{E}^{(1)}) = 0, \quad (23)$$

i.e. *the tangential component of the electric vector is continuous across the surface.*

Finally consider the behaviour of the tangential component of the magnetic vector. The analysis is similar, but there is an additional term if currents are present. In place of (21) we now have

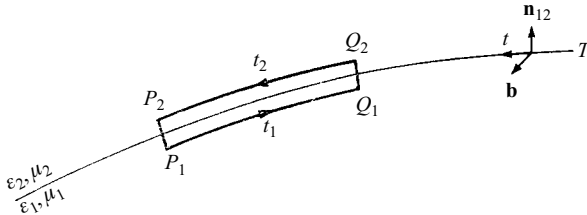


Fig. 1.2 Derivation of boundary conditions for the tangential components of \mathbf{E} and \mathbf{H} .

$$\mathbf{H}^{(1)} \cdot \mathbf{t}_1 \delta s_1 + \mathbf{H}^{(2)} \cdot \mathbf{t}_2 \delta s_2 + \text{contribution from ends} = \frac{1}{c} \dot{\mathbf{D}} \cdot \mathbf{b} \delta s \delta h + \frac{4\pi}{c} \hat{\mathbf{j}} \cdot \mathbf{b} \delta s. \quad (24)$$

On proceeding to the limit $\delta h \rightarrow 0$ as before, we obtain

$$\mathbf{n}_{12} \times (\mathbf{H}^{(2)} - \mathbf{H}^{(1)}) = \frac{4\pi}{c} \hat{\mathbf{j}}. \quad (25)$$

From (25) it follows that *in the presence of a surface current of density $\hat{\mathbf{j}}$, the tangential component (considered as a vector quantity) of the magnetic vector changes abruptly, its discontinuity being $(4\pi/c)\hat{\mathbf{j}} \times \mathbf{n}_{12}$.*

Apart from discontinuities due to the abrupt changes in the physical properties of the medium, the field vectors may also be discontinuous because of the presence of a source which begins to radiate at a particular instant of time $t = t_0$. The disturbance then spreads into the surrounding space, and at any later instant $t_1 > t_0$ will have filled a well-defined region. Across the (moving) boundary of this region, the field vectors will change abruptly from finite values on the boundary to the value zero outside it.

The various cases of discontinuity may be covered by rewriting Maxwell's equations in an integral form.* The general discontinuity conditions may also be written in the form of simple difference equations; a derivation of these equations is given in Appendix VI.

1.1.4 The energy law of the electromagnetic field

Electromagnetic theory interprets the light intensity as the energy flux of the field. It is therefore necessary to recall the energy law of Maxwell's theory.

From (1) and (2) it follows that

$$\mathbf{E} \cdot \text{curl } \mathbf{H} - \mathbf{H} \cdot \text{curl } \mathbf{E} = \frac{4\pi}{c} \mathbf{j} \cdot \mathbf{E} + \frac{1}{c} \mathbf{E} \cdot \dot{\mathbf{D}} + \frac{1}{c} \mathbf{H} \cdot \dot{\mathbf{B}}. \quad (26)$$

Also, by a well-known vector identity, the term on the left may be expressed as the divergence of the vector product of \mathbf{H} and \mathbf{E} :

$$\mathbf{E} \cdot \text{curl } \mathbf{H} - \mathbf{H} \cdot \text{curl } \mathbf{E} = -\text{div}(\mathbf{E} \times \mathbf{H}). \quad (27)$$

From (26) and (27) we have that

$$\frac{1}{c} (\mathbf{E} \cdot \dot{\mathbf{D}} + \mathbf{H} \cdot \dot{\mathbf{B}}) + \frac{4\pi}{c} \mathbf{j} \cdot \mathbf{E} + \text{div}(\mathbf{E} \times \mathbf{H}) = 0. \quad (28)$$

When we multiply this equation by $c/4\pi$, integrate throughout an arbitrary volume and apply Gauss's theorem, this gives

$$\frac{1}{4\pi} \int (\mathbf{E} \cdot \dot{\mathbf{D}} + \mathbf{H} \cdot \dot{\mathbf{B}}) dV + \int \mathbf{j} \cdot \mathbf{E} dV + \frac{c}{4\pi} \int (\mathbf{E} \times \mathbf{H}) \cdot \mathbf{n} dS = 0, \quad (29)$$

where the last integral is taken over the boundary of the volume, \mathbf{n} being the unit outward normal.

The relation (29) is a direct consequence of Maxwell's equations and is therefore

* See, for example, A. Sommerfeld, *Electrodynamics* (New York, Academic Press, 1952), p. 11; or J. A. Stratton, *Electromagnetic Theory* (New York, McGraw-Hill, 1941), p. 6.

valid whether or not the material equations (9)–(11) hold. It represents, as will be seen, the *energy law* of an electromagnetic field. We shall discuss it here only for the case where the material equations (9)–(11) are satisfied. Generalizations to anisotropic media, where the material equations are of a more complicated form, will be considered later (Chapter XV).

We have, on using the material equations,

$$\left. \begin{aligned} \frac{1}{4\pi}(\mathbf{E} \cdot \dot{\mathbf{D}}) &= \frac{1}{4\pi} \mathbf{E} \cdot \frac{\partial}{\partial t}(\epsilon \mathbf{E}) = \frac{1}{8\pi} \frac{\partial}{\partial t}(\epsilon \mathbf{E}^2) = \frac{1}{8\pi} \frac{\partial}{\partial t}(\mathbf{E} \cdot \mathbf{D}), \\ \frac{1}{4\pi}(\mathbf{H} \cdot \dot{\mathbf{B}}) &= \frac{1}{4\pi} \mathbf{H} \cdot \frac{\partial}{\partial t}(\mu \mathbf{H}) = \frac{1}{8\pi} \frac{\partial}{\partial t}(\mu \mathbf{H}^2) = \frac{1}{8\pi} \frac{\partial}{\partial t}(\mathbf{H} \cdot \mathbf{B}). \end{aligned} \right\} \quad (30)$$

Setting

$$w_e = \frac{1}{8\pi} \mathbf{E} \cdot \mathbf{D}, \quad w_m = \frac{1}{8\pi} \mathbf{H} \cdot \mathbf{B}, \quad (31)$$

and

$$W = \int (w_e + w_m) dV, \quad (32)$$

(29) becomes,

$$\frac{dW}{dt} + \int \mathbf{j} \cdot \mathbf{E} dV + \frac{c}{4\pi} \int (\mathbf{E} \times \mathbf{H}) \cdot \mathbf{n} dS = 0. \quad (33)$$

We shall show that W represents the total energy contained within the volume, so that w_e may be identified with the *electric energy density* and w_m with the *magnetic energy density* of the field.*

To justify the interpretation of W as the total energy we have to show that, for a closed system (i.e. one in which the field on the boundary surface may be neglected), the change in W as defined above is due to the work done by the field on the material charged bodies which are embedded in it. It suffices to do this for slow motion of the material bodies, which themselves may be assumed to be so small that they can be regarded as point charges e_k ($k = 1, 2, \dots$). Let the velocity of the charge e_k be \mathbf{v}_k ($|\mathbf{v}_k| \ll c$).

The force exerted by a field (\mathbf{E}, \mathbf{B}) on a charge e moving with velocity \mathbf{v} is given by the so-called *Lorentz law*,

$$\mathbf{F} = e \left(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{B} \right), \quad (34)$$

which is based on experience. It follows that if all the charges e_k are displaced by $\delta \mathbf{x}_k$ ($k = 1, 2, \dots$) in time δt , the total work done is

* In the general case the densities are defined by the expressions

$$w_e = \frac{1}{4\pi} \int \mathbf{E} \cdot d\mathbf{D}, \quad w_m = \frac{1}{4\pi} \int \mathbf{H} \cdot d\mathbf{B}.$$

When the relationship between \mathbf{E} and \mathbf{D} and between \mathbf{H} and \mathbf{B} is linear, as here assumed, these expressions reduce to (31).

$$\begin{aligned}\delta A &= \sum_k \mathbf{F}_k \cdot \delta \mathbf{x}_k = \sum_k e_k \left(\mathbf{E}_k + \frac{1}{c} \mathbf{v}_k \times \mathbf{B} \right) \cdot \delta \mathbf{x}_k \\ &= \sum_k e_k \mathbf{E}_k \cdot \delta \mathbf{x}_k = \sum_k e_k \mathbf{E}_k \cdot \mathbf{v}_k \delta t,\end{aligned}$$

since $\delta \mathbf{x}_k = \mathbf{v}_k \delta t$. If the number of charged particles is large, we can consider the distribution to be continuous. We introduce the charge density ρ (i.e. total charge per unit volume) and the last equation becomes

$$\delta A = \delta t \int \rho \mathbf{v} \cdot \mathbf{E} dV, \quad (35)$$

the integration being carried throughout an arbitrary volume. Now the velocity \mathbf{v} does not appear explicitly in Maxwell's equations, but it may be introduced by using an experimental result found by Röntgen*, according to which a *convection current* (i.e. a set of moving charges) has the same electromagnetic effect as a *conduction current* in a wire. Hence the current density \mathbf{j} appearing in Maxwell's equations can be split into two parts

$$\mathbf{j} = \mathbf{j}_c + \mathbf{j}_v, \quad (36)$$

where

$$\mathbf{j}_c = \sigma \mathbf{E}$$

is the conduction current density, and

$$\mathbf{j}_v = \rho \mathbf{v}$$

represents the convection current density. (35) may therefore be written as

$$\delta A = \delta t \int \mathbf{j}_v \cdot \mathbf{E} dV. \quad (37)$$

Let us now define a vector \mathbf{S} and a scalar Q by the relations

$$\mathbf{S} = \frac{c}{4\pi} (\mathbf{E} \times \mathbf{H}), \quad (38)$$

$$Q = \int \mathbf{j}_c \cdot \mathbf{E} dV = \int \sigma \mathbf{E}^2 dV. \quad (39)$$

Then by (35) and (36)

$$\begin{aligned}\int \mathbf{j} \cdot \mathbf{E} dV &= Q + \int \mathbf{j}_v \cdot \mathbf{E} dV \\ &= Q + \frac{\delta A}{\delta t},\end{aligned} \quad (40)$$

where the second function is not, of course, a total derivative of a space-time function. Eq. (33) now takes the form

* W. C. Röntgen, *Ann. d. Physik*, **35** (1888), 264; **40** (1890), 93.

$$\frac{dW}{dt} = -\frac{\delta A}{\delta t} - Q - \int \mathbf{S} \cdot \mathbf{n} dS. \quad (41)$$

For a nonconductor ($\sigma = 0$) we have that $Q = 0$. Assume also that the boundary surface is so far away that we can neglect the field on it, due to the electromagnetic processes inside; then $\int \mathbf{S} \cdot \mathbf{n} dS = 0$, and integration of (41) gives

$$W + A = \text{constant}. \quad (42)$$

Hence, for an isolated system, the increase of W per unit time is due to the work done on the system during this time. This result justifies our definition of electromagnetic energy by means of (32).

The term Q represents the resistive dissipation of energy (called *Joule's heat*) in a conductor ($\sigma \neq 0$). According to (41) there is a further decrease in energy if the field extends to the boundary surface. The surface integral must therefore represent the flow of energy across this boundary surface. The vector \mathbf{S} is known as the *Poynting vector* and represents the amount of energy which crosses per second a unit area normal to the directions of \mathbf{E} and \mathbf{H} .

It should be noted that the interpretation of \mathbf{S} as energy flow (more precisely as the density of the flow) is an abstraction which introduces a certain degree of arbitrariness. For the quantity which is physically significant is, according to (41), not \mathbf{S} itself, but the integral of $\mathbf{S} \cdot \mathbf{n}$ taken over a closed surface. Clearly, from the value of the integral, no unambiguous conclusion can be drawn about the detailed distribution of \mathbf{S} , and alternative definitions of the energy flux density are therefore possible. One can always add to \mathbf{S} the curl of an arbitrary vector, since such a term will not contribute to the surface integral as can be seen from Gauss' theorem and the identity $\text{div curl} \equiv 0$.* However, when the definition has been applied cautiously, in particular for averages over small but finite regions of space or time, no contradictions with experiments have been found. We shall therefore accept the above definition in terms of the Poynting vector of the density of the energy flow.

Finally we note that in a nonconducting medium ($\sigma = 0$) where no mechanical work is done ($A = 0$), the energy law may be written in the form of a hydrodynamical continuity equation for noncompressible fluids:

$$\frac{\partial w}{\partial t} + \text{div } \mathbf{S} = 0, \quad (w = w_e + w_m). \quad (43)$$

A description of propagation of light in terms of a hydrodynamical model is often helpful, particularly in the domain of geometrical optics and in connection with scalar diffraction fields, as it gives a picture of the energy transport in a simple and graphic manner. In optics, the (averaged) Poynting vector is the chief quantity of interest. The magnitude of the Poynting vector is a measure of the light intensity, and its direction represents the direction of propagation of the light.

* According to modern theories of fields the arbitrariness is even greater, allowing for alternative expressions for both the energy density and the energy flux, but consistent with the change of the Lagrangian density of the field by the addition of a four-divergence. For a discussion of this subject see, for example, G. Wentzel, *Quantum Theory of Fields* (New York, Interscience Publishers, 1949), especially §2 or J. D. Jackson, *Classical Electrodynamics* (New York, J. Wiley and Sons, 2nd ed. 1975), Sec. 12.10, especially p. 602.

1.2 The wave equation and the velocity of light

Maxwell's equations relate the field vectors by means of simultaneous differential equations. On elimination we obtain differential equations which each of the vectors must separately satisfy. We shall confine our attention to that part of the field which contains no charges or currents, i.e. where $\mathbf{j} = 0$ and $\rho = 0$.

We substitute for \mathbf{B} from the material equation §1.1 (11) into the second Maxwell equation §1.1 (2), divide both sides by μ and apply the operator curl. This gives

$$\text{curl} \left(\frac{1}{\mu} \text{curl} \mathbf{E} \right) + \frac{1}{c} \text{curl} \dot{\mathbf{H}} = 0. \quad (1)$$

Next we differentiate the first Maxwell equation §1.1 (1) with respect to time, use the material equation §1.1 (10) for \mathbf{D} , and eliminate curl $\dot{\mathbf{H}}$ between the resulting equation and (1); this gives

$$\text{curl} \left(\frac{1}{\mu} \text{curl} \mathbf{E} \right) + \frac{\varepsilon}{c^2} \ddot{\mathbf{E}} = 0. \quad (2)$$

If we use the identities $\text{curl} uv = u \text{curl} \mathbf{v} + (\text{grad} u) \times \mathbf{v}$ and $\text{curl} \text{curl} = \text{grad} \text{div} - \nabla^2$, (2) becomes

$$\nabla^2 \mathbf{E} - \frac{\varepsilon \mu}{c^2} \ddot{\mathbf{E}} + (\text{grad} \ln \mu) \times \text{curl} \mathbf{E} - \text{grad} \text{div} \mathbf{E} = 0. \quad (3)$$

Also from §1.1 (3), using again the material equation for \mathbf{D} and applying the identity $\text{div} uv = u \text{div} \mathbf{v} + \mathbf{v} \cdot \text{grad} u$ we find

$$\varepsilon \text{div} \mathbf{E} + \mathbf{E} \cdot \text{grad} \varepsilon = 0. \quad (4)$$

Hence (3) may be written in the form

$$\nabla^2 \mathbf{E} - \frac{\varepsilon \mu}{c^2} \ddot{\mathbf{E}} + (\text{grad} \ln \mu) \times \text{curl} \mathbf{E} + \text{grad} (\mathbf{E} \cdot \text{grad} \ln \varepsilon) = 0. \quad (5)$$

In a similar way we obtain an equation for \mathbf{H} alone:

$$\nabla^2 \mathbf{H} - \frac{\varepsilon \mu}{c^2} \ddot{\mathbf{H}} + (\text{grad} \ln \varepsilon) \times \text{curl} \mathbf{H} + \text{grad} (\mathbf{H} \cdot \text{grad} \ln \mu) = 0. \quad (6)$$

In particular, if the medium is homogeneous, $\text{grad} \log \varepsilon = \text{grad} \ln \mu = 0$, and (5) and (6) reduce to

$$\nabla^2 \mathbf{E} - \frac{\varepsilon \mu}{c^2} \ddot{\mathbf{E}} = 0, \quad \nabla^2 \mathbf{H} - \frac{\varepsilon \mu}{c^2} \ddot{\mathbf{H}} = 0. \quad (7)$$

These are standard equations of wave motion and suggest the existence of electromagnetic waves propagated with a velocity*

$$v = c / \sqrt{\varepsilon \mu}. \quad (8)$$

* The concept of a velocity of an electromagnetic wave has actually an unambiguous meaning only in connection with waves of very simple kind, e.g. plane waves. That v does not represent the velocity of propagation of an arbitrary solution of (7) is obvious if we bear in mind that these equations also admit standing waves as solutions.

In this introductory section it is assumed that the reader is familiar with the concept of a plane wave, and we regard v as the velocity with which such a wave advances. The mathematical representation of a plane wave will be discussed in §1.3 and §1.4.

The constant c was first determined by R. Kohlrausch and W. Weber in 1856 from the ratio of the values of the capacity of a condenser measured in electrostatic and electromagnetic units, and it was found to be identical with the velocity of light in free space. Using this result, Maxwell developed his electromagnetic theory of light, predicting the existence of electromagnetic waves; the correctness of this prediction was confirmed by the celebrated experiments of H. Hertz (see Historical introduction).

As in all wave theories of light, the elementary process which produces the optical impression is regarded as being a harmonic wave in space-time (studied in its simplest form in §1.3 and §1.4). If its frequency is in the range from $4 \times 10^{14} \text{ s}^{-1}$ to $7.5 \times 10^{14} \text{ s}^{-1}$ (approximately) it gives rise to the psychological impression of a definite colour. (The opposite, however, is not true: coloured light of a certain subjective quality may be a composition of harmonic waves of very different frequency distributions.) The actual connection between colour and frequency is very involved and will not be studied in this book.*

The first determination of the velocity of light† was made by Römer in 1675 from observations of the eclipses of the first satellite of Jupiter and later in a different way (from aberration of fixed stars) by Bradley (1728).

The first measurements of the velocity of light from terrestrial sources were carried out by Fizeau in 1849. It is necessary to employ a modulator, which marks off a portion of the beam‡ and for this purpose Fizeau used a rotation wheel. Later methods employed rotating mirrors or electronic shutters. The rotating mirror method was suggested by Wheatstone in 1834 and was used by Foucault in 1860. It was later systematically developed over a period of many years by Michelson. The average value based on about 200 measurements by Michelson gave c as 299,796 km/s. An optical shutter method employing a Kerr cell was developed by Karolus and Mittelstaedt (1928), Anderson (1937) and Hüttel (1940). The values of c obtained from these measurements are in excellent agreement with those based on indirect methods, such as determinations from the ratio of an electric charge measured in electrostatic and electromagnetic units; for example Rosa and Dorsey (1907) in this way found c as 299,784 km/s. Measurements of the velocity of electromagnetic waves on wires carried out by Mercier (1923) gave the value of c equal to 299,782 km/s. The value adopted by the Fifteenth General Conference of Weights and Measures§ is

$$c = 299,792.458 \text{ km/s.} \quad (9)$$

The close agreement between the values of c obtained from measurements of very different kinds (and in some cases using radiation whose frequencies differ by a factor of hundreds of thousands from those used in the optical measurements) gives a striking confirmation of Maxwell's theory.

The dielectric constant ϵ is usually greater than unity, and μ is practically equal to

* The sensitivity of the human eye to different colours is, however, briefly discussed in §4.8.1.

† For a description of the methods used for determination of the velocity of light, see for example, E. Bergstrand, *Encyclopedia of Physics*, ed. S. Flügge, Vol. 24 (Berlin, Springer, 1956), p. 1.

Detailed analysis of the results obtained by different methods is also given by R. T. Birge in *Rep. Progr. Phys.* (London, The Physical Society), **8** (1941), 90.

‡ Such determinations give essentially the group velocity (see §1.3.4). The difference between the group velocity and the phase velocity in air at standard temperature and pressure is about 1 part in 50,000.

§ Conférence Générale des Poids et Mesures, XV, Paris, 1975, Comptes Rendus des Séances (Paris, Bureau International des Poids et Mesures, 1976).

unity for transparent substances, so that the velocity v is then according to (8) smaller than the vacuum velocity c . This conclusion was first demonstrated experimentally for propagation of light in water in 1850 by Foucault and Fizeau.

The value of v is not as a rule determined directly, but only relative to c , with the help of the law of refraction. According to this law, if a plane electromagnetic wave falls on to a plane boundary between two homogeneous media, the sine of the angle θ_1 between the normal to the incident wave and the normal to the surface bears a constant ratio to the sine of the angle θ_2 between the normal of the refracted wave and the surface normal (Fig. 1.3), this constant ratio being equal to the ratio of the velocities v_1 and v_2 of propagation in the two media:

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{v_1}{v_2}. \quad (10)$$

This result will be derived in §1.5. Here we only note that it is equivalent to the assumption that the wave-front, though it has a kink at the boundary, is continuous, so that the line of intersection between the incident wave and the boundary travels at the same speed (v' , say) as the line of intersection between the refracted wave and the boundary. We then have

$$v_1 = v' \sin \theta_1, \quad v_2 = v' \sin \theta_2, \quad (11)$$

from which, on elimination of v' , (10) follows. This argument, in a slightly more elaborate form, is often given as an illustration of Huygens' construction (§3.3).

The value of the constant ratio in (10) is usually denoted by n_{12} and is called the *refractive index*, for refraction from the first into the second medium. We also define an '*absolute refractive index*' n of a medium; it is the refractive index for refraction from vacuum into that medium,

$$n = \frac{c}{v}. \quad (12)$$

If n_1 and n_2 are the absolute refractive indices of two media, the (relative) refractive index n_{12} for refraction from the first into the second medium then is

$$n_{12} = \frac{n_2}{n_1} = \frac{v_1}{v_2}. \quad (13)$$

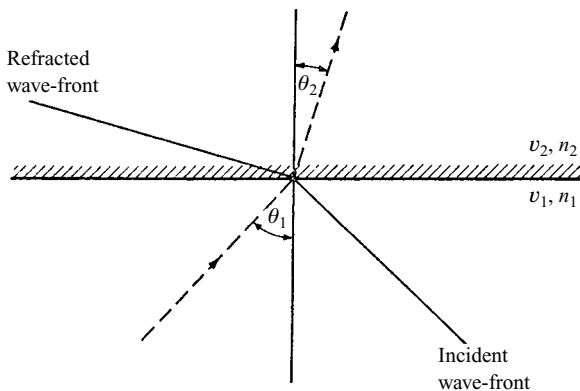


Fig. 1.3 Illustrating the refraction of a plane wave.

Table 1.1. *Refractive indices and static dielectric constants of certain gases*

	n (yellow light)	$\sqrt{\epsilon}$
Air	1.000294	1.000295
Hydrogen H ₂	1.000138	1.000132
Carbon dioxide CO ₂	1.000449	1.000473
Carbon monoxide CO	1.000340	1.000345

Table 1.2. *Refractive indices and static dielectric constants of certain liquids*

	n (yellow light)	$\sqrt{\epsilon}$
Methyl alcohol CH ₃ OH	1.34	5.7
Ethyl alcohol C ₂ H ₅ OH	1.36	5.0
Water H ₂ O	1.33	9.0

Comparison of (12) and (8) gives Maxwell's formula:

$$n = \sqrt{\epsilon\mu}. \quad (14)$$

Since for all substances with which we shall be concerned, μ is effectively unity (nonmagnetic substances), the refractive index should then be equal to the square root of the dielectric constant, which has been assumed to be a constant of the material. On the other hand, well-known experiments on prismatic colours, first carried out by Newton, show that the index of refraction depends on the colour, i.e. on the frequency of the light. If we are to retain Maxwell's formula, it must be supposed that ϵ is not a constant characteristic of the material, but is a function of the frequency of the field. The dependence of ϵ on frequency can only be treated by taking into account the atomic structure of matter, and will be briefly discussed in §2.3.

Maxwell's formula (with ϵ equal to the static dielectric constant) gives a good approximation for such substances as gases with a simple chemical structure which do not disperse light substantially, i.e. for those whose optical properties do not strongly depend on the colour of the light. Results of some early measurements for such gases, carried out by L. Boltzmann,* are given in Table 1.1. Eq. (14) also gives a good approximation for liquid hydrocarbons; for example benzene C₆H₆ has $n = 1.482$ for yellow light whilst $\sqrt{\epsilon} = 1.489$. On the other hand, there is a strong deviation from the formula for many solid bodies (e.g. glasses), and for some liquids, as illustrated in Table 1.2.

1.3 Scalar waves

In a homogeneous medium in regions free of currents and charges, each rectangular component $V(\mathbf{r}, t)$ of the field vectors satisfies, according to §1.2 (7), the homogeneous wave equation

* L. Boltzmann, *Wien. Ber.*, **69** (1874), 795; *Pogg. Ann.*, **155** (1875), 403; *Wiss. Abh. Physik-techn. Reichsanst.*, **1**, Nr. 26, 537.

$$\nabla^2 V - \frac{1}{v^2} \frac{\partial^2 V}{\partial t^2} = 0. \quad (1)$$

We shall now briefly examine the simplest solution of this equation.

1.3.1 Plane waves

Let $\mathbf{r}(x, y, z)$ be a position vector of a point P in space and $\mathbf{s}(s_x, s_y, s_z)$ a unit vector in a fixed direction. Any solution of (1) of the form

$$V = V(\mathbf{r} \cdot \mathbf{s}, t) \quad (2)$$

is said to represent a *plane wave*, since at each instant of time V is constant over each of the planes

$$\mathbf{r} \cdot \mathbf{s} = \text{constant}$$

which are perpendicular to the unit vector \mathbf{s} .

It will be convenient to choose a new set of Cartesian axes $O\xi, O\eta, O\zeta$ with $O\zeta$ in the direction of \mathbf{s} . Then (see Fig. 1.4)

$$\mathbf{r} \cdot \mathbf{s} = \zeta, \quad (3)$$

and one has

$$\frac{\partial}{\partial x} = s_x \frac{\partial}{\partial \zeta}, \quad \frac{\partial}{\partial y} = s_y \frac{\partial}{\partial \zeta}, \quad \frac{\partial}{\partial z} = s_z \frac{\partial}{\partial \zeta}.$$

From these relations one easily finds that

$$\nabla^2 V = \frac{\partial^2 V}{\partial \zeta^2}, \quad (4)$$

so that (1) becomes

$$\frac{\partial^2 V}{\partial \zeta^2} - \frac{1}{v^2} \frac{\partial^2 V}{\partial t^2} = 0. \quad (5)$$

If we set

$$\zeta - vt = p, \quad \zeta + vt = q, \quad (6)$$

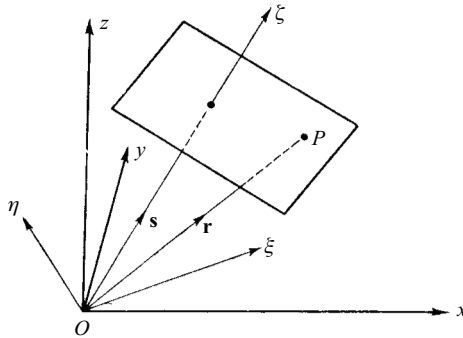


Fig. 1.4 Propagation of a plane wave.

(5) takes the form

$$\frac{\partial^2 V}{\partial p \partial q} = 0. \quad (7)$$

The general solution of this equation is

$$\begin{aligned} V &= V_1(p) + V_2(q) \\ &= V_1(\mathbf{r} \cdot \mathbf{s} - vt) + V_2(\mathbf{r} \cdot \mathbf{s} + vt), \end{aligned} \quad (8)$$

where V_1 and V_2 are arbitrary functions.

We see that the argument of V_1 is unchanged when (ζ, t) is replaced by $(\zeta + v\tau, t + \tau)$, where τ is arbitrary. Hence V_1 represents a disturbance which is propagated with velocity v in the positive ζ direction. Similarly $V_2(\zeta + vt)$ represents a disturbance which is propagated with velocity v in the negative ζ direction.

1.3.2 Spherical waves

Next we consider solutions representing spherical waves, i.e. solutions of the form

$$V = V(r, t), \quad (9)$$

where $r = |\mathbf{r}| = \sqrt{x^2 + y^2 + z^2}$.

Using the relations $\partial/\partial x = (\partial r/\partial x)(\partial/\partial r) = (x/r)(\partial/\partial r)$, etc., one finds after a straightforward calculation that

$$\nabla^2 V = \frac{1}{r} \frac{\partial^2}{\partial r^2} (rV), \quad (10)$$

so that the wave equation (1) now becomes

$$\frac{\partial^2}{\partial r^2} (rV) - \frac{1}{v^2} \frac{\partial^2}{\partial t^2} (rV) = 0. \quad (11)$$

Now this equation is identical with (5), if ζ is replaced in the latter by r and V by rV . Hence the solution of (11) can immediately be written down from (8):

$$V = \frac{V_1(r - vt)}{r} + \frac{V_2(r + vt)}{r}, \quad (12)$$

V_1 and V_2 being again arbitrary functions. The first term on the right-hand side of (12) represents a spherical wave diverging from the origin, the second a spherical wave converging towards the origin, the velocity of propagation being v in both cases.

1.3.3 Harmonic waves. The phase velocity

At a point \mathbf{r}_0 in space the wave disturbance is a function of time only:

$$V(\mathbf{r}_0, t) = F(t). \quad (13)$$

As will be evident from our earlier remarks about colour, the case when F is periodic is of particular interest. Accordingly we consider the case when F has the form

$$F(t) = a \cos(\omega t + \delta). \quad (14)$$

Here $a (> 0)$ is called the *amplitude*, and the argument $\omega t + \delta$ of the cosine term is called the *phase*. The quantity

$$\nu = \frac{\omega}{2\pi} = \frac{1}{T} \quad (15)$$

is called the *frequency* and represents the number of vibrations per second. ω is called the *angular frequency* and gives the number of vibrations in 2π seconds. Since F remains unchanged when t is replaced by $t + T$, T is the *period* of the vibrations. Wave functions (i.e. solutions of the wave equation) of the form (14) are said to be *harmonic* with respect to time.

Let us first consider a wave function which represents a *harmonic plane wave* propagated in the direction specified by a unit vector \mathbf{s} . According to §1.3.1 it is obtained on replacing t by $t - \mathbf{r} \cdot \mathbf{s}/v$ in (14):

$$V(\mathbf{r}, t) = a \cos \left[\omega \left(t - \frac{\mathbf{r} \cdot \mathbf{s}}{v} \right) + \delta \right]. \quad (16)$$

Eq. (16) remains unchanged when $\mathbf{r} \cdot \mathbf{s}$ is replaced by $\mathbf{r} \cdot \mathbf{s} + \lambda$, where

$$\lambda = v \frac{2\pi}{\omega} = vT. \quad (17)$$

The length λ is called the *wavelength*. It is also useful to define a *reduced wavelength* λ_0 as

$$\lambda_0 = cT = n\lambda; \quad (18)$$

this is the wavelength which corresponds to a harmonic wave of the same frequency propagated *in vacuo*. In spectroscopy one uses also the concept of a *wave number** κ , which is defined as the number of wavelengths *in vacuo*, per unit of length (cm):

$$\kappa = \frac{1}{\lambda_0} = \frac{\nu}{c}. \quad (19)$$

It is also convenient to define vectors \mathbf{k}_0 and \mathbf{k} in the direction \mathbf{s} of propagation, whose lengths are respectively

$$k_0 = 2\pi\kappa = \frac{2\pi}{\lambda_0} = \frac{\omega}{c}, \quad (20)$$

and

$$k = nk_0 = \frac{2\pi}{\lambda} = \frac{n\omega}{c} = \frac{\omega}{v}. \quad (21)$$

The vector $\mathbf{k} = k\mathbf{s}$ is called the *wave vector* or the *propagation vector* in the medium, $\mathbf{k}_0 = k_0\mathbf{s}$ being the corresponding vector in the vacuum.

Instead of the constant δ one also uses the concept of *path length* l , which is the distance through which a wave-front recedes when the phase increases by δ :

$$l = \frac{v}{\omega} \delta = \frac{\lambda}{2\pi} \delta = \frac{\lambda_0}{2\pi n} \delta. \quad (22)$$

* We shall refer to κ as the 'spectroscopic wave number' and reserve the term 'wave number' for k_0 or k , defined by (20) and (21), as customary in optics.

Let us now consider time-harmonic waves of more complicated form. A general time-harmonic, real, scalar wave of frequency ω may be defined as a real solution of the wave equation, of the form

$$V(\mathbf{r}, t) = a(\mathbf{r}) \cos[\omega t - g(\mathbf{r})], \quad (23)$$

$a (> 0)$ and g being real scalar functions of positions. The surfaces

$$g(\mathbf{r}) = \text{constant} \quad (24)$$

are called *cophasal surfaces* or *wave surfaces*. In contrast with the previous case, the surfaces of constant amplitude of the wave (23) do not, in general, coincide with the surfaces of constant phase. Such a wave is said to be *inhomogeneous*.

Calculations with harmonic waves are simplified by the use of exponential instead of trigonometric functions. Eq. (23) may be written as

$$V(\mathbf{r}, t) = \mathcal{R}\{U(\mathbf{r})e^{-i\omega t}\}, \quad (25)$$

where

$$U(\mathbf{r}) = a(\mathbf{r})e^{ig(\mathbf{r})}, \quad (26)$$

and \mathcal{R} denotes the real part. On substitution from (26) into the wave equation (1), one finds that U must satisfy the equation

$$\nabla^2 U + n^2 k_0^2 U = 0. \quad (27)$$

U is called the *complex amplitude** of the wave. In particular, for a plane wave one has

$$g(\mathbf{r}) = \omega \left(\frac{\mathbf{r} \cdot \mathbf{s}}{v} \right) - \delta = k(\mathbf{r} \cdot \mathbf{s}) - \delta = \mathbf{k} \cdot \mathbf{r} - \delta. \quad (28)$$

If the operations on V are linear, one may drop the symbol \mathcal{R} in (25) and operate directly with the complex function, the real part of the final expression being then understood to represent the physical quantity in question. However, when dealing with expressions which involve nonlinear operations such as squaring, etc. (e.g. in calculations of the electric or magnetic energy densities), one must in general take the real parts first and operate with these alone.†

Unlike a plane harmonic wave, the more general wave (25) is not periodic with respect to space. The phase $\omega t - g(\mathbf{r})$ is, however, seen to be the same for (\mathbf{r}, t) and $(\mathbf{r} + d\mathbf{r}, t + dt)$, provided that

$$\omega dt - (\text{grad } g) \cdot d\mathbf{r} = 0. \quad (29)$$

If we denote by \mathbf{q} the unit vector in the direction of $d\mathbf{r}$, and write $d\mathbf{r} = \mathbf{q} ds$, then (29) gives

$$\frac{ds}{dt} = \frac{\omega}{\mathbf{q} \cdot \text{grad } g}. \quad (30)$$

This expression will be numerically smallest when \mathbf{q} is the normal to the cophasal surface, i.e. when $\mathbf{q} = \text{grad } g / |\text{grad } g|$, the value then being

* In the case of a plane wave, one often separates the constant factor $e^{-i\delta}$ and implies by complex amplitude only the variable part $ae^{ik \cdot r}$.

† This is not necessary when only a time average of a quadratic expression is required [see §1.4, (54)–(56)].

$$v^{(p)}(\mathbf{r}) = \frac{\omega}{|\text{grad } g|}. \quad (31)$$

$v^{(p)}(\mathbf{r})$ is called the *phase velocity* and is the speed with which each of the cophasal surfaces advances. For a plane electromagnetic wave one has from (28) that $\text{grad } g = \mathbf{k}$, and therefore

$$v^{(p)} = \frac{\omega}{\mathbf{k}} = \frac{c}{\sqrt{\varepsilon\mu}},$$

because of (21). For waves of more complicated form, the phase velocity $v^{(p)}$ will in general differ from $c/\sqrt{\varepsilon\mu}$ and will vary from point to point even in a homogeneous medium. However, it will be seen later (§3.1.2) that, when the frequency is sufficiently large, the phase velocity is *approximately* equal to $c/\sqrt{\varepsilon\mu}$, even for waves whose cophasal surfaces are not plane.

It must be noted that the expression for ds/dt given by (30) is not the resolute of the phase velocity in the \mathbf{q} direction, i.e. the phase velocity does not behave as a vector. On the other hand its reciprocal, i.e. the quantity

$$\frac{dt}{ds} = \frac{\mathbf{q} \cdot \text{grad } g}{\omega}, \quad (32)$$

is seen to be the component of the vector $(\text{grad } g)/\omega$ in the \mathbf{q} direction. The vector $(\text{grad } g)/\omega$ is sometimes called *phase slowness*.

The phase velocity may in certain cases be greater than c . For plane waves this will be so when $n = \sqrt{\varepsilon\mu}$ is smaller than unity, as in the case of dispersing media in regions of the so-called anomalous dispersion* (see §2.3.4). Now according to the theory of relativity, signals can never travel faster than c . This implies that the phase velocity cannot correspond to a velocity with which a signal is propagated. It is, in fact, easy to see that the phase velocity cannot be determined experimentally and must therefore be considered to be void of any direct physical significance. For in order to measure this velocity, it would be necessary to affix a mark to the infinitely extended smooth wave and to measure the velocity of the mark. This would, however, mean the replacement of the infinite harmonic wave train by another function of space and time.

1.3.4 Wave packets. The group velocity

The monochromatic waves considered in the preceding section are idealizations never strictly realized in practice. It follows from Fourier's theorem that any wave $V(\mathbf{r}, t)$ (provided it satisfies certain very general conditions) may be regarded as a superposition of monochromatic waves of different frequencies:

$$V(\mathbf{r}, t) = \int_0^\infty a_\omega(\mathbf{r}) \cos[\omega t - g_\omega(\mathbf{r})] d\omega. \quad (33)$$

* The problem of propagation of electromagnetic signals in dispersive media has been investigated in classic papers by A. Sommerfeld, *Ann. d. Physik*, **44** (1914), 177 and by L. Brillouin, *ibid.*, **44** (1914), 203. English translations of these papers are included in L. Brillouin, *Wave Propagation and Group Velocity* (New York, Academic Press, 1960), pp. 17, 43. A systematic treatment of propagation of transient electromagnetic fields through dielectric media which exhibit both dispersion and absorption is given in K. E. Oughstun and G. C. Sherman, *Electromagnetic Pulse Propagation in Causal Dielectrics* (Berlin and New York, Springer, 1994).

It will again be convenient to use a complex representation, in which V is regarded as the real part of an associated complex wave:*

$$V(\mathbf{r}, t) = \mathcal{R} \int_0^\infty a_\omega(\mathbf{r}) e^{-i[\omega t - g_\omega(\mathbf{r})]} d\omega. \quad (33a)$$

A wave may be said to be ‘almost monochromatic,’ if the Fourier amplitudes a_ω differ appreciably from zero only within a narrow range

$$\bar{\omega} - \frac{1}{2}\Delta\omega \leq \omega \leq \bar{\omega} + \frac{1}{2}\Delta\omega \quad (\Delta\omega/\bar{\omega} \ll 1)$$

around a mean frequency $\bar{\omega}$. In such a case one usually speaks of a *wave group* or a *wave packet*.†

To illustrate some of the main properties of a wave group, consider first a wave formed by the superposition of two plane monochromatic waves of the same amplitudes and slightly different frequencies and wave numbers, propagated in the direction of the z -axis:

$$V(z, t) = ae^{-i(\omega t - kz)} + ae^{-i[(\omega + \delta\omega)t - (k + \delta k)z]}. \quad (34)$$

The symbol \mathcal{R} is omitted here in accordance with the convention explained earlier. Eq. (34) may be written in the form

$$\begin{aligned} V(z, t) &= a[e^{\frac{1}{2}i(t\delta\omega + z\delta k)} + e^{-\frac{1}{2}i(t\delta\omega - z\delta k)}]e^{-i(\bar{\omega}t - \bar{k}z)} \\ &= 2a \cos\left[\frac{1}{2}(t\delta\omega - z\delta k)\right]e^{-i(\bar{\omega}t - \bar{k}z)}, \end{aligned} \quad (35)$$

where

$$\bar{\omega} = \omega + \frac{1}{2}\delta\omega, \quad \bar{k} = k + \frac{1}{2}\delta k \quad (36)$$

are the mean frequency and the mean wave number respectively. Eq. (35) may be interpreted as representing a plane wave of frequency $\bar{\omega}$ and wavelength $2\pi/\bar{k}$ propagated in the z direction. The amplitude of this wave is, however, not constant, but varies with time and position, between the values $2a$ and 0 (Fig. 1.5), giving rise to the well-known phenomenon of beats. The successive maxima of the amplitude function are at intervals

$$\delta t = \frac{4\pi}{\delta\omega} \quad (\text{with } z \text{ fixed}) \quad \text{or} \quad \delta z = \frac{4\pi}{\delta k} \quad (\text{with } t \text{ fixed}) \quad (37)$$

from each other, whilst the maxima of the phase function are at intervals

$$\delta t = \frac{2\pi}{\bar{\omega}} \quad (\text{with } z \text{ fixed}) \quad \text{or} \quad \delta z = \frac{2\pi}{\bar{k}} \quad (\text{with } t \text{ fixed}). \quad (38)$$

Hence, since $\delta\omega/\bar{\omega}$ and $\delta k/\bar{k}$ are assumed to be small compared with unity, the amplitude will vary slowly in comparison with the other term.

From (35) it follows that the planes of constant amplitude and, in particular, the maxima of the amplitude are propagated with the velocity

* For a fuller discussion of the complex representation of real polychromatic waves see §10.2.

† Strictly speaking, in order that V should exhibit properties commonly attributed to a wave group, one should also assume that over the effective frequency range the phase function g_ω can be approximated by a linear function of ω .