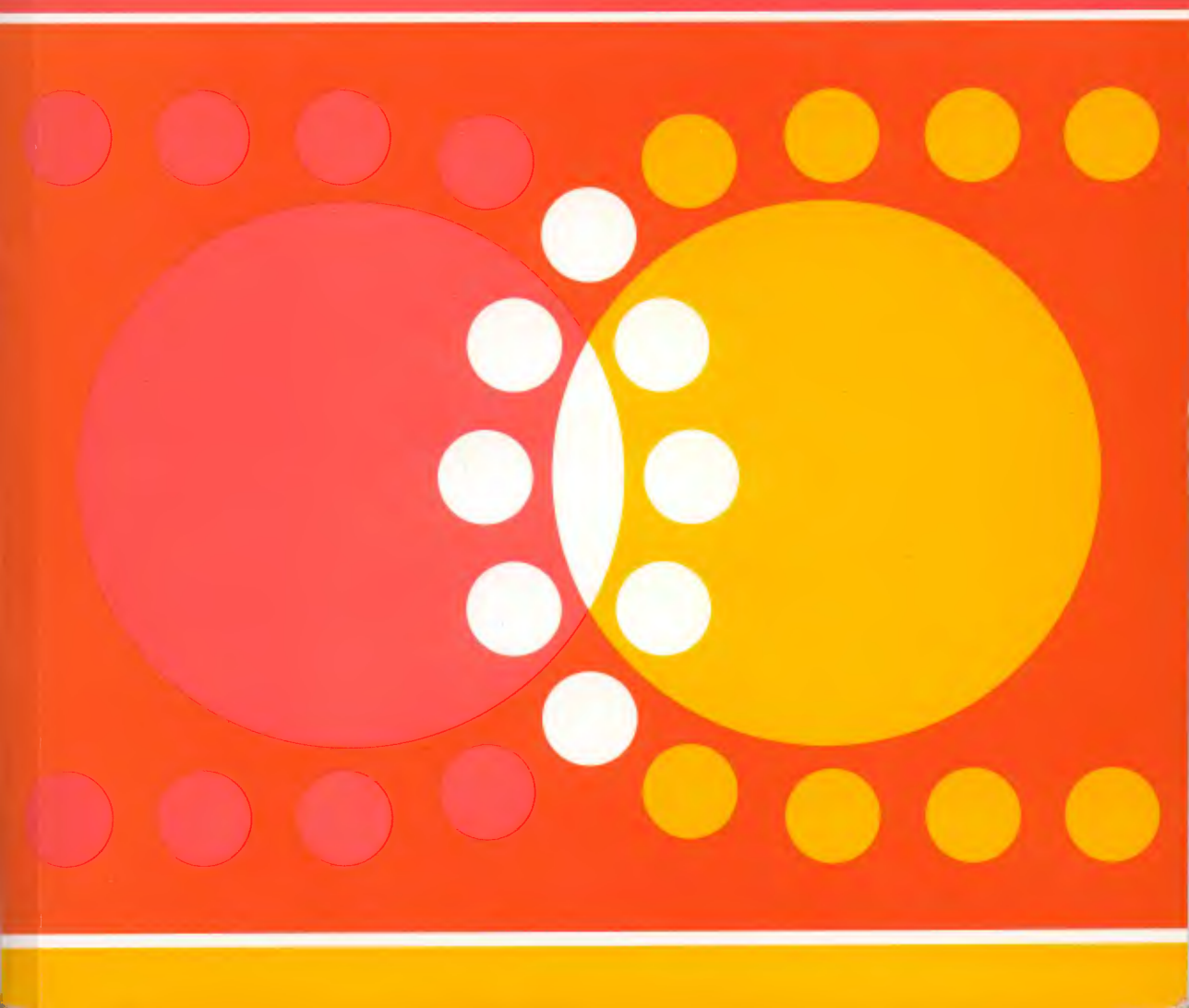
The Classical Theory of Fields

Fourth Revised English Edition

Course of Theoretical Physics
Volume 2

L.D. Landau and E.M. Lifshitz



THE CLASSICAL THEORY OF FIELDS

Fourth Revised English Edition

L. D. LANDAU AND E. M. LIFSHITZ
Institute for Physical Problems, Academy of Sciences of the U.S.S.R.

Translated from the Russian
by
MORTON HAMERMESH
University of Minnesota

AMSTERDAM BOSTON HEIDELBERG LONDON NEW YORK OXFORD
PARIS SAN DIEGO SAN FRANCISCO SINGAPORE SYDNEY TOKYO

CONTENTS

Excerpts from the Prefaces to the First and Second Editions	ix
Preface to the Fourth English Edition	x
Editor's Preface to the Seventh Russian Edition	xi
Notation	xiii
Chapter 1. The Principle of Relativity	1
1 Velocity of propagation of interaction	1
2 Intervals	3
3 Proper time	7
4 The Lorentz transformation	9
5 Transformation of velocities	12
6 Four-vectors	14
7 Four-dimensional velocity	23
Chapter 2. Relativistic Mechanics	25
8 The principle of least action	25
9 Energy and momentum	26
10 Transformation of distribution functions	30
11 Decay of particles	32
12 Invariant cross-section	36
13 Elastic collisions of particles	38
14 Angular momentum	42
Chapter 3. Charges in Electromagnetic Fields	46
15 Elementary particles in the theory of relativity	46
16 Four-potential of a field	47
17 Equations of motion of a charge in a field	49
18 Gauge invariance	52
19 Constant electromagnetic field	53
20 Motion in a constant uniform electric field	55
21 Motion in a constant uniform magnetic field	56
22 Motion of a charge in constant uniform electric and magnetic fields	59
23 The electromagnetic field tensor	64
24 Lorentz transformation of the field	66
25 Invariants of the field	67

CHAPTER 4. THE ELECTROMAGNETIC FIELD EQUATIONS 70

26 The first pair of Maxwell's equations 70
27 The action function of the electromagnetic field 71
28 The four-dimensional current vector 73
29 The equation of continuity 76
30 The second pair of Maxwell equations 78
31 Energy density and energy flux 80
32 The energy-momentum tensor 82
33 Energy-momentum tensor of the electromagnetic field 86
34 The virial theorem 90
35 The energy-momentum tensor for macroscopic bodies 92

CHAPTER 5. CONSTANT ELECTROMAGNETIC FIELDS 95

36 Coulomb's law 95
37 Electrostatic energy of charges 96
38 The field of a uniformly moving charge 98
39 Motion in the Coulomb field 100
40 The dipole moment 103
41 Multipole moments 105
42 System of charges in an external field 108
43 Constant magnetic field 110
44 Magnetic moments 111
45 Larmor's theorem 113

CHAPTER 6. ELECTROMAGNETIC WAVES 116

46 The wave equation 116
47 Plane waves 118
48 Monochromatic plane waves 123
49 Spectral resolution 128
50 Partially polarized light 129
51 The Fourier resolution of the electrostatic field 134
52 Characteristic vibrations of the field 135

CHAPTER 7. THE PROPAGATION OF LIGHT 140

53 Geometrical optics 140
54 Intensity 143
55 The angular eikonal 145
56 Narrow bundles of rays 147
57 Image formation with broad bundles of rays 153
58 The limits of geometrical optics 154
59 Diffraction 156
60 Fresnel diffraction 162
61 Fraunhofer diffraction 165

CHAPTER 8. THE FIELD OF MOVING CHARGES 171

62 The retarded potentials 171
63 The Lienard–Wiechert potentials 173
64 Spectral resolution of the retarded potentials 176
65 The Lagrangian to terms of second order 179

CHAPTER 9. RADIATION OF ELECTROMAGNETIC WAVES 184

66 The field of a system of charges at large distances 184
67 Dipole radiation 187
68 Dipole radiation during collisions 191
69 Radiation of low frequency in collisions 193
70 Radiation in the case of Coulomb interaction 195
71 Quadrupole and magnetic dipole radiation 203
72 The field of the radiation at near distances 206
73 Radiation from a rapidly moving charge 210
74 Synchrotron radiation (magnetic bremsstrahlung) 215
75 Radiation damping 222
76 Radiation damping in the relativistic case 226
77 Spectral resolution of the radiation in the ultrarelativistic case 230
78 Scattering by free charges 233
79 Scattering of low-frequency waves 238
80 Scattering of high-frequency waves 240

CHAPTER 10. PARTICLE IN A GRAVITATIONAL FIELD 243

81 Gravitational fields in nonrelativistic mechanics 243
82 The gravitational field in relativistic mechanics 244
83 Curvilinear coordinates 247
84 Distances and time intervals 251
85 Covariant differentiation 255
86 The relation of the Christoffel symbols to the metric tensor 260
87 Motion of a particle in a gravitational field 263
88 The constant gravitational field 266
89 Rotation 273
90 The equations of electrodynamics in the presence of a gravitational field 275

CHAPTER 11. THE GRAVITATIONAL FIELD EQUATIONS 278

91 The curvature tensor 278
92 Properties of the curvature tensor 281
93 The action function for the gravitational field 287
94 The energy-momentum tensor 290
95 The Einstein equations 295
96 The energy-momentum pseudotensor of the gravitational field 301
97 The synchronous reference system 307
98 The tetrad representation of the Einstein equations 313

CHAPTER 12. THE FIELD OF GRAVITATING BODIES 316

99 Newton's law 316
100 The centrally symmetric gravitational field 320
101 Motion in a centrally symmetric gravitational field 328
102 Gravitational collapse of a spherical body 331
103 Gravitational collapse of a dustlike sphere 338
104 Gravitational collapse of nonspherical and rotating bodies 344
105 Gravitational fields at large distances from bodies 353
106 The equations of motion of a system of bodies in the second approximation 360

CHAPTER 13. GRAVITATIONAL WAVES 368

107 Weak gravitational waves 368
108 Gravitational waves in curved space-time 370
109 Strong gravitational waves 373
110 Radiation of gravitational waves 376

CHAPTER 14. RELATIVISTIC COSMOLOGY 382

111 Isotropic space 382
112 The closed isotropic model 386
113 The open isotropic model 390
114 The red shift 394
115 Gravitational stability of an isotropic universe 400
116 Homogeneous spaces 406
117 The flat anisotropic model 412
118 Oscillating regime of approach to a singular point 416
119 The time singularity in the general cosmological solution of the Einstein equations 420

INDEX 425

EXCERPTS FROM THE PREFACES TO THE FIRST AND SECOND EDITIONS

THIS book is devoted to the presentation of the theory of the electromagnetic and gravitational fields, i.e. electrodynamics and general relativity. A complete, logically connected theory of the electromagnetic field includes the special theory of relativity, so the latter has been taken as the basis of the presentation. As the starting point of the derivation of the fundamental relations we take the variational principles, which make possible the attainment of maximum generality, unity and simplicity of presentation.

In accordance with the overall plan of our Course of Theoretical Physics (of which this book is a part), we have not considered questions concerning the electrodynamics of continuous media, but restricted the discussion to "microscopic electrodynamics"—the electrodynamics of point charges *in vacuo*.

The reader is assumed to be familiar with electromagnetic phenomena as discussed in general physics courses. A knowledge of vector analysis is also necessary. The reader is not assumed to have any previous knowledge of tensor analysis, which is presented in parallel with the development of the theory of gravitational fields.

Moscow, December 1939
Moscow, June 1947

L. LANDAU, E. LIFSHITZ

PREFACE TO THE FOURTH ENGLISH EDITION

THE first edition of this book appeared more than thirty years ago. In the course of reissues over these decades the book has been revised and expanded; its volume has almost doubled since the first edition. But at no time has there been any need to change the method proposed by Landau for developing the theory, or his style of presentation, whose main feature was a striving for clarity and simplicity. I have made every effort to preserve this style in the revisions that I have had to make on my own.

As compared with the preceding edition, the first nine chapters, devoted to electrodynamics, have remained almost without changes. The chapters concerning the theory of the gravitational field have been revised and expanded. The material in these chapters has increased from edition to edition, and it was finally necessary to redistribute and rearrange it.

I should like to express here my deep gratitude to all of my helpers in this work—too many to be enumerated—who, by their comments and advice, helped me to eliminate errors and introduce improvements. Without their advice, without the willingness to help which has met all my requests, the work to continue the editions of this course would have been much more difficult. A special debt of gratitude is due to L. P. Pitaevskii, with whom I have constantly discussed all the vexing questions.

The English translation of the book was done from the last Russian edition, which appeared in 1973. No further changes in the book have been made. The 1994 corrected reprint includes the changes made by E. M. Lifshitz in the Seventh Russian Edition published in 1987.

I should also like to use this occasion to sincerely thank Prof. Hamermesh, who has translated this book in all its editions, starting with the first English edition in 1951. The success of this book among English-speaking readers is to a large extent the result of his labour and careful attention.

E. M. LIFSHITZ

PUBLISHER'S NOTE

As with the other volumes in the *Course of Theoretical Physics,* the authors do not, as a rule, give references to original papers, but simply name their authors (with dates). Full bibliographic references are only given to works which contain matters not fully expounded in the text.

EDITOR'S PREFACE TO THE SEVENTH RUSSIAN EDITION

E. M. Lifshitz began to prepare a new edition of *Teoria Polia* in 1985 and continued his work on it even in hospital during the period of his last illness. The changes that he proposed are made in the present edition. Of these we should mention some revision of the proof of the law of conservation of angular momentum in relativistic mechanics, and also a more detailed discussion of the question of symmetry of the Christoffel symbols in the theory of gravitation. The sign has been changed in the definition of the electromagnetic field stress tensor. (In the present edition this tensor was defined differently than in the other volumes of the Course.)

June 1987

L. P. PITAEVSKII

NOTATION

Three-dimensional quantities

Three-dimensional tensor indices are denoted by Greek letters
Element of volume, area and length: dV, $d\mathbf{f}$, $d\mathbf{l}$
Momentum and energy of a particle: $\mathbf{p}$ and $\mathscr{E}$
Hamiltonian function: $\mathscr{H}$
Scalar and vector potentials of the electromagnetic field: ϕ and $\mathbf{A}$
Electric and magnetic field intensities: $\mathbf{E}$ and $\mathbf{H}$
Charge and current density: ρ and $\mathbf{j}$
Electric dipole moment: $\mathbf{d}$
Magnetic dipole moment: m

Four-dimensional quantities

Four-dimensional tensor indices are denoted by Latin letters $i, k, l, \ldots$ and take on the values 0, 1, 2, 3
We use the metric with signature $(+ - - -)$
Rule for raising and lowering indices—see p. 14
Components of four-vectors are enumerated in the form $A^i = (A^0, \mathbf{A})$
Antisymmetric unit tensor of rank four is e^{iklm}, where $e^{0123} = 1$ (for the definition, see p. 17)
Element of four-volume $d\Omega = dx^0 dx^1 dx^2 dx^3$
Element of hypersurface dS^i (defined on pp. 20–21)
Radius four-vector: $x^i = (ct, \mathbf{r})$
Velocity four-vector: $u^i = dx^i/ds$
Momentum four-vector: $p = (\mathscr{E}/c, \mathbf{p})$
Current four-vector: $j^i = (c\rho, \rho\mathbf{v})$
Four-potential of the electromagnetic field: $A^i = (\phi, \mathbf{A})$
Electromagnetic field four-tensor $F_{ik} = \dfrac{\partial A_k}{\partial x^i} - \dfrac{\partial A_i}{\partial x^k}$ (for the relation of the components of F_{ik} to the components of $\mathbf{E}$ and $\mathbf{H}$, see p. 65)
Energy-momentum four-tensor T^{ik}(for the definition of its components, see p. 83)

CHAPTER 1

THE PRINCIPLE OF RELATIVITY

§ 1. Velocity of propagation of interaction

For the description of processes taking place in nature, one must have a *system of reference.* By a system of reference we understand a system of coordinates serving to indicate the position of a particle in space, as well as clocks fixed in this system serving to indicate the time.

There exist systems of reference in which a freely moving body, i.e. a moving body which is not acted upon by external forces, proceeds with constant velocity. Such reference systems are said to be *inertial.*

If two reference systems move uniformly relative to each other, and if one of them is an inertial system, then clearly the other is also inertial (in this system too every free motion will be linear and uniform). In this way one can obtain arbitrarily many inertial systems of reference, moving uniformly relative to one another.

Experiment shows that the so-called *principle of relativity* is valid. According to this principle all the laws of nature are identical in all inertial systems of reference. In other words, the equations expressing the laws of nature are invariant with respect to transformations of coordinates and time from one inertial system to another. This means that the equation describing any law of nature, when written in terms of coordinates and time in different inertial reference systems, has one and the same form.

The interaction of material particles is described in ordinary mechanics by means of a potential energy of interaction, which appears as a function of the coordinates of the interacting particles. It is easy to see that this manner of describing interactions contains the assumption of instantaneous propagation of interactions. For the forces exerted on each of the particles by the other particles at a particular instant of time depend, according to this description, only on the positions of the particles at this one instant. A change in the position of any of the interacting particles influences the other particles immediately.

However, experiment shows that instantaneous interactions do not exist in nature. Thus a mechanics based on the assumption of instantaneous propagation of interactions contains within itself a certain inaccuracy. In actuality, if any change takes place in one of the interacting bodies, it will influence the other bodies only after the lapse of a certain interval of time. It is only after this time interval that processes caused by the initial change begin to take place in the second body. Dividing the distance between the two bodies by this time interval, we obtain the *velocity of propagation of the interaction.*

We note that this velocity should, strictly speaking, be called the *maximum* velocity of propagation of interaction. It determines only that interval of time after which a change occurring in one body *begins* to manifest itself in another. It is clear that the existence of a

maximum velocity of propagation of interactions implies, at the same time, that motions of bodies with greater velocity than this are in general impossible in nature. For if such a motion could occur, then by means of it one could realize an interaction with a velocity exceeding the maximum possible velocity of propagation of interactions.

Interactions propagating from one particle to another are frequently called "signals", sent out from the first particle and "informing" the second particle of changes which the first has experienced. The velocity of propagation of interaction is then referred to as the *signal velocity.*

From the principle of relativity it follows in particular that the velocity of propagation of interactions is the *same* in *all* inertial systems of reference. Thus the velocity of propagation of interactions is a universal constant. This constant velocity (as we shall show later) is also the velocity of light in empty space. The velocity of light is usually designated by the letter c, and its numerical value is

$$c = 2.998 \times 10^{10} \text{ cm/sec.} \tag{1.1}$$

The large value of this velocity explains the fact that in practice classical mechanics appears to be sufficiently accurate in most cases. The velocities with which we have occasion to deal are usually so small compared with the velocity of light that the assumption that the latter is infinite does not materially affect the accuracy of the results.

The combination of the principle of relativity with the finiteness of the velocity of propagation of interactions is called the *principle of relativity of Einstein* (it was formulated by Einstein in 1905) in contrast to the principle of relativity of Galileo, which was based on an infinite velocity of propagation of interactions.

The mechanics based on the Einsteinian principle of relativity (we shall usually refer to it simply as the principle of relativity) is called *relativistic.* In the limiting case when the velocities of the moving bodies are small compared with the velocity of light we can neglect the effect on the motion of the finiteness of the velocity of propagation. Then relativistic mechanics goes over into the usual mechanics, based on the assumption of instantaneous propagation of interactions; this mechanics is called *Newtonian* or *classical.* The limiting transition from relativistic to classical mechanics can be produced formally by the transition to the limit $c \to \infty$ in the formulas of relativistic mechanics.

In classical mechanics distance is already relative, i.e. the spatial relations between different events depend on the system of reference in which they are described. The statement that two nonsimultaneous events occur at one and the same point in space or, in general, at a definite distance from each other, acquires a meaning only when we indicate the system of reference which is used.

On the other hand, time is absolute in classical mechanics; in other words, the properties of time are assumed to be independent of the system of reference; there is one time for all reference frames. This means that if any two phenomena occur simultaneously for any one observer, then they occur simultaneously also for all others. In general, the interval of time between two given events must be identical for all systems of reference.

It is easy to show, however, that the idea of an absolute time is in complete contradiction to the Einstein principle of relativity. For this it is suffcient to recall that in classical mechanics, based on the concept of an absolute time, a general law of combination of velocities is valid, according to which the velocity of a composite motion is simply equal to the (vector) sum of the velocities which constitute this motion. This law, being universal, should also be applicable to the propagation of interactions. From this it would follow that the velocity of

propagation must be different in different inertial systems of reference, in contradiction to the principle of relativity. In this matter experiment completely confirms the principle of relativity. Measurements first performed by Michelson (1881) showed complete lack of dependence of the velocity of light on its direction of propagation; whereas according to classical mechanics the velocity of light should be smaller in the direction of the earth's motion than in the opposite direction.

Thus the principle of relativity leads to the result that time is not absolute. Time elapses differently in different systems of reference. Consequently the statement that a definite time interval has elapsed between two given events acquires meaning only when the reference frame to which this statement applies is indicated. In particular, events which are simultaneous in one reference frame will not be simultaneous in other frames.

To clarify this, it is instructive to consider the following simple example.

Let us look at two inertial reference systems K and K' with coordinate axes XYZ and $X'Y'Z'$ respectively, where the system K' moves relative to K along the $X(X')$ axis (Fig. 1).

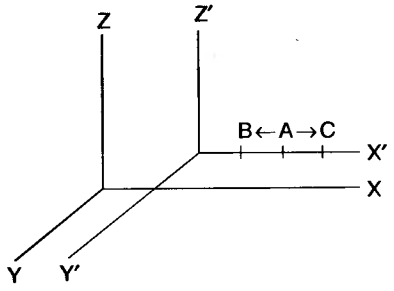


FIG. 1.

Suppose signals start out from some point A on the X' axis in two opposite directions. Since the velocity of propagation of a signal in the K' system, as in all inertial systems, is equal (for both directions) to c, the signals will reach points B and C, equidistant from A, at one and the same time (in the K' system)

But it is easy to see that the same two events (arrival of the signal at B and C) can by no means be simultaneous for an observer in the K system. In fact, the velocity of a signal relative to the K system has, according to the principle of relativity, the same value c, and since the point B moves (relative to the K system) toward the source of its signal, while the point C moves in the direction away from the signal (sent from A to C), in the K system the signal will reach point B earlier than point C.

Thus the principle of relativity of Einstein introduces very drastic and fundamental changes in basic physical concepts. The notions of space and time derived by us from our daily experiences are only approximations linked to the fact that in daily life we happen to deal only with velocities which are very small compared with the velocity of light.

§ 2. Intervals

In what follows we shall frequently use the concept of an *event*. An event is described by the place where it occurred and the time when it occurred. Thus an event occurring in a certain material particle is defined by the three coordinates of that particle and the time when the event occurs.

It is frequently useful for reasons of presentation to use a fictitious four-dimensional

space, on the axes of which are marked three space coordinates and the time. In this space events are represented by points, called *world points.* In this fictitious four-dimensional space there corresponds to each particle a cetain line, called a *world line.* The points of this line determine the coordinates of the particle at all moments of time. It is easy to show that to a particle in uniform rectilinear motion there corresponds a straight world line.

We now express the principle of the invariance of the velocity of light in mathematical form. For this purpose we consider two reference systems K and K' moving relative to each other with constant velocity. We choose the coordinate axes so that the axes X and X' coincide, while the Y and Z axes are parallel to Y' and Z'; we designate the time in the systems K and K' by t and t'.

Let the first event consist of sending out a signal, propagating with light velocity, from a point having coordinates $x_1y_1z_1$ in the K system, at time t_1 in this system. We observe the propagation of this signal in the K system. Let the second event consist of the arrival of the signal at point $x_2y_2z_2$ at the moment of time t_2. The signal propagates with velocity c; the distance covered by it is therefore $c(t_1 - t_2)$. On the other hand, this same distance equals $[(x_2-x_1)^2 + (y_2-y_1)^2 + (z_2 - z_1)^2]^{\frac{1}{2}}$. Thus we can write the following relation between the coordinates of the two events in the K system:

$$(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 - c^2(t_2 - t_1)^2 = 0. \tag{2.1}$$

The same two events, i.e. the propagation of the signal, can be observed from the K' system:

Let the coordinates of the first event in the K' system be $x_1'y_1'z_1't_1'$, and of the second: $x_2'y_2'z_2't_2'$. Since the velocity of light is the same in the K and K' systems, we have, similarly to (2.1):

$$(x_2' - x_1')^2 + (y_2' - y_1')^2 + (z_2' - z_1')^2 - c^2(t_2' - t_1')^2 = 0. \tag{2.2}$$

If $x_1y_1z_1t_1$ and $x_2y_2z_2t_2$ are the coordinates of *any* two events, then the quantity

$$s_{12} = [c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2]^{\frac{1}{2}} \tag{2.3}$$

is called the *interval* between these two events.

Thus it follows from the principle of invariance of the velocity of light that if the interval between two events is zero in one coordinate system, then it is equal to zero in all other systems.

If two events are infinitely close to each other, then the interval ds between them is

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2. \tag{2.4}$$

The form of expressions (2.3) and (2.4) permits us to regard the interval, from the formal point of view, as the distance between two points in a fictitious four-dimensional space (whose axes are labelled by x, y, z, and the product ct). But there is a basic difference between the rule for forming this quantity and the rule in ordinary geometry: in forming the square of the interval, the squares of the coordinate differences along the different axes are summed, not with the same sign, but rather with varying signs.†

As already shown, if $ds = 0$ in one inertial system, then $ds' = 0$ in any other system. On

† The four-dimensional geometry described by the quadratic form (2.4) was introduced by H. Minkowski, in connection with the theory of relativity. This geometry is called *pseudo-euclidean,* in contrast to ordinary euclidean geometry.

the other hand, ds and ds' are infinitesimals of the same order. From these two conditions it follows that ds^2 and ds'^2 must be proportional to each other:

$$ds^2 = ads'^2$$

where the coefficient a can depend only on the absolute value of the relative velocity of the two inertial systems. It cannot depend on the coordinates or the time, since then different points in space and different moments in time would not be equivalent, which would be in contradiction to the homogeneity of space and time. Similarly, it cannot depend on the direction of the relative velocity, since that would contradict the isotropy of space.

Let us consider three reference systems K, K_1, K_2, and let V_1 and V_2 be the velocities of systems K_1 and K_2 relative to K. We then have:

$$ds^2 = a(V_1)ds_1^2, \qquad ds^2 = a(V_2)ds_2^2.$$

Similarly we can write

$$ds_1^2 = a(V_{12})ds_2^2,$$

where V_{12} is the absolute value of the velocity of K_2 relative to K_1. Comparing these relations with one another, we find that we must have

$$\frac{a(V_2)}{a(V_1)} = a(V_{12}). \tag{2.5}$$

But V_{12} depends not only on the absolute values of the vectors $\mathbf{V}_1$ and $\mathbf{V}_2$, but also on the angle between them. However, this angle does not appear on the left side of formula (2.5). It is therefore clear that this formula can be correct only if the function $a(V)$ reduces to a constant, which is equal to unity according to this same formula.

Thus,

$$ds^2 = ds'^2, \tag{2.6}$$

and from the equality of the infinitesimal intervals there follows the equality of finite intervals: $s = s'$.

Thus we arrive at a very important result: the interval between two events is the same in all inertial systems of reference, i.e. it is invariant under transformation from one inertial system to any other. This invariance is the mathematical expression of the constancy of the velocity of light.

Again let $x_1y_1z_1t_1$ and $x_2y_2z_2t_2$ be the coordinates of two events in a certain reference system K. Does there exist a coordinate system K', in which these two events occur at one and the same point in space?

We introduce the notation

$$t_2 - t_1 = t_{12}, \qquad (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 = l_{12}^2.$$

Then the interval between events in the K system is:

$$s_{12}^2 = c^2t_{12}^2 - l_{12}^2$$

and in the K' system

$$s_{12}'^2 = c^2t_{12}'^2 - l_{12}'^2,$$

whereupon, because of the invariance of intervals,

$$c^2t_{12}^2 - l_{12}^2 = c^2t_{12}'^2 - l_{12}'^2.$$

We want the two events to occur at the same point in the K' system, that is, we require $l_{12}' = 0$. Then

$$s_{12}^2 = c^2t_{12}^2 - l_{12}^2 = c^2t_{12}'^2 > 0.$$

Consequently a system of reference with the required property exists if $s_{12}^2 > 0$, that is, if the interval between the two events is a real number. Real intervals are said to be *timelike*.

Thus, if the interval between two events is timelike, then there exists a system of reference in which the two events occur at one and the same place. The time which elapses between the two events in this system is

$$t_{12}' = \frac{1}{c}\sqrt{c^2t_{12}^2 - l_{12}^2} = \frac{s_{12}}{c}. \tag{2.7}$$

If two events occur in one and the same body, then the interval between them is always timelike, for the distance which the body moves between the two events cannot be greater than ct_{12}, since the velocity of the body cannot exceed c. So we have always

$$l_{12} < ct_{12}.$$

Let us now ask whether or not we can find a system of reference in which the two events occur at one and the same time. As before, we have for the K and K' systems $c^2t_{12}^2 - l_{12}^2 = c^2t_{12}'^2 - l_{12}'^2$. We want to have $t_{12}' = 0$, so that

$$s_{12}^2 = -l_{12}'^2 < 0.$$

Consequently the required system can be found only for the case when the interval s_{12} between the two events is an imaginary number. Imaginary intervals are said to be *spacelike*.

Thus if the interval between two events is spacelike, there exists a reference system in which the two events occur simultaneously. The distance between the points where the events occur in this system is

$$l_{12}' = \sqrt{l_{12}^2 - c^2t_{12}^2} = is_{12}. \tag{2.8}$$

The division of intervals into space- and timelike intervals is, because of their invariance, an absolute concept. This means that the timelike or spacelike character of an interval is independent of the reference system.

Let us take some event O as our origin of time and space coordinates. In other words, in the four-dimensional system of coordinates, the axes of which are marked x, y, z, t, the world point of the event O is the origin of coordinates. Let us now consider what relation other events bear to the given event O. For visualization, we shall consider only one space dimension and the time, marking them on two axes (Fig. 2). Uniform rectilinear motion of a particle, passing through $x = 0$ at $t = 0$, is represented by a straight line going through O and inclined to the t axis at an angle whose tangent is the velocity of the particle. Since the maximum possible velocity is c, there is a maximum angle which this line can subtend with the t axis. In Fig. 2 are shown the two lines representing the propagation of two signals (with the velocity of light) in opposite directions passing through the event O (i.e. going through $x = 0$ at $t = 0$). All lines representing the motion of particles can lie only in the regions aOc and dOb. On the lines ab and cd, $x = \pm ct$. First consider events whose world points lie within the region aOc. It is easy to show that for all the points of this region $c^2t^2 - x^2 > 0$.

In other words, the interval between any event in this region and the event O is timelike. In this region $t > 0$, i.e. all the events in this region occur "after" the event O. But two events which are separated by a timelike interval cannot occur simultaneously in any reference system. Consequently it is impossible to find a reference system in which any of the events in region aOc occurred "before" the event O, i.e. at time $t < 0$. Thus all the events in region aOc are future events relative to O in *all* reference systems. Therefore this region can be called the *absolute future* relative to O.

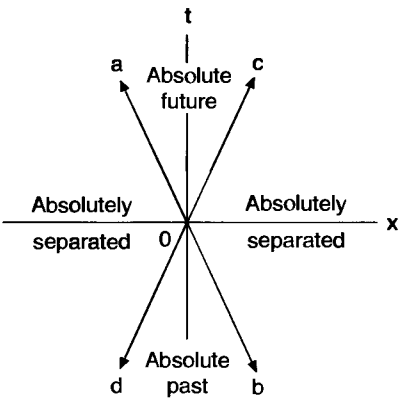


FIG. 2

In exactly the same way, all events in the region bOd are in the *absolute past* relative to O; i.e. events in this region occur before the event O in all systems of reference.

Next consider regions dOa and cOb. The interval between any event in this region and the event O is spacelike. These events occur at different points in space in every reference system. Therefore these regions can be said to be *absolutely remote* relative to O. However, the concepts "simultaneous", "earlier", and "later" are relative for these regions. For any event in these regions there exist systems of reference in which it occurs after the event O, systems in which it occurs earlier than O, and finally one reference system in which it occurs simultaneously with O.

Note that if we consider all three space coordinates instead of just one, then instead of the two intersecting lines of Fig. 2 we would have a "cone" $x^2 + y^2 + z^2 - c^2t^2 = 0$ in the four-dimensional coordinate system x, y, z, t, the axis of the cone coinciding with the t axis. (This cone is called the *light cone*.) The regions of absolute future and absolute past are then represented by the two interior portions of this cone.

Two events can be related causally to each other only if the interval between them is timelike; this follows immediately from the fact that no interaction can propagate with a velocity greater than the velocity of light. As we have just seen, it is precisely for these events that the concepts "earlier" and "later" have an absolute significance, which is a necessary condition for the concepts of cause and effect to have meaning.

§ 3. Proper time

Suppose that in a certain inertial reference system we observe clocks which are moving relative to us in an arbitrary manner. At each different moment of time this motion can be considered as uniform. Thus at each moment of time we can introduce a coordinate system

rigidly linked to the moving clocks, which with the clocks constitutes an inertial reference system.

In the course of an infinitesimal time interval dt (as read by a clock in our rest frame) the moving clocks go a distance $\sqrt{dx^2 + dy^2 + dz^2}$. Let us ask what time interval dt' is indicated for this period by the moving clocks. In a system of coordinates linked to the moving clocks, the latter are at rest, i.e., $dx' = dy' = dz' = 0$. Because of the invariance of intervals

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2 = c^2dt'^2,$$

from which

$$dt' = dt\sqrt{1 - \frac{dx^2 + dy^2 + dz^2}{c^2dt^2}}.$$

But

$$\frac{dx^2 + dy^2 + dz^2}{dt^2} = v^2,$$

where v is the velocity of the moving clocks; therefore

$$dt' = \frac{ds}{c} = dt\sqrt{1 - \frac{v^2}{c^2}}. \tag{3.1}$$

Integrating this expression, we can obtain the time interval indicated by the moving clocks when the elapsed time according to a clock at rest is $t_2 - t_1$:

$$t_2' - t_1' = \int_{t_1}^{t_2} dt\sqrt{1 - \frac{v^2}{c^2}}. \tag{3.2}$$

The time read by a clock moving with a given object is called the *proper time* for this object. Formulas (3.1) and (3.2) express the proper time in terms of the time for a system of reference from which the motion is observed.

As we see from (3.1) or (3.2), the proper time of a moving object is always less than the corresponding interval in the rest system. In other words, moving clocks go more slowly than those at rest.

Suppose some clocks are moving in uniform rectilinear motion relative to an inertial system K. A reference frame K' linked to the latter is also inertial. Then from the point of view of an observer in the K system the clocks in the K' system fall behind. And conversely, from the point of view of the K' system, the clocks in K lag. To convince ourselves that there is no contradiction, let us note the following. In order to establish that the clocks in the K' system lag behind those in the K system, we must proceed in the following fashion. Suppose that at a certain moment the clock in K' passes by the clock in K, and at that moment the readings of the two clocks coincide. To compare the rates of the two clocks in K and K' we must once more compare the readings of the same moving clock in K' with the clocks in K. But now we compare this clock with *different* clocks in K—with those past which the clock in K' goes at ths new time. Then we find that the clock in K' lags behind the clocks in K with which it is being compared. We see that to compare the rates of clocks in two reference

frames we require several clocks in one frame and one in the other, and that therefore this process is not symmetric with respect to the two systems. The clock that appears to lag is always the one which is being compared with different clocks in the other system.

If we have two clocks, one of which describes a closed path returning to the starting point (the position of the clock which remained at rest), then clearly the moving clock appears to lag relative to the one at rest. The converse reasoning, in which the moving clock would be considered to be at rest (and vice versa) is now impossible, since the clock describing a closed trajectory does not carry out a uniform rectilinear motion, so that a coordinate system linked to it will not be inertial.

Since the laws of nature are the same only for inertial reference frames, the frames linked to the clock at rest (inertial frame) and to the moving clock (non-inertial) have different properties, and the argument which leads to the result that the clock at rest must lag is not valid.

The time interval read by a clock is equal to the integral

$$\frac{1}{c}\int_a^b ds,$$

taken along the world line of the clock. If the clock is at rest then its world line is clearly a line parallel to the t axis; if the clock carries out a nonuniform motion in a closed path and returns to its starting point, then its world line will be a curve passing through the two points, on the straight world line of a clock at rest, corresponding to the beginning and end of the motion. On the other hand, we saw that the clock at rest always indicates a greater time interval than the moving one. Thus we arrive at the result that the integral

$$\int_a^b ds,$$

taken between a given pair of world points, has its maximum value if it is taken along the straight world line joining these two points.†

§ 4. The Lorentz transformation

Our purpose is now to obtain the formula of transformation from one inertial reference system to another, that is, a formula by means of which, knowing the coordinates x, y, z, t, of a certain event in the K system, we can find the coordinates x', y', z', t' of the same event in another inertial system K'.

In classical mechanics this question is resolved very simply. Because of the absolute nature of time we there have $t = t'$; if, furthermore, the coordinate axes are chosen as usual (axes X, X' coincident, Y, Z axes parallel to Y', Z', motion along X, X') then the coordinates y, z clearly are equal to y', z', while the coordinates x and x' differ by the distance traversed by one system relative to the other. If the time origin is chosen as the moment when the two coordinate systems coincide, and if the velocity of the K' system relative to K is V, then this distance is Vt. Thus

† It is assumed, of course, that the points a and b and the curves joining them are such that all elements ds along the curves are timelike.

This property of the integral is connected with the pseudo-euclidean character of the four-dimensional geometry. In euclidean space the integral would, of course, be a minimum along the straight line.

$$x = x' + Vt, \quad y = y', \quad z = z', \quad t = t'. \tag{4.1}$$

This formula is called the *Galileo transformation.* It is easy to verify that this transformation, as was to be expected, does not satisfy the requirements of the theory of relativity; it does not leave the interval between events invariant.

We shall obtain the relativistic transformation precisely as a consequence of the requirement that it leaves the interval between events invariant.

As we saw in § 2, the interval between events can be looked on as the distance between the corresponding pair of world points in a four-dimensional system of coordinates. Consequently we may say that the required transformation must leave unchanged all distances in the four-dimensional x, y, z, ct, space. But such transformations consist only of parallel displacements, and rotations of the coordinate system. Of these the displacement of the coordinate system parallel to itself is of no interest, since it leads only to a shift in the origin of the space coordinates and a change in the time reference point. Thus the required transformation must be expressible mathematically as a rotation of the four-dimensional x, y, z, ct, coordinate system.

Every rotation in the four-dimensional space can be resolved into six rotations, in the planes xy, zy, xz, tx, ty, tz (just as every rotation in ordinary space can be resolved into three rotations in the planes xy, zy and xz). The first three of these rotations transform only the space coordinates; they correspond to the usual space rotations.

Let us consider a rotation in the tx plane; under this, the y and z coordinates do not change. In particular, this transformation must leave unchanged the difference $(ct)^2 - x^2$, the square of the "distance" of the point (ct, x) from the origin. The relation between the old and the new coordinates is given in most general form by the formulas:

$$x = x' \cosh \psi + ct' \sinh \psi, \quad ct = x' \sinh \psi + ct' \cosh \psi, \tag{4.2}$$

where ψ is the "angle of rotation"; a simple check shows that in fact $c^2t^2 - x^2 = c^2t'^2 - x'^2$. Formula (4.2) differs from the usual formulas for transformation under rotation of the coordinate axes in having hyperbolic functions in place of trigonometric functions. This is the difference between pseudo-euclidean and euclidean geometry.

We try to find the formula of transformation from an inertial reference frame K to a system K' moving relative to K with velocity V along the x axis. In this case clearly only the coordinate x and the time t are subject to change. Therefore this transformation must have the form (4.2). Now it remains only to determine the angle ψ, which can depend only on the relative velocity V.†

Let us consider the motion, in the K system, of the origin of the K' system. Then $x' = 0$ and formulas (4.2) take the form:

$$x = ct' \sinh \psi, \quad ct = ct' \cosh \psi,$$

or dividing one by the other,

$$\frac{x}{ct} = \tanh \psi.$$

But x/t is clearly the velocity V of the K' system relative to K. So

† Note that to avoid confusion we shall always use V to signify the constant relative velocity of two inertial systems, and v for the velocity of a moving particle, not necessarily constant.

$$\tanh \psi = \frac{V}{c}.$$

From this

$$\sinh \psi = \frac{\frac{V}{c}}{\sqrt{1 - \frac{V^2}{c^2}}}, \qquad \cosh \psi = \frac{1}{\sqrt{1 - \frac{V^2}{c^2}}}.$$

Substituting in (4.2), we find:

$$x = \frac{x' + Vt'}{\sqrt{1 - \frac{V^2}{c^2}}}, \quad y = y', \quad z = z', \quad t = \frac{t' + \frac{V}{c^2}x'}{\sqrt{1 - \frac{V^2}{c^2}}}. \tag{4.3}$$

This is the required transformation formula. It is called the *Lorentz transformation,* and is of fundamental importance for what follows.

The inverse formulas, expressing x', y', z', t' in terms of x, y, z, t, are most easily obtained by changing V to $-V$ (since the K system moves with velocity $-V$ relative to the K' system). The same formulas can be obtained directly by solving equations (4.3) for x', y', z', t'.

It is easy to see from (4.3) that on making the transition to the limit $c \to \infty$ and classical mechanics, the formula for the Lorentz transformation actually goes over into the Galileo transformation.

For $V > c$ in formula (4.3) the coordinates x, t are imaginary; this corresponds to the fact that motion with a velocity greater than the velocity of light is impossible. Moreover, one cannot use a reference system moving with the velocity of light—in that case the denominators in (4.3) would go to zero.

For velocities V small compared with the velocity of light, we can use in place of (4.3) the approximate formulas:

$$x = x' + Vt', \quad y = y', \quad z = z', \quad t = t' + \frac{V}{c^2}x'. \tag{4.4}$$

Suppose there is a rod at rest in the K system, parallel to the X axis. Let its length, measured in this system, be $\Delta x = x_2 - x_1$ (x_2 and x_1 are the coordinates of the two ends of the rod in the K system). We now determine the length of this rod as measured in the K' system. To do this we must find the coordinates of the two ends of the rod (x_2' and x_1') in this system at one and the same time t'. From (4.3) we find:

$$x_1 = \frac{x_1' + Vt'}{\sqrt{1 - \frac{V^2}{c^2}}}, \qquad x_2 = \frac{x_2' + Vt'}{\sqrt{1 - \frac{V^2}{c^2}}}.$$

The length of the rod in the K' system is $\Delta x' = x_2' - x_1'$; subtracting x_1 from x_2, we find

$$\Delta x = \frac{\Delta x'}{\sqrt{1 - \frac{V^2}{c^2}}}.$$

The *proper length* of a rod is its length in a reference system in which it is at rest. Let

us denote it by $l_0 = \Delta x$, and the length of the rod in any other reference frame K' by l. Then

$$l = l_0 \sqrt{1 - \frac{V^2}{c^2}}. \tag{4.5}$$

Thus a rod has its greatest length in the reference system in which it is at rest. Its length in a system in which it moves with velocity V is decreased by the factor $\sqrt{1 - V^2/c^2}$. This result of the theory of relativity is called the *Lorentz contraction.*

Since the transverse dimensions do not change because of its motion, the volume $\mathscr{V}$ of a body decreases according to the similar formula

$$\mathscr{V} = \mathscr{V}_0 \sqrt{1 - \frac{V^2}{c^2}}, \tag{4.6}$$

where $\mathscr{V}_0$ is the *proper volume* of the body.

From the Lorentz transformation we can obtain anew the results already known to us concerning the proper time (§ 3). Suppose a clock to be at rest in the K' system. We take two events occurring at one and the same point x', y', z' in space in the K' system. The time between these events in the K' system is $\Delta t' = t_2' - t_1'$. Now we find the time Δt which elapses between these two events in the K system. From (4.3), we have

$$t_1 = \frac{t_1' + \frac{V}{c^2}x'}{\sqrt{1 - \frac{V^2}{c^2}}}, \quad t_2 = \frac{t_2' + \frac{V}{c^2}x'}{\sqrt{1 - \frac{V^2}{c^2}}},$$

or, subtracting one from the other,

$$t_2 - t_1 = \Delta t = \frac{\Delta t'}{\sqrt{1 - \frac{V^2}{c^2}}},$$

in complete agreement with (3.1).

Finally we mention another general property of Lorentz transformations which distinguishes them from Galilean transformations. The latter have the general property of commutativity, i.e. the combined result of two successive Galilean transformations (with different velocities $\mathbf{V}_1$ and $\mathbf{V}_2$) does not depend on the order in which the transformations are performed. On the other hand, the result of two successive Lorentz transformations does depend, in general, on their order. This is already apparent purely mathematically from our formal description of these transformations as rotations of the four-dimensional coordinate system: we know that the result of two rotations (about different axes) depends on the order in which they are carried out. The sole exception is the case of transformations with parallel vectors $\mathbf{V}_1$ and $\mathbf{V}_2$ (which are equivalent to two rotations of the four-dimensional coordinate system about the same axis).

§ 5. Transformation of velocities

In the preceding section we obtained formulas which enable us to find from the coordinates of an event in one reference frame, the coordinates of the same event in a second reference

frame. Now we find formulas relating the velocity of a material particle in one reference system to its velocity in a second reference system.

Let us suppose once again that the K' system moves relative to the K system with velocity V along the x axis. Let $v_x = dx/dt$ be the component of the particle velocity in the K system and $v'_x = dx'/dt'$ the velocity component of the same particle in the K' system. From (4.3), we have

$$dx = \frac{dx' + Vdt'}{\sqrt{1 - \frac{V^2}{c^2}}}, \quad dy = dy', \quad dz = dz', \quad dt = \frac{dt' + \frac{V}{c^2}dx'}{\sqrt{1 - \frac{V^2}{c^2}}}.$$

Dividing the first three equations by the fourth and introducing the velocities

$$\mathbf{v} = \frac{d\mathbf{r}}{dt}, \qquad \mathbf{v}' = \frac{d\mathbf{r}'}{dt'},$$

we find

$$v_x = \frac{v'_x + V}{1 + v'_x \frac{V}{c^2}}, \qquad v_y = \frac{v'_y \sqrt{1 - \frac{V^2}{c^2}}}{1 + v'_x \frac{V}{c^2}}, \qquad v_z = \frac{v'_z \sqrt{1 - \frac{V^2}{c^2}}}{1 + v'_x \frac{V}{c^2}}. \tag{5.1}$$

These formulas determine the transformation of velocities. They describe the law of composition of velocities in the theory of relativity. In the limiting case of $c \to \infty$, they go over into the formulas $v_x = v'_x + V$, $v_y = v'_y$, $v_z = v'_z$ of classical mechanics.

In the special case of motion of a particle parallel to the X axis, $v_x = v$, $v_y = v_z = 0$. Then $v'_y = v'_z = 0$, $v'_x = v'$, so that

$$v = \frac{v' + V}{1 + v' \frac{V}{c^2}}. \tag{5.2}$$

It is easy to convince oneself that the sum of two velocities each smaller than the velocity of light is again not greater than the light velocity.

For a velocity V significantly smaller than the velocity of light (the velocity v can be arbitrary), we have approximately, to terms of order V/c:

$$v_x = v'_x + V\left(1 - \frac{v'^2_x}{c^2}\right), \quad v_y = v'_y - v'_x v'_y \frac{V}{c^2}, \quad v_z = v'_z - v'_x v'_z \frac{V}{c^2}.$$

These three formulas can be written as a single vector formula

$$\mathbf{v} = \mathbf{v}' + \mathbf{V} - \frac{1}{c^2}(\mathbf{V} \cdot \mathbf{v}')\mathbf{v}'. \tag{5.3}$$

We may point out that in the relativistic-law of addition of velocities (5.1) the two velocities $\mathbf{v}'$ and $\mathbf{V}$ which are combined enter unsymmetrically (provided they are not both directed along the x axis). This fact is related to the noncommutativity of Lorentz transformations which we mentioned in the preceding section.

Let us choose our coordinate axes so that the velocity of the particle at the given moment

lies in the XY plane. Then the velocity of the particle in the K system has components $v_x = v\cos\theta$, $v_y = v\sin\theta$, and in the K' system $v'_x = v'\cos\theta'$, $v'_y = v'\sin\theta'$ (v, v', θ, θ' are the absolute values and the angles subtended with the X, X' axes respectively in the K, K' systems), With the help of formula (5.1), we then find

$$\tan\theta = \frac{v'\sqrt{1-\frac{V^2}{c^2}}\sin\theta'}{v'\cos\theta' + V}. \tag{5.4}$$

This formula describes the change in the direction of the velocity on transforming from one reference system to another.

Let us consider a very important special case of this formula, namely, the deviation of light in transforming to a new reference system—a phenomenon known as the *aberration of light*. In this case $v = v' = c$, so that the preceding formula goes over into

$$\tan\theta = \frac{\sqrt{1-\frac{V^2}{c^2}}}{\frac{V}{c}+\cos\theta'}\sin\theta'. \tag{5.5}$$

From the same transformation formulas (5.1) it is easy to obtain for $\sin\theta$ and $\cos\theta$:

$$\sin\theta = \frac{\sqrt{1-\frac{V^2}{c^2}}}{1+\frac{V}{c}\cos\theta'}\sin\theta', \qquad \cos\theta = \frac{\cos\theta'+\frac{V}{c}}{1+\frac{V}{c}\cos\theta'}. \tag{5.6}$$

In case $V \ll c$, we find from this formula, correct to terms of order V/c:

$$\sin\theta - \sin\theta' = -\frac{V}{c}\sin\theta'\cos\theta'.$$

Introducing the angle $\Delta\theta = \theta' - \theta$ (the aberration angle), we find to the same order of accuracy

$$\Delta\theta = \frac{V}{c}\sin\theta', \tag{5.7}$$

which is the well-known elementary formula for the aberration of light.

§ 6. Four-vectors

The coordinates of an event (ct, x, y, z) can be considered as the components of a four-dimensional radius vector (or, for short, a four-radius vector) in a four-dimensional space. We shall denote its components by x^i, where the index i takes on the values 0, 1, 2, 3, and

$$x^0 = ct, \qquad x^1 = x, \qquad x^2 = y, \qquad x^3 = z.$$

The square of the "length" of the radius four-vector is given by

$$(x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2.$$

It does not change under any rotations of the four-dimensional coordinate system, in particular under Lorentz transformations.

In general a set of four quantities A^0, A^1, A^2, A^3 which transform like the components of the radius four-vector x^i under transformations of the four-dimensional coordinate system is called a *four-dimensional vector* (*four-vector*) A^i. Under Lorentz transformations,

$$A^0 = \frac{A'^0 + \frac{V}{c}A'^1}{\sqrt{1-\frac{V^2}{c^2}}}, \quad A^1 = \frac{A'^1 + \frac{V}{c}A'^0}{\sqrt{1-\frac{V^2}{c^2}}}, \quad A^2 = A'^2, \quad A^3 = A'^3. \tag{6.1}$$

The square magnitude of any four-vector is defined analogously to the square of the radius four-vector:

$$(A^0)^2 - (A^1)^2 - (A^2)^2 - (A^3)^2.$$

For convenience of notation, we introduce two "types" of components of four-vectors, denoting them by the symbols A^i and A_i, with superscripts and subscripts. These are related by

$$A_0 = A^0, \quad A_1 = -A^1, \quad A_2 = -A^2, \quad A_3 = -A^3. \tag{6.2}$$

The quantities A^i are called the *contravariant*, and the A_i the *covariant* components of the four-vector. The square of the four-vector then appears in the form

$$\sum_{i=0}^{3} A^i A_i = A^0 A_0 + A^1 A_1 + A^2 A_2 + A^3 A_3.$$

Such sums are customarily written simply as $A^i A_i$, omitting the summation sign. One agrees that one sums over any repeated index, and omits the summation sign. Of the pair of indices, one must be a superscript and the other a subscript. This convention for summation over "dummy" indices is very convenient and considerably simplifies the writing of formulas.

We shall use Latin letters $i, k, l, \ldots$, for four-dimensional indices, taking on the values 0, 1, 2, 3.

In analogy to the square of a four-vector, one forms the *scalar product* of two different four-vectors:

$$A^i B_i = A^0 B_0 + A^1 B_1 + A^2 B_2 + A^3 B_3.$$

It is clear that this can be written either as $A^i B_i$ or $A_i B^i$—the result is the same. In general one can switch upper and lower indices in any pair of dummy indices.†

The product $A^i B_i$ is a *four-scalar*—it is invariant under rotations of the four-dimensional coordinate system. This is easily verified directly,‡ but it is also apparent beforehand (from the analogy with the square $A^i A_i$) from the fact that all four-vectors transform according to the same rule.

† In the literature the indices are often omitted on four-vectors, and their squares and scalar products are written as A^2, AB. We shall not use this notation in the present text.

‡ One should remember that the law for transformation of a four-vector expressed in covariant components differs (in signs) from the same law expressed for contravariant components. Thus, instead of (6.1), one will have:

$$A_0 = \frac{A'_0 - \frac{V}{c}A'_1}{\sqrt{1-\frac{V^2}{c^2}}}, \quad A_1 = \frac{A'_1 \frac{V}{c} A'_0}{\sqrt{1-\frac{V^2}{c^2}}}, \quad A_2 = A'_2, \quad A_3 = A'_3.$$

The component A^0 is called the *time component,* and A^1, A^2, A^3 the *space components* of the four-vector (in analogy to the radius four-vector). The square of a four-vector can be positive, negative, or zero; such vectors are called, *timelike, spacelike,* and *null-vectors,* respectively (again in analogy to the terminology for intervals).†

Under purely spatial rotations (i.e. transformations not affecting the time axis) the three space components of the four-vector A^i form a three-dimensional vector **A**. The time component of the four-vector is a three-dimensional scalar (with respect to these transformations). In enumerating the components of a four-vector, we shall often write them as

$$A^i = (A^0, \mathbf{A}).$$

The covariant components of the same four-vector are $A_i = (A^0, -\mathbf{A})$, and the square of the four-vector is $A^iA_i = (A^0)^2 - \mathbf{A}^2$. Thus, for the radius four-vector:

$$x^i = (ct, \mathbf{r}), \qquad x_i = (ct, -\mathbf{r}), \qquad x^ix_i = c^2t^2 - \mathbf{r}^2.$$

For three-dimensional vectors (with coordinates x, y, z) there is no need to distinguish between contra- and covariant components. Whenever this can be done without causing confusion, we shall write their components as $A_\alpha (\alpha = x, y, z)$ using Greek letters for subscripts. In particular we shall assume a summation over x, y, z for any repeated index (for example, $\mathbf{A} \cdot \mathbf{B} = A_\alpha B_\alpha$).

A *four-dimensional tensor* (*four-tensor*) of the second rank is a set of sixteen quantities A^{ik}, which under coordinate transformations transform like the products of components of two four-vectors. We similarly define four-tensors of higher rank.

The components of a second-rank tensor can be written in three forms: covariant, A_{ik}, contravariant, A^{ik}, and mixed, A_k^i (where, in the last case, one should distinguish between $A_k{}^i$ and $A_i{}^k$, i.e. one should be careful about which of the two is superscript and which a subscript). The connection between the different types of components is determined from the general rule: raising or lowering a space index (1, 2, 3) changes the sign of the component, while raising or lowering the time index (0) does not. Thus:

$$A_{00} = A^{00}, \quad A_{01} = -A^{01}, \quad A_{11} = A^{11}, \ldots,$$

$$A_0{}^0 = A^{00}, A_0{}^1 = A^{01}, A_1{}^0 = -A^{01}, A_1{}^1 = -A^{11}, \ldots.$$

Under purely spatial transformations, the nine quantities A^{11}, A^{12}, ... form a three-tensor. The three components A^{01}, A^{02}, A^{03} and the three components A^{10}, A^{20}, A^{30} constitute three-dimensional vectors, while the component A^{00} is a three-dimensional scalar.

A tensor A^{ik} is said to be *symmetric* if $A^{ik} = A^{ki}$, and *antisymmetric* if $A^{ik} = -A^{ki}$. In an antisymmetric tensor, all the diagonal components (i.e. the components A^{00}, A^{11}, . . .) are zero, since, for example, we must have $A^{00} = -A^{00}$. For a symmetric tensor A^{ik}, the mixed components $A^i{}_k$ and $A_k{}^i$ obviously coincide; in such cases we shall simply write A_k^i, putting the indices one above the other.

In every tensor equation, the two sides must contain identical and identically placed (i.e. above or below) free indices (as distinguished from dummy indices). The free indices in tensor equations can be shifted up or down, but this must be done simultaneously in all terms in the equation. Equating covariant and contravariant components of different tensors is "illegal"; such an equation, even if it happened by chance to be valid in a particular reference system, would be violated on going to another frame.

† Null vectors are also said to be *isotropic.*

From the tensor components A^{ik} one can form a scalar by taking the sum

$$A^{i}_{\ i} = A^{0}_{\ 0} + A^{1}_{\ 1} + A^{2}_{\ 2} + A^{3}_{\ 3}$$

(where, of course, $A_i^{\ i} = A^i_{\ i}$). This sum is called the *trace* of the tensor, and the operation for obtaining it is called *contraction.*

The formation of the scalar product of two vectors, considered earlier, is a contraction operation: it is the formation of the scalar A^iB_i from the tensor A^iB_k. In general, contracting on any pair of indices reduces the rank of the tensor by 2. For example, $A^i_{\ kli}$ is a tensor of second rank $A^i_{\ k}B^k$ is a four-vector, $A^{ik}_{\ \ ik}$ is a scalar, etc.

The unit four-tensor δ^i_k satisfies the condition that for any four-vector A^i,

$$\delta^k_i A^i = A^k. \tag{6.3}$$

It is clear that the components of this tensor are

$$\delta^k_i = \begin{cases} 1, & \text{if} \quad i = k \\ 0, & \text{if} \quad i \neq k \end{cases} \tag{6.4}$$

Its trace is $\delta^i_i = 4$.

By raising the one index or lowering the other in δ^k_i, we can obtain the contra- or covariant tensor g^{ik} or g_{ik}, which is called the *metric tensor.* The tensors g^{ik} and g_{ik} have identical components, which can be written as a matrix:

$$(g^{ik}) = (g_{ik}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{6.5}$$

(the index i labels the rows, and k the columns, in the order 0, 1, 2, 3). It is clear that

$$g_{ik}A^k = A_i, \quad g^{ik}A_k = A^i. \tag{6.6}$$

The scalar product of two four-vectors can therefore be written in the form:

$$A^iA_i = g_{ik}A^iA^k = g^{ik}A_iA_k. \tag{6.7}$$

The tensors δ^i_k, g_{ik}, g^{ik} are special in the sense that their components are the same in all coordinate systems. The *completely antisymmetric unit tensor* of fourth rank, e^{iklm}, has the same property. This is the tensor whose components change sign under interchange of any pair of indices, and whose nonzero components are ± 1. From the antisymmetry it follows that all components in which two indices are the same are zero, so that the only non-vanishing components are those for which all four indices are different. We set

$$e^{0123} = +1 \tag{6.8}$$

(hence $e_{0123} = -1$). Then all the other nonvanishing components e^{iklm} are equal to +1 or −1, according as the numbers i, k, l, m can be brought to the arrangement 0, 1, 2, 3 by an even or an odd number of transpositions. The number of such components is 4! = 24. Thus,

$$e^{iklm}e_{iklm} = -24. \tag{6.9}$$

With respect to rotations of the coordinate system, the quantities e^{iklm} behave like the components of a tensor; but if we change the sign of one or three of the coordinates the components e^{iklm}, being defined as the same in all coordinate systems, do not change, whereas some of the components of a tensor should change sign. Thus e^{iklm} is, strictly speaking, not a tensor, but rather a *pseudotensor.* Pseudotensors of any rank, in particular *pseudoscalars,* behave like tensors under all coordinate transformations except those that cannot be reduced to rotations, i.e. reflections, which are changes in sign of the coordinates that are not reducible to a rotation.

The products $e^{iklm}e^{prst}$ form a four-tensor of rank 8, which is a true tensor; by contracting on one or more pairs of indices, one obtains tensors of rank 6, 4, and 2. All these tensors have the same form in all coordinate systems. Thus their components must be expressed as combinations of products of components of the unit tensor δ^i_k — the only true tensor whose components are the same in all coordinate systems. These combinations can easily be found by starting from the symmetries that they must possess under permutation of indices.†

If A^{ik} is an antisymmetric tensor, the tensor A^{ik} and the pseudotensor $A^{*ik} = \frac{1}{2}e^{iklm}A_{lm}$ are said to be *dual* to one another. Similarly, $e^{iklm}A_m$ is an antisymmetric pseudotensor of rank 3, dual to the vector A^i. The product $A^{ik}A^*_{ik}$ of dual tensors is obviously a pseudoscalar.

In this connection we note some analogous properties of three-dimensional vectors and tensors. The completely antisymmetric unit pseudotensor of rank 3 is the set of quantities $e_{\alpha\beta\gamma}$ which change sign under any transposition of a pair of indices. The only nonvanishing components of $e_{\alpha\beta\gamma}$ are those with three different indices. We set $e_{xyz} = 1$; the others are 1 or -1, depending on whether the sequence α, β, γ can be brought to the order x, y, z by an even or an odd number of transpositions.‡

† For reference we give the following formulas:

$$e^{iklm}e_{prst} = -\begin{vmatrix} \delta^i_p & \delta^i_r & \delta^i_s & \delta^i_t \\ \delta^k_p & \delta^k_r & \delta^k_s & \delta^k_t \\ \delta^l_p & \delta^l_r & \delta^l_s & \delta^l_t \\ \delta^m_p & \delta^m_r & \delta^m_s & \delta^m_t \end{vmatrix}, \qquad e^{iklm}e_{prsm} = -\begin{vmatrix} \delta^i_p & \delta^i_r & \delta^i_s \\ \delta^k_p & \delta^k_r & \delta^k_s \\ \delta^l_p & \delta^l_r & \delta^l_s \end{vmatrix}$$

$$e^{iklm}e_{prlm} = -2(\delta^i_p\delta^k_r - \delta^i_r\delta^k_p), \qquad e^{iklm}e_{prlm} = -6\delta^i_p.$$

The overall coefficient in these formulas can be checked using the result of a complete contraction, which should give (6.9).

As a consequence of these formulas we have:

$$e^{prst}A_{lp}A_{kr}A_{is}A_{mt} = -Ae_{iklm},$$

$$e^{iklm}e^{prst}A_{tp}A_{kr}A_{ls}A_{mt} = 24A.$$

where A is the determinant formed from the quantities A_{ik}.

‡ The fact that the components of the four-tensor e^{iklm} are unchanged under rotations of the four-dimensional coordinate system, and that the components of the three-tensor $e_{\alpha\beta\gamma}$ are unchanged by rotations of the space axes are special cases of a general rule: any completely antisymmetric tensor of rank equal to the number of dimensions of the space in which it is defined is invariant under rotations of the coordinate system in the space.

The products $e_{\alpha\beta\gamma}e_{\lambda\mu\nu}$ form a true three-dimensional tensor of rank 6, and are therefore expressible as combinations of products of components of the unit three-tensor $\delta_{\alpha\beta}$.†

Under a reflection of the coordinate system, i.e. under a change in sign of all the coordinates, the components of an ordinary vector also change sign. Such vectors are said to be *polar*. The components of a vector that can be written as the cross product of two polar vectors do not change sign under inversion. Such vectors are said to be *axial*. The scalar product of a polar and an axial vector is not a true scalar, but rather a pseudoscalar; it changes sign under a coordinate inversion. An axial vector is a pseudovector, dual to some antisymmetric tensor. Thus, if $\mathbf{C} = \mathbf{A} \times \mathbf{B}$, then

$$C_\alpha = \tfrac{1}{2}e_{\alpha\beta\gamma}C_{\beta\gamma}, \quad \text{where} \quad C_{\beta\gamma} = A_\beta B_\gamma - A_\gamma B_\beta .$$

Now consider four-tensors. The space components ($i, k, = 1, 2, 3$) of the antisymmetric tensor A^{ik} form a three-dimensional antisymmetric tensor with respect to purely spatial transformations; according to our statement its components can be expressed in terms of the components of a three-dimensional axial vector. With respect to these same transformations the components A^{01}, A^{02}, A^{03} form a three-dimensional polar vector. Thus the components of an antisymmetric four-tensor can be written as a matrix:

$$(A^{ik}) = \begin{vmatrix} 0 & p_x & p_y & p_z \\ -p_x & 0 & -a_z & a_y \\ -p_y & a_z & 0 & -a_x \\ -p_z & -a_y & a_x & 0 \end{vmatrix}, \tag{6.10}$$

where, with respect to spatial transformations, **p** and **a** are polar and axial vectors, respectively. In enumerating the components of an antisymmetric four-tensor, we shall write them in the form

$$A^{ik} = (\mathbf{p}, \mathbf{a});$$

then the covariant components of the same tensor are

$$A_{ik} = (-\mathbf{p}, \mathbf{a}).$$

Finally we consider certain differential and integral operations of four-dimensional tensor analysis.

The four-gradient of a scalar ϕ is the four-vector

† For reference, we give the appropriate formulas:

$$e_{\alpha\beta\gamma}e_{\lambda\mu\nu} = \begin{vmatrix} \delta_{\alpha\lambda} & \delta_{\alpha\mu} & \delta_{\alpha\nu} \\ \delta_{\beta\lambda} & \delta_{\beta\mu} & \delta_{\beta\nu} \\ \delta_{\gamma\lambda} & \delta_{\gamma\mu} & \delta_{\gamma\nu} \end{vmatrix}.$$

Contracting this tensor on one, two and three pairs of indices, we get:

$$e_{\alpha\beta\gamma}e_{\lambda\mu\gamma} = \delta_{\alpha\lambda}\delta_{\beta\mu} - \delta_{\alpha\mu}\delta_{\beta\lambda},$$

$$e_{\alpha\beta\gamma}e_{\lambda\beta\gamma} = 2\delta_{\alpha\lambda},$$

$$e_{\alpha\beta\gamma}e_{\alpha\beta\gamma} = 6.$$

$$\frac{\partial\phi}{\partial x^i} = \left(\frac{1}{c}\frac{\partial\phi}{\partial t}, \nabla\phi\right).$$

We must remember that these derivatives are to be regarded as the covariant components of the four-vector. In fact, the differential of the scalar

$$d\phi = \frac{\partial\phi}{\partial x^i}dx^i$$

is also a scalar; from its form (scalar product of two four-vectors) our assertion is obvious.

In general, the operators of differentiation with respect to the coordinates x^i, $\partial/\partial x^i$, should be regarded as the covariant components of the operator four-vector. Thus, for example, the divergence of a four-vector, the expression $\partial A^i/\partial x^i$, in which we differentiate the contravariant components A^i, is a scalar.†

In three-dimensional space one can extend integrals over a volume, a surface or a curve. In four-dimensional space there are four types of integrations:

(1) Integral over a curve in four-space. The element of integration is the line element, i.e. the four-vector dx^i.

(2) Integral over a (two-dimensional) surface in four-space. As we know, in three-space the projections of the area of the parallelogram formed from the vectors $d\mathbf{r}$ and $d\mathbf{r}'$ on the coordinate planes $x_\alpha x_\beta$ are $dx_\alpha dx'_\beta - dx_\beta dx'_\alpha$. Analogously, in four-space the infinitesimal element of surface is given by the antisymmetric tensor of second rank $df^{ik} = dx^i dx'^k - dx^k dx'^i$; its components are the projections of the element of area on the coordinate planes. In three-dimensional space, as we know, one uses as surface element in place of the tensor $df_{\alpha\beta}$ the vector df_α dual to the tensor $df_{\alpha\beta}$: $df_\alpha = \frac{1}{2}e_{\alpha\beta\gamma}df_{\beta\gamma}$. Geometrically this is a vector normal to the surface element and equal in absolute magnitude to the area of the element. In four-space we cannot construct such a vector, but we can construct the tensor df^{*ik} dual to the tensor df^{ik},

$$df^{*ik} = \frac{1}{2}e^{iklm}df_{lm}. \tag{6.11}$$

Geometrically it describes an element of surface equal to and "normal" to the element of

† If we differentiate with respect to the "covariant coordinates" x_i, then the derivatives

$$\frac{\partial\phi}{\partial x_i} = \left(\frac{1}{c}\frac{\partial\phi}{\partial t}, -\nabla\phi\right)$$

form the contravariant components of a four-vector. We shall use this form only in exceptional cases [for example, for writing the square of the four-gradient $(\partial\phi/\partial x^i)/(\partial\phi/\partial x_i)$].

We note that in the literature partial derivatives with respect to the coordinates are often abbreviated using the symbols.

$$\partial^i = \frac{\partial}{\partial x_i}, \quad \partial_i = \frac{\partial}{\partial x^i}.$$

In this form of writing of the differentiation operators, the co- or contravariant character of quantities formed with them is explicit. This same advantage exists for another abbreviated form for writing derivatives, using the index preceded by a comma:

$$\phi_{,i} = \frac{\partial\phi}{\partial x^i}, \quad \phi^{,i} = \frac{\partial\phi}{\partial x_i}.$$

surface df^{ik}; all segments lying in it are orthogonal to all segments in the element df^{ik}. It is obvious that $df^{ik}df^{*}_{ik} = 0$.

(3) Integral over a hypersurface, i.e. over a three-dimensional manifold. In three-dimensional space the volume of the parallelepiped spanned by three vectors is equal to the determinant of the third rank formed from the components of the vectors. One obtains analogously the projections of the volume of the parallelepiped (i.e. the "areas" of the hypersurface) spanned by three four-vectors dx^i, dx'^i, dx''^i; they are given by the determinants

$$dS^{ikl} = \begin{vmatrix} dx^i & dx'^i & dx''^i \\ dx^k & dx'^k & dx''^k \\ dx^l & dx'^l & dx''^l \end{vmatrix},$$

which form a tensor of rank 3, antisymmetric in all three indices. As element of integration over the hypersurface, it is more convenient to use the four-vector dS^i, dual to the tensor dS^{ikl}:

$$dS^i = -\frac{1}{6}e^{iklm}dS_{klm}, \quad dS_{klm} = e_{nklm}dS^n. \tag{6.12}$$

Here

$$dS^0 = dS^{123}, \quad dS^1 = dS^{023}, \ldots$$

Geometrically dS^i is a four-vector equal in magnitude to the "areas" of the hypersurface element, and normal to this element (i.e. perpendicular to all lines drawn in the hypersurface element). In particular, $dS^0 = dx\,dy\,dz$, i.e. it is the element of three-dimensional volume dV, the projection of the hypersurface element on the hyperplane $x^0 = \text{const}$.

(4) Integral over a four-dimensional volume; the element of integration is the scalar

$$d\Omega = dx^0dx^1dx^2\,dx^3 = cdtdV. \tag{6.13}$$

The element is a scalar: it is obvious that the volume of a portion of four-space is unchanged by a rotation of the coordinate system.†

Analogous to the theorems of Gauss and Stokes in three-dimensional vector analysis, there are theorems that enable us to transform four-dimensional integrals.

The integral over a closed hypersurface can be transformed into an integral over the four-volume contained within it by replacing the element of integration dS_i by the operator

$$dS_i \to d\Omega\frac{\partial}{\partial x^i}. \tag{6.14}$$

For example, for the integral of a vector A^i we have:

† Under a transformation from the integration variables x^0, x^1, x^2, x^3 to new variables x'^0, x'^1, x'^2, x'^3, the element of integration changes to $J\,d\Omega'$, where $d\Omega' = dx'^0\,dx'^1\,dx'^2\,dx'^3$

$$J = \frac{\partial(x'^0, x'^1, x'^2, x'^3)}{\partial(x^0, x^1, x^2, x^3)}$$

is the Jacobian of the transformation. For a linear transformation of the form $x'^i = a^i_k x^k$, the Jacobian J coincides with the determinant $|\,a^i_k\,|$ and is equal to unity for rotations of the coordinate system; this shows the invariance of $d\Omega$.

$$\oint A^i dS_i = \int \frac{\partial A^i}{\partial x^i} d\Omega. \tag{6.15}$$

This formula is the generalization of Gauss' theorem.

An integral over a two-dimensional surface is transformed into an integral over the hypersurface "spanning" it by replacing the element of integration df_{ik}^* by the operator

$$df_{ik}^* \to dS_i \frac{\partial}{\partial x^k} - dS_k \frac{\partial}{\partial x^i}. \tag{6.16}$$

For example, for the integral of an antisymmetric tensor A^{ik} we have:

$$\frac{1}{2}\int A^{ik} df_{ik}^* = \frac{1}{2}\int \left(dS_i \frac{\partial A^{ik}}{\partial x^k} - dS_k \frac{\partial A^{ik}}{\partial x^i} \right) = \int dS_i \frac{\partial A^{ik}}{\partial x^k}. \tag{6.17}$$

The integral over a four-dimensional closed curve is transformed into an integral over the surface spanning it by the substitution:

$$dx^i \to df^{ki} \frac{\partial}{\partial x^k}. \tag{6.18}$$

Thus for the integral of a vector, we have:

$$\oint A_i dx^i = \int df^{ki} \frac{\partial A_i}{\partial x^k} = \frac{1}{2}\int df^{ik} \left(\frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k} \right), \tag{6.19}$$

which is the generalization of Stokes' theorem.

PROBLEMS

1. Find the law of transformation of the components of a symmetric four-tensor A^{ik} under Lorentz transformations (6.1).

Solution: Considering the components of the tensor as products of components of two four-vectors, we get:

$$A^{00} = \frac{1}{1-\frac{V^2}{c^2}} \left(A'^{00} + 2\frac{V}{c} A'^{01} + \frac{V^2}{c^2} A'^{11} \right), \quad A^{11} = \frac{1}{1-\frac{V^2}{c^2}} \left(A'^{11} + 2\frac{V}{c} A'^{01} + \frac{V^2}{c^2} A'^{00} \right),$$

$$A^{22} = A'^{22}, \quad A^{23} = A'^{23}, \quad A^{12} = \frac{1}{\sqrt{1-\frac{V^2}{c^2}}} \left(A'^{12} + \frac{V}{c} A'^{02} \right),$$

$$A^{01} = \frac{1}{1-\frac{V^2}{c^2}} \left[A'^{01} \left(1 + \frac{V^2}{c^2} \right) + \frac{V}{c} A'^{00} + \frac{V}{c} + A'^{11} \right],$$

$$A^{02} = \frac{1}{\sqrt{1-\frac{V^2}{c^2}}} \left(A'^{02} + \frac{V}{c} A'^{12} \right),$$

and analogous formulas for A^{33}, A^{13} and A^{03}.

2. The same for the antisymmetric tensor A^{ik}.

Solution: Since the coordinates x^2 and x^3 do not change, the tensor component A^{23} does not change, while the components A^{12}, A^{13} and A^{02}, A^{03} transform like x^1 and x^0:

$$A^{23} = A'^{23}, \quad A^{12} = \frac{A'^{12} + \frac{V}{c}A'^{02}}{\sqrt{1 - \frac{V^2}{c^2}}}, \quad A^{02} = \frac{A'^{02} + \frac{V}{c}A'^{12}}{\sqrt{1 - \frac{V^2}{c^2}}}$$

and similarly for A^{13}, A^{03}.

With respect to rotations of the two-dimensional coordinate system in the plane x^0x^1 (which are the transformations we are considering) the components $A^{01} = -A^{10}$, $A^{00} = A^{11} = 0$, form an antisymmetric of tensor of rank two, equal to the number of dimensions of the space. Thus, (see the remark on p. 19) these components are not changed by the transformations:

$$A^{01} = A'^{01}.$$

§ 7. Four-dimensional velocity

From the ordinary three-dimensional velocity vector one can form a four-vector. This four-dimensional velocity (*four-velocity*) of a particle is the vector

$$u^i = \frac{dx^i}{ds}. \tag{7.1}$$

To find its components, we note that according to (3.1),

$$ds = cdt\sqrt{1 - \frac{v^2}{c^2}},$$

where v is the ordinary three-dimensional velocity of the particle. Thus

$$u^1 = \frac{dx^1}{ds} = \frac{dx}{cdt\sqrt{1 - \frac{v^2}{c^2}}} = \frac{v_x}{c\sqrt{1 - \frac{v^2}{c^2}}},$$

etc. Thus

$$u^i = \left(\frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}, \frac{\mathbf{v}}{c\sqrt{1 - \frac{v^2}{c^2}}}\right). \tag{7.2}$$

Note that the four-velocity is a dimensionless quantity.

The components of the four-velocity are not independent. Noting that $dx_i dx^i = ds^2$, we have

$$u^i u_i = 1. \tag{7.3}$$

Geometrically, u^i is a unit four-vector tangent to the world line of the particle.

Similarly to the definition of the four-velocity, the second derivative

$$w^i = \frac{d^2x^i}{ds^2} = \frac{du^i}{ds}$$

may be called the four-acceleration. Differentiating formula (7.3), we find:

$$u_i w^i = 0, \tag{7.4}$$

i.e. the four-vectors of velocity and acceleration are "mutually perpendicular".

PROBLEM

Determine the relativistic uniformly accelerated motion, i.e. the rectilinear motion for which the acceleration w in the proper reference frame (at each instant of time) remains constant.

Solution: In the reference frame in which the particle velocity is $v = 0$, the components of the four-acceleration $w^i = (0, w/c^2, 0, 0)$ (where w is the ordinary three-dimensional acceleration, which is directed along the x axis). The relativistically invariant condition for uniform acceleration must be expressed by the constancy of the four-scalar which coincides with w^2 in the proper reference frame:

$$w^i w_i = \text{const} \equiv -\frac{w^2}{c^4}.$$

In the "fixed" frame, with respect to which the motion is observed, writing out the expression for $w^i w_i$ gives the equation

$$\frac{d}{dt}\frac{v}{\sqrt{1-\frac{v^2}{c^2}}} = w, \quad \text{or} \quad \frac{v}{\sqrt{1-\frac{v^2}{c^2}}} = wt + \text{const}.$$

Setting $v = 0$ for $t = 0$, we find that const = 0, so that

$$v = \frac{wt}{\sqrt{1+\frac{w^2t^2}{c^2}}}.$$

Integrating once more and setting $x = 0$ for $t = 0$, we find:

$$x = \frac{c^2}{w}\left(\sqrt{1+\frac{w^2t^2}{c^2}} - 1\right).$$

For $wt \ll c$, these formulas go over the classical expressions $v = wt$, $x = wt^2/2$. For $wt \to \infty$, the velocity tends toward the constant value c.

The proper time of a uniformly accelerated particle is given by the integral

$$\int_0^t \sqrt{1-\frac{v^2}{c^2}}\,dt = \frac{c}{w}\sinh^{-1}\left(\frac{wt}{c}\right).$$

As $t \to \infty$, it increases much more slowly than t, according to the law $c/w \ln(2wt/c)$.

CHAPTER 2

RELATIVISTIC MECHANICS

§ 8. The principle of least action

In studying the motion of material particles, we shall start from the Principle of Least Action. The *principle of least action* is defined, as we know, by the statement that for each mechanical system there exists a certain integral S, called the *action*, which has a minimum value for the actual motion, so that its variation δS is zero.†

To determine the action integral for a free material particle (a particle not under the influence of any external force), we note that this integral must not depend on our choice of reference system, that is, it must be invariant under Lorentz transformations. Then it follows that it must depend on a scalar. Furthermore, it is clear that the integrand must be a differential of the first order. But the only scalar of this kind that one can construct for a free particle is the interval ds, or $\alpha\, ds$, where α is some constant. So for a free particle the action must have the form

$$S = -\alpha \int_a^b ds,$$

where ${}_a\int^b$ is an integral along the world line of the particle between the two particular events of the arrival of the particle at the initial position and at the final position at definite times t_1 and t_2, i.e. between two given world points; and α is some constant characterizing the particle. It is easy to see that α must be a positive quantity for all particles. In fact, as we saw in § 3, ${}_a\int^b ds$ has its maximum value along a straight world line; by integrating along a curved world line we can make the integral arbitrarily small. Thus the integral ${}_a\int^b ds$ with the positive sign cannot have a minimum; with the opposite sign it clearly has a minimum, along the straight world line.

The action integral can be represented as an integral with respect to the time

$$S = \int_{t_1}^{t_2} L dt.$$

The coefficient L of dt represents the *Lagrange function* of the mechanical system. With the aid of (3.1), we find:

† Strictly speaking, the principle of least action asserts that the integral S must be a minimum only for infinitesimal lengths of the path of integration. For paths of arbitrary length we can say only that S must be an extremum, not necessarily a minimum. (See *Mechanics*, § 2.)

$$S = -\int_{t_1}^{t_2} \alpha c \sqrt{1 - \frac{v^2}{c^2}}\, dt,$$

where v is the velocity of the material particle. Consequently the Lagrangian for the particle is

$$L = -\alpha c \sqrt{1 - v^2/c^2}\,.$$

The quantity α, as already mentioned, characterizes the particle. In classical mechanics each particle is characterized by its mass m. Let us find the relation between α and m. It can be determined from the fact that in the limit as $c \to \infty$, our expression for L must go over into the classical expression $L = mv^2/2$. To carry out this transition we expand L in powers of v/c. Then, neglecting terms of higher order, we find

$$L = -\alpha c \sqrt{1 - \frac{v^2}{c^2}} \approx -\alpha c + \frac{\alpha v^2}{2c}.$$

Constant terms in the Lagrangian do not affect the equation of motion and can be omitted. Omitting the constant αc in L and comparing with the classical expression $L = mv^2/2$, we find that $\alpha = mc$.

Thus the action for a free material point is

$$S = -mc \int_a^b ds \tag{8.1}$$

and the Lagrangian is

$$L = -mc^2 \sqrt{1 - \frac{v^2}{c^2}}. \tag{8.2}$$

§ 9. Energy and momentum

By the *momentum* of a particle we can mean the vector $\mathbf{p} = \partial L/\partial \mathbf{v}$ ($\partial L/\partial \mathbf{v}$ is the symbolic representation of the vector whose components are the derivatives of L with respect to the corresponding components of $\mathbf{v}$). Using (8.2), we find;

$$\mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{9.1}$$

For small velocities ($v << c$) or, in the limit as $c \to \infty$, this expression goes over into the classical $\mathbf{p} = m\mathbf{v}$. For $v = c$, the momentum becomes infinite.

The time derivative of the momentum is the force acting on the particle. Suppose the velocity of the particle changes only in direction, that is, suppose the force is directed perpendicular to the velocity. Then

$$\frac{d\mathbf{p}}{dt} = \frac{m}{\sqrt{1 - \frac{v^2}{c^2}}} \frac{d\mathbf{v}}{dt}. \tag{9.2}$$

If the velocity changes only in magnitude, that is, if the force is parallel to the velocity, then

$$\frac{d\mathbf{p}}{dt} = \frac{m}{\left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}}} + \frac{d\mathbf{v}}{dt}. \tag{9.3}$$

We see that the ratio of force to acceleration is different in the two cases.

The *energy* $\mathscr{E}$ of the particle is defined as the quantity †

$$\mathscr{E} = \mathbf{p} \cdot \mathbf{v} - L.$$

Substituting the expressions (8.2) and (9.1) for L and $\mathbf{p}$, we find

$$\mathscr{E} = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{9.4}$$

This very important formula shows, in particular, that in relativistic mechanics the energy of a free particle does not go to zero for $v = 0$, but rather takes on a finite value

$$\mathscr{E} = mc^2. \tag{9.5}$$

This quantity is called the *rest energy* of the particle.

For small velocities ($v/c << 1$), we have, expanding (9.4) in series in powers of v/c,

$$\mathscr{E} \approx mc^2 + \frac{mv^2}{2},$$

which, except for the rest energy, is the classical expression for the kinetic energy of a particle.

We emphasize that, although we speak of a "particle", we have nowhere made use of the fact that it is "elementary". Thus the formulas are equally applicable to any composite body consisting of many particles, where by m we mean the total mass of the body and by v the velocity of its motion as a whole. In particular, formula (9.5) is valid for any body which is at rest as a whole. We call attention to the fact that in relativistic mechanics the energy of a free body (i.e. the energy of any closed system) is a completely definite quantity which is always positive and is directly related to the mass of the body. In this connection we recall that in classical mechanics the energy of a body is defined only to within an arbitrary constant, and can be either positive or negative.

The energy of a body at rest contains, in addition to the rest energies of its constituent particles, the kinetic energy of the particles and the energy of their interactions with one another. In other words, mc^2 is not equal to $\sum m_a c^2$ (where m_a are the masses of the particles), and so m is not equal to $\sum m_a$. Thus in relativistic mechanics the law of conservation of mass does not hold: the mass of a composite body is not equal to the sum of the masses of its parts. Instead only the law of conservation of energy, in which the rest energies of the particles are included, is valid.

Squaring (9.1) and (9.4) and comparing the results, we get the following relation between the energy and momentum of particle:

† *See Mechanics*, § 6.

$$\frac{\mathscr{E}^2}{c^2} = p^2 + m^2c^2. \tag{9.6}$$

The energy expressed in terms of the momentum is called the Hamiltonian function $\mathscr{H}$:

$$\mathscr{H} = c\sqrt{p^2 + m^2c^2}. \tag{9.7}$$

For low velocities, $p << mc$, and we have approximately

$$\mathscr{H} \approx mc^2 + \frac{p^2}{2m},$$

i.e., except for the rest energy we get the familiar classical expression for the Hamiltonian.

From (9.1) and (9.4) we get the following relation between the energy, momentum, and velocity of a free particle:

$$\mathbf{p} = \mathscr{E}\frac{\mathbf{v}}{c^2}. \tag{9.8}$$

For $v = c$, the momentum and energy of the particle become infinite. This means that a particle with mass m different from zero cannot move with the velocity of light. Nevertheless, in relativistic mechanics, particles of zero mass moving with the velocity of light can exist.† From (9.8) we have for such particles:

$$P = \frac{\mathscr{E}}{c}. \tag{9.9}$$

The same formula also holds approximately for particles with nonzero mass in the so-called *ultrarelativistic* case, when the particle energy $\mathscr{E}$ is large compared to its rest energy mc^2.

We now write all our formulas in four-dimensional form. According to the principle of least action,

$$\delta S = -mc\delta\int_a^b ds = 0.$$

To set up the expression for δS, we note that $ds = \sqrt{dx_i dx^i}$ and therefore

$$\delta S = -mc\int_a^b \frac{dx_i \delta dx^i}{ds} = -mc\int_a^b u_i d\delta x^i.$$

Integrating by parts, we obtain

$$\delta S = -mcu_i\delta x^i\Big|_a^b + mc\int_a^b \delta x^i \frac{du_i}{ds}ds. \tag{9.10}$$

As we know, to get the equations of motion we compare different trajectories between the same two points, i.e. at the limits $(\delta x^i)_a = (\delta x^i)_b = 0$. The actual trajectory is then determined

† For example, light quanta and neutrinos.

from the condition $\delta S = 0$. From (9.10) we thus obtain the equations $du_i/ds = 0$; that is, a constant velocity for the free particle in four-dimensional form.

To determine the variation of the action as a function of the coordinates, one must consider the point a as fixed, so that $(\delta x^i)_a = 0$. The second point is to be considered as variable, but only actual trajectories are admissible, i.e., those which satisfy the equations of motion. Therefore the integral in expression (9.10) for δS is zero. In place of $(\delta x^i)_b$ we may write simply δx^i, and thus obtain

$$\delta S = - mcu_i \delta x^i. \tag{9.11}$$

The four-vector

$$p_i = - \frac{\partial S}{\partial x^i} \tag{9.12}$$

is called the *momentum four-vector.* As we know from mechanics, the derivatives $\partial S/\partial x$, $\partial S/\partial y$, $\partial S/\partial z$ are the three components of the momentum vector $\mathbf{p}$ of the particle, while the derivative $-\partial S/\partial t$ is the particle energy $\mathscr{E}$. Thus the covariant components of the four-mementum-are $p_i = (\mathscr{E}/c, -\mathbf{p})$, while the contravariant components are†

$$p^i = (\mathscr{E}/c, \mathbf{p}). \tag{9.13}$$

From (9.11) we see that the components of the four-momentum of a free particle are:

$$p^i = mcu^i. \tag{9.14}$$

Substituting the components of the four-velocity from (7.2), we see that we actually get expressions (9.1) and (9.4) for $\mathbf{p}$ and $\mathscr{E}$.

Thus, in relativistic mechanics, momentum and energy are the components of a single four-vector. From this we immediately get the formulas for transformation of momentum and energy from one inertial system to another. Substituting (9.13) in the general formulas (6.1) for transformation of four-vectors, we find:

$$p_x = \frac{p'_x + \frac{V}{c^2}\mathscr{E}'}{\sqrt{1 - \frac{V^2}{c^2}}}, \quad p_y = p'_y, \quad p_z = p'_z, \quad \mathscr{E} = \frac{\mathscr{E}' + Vp'_x}{\sqrt{1 - \frac{V^2}{c^2}}}, \tag{9.15}$$

where p_x, p_y, p_z are the components of the three-dimensional vector $\mathbf{p}$.

From the definition (9.14) of the four-momentum, and the identity $u^i u_i = 1$, we have, for the square of the four-momentum of a free particle:

$$p_i p^i = m^2 c^2. \tag{9.16}$$

Substituting the expressions (9.13), we get back (9.6).

By analogy with the usual definition of the force, the force four-vector is defined as the derivative:

$$g^i = \frac{dp^i}{ds} = mc \frac{du^i}{ds}. \tag{9.17}$$

† We call attention to a mnemonic for remembering the definition of the physical four-vectors: the *contravariant* components are related to the corresponding three-dimensional vectors ($\mathbf{r}$ for x^i, $\mathbf{p}$ for p^i) with the "right", positive sign.

Its components satisfy the identity $g_i u^i = 0$. The components of this four-vector are expressed in terms of the usual three-dimensional force vector $\mathbf{f} = d\mathbf{p}/dt$:

$$g^i = \left(\frac{\mathbf{f} \cdot \mathbf{v}}{c^2 \sqrt{1 - \frac{v^2}{c^2}}}, \frac{\mathbf{f}}{c\sqrt{1 - \frac{v^2}{c^2}}} \right). \tag{9.18}$$

The time component is related to the work done by the force.

The relativistic Hamilton–Jacobi equation is obtained by substituting the derivatives $-\partial S/\partial x^i$ for p_i in (9.16):

$$\frac{\partial S}{\partial x_i} \frac{\partial S}{\partial x^i} \equiv g^{ik} \frac{\partial S}{\partial x^i} \frac{\partial S}{\partial x^k} = m^2 c^2, \tag{9.19}$$

or, writing the sum explicitly:

$$\frac{1}{c^2}\left(\frac{\partial S}{\partial t}\right)^2 - \left(\frac{\partial S}{\partial x}\right)^2 - \left(\frac{\partial S}{\partial y}\right)^2 - \left(\frac{\partial S}{\partial z}\right)^2 = m^2 c^2. \tag{9.20}$$

The transition to the limiting case of classical mechanics in equation (9.19) is made as follows. First of all we must notice that just as in the corresponding transition with (9.7), the energy of a particle in relativistic mechanics contains the term mc^2, which it does not in classical mechanics. Inasmuch as the action S is related to the energy by $\mathscr{E} = -(\partial S/\partial t)$, in making the transition to classical mechanics we must in place of S substitute a new action S' according to the relation:

$$S = S' - mc^2 t.$$

Substituting this in (9.20), we find

$$\frac{1}{2m}\left[\left(\frac{\partial S'}{\partial x}\right)^2 + \left(\frac{\partial S'}{\partial y}\right)^2 + \left(\frac{\partial S'}{\partial z}\right)^2\right] - \frac{1}{2mc^2}\left(\frac{\partial S'}{\partial t}\right)^2 + \frac{\partial S'}{\partial t} = 0.$$

In the limit as $c \to \infty$, this equation goes over into the classical Hamilton–Jacobi equation.

§ 10. Transformation of distribution functions

In various physical problems we have to deal with distribution functions for the momenta of particles: $f(\mathbf{p})dp_x dp_y dp_z$ is the number of particles having momenta with components in given intervals dp_x, dp_y, dp_z (or, as we say for brevity, the number of particles in a given volume element $d^3p \equiv dp_x dp_y dp_z$ in "momentum space"). We are then faced with the problem of finding the law of transformation of the distribution function $f(\mathbf{p})$ when we transform from one reference system to another.

To solve this problem, we first determine the properties of the "volume element" $dp_x dp_y dp_z$ with respect to Lorentz transformations. If we introduce a four-dimensional coordinate system, on whose axes are marked the components of the four-momentum of a particle, then $dp_x dp_y dp_z$ can be considered as the zeroth component of an element of the hypersurface defined by the equation $p^i p_i = m^2 c^2$. The element of hypersurface is a four-vector directed

along the normal to the hypersurface; in our case the direction of the normal obviously coincides with the direction of the four-vector p_i. From this it follows that the ratio

$$\frac{dp_x dp_y dp_z}{\mathscr{E}} \tag{10.1}$$

is an invariant quantity, since it is the ratio of corresponding components of two parallel four-vectors.†

The number of particles, $f dp_x dp_y dp_z$, is also obviously an invariant, since it does not depend on the choice of reference frame. Writing it in the form

$$f(\mathbf{p})\mathscr{E}\frac{dp_x dp_y dp_z}{\mathscr{E}}$$

and using the invariance of the ratio (10.1), we conclude that the product $f(\mathbf{p})\mathscr{E}$ is invariant. Thus the distribution function in the K' system is related to the distribution function in the K system by the formula

$$f'(\mathbf{p}') = \frac{f(\mathbf{p})\mathscr{E}}{\mathscr{E}'}, \tag{10.2}$$

where $\mathbf{p}$ and $\mathscr{E}$ must be expressed in terms of $\mathbf{p}'$ and $\mathscr{E}'$ by using the transformation formulas (9.15).

Let us now return to the invariant expression (10.1). If we introduce "spherical coordinates" in momentum space, the volume element $dp_x dp_y dp_z$ becomes $p^2 dp\, do$, where do is the element of solid angle around the direction of the vector $\mathbf{p}$. Noting that $pdp = \mathscr{E}d\mathscr{E}/c^2$ [from (9.6)], we have:

$$\frac{p^2 dp do}{\mathscr{E}} = \frac{pd\mathscr{E}\, do}{c^2}.$$

Thus we find that the expression

$$pd\mathscr{E}\, do \tag{10.3}$$

is also invariant.

The notion of a distribution function appears in a different aspect in the kinetic theory of gases: the product $f(\mathbf{r}, \mathbf{p})dp_x dp_y dp_z dV$ is the number of particles lying in a given volume element dV and having momenta in definite intervals dp_x, dp_y, dp_z. The function $f(\mathbf{r}, \mathbf{p})$ is

† The integration with respect to the element (10.1) can be expressed in four-dimensional form by means of the δ-function (cf. the footnote on p. 74) as an integration with respect to

$$\frac{2}{c}\delta(p_i p^i - m^2c^2)d^4p, \qquad d^4p = dp^0 dp^1 dp^2 dp^3. \tag{10.1a}$$

The four components p^i are treated as independent variables (with p^0 taking on only positive values). Formula (10.1a) is obvious from the following representation of the delta function appearing in it:

$$\delta(p^i p_i - m^2c^2) = \delta\left((p_0)^2 - \frac{\mathscr{E}^2}{c^2}\right) = \frac{c}{2\mathscr{E}}\left[\delta\left(p_0 + \frac{\mathscr{E}}{c}\right) + \delta\left(p_0 - \frac{\mathscr{E}}{c}\right)\right], \tag{10.1b}$$

where $\mathscr{E} = c\sqrt{p^2 + m^2c^2}$. This formula in turn follows from formula (V) of the footnote on p. 74.

called the distribution function in *phase space* (the space of the coordinates and momenta of the particle), and the product of differentials $d\tau = d^3p\, dV$ is the element of volume of this space. We shall find the law of transformation of this function.

In addition to the two reference systems K and K', we also introduce the frame K_0 in which the particles with the given momentum are at rest; the proper volume dV_0 of the element occupied by the particles is defined relative to this system. The velocities of the systems K and K' relative to the system K_0 coincide, by definition, with the velocities v and v' which these particles have in the systems K and K'. Thus, according to (4.6), we have

$$dV = dV_0\sqrt{1 - \frac{v^2}{c^2}}, \quad dV' = dV_0\sqrt{1 - \frac{v'^2}{c^2}},$$

from which

$$\frac{dV}{dV'} = \frac{\mathscr{E}'}{\mathscr{E}}.$$

Multiplying this equation by the equation $d^3p/d^3p' = \mathscr{E}/\mathscr{E}'$, we find that

$$d\tau = d\tau', \tag{10.4}$$

i.e. the element of phase volume is invariant. Since the number of particles $f\, d\tau$ is also invariant, by definition, we conclude that the distribution function in phase space is an invariant:

$$f'(\mathbf{r}', \mathbf{p}') = f(\mathbf{r}, \mathbf{p}), \tag{10.5}$$

where $\mathbf{r}'$, $\mathbf{p}'$ are related to $\mathbf{r}$, $\mathbf{p}$ by the formulas for the Lorentz transformation.

§ 11. Decay of particles

Let us consider the spontaneous decay of a body of mass M into two parts with masses m_1 and m_2. The law of conservation of energy in the decay, applied in the system of reference in which the body is at rest, gives†

$$M = \mathscr{E}_{10} + \mathscr{E}_{20}. \tag{11.1}$$

where $\mathscr{E}_{10}$ and $\mathscr{E}_{20}$ are the energies of the emerging particles. Since $\mathscr{E}_{10} > m_1$ and $\mathscr{E}_{20} > m_2$, the equality (11.1) can be satisfied only if $M > m_1 + m_2$, i.e. a body can disintegrate spontaneously into parts the sum of whose masses is less than the mass of the body. On the other hand, if $M < m_1 + m_2$, the body is stable (with respect to the particular decay) and does not decay spontaneously. To cause the decay in this case, we would have to supply to the body from outside an amount of energy at least equal to its "binding energy" $(m_1 + m_2 - M)$.

Momentum as well as energy must be conserved in the decay process. Since the initial momentum of the body was zero, the sum of the momenta of the emerging particles must be zero: $\mathbf{p}_{10} + \mathbf{p}_{20} = 0$. Consequently $p_{10}^2 = p_{20}^2$, or

† In §§ 11–13 we set $c = 1$. In other words the velocity of light is taken as the unit of velocity (so that the dimensions of length and time become the same). This choice is a natural one in relativistic mechanics and greatly simplifies the writing of formulas. However, in this book (which also contains a considerable amount of nonrelativistic theory) we shall not usually use this system of units, and will explicitly indicate when we do.

If c has been set equal to unity in formulas, it is easy to convert back to ordinary units: the velocity is introduced to assure correct dimensions.

$$\mathscr{E}_{10}^2 - m_1^2 = \mathscr{E}_{20}^2 - m_2^2. \tag{11.2}$$

The two equations (11.1) and (11.2) uniquely determine the energies of the emerging particles:

$$\mathscr{E}_{10} = \frac{M^2 + m_1^2 - m_2^2}{2M}, \qquad \mathscr{E}_{20} = \frac{M^2 - m_1^2 + m_2^2}{2M}. \tag{11.3}$$

In a certain sense the inverse of this problem is the calculation of the total energy M of two colliding particles in the system of reference in which their total momentum is zero. (This is abbreviated as the "system of the centre of inertia" or the "C-system".) The computation of this quantity gives a criterion for the possible occurrence of various inelastic collision processes, accompanied by a change in state of the colliding particles, or the "creation" of new particles. A process of this type can occur only if the sum of the masses of the "reaction products" does not exceed M.

Suppose that in the initial reference system (the "laboratory" system) a particle with mass m_1 and energy $\mathscr{E}_1$ collides with a particle of mass m_2 which is at rest. The total energy of the two particles is

$$\mathscr{E} = \mathscr{E}_1 + \mathscr{E}_2 = \mathscr{E}_1 + m_2,$$

and their total momentum is $\mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}_1$. Considering the two particles together as a single composite system, we find the velocity of its motion as a whole from (9.8):

$$\mathbf{V} = \frac{\mathbf{p}}{\mathscr{E}} = \frac{\mathbf{p}_1}{\mathscr{E}_1 + m_2}. \tag{11.4}$$

This quantity is the velocity of the C-system with respect to the laboratory system (the L-system).

However, in determining the mass M, there is no need to transform from one reference frame to the other. Instead we can make direct use of formula (9.6), which is applicable to the composite system just as it is to each particle individually. We thus have

$$M^2 = \mathscr{E}^2 - p^2 = (\mathscr{E}_1 + m_2)^2 - (\mathscr{E}_1^2 - m_1^2),$$

from which

$$M^2 = m_1^2 + m_2^2 + 2m_2\mathscr{E}_1. \tag{11.5}$$

PROBLEMS

1. A particle moving with velocity V dissociates "in flight" into two particles. Determine the relation between the angles of emergence of these particles and their energies.

Solution: Let $\mathscr{E}_0$ be the energy of one of the decay particles in the C-system [i.e. $\mathscr{E}_{10}$ or $\mathscr{E}_{20}$ in (11.3)], $\mathscr{E}$ the energy of this same particle in the L-system, and θ its angle of emergence in the L-system (with respect to the direction of $\mathbf{V}$). By using the transformation formulas we find:

$$\mathscr{E}_0 = \frac{\mathscr{E} - Vp\cos\theta}{\sqrt{1 - V^2}},$$

so that

$$\cos\theta = \frac{\mathscr{E} - \mathscr{E}_0\sqrt{1 - V^2}}{V\sqrt{\mathscr{E}^2 - m^2}}. \tag{1}$$

For the determination of $\mathscr{E}$ from cos θ we then get the quadratic equation

$$\mathscr{E}^2(1 - V^2\cos^2\theta) - 2\mathscr{E}\mathscr{E}_0\sqrt{1 - V^2} + \mathscr{E}_0^2(1 - V^2) + V^2m^2\cos^2\theta = 0, \quad (2)$$

which has one positive root (if the velocity v_0 of the decay particle in the C-system satisfies $v_0 > V$) or two positive roots (if $v_0 < V$).

The source of this ambiguity is clear from the following graphical construction. According to (9.15), the momentum components in the L-system are expressed in terms of quantities referring to the C-system by the formulas

$$p_z = \frac{p_0\cos\theta_0 + \mathscr{E}_0V}{\sqrt{1 - V^2}}, \quad p_y = p_0\sin\theta_0.$$

Eliminating θ_0, we get

$$p_y^2 + (p_x\sqrt{1 - V^2} - \mathscr{E}_0V)^2 = p_0^2.$$

With respect to the variables p_x, p_y, this is the equation of an ellipse with semiaxes $p_0/\sqrt{1 - V^2}, p_0$, whose centre (the point O in Fig. 3) has been shifted a distance $\mathscr{E}_0V/\sqrt{1 - V^2}$ from the point $\mathbf{p} = 0$ (point A in Fig. 3).†

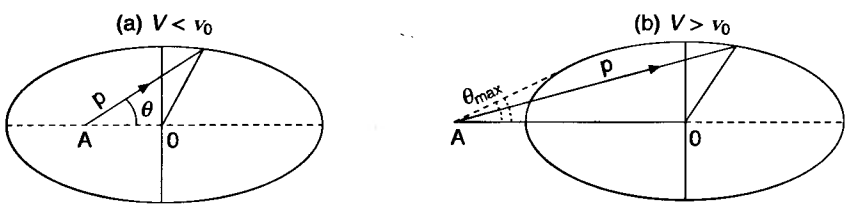


FIG. 3.

If $V > p_0/\mathscr{E}_0 = v_0$, the point A lies outside the ellipse (Fig. 3b), so that for a fixed angle θ the vector $\mathbf{p}$ (and consequently the energy $\mathscr{E}$) can have two different values. It is also clear from the construction that in this case the angle θ cannot exceed a definite value θ_{max} (corresponding to the position of the vector $\mathbf{p}$ in which it is tangent to the ellipse). The value of θ_{max} is most easily determined analytically from the condition that the discriminant of the quadratic equation (2) go to zero:

$$\sin\theta_{max} = \frac{p_0\sqrt{1 - V^2}}{mV}.$$

2. Find the energy distribution of the decay particles in the L-system.

Solution: In the C-system the decay particles are distributed isotropically in direction, i.e. the number of particles within the element of solid angle $do_0 = 2\pi\sin\theta_0\,d\theta_0$ is

$$dN = \frac{1}{4\pi}do_0 = \frac{1}{2}|\,d\cos\theta_0\,|. \quad (1)$$

The energy in the L-system is given in terms of quantities referring to the C-system by

$$\mathscr{E} = \frac{\mathscr{E}_0 + p_0V\cos\theta_0}{\sqrt{1 - V^2}}$$

and runs through the range of values from

$$\frac{\mathscr{E}_0 - Vp_0}{\sqrt{1 - V^2}} \text{ to } \frac{\mathscr{E}_0 + Vp_0}{\sqrt{1 - V^2}}.$$

† In the classical limit, the ellipse reduces to a circle. (See *Mechanics*, § 16.)

Expressing $d|\cos\theta_0|$ in terms of $d\sigma$, we obtain the normalized energy distribution (for each of the two types of decay particles):

$$dN = \frac{1}{2Vp_0}\sqrt{1-V^2}\,d\sigma.$$

3. Determine the range of values in the L-system for the angle between the two decay particles (their separation angle) for the case of decay into two identical particles.

Solution: In the C-system, the particles fly off in opposite directions, so that $\theta_{10} = \pi - \theta_{20} \equiv \theta_0$. According to (5.4), the connection between angles in the C- and L-systems is given by the formulas:

$$\cot\theta_1 = \frac{v_0\cos\theta_0 + V}{v_0\sin\theta_0\sqrt{1-V^2}}, \qquad \cot\theta_2 = \frac{-v_0\cos\theta_0 + V}{v_0\sin\theta_0\sqrt{1-V^2}}$$

(since $v_{10} = v_{20} = v_0$ in the present case). The required separation angle is $\Theta = \theta_1 + \theta_2$, and a simple calculation gives:

$$\cot\Theta = \frac{V^2 - v_0^2 + V^2 v_0^2\sin^2\theta_0}{2Vv_0\sqrt{1-V^2}\sin\theta_0}.$$

An examination of the extreme for this expression gives the following ranges of possible values of Θ:

$$\text{for } V < v_0\text{: } 2\tan^{-1}\left(\frac{v_0}{V}\sqrt{1-V^2}\right) < \Theta < \pi;$$

$$\text{for } v_0 < V < \frac{v_0}{\sqrt{1-v_0^2}}\text{: } 0 < \Theta < \sin^{-1}\sqrt{\frac{1-V^2}{1-v_0^2}} < \frac{\pi}{2};$$

$$\text{for } V > \frac{v_0}{\sqrt{1-v_0^2}}\text{: } 0 < \Theta < 2\tan^{-1}\left(\frac{v_0}{V}\sqrt{1-V^2}\right) < \frac{\pi}{2}.$$

4. Find the angular distribution in the L-system for decay particles of zero mass.

Solution: According to (5.6) the connection between the angles of emergence in the C- and L-systems for particles with $m = 0$ is

$$\cos\theta_0 = \frac{\cos\theta - V}{1 - V\cos\theta}.$$

Substituting this expression in formula (1) of Problem 2, we find:

$$dN = \frac{(1-V^2)\,do}{4\pi(1-V\cos\theta)^2}.$$

5. Find the distribution of separation angles in the L-system for a decay into two particles of zero mass.

Solution: The relation between the angles of emergence, θ_1, θ_2 in the L-system and the angles $\theta_{10} \equiv \theta_0$, $\theta_{20} = \pi - \theta_0$ in the C-system is given by (5.6), so that we have for the separation angle $\Theta = \theta_1 + \theta_2$:

$$\cos\Theta = \frac{2V^2 - 1 - V^2\cos^2\theta_0}{1 - V^2\cos^2\theta_0}$$

and conversely,

$$\cos\theta_0 = \sqrt{1 - \frac{1-V^2}{V^2}\cot^2\frac{\Theta}{2}}.$$

Substituting this expression in formula (1) of problem 2, we find:

$$dN = \frac{1 - V^2}{16\pi V \sin^3 \frac{\Theta}{2} \sqrt{V^2 - \cos^2 \frac{\Theta}{2}}} do.$$

The angle Θ takes on values from π to $\Theta_{\min} = 2\cos^{-1} V$.

6. Determine the maximum energy which can be carried off by one of the decay particles, when a particle of mass M at rest decays into three particles with masses m_1, m_2, and m_3.

Solution: The particle m_1 has its maximum energy if the system of the other two particles m_2 and m_3 has the least possible mass; the latter is equal to the sum $m_2 + m_3$ (and corresponds to the case where the two particles move together with the same velocity). Having thus reduced the problem to the decay of a body into two parts, we obtain from (11.3):

$$\mathscr{E}_{1\max} = \frac{M^2 + m_1^2 - (m_2 + m_3)^2}{2M}.$$

§ 12. Invariant cross-section

Collision processes are characterized by their *invariant cross-sections,* which determine the number of collisions (of the particular type) occurring between beams of colliding particles.

Suppose that we have two colliding beams; we denote by n_1 and n_2 the particle densities in them (i.e. the numbers of particles per unit volume) and by $\mathbf{v}_1$ and $\mathbf{v}_2$ the velocities of the particles. In the reference system in which particle 2 is at rest (or, as one says, in the *rest frame* of particle 2), we are dealing with the collision of the beam of particles 1 with a stationary target. Then according to the usual definition of the cross-section σ, the number of collisions occurring in volume dV in time dt is

$$dv = \sigma v_{\text{rel}} n_1 n_2 dV dt.$$

where v_{rel} is the velocity of particle 1 in the rest system of particle 2 (which is just the definition of the relative velocity of two particles in relativistic mechanics).

The number dv is by its very nature an invariant quantity. Let us try to express it in a form which is applicable in any reference system:

$$dv = An_1 n_2 dV dt, \tag{12.1}$$

where A is a number to be determined, for which we know that its value in the rest frame of one of the particles is $v_{\text{rel}}\sigma$. We shall always mean by σ precisely the cross-section in the rest frame of one of the particles, i.e. by definition, an invariant quantity. From its definition, the relative velocity v_{rel} is also invariant.

In the expression (12.1) the product $dVdt$ is an invariant. Therefore the product An_1n_2 must also be an invariant.

The law of transformation of the particle density n is easily found by noting that the number of particles in a given volume element dV, ndV, is invariant. Writing $ndV = n_0 dV_0$ (the index 0 refers to the rest frame of the particles) and using formula (4.6) for the transformation of the volume, we find:

$$n = \frac{n_0}{\sqrt{1 - v^2}} \tag{12.2}$$

or $n = n_0 \mathscr{E}/m$, where $\mathscr{E}$ is the energy and m the mass of the particles.

Thus the statement that An_1n_2 is invariant is equivalent to the invariance of the expression $A\mathscr{E}_1\mathscr{E}_2$. This condition is more conveniently represented in the form

$$A\frac{\mathscr{E}_1\mathscr{E}_2}{p_{1i}p_2^i} = A\frac{\mathscr{E}_1\mathscr{E}_2}{\mathscr{E}_1\mathscr{E}_2 - \mathbf{p}_1\cdot\mathbf{p}_2} = \text{inv}, \tag{12.3}$$

where the denominator is an invariant—the product of the four-momenta of the two particles.

In the rest frame of particle 2, we have $\mathscr{E}_2 = m_2$, $\mathbf{p}_2 = 0$, so that the invariant quantity (12.3) reduces to A. On the other hand, in this frame $A = \sigma v_{\text{rel}}$. Thus in an arbitrary reference system,

$$A = \sigma v_{\text{rel}}\frac{p_{1i}p_2^i}{\mathscr{E}_1\mathscr{E}_2}. \tag{12.4}$$

To give this expression its final form, we express v_{rel} in terms of the momenta or velocities of the particles in an arbitrary reference frame. To do this we note that in the rest frame of particle 2,

$$p_{1i}p_2^i = \frac{m_1}{\sqrt{1-v_{\text{rel}}^2}}m_2.$$

Then

$$v_{\text{rel}} = \sqrt{1-\frac{m_1^2m_2^2}{(p_{1i}p_2^i)^2}}. \tag{12.5}$$

Expressing the quantity $p_{1i}p_2^i = \mathscr{E}_1\mathscr{E}_2 - \mathbf{p}_1\cdot\mathbf{p}_2$ in terms of the velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ by using formulas (9.1) and (9.4):

$$p_{1i}p_2^i = m_1m_2\frac{1-\mathbf{v}_1\cdot\mathbf{v}_2}{\sqrt{(1-v_1^2)(1-v_2^2)}},$$

and substituting in (12.5), after some simple transformations we get the following expression for the relative velocity:

$$v_{\text{rel}} = \frac{\sqrt{(\mathbf{v}_1-\mathbf{v}_2)^2 - (\mathbf{v}_1\times\mathbf{v}_2)^2}}{1-\mathbf{v}_1\cdot\mathbf{v}_2} \tag{12.6}$$

(we note that this expression is symmetric in $\mathbf{v}_1$ and $\mathbf{v}_2$, i.e. the magnitude of the relative velocity is independent of the choice of particle used in defining it).

Substituting (12.5) or (12.6) in (12.4) and then in (12.1), we get the final formulas for solving our problem:

$$dv = \sigma\frac{\sqrt{(p_{1i}p_2^i)^2 - m_1^2m_2^2}}{\mathscr{E}_1\mathscr{E}_2}n_1n_2dVdt \tag{12.7}$$

or

$$dv = \sigma\sqrt{(\mathbf{v}_1-\mathbf{v}_2)^2 - (\mathbf{v}_1\times\mathbf{v}_2)^2}\,n_1n_2dVdt \tag{12.8}$$

(W. Pauli, 1933).

If the velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ are collinear, then $\mathbf{v}_1\times\mathbf{v}_2 = 0$, so that formula (12.8) takes the form:

$$dv = \sigma|\,\mathbf{v}_1-\mathbf{v}_2\,|\,n_1n_2dVdt. \tag{12.9}$$

PROBLEM

Find the "element of length" in relativistic "velocity space".

Solution: The required line element dl_v is the relative velocity of two points with velocities $\mathbf{v}$ and $\mathbf{v} + d\mathbf{v}$. We therefore find from (12.6)

$$dl_v^2 = \frac{(d\mathbf{v})^2 - (\mathbf{v} \times d\mathbf{v})^2}{(1 - v^2)^2} = \frac{dv^2}{(1 - v^2)^2} + \frac{v^2}{1 - v^2}(d\theta^2 + \sin^2\theta \cdot d\phi^2),$$

where θ, ϕ are the polar angle and azimuth of the direction of $\mathbf{v}$. If in place of v we introduce the new variable χ through the equation $v = \tanh\chi$, the line element is expressed as:

$$dl_v^2 = d\chi^2 + \sinh^2\chi(d\theta^2 + \sin^2\theta \cdot d\phi^2).$$

From the geometrical point of view this is the line element in three-dimensional Lobachevskii space—the space of constant negative curvature (see (111.12)).

§ 13. Elastic collisions of particles

Let us consider, from the point of view of relativistic mechanics, the *elastic collision* of particles. We denote the momenta and energies of the two colliding particles (with masses m_1 and m_2) by $\mathbf{p}_1$, $\mathscr{E}_1$ and $\mathbf{p}_2$, $\mathscr{E}_2$; we use primes for the corresponding quantities after collision. The laws of conservation of momentum and energy in the collision can be written together as the equation for conservation of the four-momentum:

$$p_1^i + p_2^i = p_1'^i + p_2'^i. \tag{13.1}$$

From this four-vector equation we construct invariant relations which will be helpful in further computations. To do this we rewrite (13.1) in the form:

$$p_1^i + p_2^i - p_1'^i + p_2'^i,$$

and square both sides (i.e. we write the scalar product of each side with itself). Noting that the squares of the four-momenta p_1^i and $p_1'^i$ are equal to m_1^2, and the squares of p_2^i and $p_2'^i$ are equal to m_2^2, we get:

$$m_1^2 + p_{1i}p_2^i - p_{1i}p_1'^i - p_{2i}p_1'^i = 0. \tag{13.2}$$

Similarly, squaring the equation $p_1^i + p_2^i - p_2'^i = p_1'^i$, we get:

$$m_2^2 + p_{1i}p_2^i - p_2^i p_2'^i - p_{1i}p_2'^i = 0. \tag{13.3}$$

Let us consider the collision in a reference system (the L-system) in which one of the particles (m_2) was at rest before the collision. Then $\mathbf{p}_2 = 0$, $\mathscr{E}_2 = m_2$, and the scalar products appearing in (13.2) are:

$$\begin{aligned} p_{1i}p_2^i &= \mathscr{E}_1 m_2, \\ p_{2i}p_1'^i &= m_2\mathscr{E}_1', \\ p_{1i}p_1'^i &= \mathscr{E}_1\mathscr{E}_1' - \mathbf{p}_1 \cdot \mathbf{p}_1' = \mathscr{E}_1\mathscr{E}_1' - p_1 p_1' \cos\theta_1, \end{aligned} \tag{13.4}$$

where θ_1 is the angle of scattering of the incident particle m_1. Substituting these expressions in (13.2) we get:

$$\cos\theta_1 = \frac{\mathscr{E}_1'(\mathscr{E}_1 + m_2) - \mathscr{E}_1 m_2 - m_1^2}{p_1 p_1'}. \tag{13.5}$$

Similarly, we find from (13.3):

$$\cos\theta_2 = \frac{(\mathscr{E}_1 + m_2)(\mathscr{E}_2' - m_2)}{p_1 p_2'}, \tag{13.6}$$

where θ_2 is the angle between the transferred momentum $\mathbf{p}_2'$ and the momentum of the incident particle $\mathbf{p}_1$.

The formulas (13.5)–(13.6) relate the angles of scattering of the two particles in the L-system to the changes in their energy in the collision. Inverting these formulas, we can express the energies $\mathscr{E}_1', \mathscr{E}_2'$ in terms of the angles θ_1 or θ_2. Thus, substituting in (13.6) $p_1 = \sqrt{\mathscr{E}_1^2 - m_1^2}$, $p_2' = \sqrt{(\mathscr{E}_2')^2 - m_2^2}$ and squaring both sides, we find after a simple computation:

$$\mathscr{E}_2' = m_2 \frac{(\mathscr{E}_1 + m_2)^2 + (\mathscr{E}_1^2 - m_1^2)\cos^2\theta_2}{(\mathscr{E}_1 + m_2)^2 - (\mathscr{E}_1^2 - m_1^2)\cos^2\theta_2}. \tag{13.7}$$

Inversion of formula (13.5) leads in the general case to a very complicated formula for $\mathscr{E}_1'$ in terms of θ_1.

We note that if $m_1 > m_2$, i.e. if the incident particle is heavier than the target particle, the scattering angle θ_1 cannot exceed a certain maximum value. It is easy to find by elementary computations that this value is given by the equation

$$\sin\theta_{1\,\max} = \frac{m_2}{m_1}, \tag{13.8}$$

which coincides with the familiar classical result.

Formulas (13.5)–(13.6) simplify in the case when the incident particle has zero mass: $m_1 = 0$, and correspondingly $p_1 = \mathscr{E}_1$, $p_1' = \mathscr{E}_1'$. For this case let us write the formula for the energy of the incident particle after the collision, expressed in terms of its angle of deflection:

$$\mathscr{E}_1' = \frac{m_2}{1 - \cos\theta_1 + \dfrac{m_2}{\mathscr{E}_1}}. \tag{13.9}$$

Let us now turn once again to the general case of collision of particles of arbitrary mass. The collision is most simply treated in the C-system. Designating quantities in this system by the additional subscript 0, we have $\mathbf{p}_{10} = -\mathbf{p}_{20} \equiv \mathbf{p}_0$. From the conservation of momentum, during the collision the momenta of the two particles merely rotate, remaining equal in magnitude and opposite in direction. From the conservation of energy, the value of each of the momenta remains unchanged.

Let χ be the angle of scattering in the C-system—the angle through which the momenta $\mathbf{p}_{10}$ and $\mathbf{p}_{20}$ are rotated by the collision. This quantity completely determines the scattering process in the C-system, and therefore also in any other reference system. It is also convenient in describing the collision in the L-system and serves as the single parameter which remains undetermined after the conservation of momentum and energy are applied.

We express the final energies of the two particles in the L-system in terms of this parameter. To do this we return to (13.2), but this time write out the product $p_{1i}p_1'^{i}$ in the C-system:

$$p_{1i}p_1'^i = \mathscr{E}_{10}\mathscr{E}_{10}' - \mathbf{p}_{10}\cdot\mathbf{p}_{10}' = \mathscr{E}_{10}^2 - p_0^2\cos\chi = p_0^2(1-\cos\chi) + m_1^2$$

(in the C-system the energies of the particles do not change in the collision: $\mathscr{E}_{10}' = \mathscr{E}_{10}$). We write out the other two products in the L-system, i.e we use (13.4). As a result we get: $\mathscr{E}_1' - \mathscr{E}_1 = -(p_0^2/m_2)(1-\cos\chi)$. We must still express p_0^2 in terms of quantities referring to the L-system. This is easily done by equating the values of the invariant $p_{1i}p_2^i$ in the L- and C-systems:

$$\mathscr{E}_{10}\mathscr{E}_{20} - \mathbf{p}_{10}\cdot\mathbf{p}_{20} = \mathscr{E}_1 m_2,$$

or

$$\sqrt{(p_0^2+m_1^2)(p_0^2+m_2^2)} = \mathscr{E}_1 m_2 - p_0^2.$$

Solving the equation for p_0^2, we get:

$$p_0^2 = \frac{m_2^2(\mathscr{E}_1^2 - m_1^2)}{m_1^2 + m_2^2 + 2m_2\mathscr{E}_1}. \tag{13.10}$$

Thus, we finally have:

$$\mathscr{E}_1' = \mathscr{E}_1 - \frac{m_2(\mathscr{E}_1^2 - m_1^2)}{m_1^2 + m_2^2 + 2m_2\mathscr{E}_1}(1-\cos\chi). \tag{13.11}$$

The energy of the second particle is obtained from the conservation law: $\mathscr{E}_1 + m_2 = \mathscr{E}_1' + \mathscr{E}_2'$. Therefore

$$\mathscr{E}_2' = m_2 + \frac{m_2(\mathscr{E}_1^2 - m_1^2)}{m_1^2 + m_2^2 + 2m_2\mathscr{E}_1}(1-\cos\chi). \tag{13.12}$$

The second terms in these formulas represent the energy lost by the first particle and transferred to the second particle. The maximum energy transfer occurs for $\chi = \pi$, and is equal to

$$\mathscr{E}_{2\max}' - m_2 = \mathscr{E}_1 - \mathscr{E}_{1\min}' = \frac{2m_2(\mathscr{E}_1^2 - m_1^2)}{m_1^2 + m_2^2 + 2m_2\mathscr{E}_1}. \tag{13.13}$$

The ratio of the minimum kinetic energy of the incident particle after collision to its initial energy is:

$$\frac{\mathscr{E}_{1\min}' - m_1}{\mathscr{E}_1 - m_1} = \frac{(m_1 - m_2)^2}{m_1^2 + m_2^2 + 2m_2\mathscr{E}_1}. \tag{13.14}$$

In the limiting case of low velocities (when $\mathscr{E} \approx m + mv^2/2$), this relation tends to a constant limit, equal to

$$\left(\frac{m_1 - m_2}{m_1 + m_2}\right)^2.$$

In the opposite limit of large energies $\mathscr{E}_1$, relation (13.14) tends to zero; the quantity $\mathscr{E}_{1\min}'$ tends to a constant limit. This limit is

$$\mathscr{E}_{1\min}' = \frac{m_1^2 + m_2^2}{2m_2}.$$

Let us assume that $m_2 >> m_1$, i.e. the mass of the incident particle is small compared to the mass of the particle at rest. According to classical mechanics the light particle could transfer only a negligible part of its energy (see *Mechanics*, § 17). This is not the case in relativistic mechanics. From formula (13.14) we see that for sufficiently large energies $\mathscr{E}_1$ the fraction of the energy transferred can reach the order of unity. For this it is not sufficient that the velocity of m_1 be of order 1, but one must have $\mathscr{E}_1 \sim m_2$, i.e. the light particle must have an energy of the order of the rest energy of the heavy particle.

A similar situation occurs for $m_2 << m_1$, i.e. when a heavy particle is incident on a light one. Here too, according to classical mechanics, the energy transfer would be insignificant. The fraction of the energy transferred begins to be significant only for energies $\mathscr{E}_1 \sim m_1^2/m_2$. We note that we are not taking simply of velocities of the order of the light velocity, but of energies large compared to m_1, i.e. we are dealing with the ultrarelativistic case.

PROBLEMS

1. The triangle ABC in Fig. 4 is formed by the momentum vector $\mathbf{p}$ of the impinging particle and the momenta $\mathbf{p}_1'$, $\mathbf{p}_2'$ of the two particles after the collision. Find the locus of the points C corresponding to all possible values of $\mathbf{p}_1'$, $\mathbf{p}_2'$.

Solution: The required curve is an ellipse whose semiaxes can be found by using the formulas obtained in problem 1 of § 11. In fact, the construction given there determined the locus of the vectors $\mathbf{p}$ in the L-system which are obtained from arbitrarily directed vectors $\mathbf{p}_0$ with given length p_0 in the C-system.

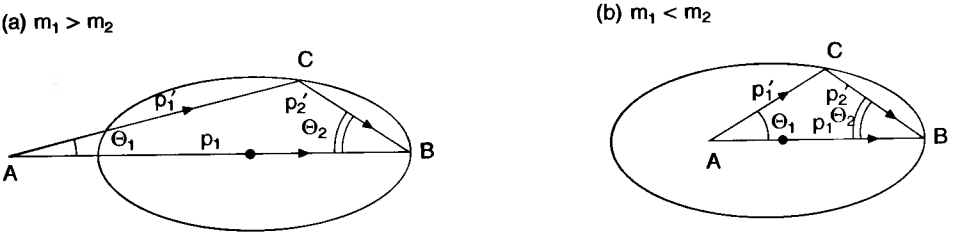


FIG. 4.

Since the absolute values of the momenta of the colliding particles are identical in the C-system, and do not change in the collision, we are dealing with a similar construction for the vector $\mathbf{p}_1'$, for which

$$p_0 \equiv p_{10} = p_{20} = \frac{m_2 V}{\sqrt{1 - V^2}}$$

in the C-system where V is the velocity of particle m_2 in the C-system, coincides in magnitude with the velocity of the centre of inertia, and is equal to $V = p_1/(\mathscr{E}_1 + m_2)$ (see (11.4)). As a result we find that the minor and major semiaxes of the ellipse are

$$p_0 = \frac{m_2 p_1}{\sqrt{m_1^2 + m_2^2 + 2m_2\mathscr{E}_1}}, \quad \frac{p_0}{\sqrt{1 - V^2}} = \frac{m_2 p_1 (\mathscr{E}_1 + m_2)}{m_1^2 + m_2^2 + 2m_2\mathscr{E}_1}$$

(the first of these is, of course, the same as (13.10)).

For $\theta_1 = 0$, the vector $\mathbf{p}_1'$ coincides with $\mathbf{p}_1$, so that the distance AB is equal to p_1. Comparing p_1 with the length of the major axis of the ellipse, it is easily shown that the point A lies outside the ellipse if $m_1 > m_2$ (Fig. 4a), and inside it if $m_1 < m_2$ (Fig. 4b).

2. Determine the minimum separation angle $\Theta_{\min}$ of two particles after collision of the masses of the two particles are the same ($m_1 = m_2 \equiv m$).

Solution: If $m_1 = m_2$, the point A of the diagram lies on the ellipse, while the minimum separation angle corresponds to the situation where point C is at the end of the minor axis (Fig. 5). From the construction it is clear that $\tan(\Theta_{\min}/2)$ is the ratio of the lengths of the semiaxes, and we find:

$$\tan\frac{\Theta_{\min}}{2} = \sqrt{\frac{2m}{\mathscr{E}_1 + m}},$$

or

$$\cos\Theta_{\min} = \frac{\mathscr{E}_1 - m}{\mathscr{E}_1 + 3m}.$$

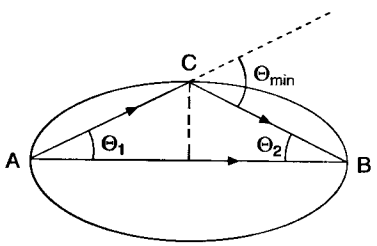


FIG. 5.

3. For the collision of two particles of equal mass m, express $\mathscr{E}_1', \mathscr{E}_2', \chi$ in terms of the angle θ_1 of scattering in the L-system.

Solution: Inversion of formula (13.5) in this case gives:

$$\mathscr{E}_1' = m\frac{(\mathscr{E}_1 + m) + (\mathscr{E}_1 - m)\cos^2\theta_1}{(\mathscr{E}_1 + m) - (\mathscr{E}_1 - m)\cos^2\theta_1}, \qquad \mathscr{E}_2' = m + \frac{(\mathscr{E}_1^2 - m^2)\sin^2\theta_1}{2m + (\mathscr{E}_1 - m)\sin^2\theta_1}.$$

Comparing with the expression for $\mathscr{E}_1'$ in terms of χ:

$$\mathscr{E}_1' = \mathscr{E}_1 - \frac{\mathscr{E}_1 - m}{2}(1 - \cos\chi),$$

we find the angle of scattering in the C-system:

$$\cos\chi = \frac{2m - (\mathscr{E}_1 + 3m)\sin^2\theta_1}{2m + (\mathscr{E}_1 - m)\sin^2\theta_1}.$$

§ 14. Angular momentum

As is well known from classical mechanics, for a closed system, in addition to conservation of energy and momentum, there is conservation of angular momentum, that is, of the vector

$$\mathbf{M} = \sum \mathbf{r} \times \mathbf{p}$$

where $\mathbf{r}$ and $\mathbf{p}$ are the radius vector and momentum of the particle; the summation runs over all the particles making up the system. The conservation of angular momentum is a consequence of the fact that because of the isotropy of space, the Lagrangian of a closed system does not change under a rotation of the system as a whole.

By carrying through a similar derivation in four-dimensional form, we obtain the relativistic expression for the angular momentum. Let x^i be the coordinates of one of the particles of the system. We make an infinitesimal rotation in the four-dimensional space. Under such a

transformation, the coordinates x^i take on new values x'^i such that the differences $x'^i - x^i$ are linear functions

$$x'^i - x^i = x_k \delta\Omega^{ik} \tag{14.1}$$

with infinitesimal coefficients $\delta\Omega_{ik}$. The components of the four-tensor $\delta\Omega_{ik}$ are connected to one another by the relations resulting from the requirement that, under a rotation, the length of the radius vector must remain unchanged, that is, $x'_i x'^i = x_i x^i$. Substituting for x'^i from (14.1) and dropping terms quadratic in $\delta\Omega_{ik}$, as infinitesimals of higher order, we find

$$x^i x^k \delta\Omega_{ik} = 0.$$

This equation must be fulfilled for arbitrary x^i. Since $x^i x^k$ is a symmetric tensor, $\delta\Omega_{ik}$ must be an antisymmetric tensor (the product of a symmetrical and an antisymmetrical tensor is clearly identically zero). Thus we find that

$$\delta\Omega_{ki} = -\delta\Omega_{ik}. \tag{14.2}$$

The change in the action for an infinitesimal change of coordinates of the initial point a and the final point b of the trajectory has the form (see 9.11):

$$\delta S = -\Sigma p^i \delta x_i \Big|_a^b$$

(the summation extends over all the particles of the system). In the case of rotation which we are now considering, $\delta x_i = \delta\Omega_{ik} x^k$, and so

$$\delta S = -\delta\Omega_{ik} \Sigma p^i x^k \Big|_a^b.$$

If we resolve the tensor $\Sigma p^i x^k$ into symmetric and antisymmetric parts, then the first of these when multiplied by an antisymmetric tensor gives identically zero. Therefore, taking the antisymmetric part of $\Sigma p^i x^k$, we can write the preceding equality in the form

$$\delta S = -\delta\Omega_{ik} \tfrac{1}{2} \Sigma (p^i x^k - p^k x^i) \Big|_a^b . \tag{14.3}$$

For a closed system the action, being an invariant, is not changed by a rotation in 4-space. This means that the coefficients of $\delta\Omega_{ik}$ in (14.3) must vanish:

$$\Sigma(p^i x^k - p^k x^i)_b = \Sigma(p^i x^k - p^k x^i)_a .$$

Consequently we see that for a closed system the tensor

$$M^{ik} = \Sigma(x^i p^k - x^k p^i) . \tag{14.4}$$

This antisymmetric tensor is called the *four-tensor of angular momentum*. The space components of this tensor are the components of the three-dimensional angular momentum vector $\mathbf{M} = \Sigma \mathbf{r} \times \mathbf{p}$:

$$M^{23} = M_x, \quad -M^{13} = M_y, \quad M^{12} = M_z.$$

The components M^{01}, M^{02}, M^{03} form a vector $\Sigma(t\mathbf{p} - \mathscr{E}\mathbf{r}/c^2)$. Thus, we can write the components of the tensor M^{ik} in the form:

$$M^{ik} = \left[c\Sigma \left(t\mathbf{p} - \frac{\mathscr{E}\mathbf{r}}{c^2} \right), -\mathbf{M} \right]. \tag{14.5}$$

(Compare (6.10).)

Because of the conservation of M^{ik} for a closed system, we have, in particular,

$$\Sigma\left(t\mathbf{p}-\frac{\mathscr{E}\mathbf{r}}{c^2}\right)=\text{const.}$$

Since, on the other hand, the total energy $\Sigma\mathscr{E}$ is also conserved, this equality can be written in the form

$$\frac{\Sigma\mathscr{E}\mathbf{r}}{\Sigma\mathscr{E}}-\frac{c^2\,\Sigma\mathbf{p}}{\Sigma\mathscr{E}}t=\text{const.}$$

(Quantities referring to different particles are taken at the same time t).

From this we see that the point with the radius vector

$$\mathbf{R}=\frac{\Sigma\mathscr{E}\mathbf{r}}{\Sigma\mathscr{E}} \tag{14.6}$$

moves uniformly with the velocity

$$\mathbf{V}=\frac{c^2\,\Sigma\mathbf{p}}{\Sigma\mathscr{E}}, \tag{14.7}$$

which is none other than the velocity of motion of the system as a whole. [It relates the total energy and momentum, according to formula (9.8).] Formula (14.6) gives the relativistic definition of the coordinates of the *centre of inertia* of the system. If the velocities of all the particles are small compared to c, we can approximately set $\mathscr{E}\approx mc^2$ so that (14.6) goes over into the usual classical expression

$$\mathbf{R}=\frac{\Sigma m\mathbf{r}}{\Sigma m}.\ \dagger$$

We note that the components of the vector (14.6) do not constitute the space components of any four-vector, and therefore under a transformation of reference frame they do not transform like the coordinates of a point. Thus we get different points for the centre of inertia of a given system with respect to different reference frames.

PROBLEM

Find the connection between the angular momentum $\mathbf{M}$ of a body (system of particles) in the reference frame K in which the body moves with velocity $\mathbf{V}$, and its angular momentum $\mathbf{M}^{(0)}$ in the frame K_0 in which the body is at rest as a whole; in both cases the angular momentum is defined with respect to the same point—the centre of inertia of the body in the system K_0.‡

† We note that whereas the classical formula for the centre of inertia applies equally well to interacting and non-interacting particles, formula (14.6) is valid only if we neglect interaction. In relativistic mechanics, the definition of the centre of inertia of a system of interacting particles requires us to include explicitly the momentum and energy of the field produced by the particles.

‡ We remind the reader that although in the system K_0 (in which $\Sigma\mathbf{p}=0$) the angular momentum is independent of the choice of the point with respect to which it is defined, in the K system (in which $\Sigma\mathbf{p}\neq 0$) the angular momentum does depend on this choice (see *Mechanics*, § 9).

Solution: The K_0 system moves relative to the K system with velocity $\mathbf{V}$; we choose its direction for the x axis. The components of M^{ik} that we want transform according to the formulas (see problem 2 in § 6):

$$M^{12} = \frac{M^{(0)12} + \frac{V}{c} M^{(0)02}}{\sqrt{1 - \frac{V^2}{c^2}}}, \quad M^{13} = \frac{M^{(0)13} + \frac{V}{c} M^{(0)03}}{\sqrt{1 - \frac{V^2}{c^2}}}, \quad M^{23} = M^{(0)23}.$$

Since the origin of coordinates was chosen at the centre of inertia of the body (in the K_0 system), in that system $\sum \varepsilon \mathbf{r} = 0$, and since in that system $\sum \mathbf{p} = 0$, $M^{(0)02} = M^{(0)03} = 0$. Using the connection between the components of M^{ik} and the vector $\mathbf{M}$, we find for the latter:

$$M_z = M_x^{(0)}, \quad M_y = \frac{M_y^{(0)}}{\sqrt{1 - \frac{V^2}{c^2}}}, \quad M_z = \frac{M_z^{(0)}}{\sqrt{1 - \frac{V^2}{c^2}}}.$$

CHAPTER 3

CHARGES IN ELECTROMAGNETIC FIELDS

§ 15. Elementary particles in the theory of relativity

The interaction of particles can be described with the help of the concept of a *field* of force. Namely, instead of saying that one particle acts on another, we may say that the particle creates a field around itself; a certain force then acts on every other particle located in this field. In classical mechanics, the field is merely a mode of description of the physical phenomenon—the interaction of particles. In the theory of relativity, because of the finite velocity of propagation of interactions, the situation is changed fundamentally. The forces acting on a particle at a given moment are not determined by the positions at that same moment. A change in the position of one of the particles influences other particles only after the lapse of a certain time interval. This means that the field itself acquires physical reality. We cannot speak of a direct interaction of particles located at a distance from one another. Interactions can occur at any one moment only between neighbouring points in space (contact interaction). Therefore we must speak of the interaction of the one particle with the field, and of the subsequent interaction of the field with the second particle.

We shall consider two types of fields, gravitational and electromagnetic. The study of gravitational fields is left to Chapters 10 to 14 and in the other chapters we consider only electromagnetic fields.

Before considering the interactions of particles with the electromagnetic field, we shall make some remarks concerning the concept of a "particle" in relativistic mechanics.

In classical mechanics one can introduce the concept of a rigid body, i.e., a body which is not deformable under any conditions. In the theory of relativity it should follow similarly that we would consider as rigid those bodies whose dimensions all remain unchanged in the reference system in which they are at rest. However, it is easy to see that the theory of relativity makes the existence of rigid bodies impossible in general.

Consider, for example, a circular disk rotating around its axis, and let us assume that it is rigid. A reference frame fixed in the disk is clearly not inertial. It is possible, however, to introduce for each of the infinitesimal elements of the disk an inertial system in which this element would be at rest at the moment; for different elements of the disk, having different velocities, these systems will, of course, also be different. Let us consider a series of line elements, lying along a particular radius vector. Because of the rigidity of the disk, the length of each of these segments (in the corresponding inertial system of reference) will be the same as it was when the disk was at rest. This same length would be measured by an observer at rest, past whom this radius swings at the given moment, since each of its segments is perpendicular to its velocity and consequently a Lorentz contraction does not occur. Therefore the total length of the radius as measured by the observer at rest, being the

sum of its segments, will be the same as when the disk was at rest. On the other hand, the length of each element of the circumference of the disk, passing by the observer at rest at a given moment, undergoes a Lorentz contraction, so that the length of the whole circumference (measured by the observer at rest as the sum of the lengths of its various segments) turns out to be smaller than the length of the circumference of the disk at rest. Thus we arrive at the result that due to the rotation of the disk, the ratio of circumference to radius (as measured by an observer at rest) must change, and not remain equal to 2π. The absurdity of this result shows that actually the disk cannot be rigid, and that in rotation it must necessarily undergo some complex deformation depending on the elastic properties of the material of the disk.

The impossibility of the existence of rigid bodies can be demonstrated in another way. Suppose some solid body is set in motion by an external force acting at one of its points. If the body were rigid, all of its points would have to be set in motion at the same time as the point to which the force is applied; if this were not so the body would be deformed. However, the theory of relativity makes this impossible, since the force at the particular point is transmitted to the others with a finite velocity, so that all the points cannot begin moving simultaneously.

From this discussion we can draw certain conclusions concerning the treatment of "*elementary*" particles, i.e. particles whose state we assume to be described completely by giving its three coordinates and the three components of its velocity as a whole. It is obvious that if an elementary particle had finite dimensions, i.e. if it were extended in space, it could not be deformable, since the concept of deformability is related to the possibility of independent motion of individual parts of the body. But, as we have seen, the theory of relativity shows that it is impossible for absolutely rigid bodies to exist.

Thus we come to the conclusion that in classical (non-quantum) relativistic mechanics, we cannot ascribe finite dimensions to particles which we regard as elementary. In other words, within the framework of classical theory elementary particles must be treated as points.†

§ 16. Four-potential of a field

For a particle moving in a given electromagnetic field, the action is made up of two parts: the action (8.1) for the free particle, and a term describing the interaction of the particle with the field. The latter term must contain quantities characterizing the particle and quantities characterizing the field.

It turns out‡ that the properties of a particle with respect to interaction with the electromagnetic field are determined by a single parameter—the *charge e* of the particle, which can be either positive or negative (or equal to zero). The properties of the field are characterized by a four-vector A_i, the *four-potential,* whose components are functions of the coordinates and time. These quantities appear in the action function in the term

† Quantum mechanics makes a fundamental change in this situation, but here again relativity theory makes it extremely difficult to introduce anything other than point interactions.

‡ The assertions which follow should be regarded as being, to a certain extent, the consequence of experimental data. The form of the action for a particle in an electromagnetic field cannot be fixed on the basis of general considerations alone (such as, for example, the requirement of relativistic invariance). The latter would permit the occurrence in formula (16.1) of terms of the form $\int A\,ds$, where A is a scalar function.

To avoid any misunderstanding, we repeat that we are considering classical (and not quantum) theory, and therefore do not include effects which are related to the spins of particles.

$$-\frac{e}{c}\int_a^b A_i dx^i,$$

where the functions A_i are taken at points on the world line of the particle. The factor $1/c$ has been introduced for convenience. It should be pointed out that, so long as we have no formulas relating the charge or the potentials with already known quantities, the units for measuring these new quantities can be chosen arbitrarily.†

Thus the action function for a charge in an electromagnetic field has the form

$$S = \int_a^b \left(-mcds - \frac{e}{c}A_i dx^i\right). \tag{16.1}$$

The three space components of the four-vector A^i form a three-dimensional vector $\mathbf{A}$ called the *vector potential* of the field. The time component is called the *scalar potential;* we denote it by $A^0 = \phi$. Thus

$$A^i = (\phi, \mathbf{A}). \tag{16.2}$$

Therefore the action integral can be written in the form

$$S = \int_a^b \left(-mcds + \frac{e}{c}\mathbf{A}\cdot d\mathbf{r} - e\phi dt\right).$$

Introducing $d\mathbf{r}/dt = \mathbf{v}$, and changing to an integration over t,

$$S = \int_{t_1}^{t_2} \left(-mc^2\sqrt{1-\frac{v^2}{c^2}} + \frac{e}{c}\mathbf{A}\cdot\mathbf{v} - e\phi\right)dt. \tag{16.3}$$

The integrand is just the Lagrangian for a charge in an electromagnetic field:

$$L = -mc^2\sqrt{1-\frac{v^2}{c^2}} + \frac{e}{c}\mathbf{A}\cdot\mathbf{v} - e\phi. \tag{16.4}$$

This function differs from the Lagrangian for a free particle (8.2) by the terms $(e/c)\,\mathbf{A}\cdot\mathbf{v} - e\phi$, which describe the interaction of the charge with the field.

The derivative $\partial L/\partial\mathbf{v}$ is the generalized momentum of the particle; we denote it by $\mathbf{P}$. Carrying out the differentiation, we find

$$\mathbf{P} = \frac{m\mathbf{v}}{\sqrt{1-\frac{v^2}{c^2}}} + \frac{e}{c}\mathbf{A} = \mathbf{p} + \frac{e}{c}\mathbf{A}. \tag{16.5}$$

Here we have denoted by $\mathbf{p}$ the ordinary momentum of the particle, which we shall refer to simply as its momentum.

From the Lagrangian we can find the Hamiltonian function for a particle in a field from the general formula

† Concerning the establishment of these units, see § 27.

$$\mathscr{H} = \mathbf{v} \cdot \frac{\partial L}{\partial \mathbf{v}} - L.$$

Substituting (16.4), we get

$$\mathscr{H} = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} + e\phi. \tag{16.6}$$

However, the Hamiltonian must be expressed not in terms of the velocity, but rather in terms of the generalized momentum of the particle.

From (16.5) and (16.6) it is clear that the relation between $\mathscr{H} - e\phi$ and $\mathbf{P} - (e/c)\mathbf{A}$ is the same as the relation between $\mathscr{H}$ and $\mathbf{p}$ in the absence of the field, i.e.

$$\left(\frac{\mathscr{H} - e\phi}{c}\right)^2 = m^2c^2 + \left(\mathbf{P} - \frac{e}{c}\mathbf{A}\right)^2, \tag{16.7}$$

or else

$$\mathscr{H} = \sqrt{m^2c^4 + c^2\left(\mathbf{P} - \frac{e}{c}\mathbf{A}\right)^2} + e\phi. \tag{16.8}$$

For low velocities, i.e. for classical mechanics, the Lagrangian (16.4) goes over into

$$L = \frac{mv^2}{2} + \frac{e}{c}\mathbf{A} \cdot \mathbf{v} - e\phi. \tag{16.9}$$

In this approximation

$$\mathbf{p} = m\mathbf{v} = \mathbf{P} - \frac{e}{c}\mathbf{A},$$

and we find the following expression for the Hamiltonian:

$$\mathscr{H} = \frac{1}{2m}\left(\mathbf{P} - \frac{e}{c}\mathbf{A}\right)^2 + e\phi. \tag{16.10}$$

Finally we write the Hamilton–Jacobi equation for a particle in an electromagnetic field. It is obtained by replacing, in the equation for the Hamiltonian, $\mathbf{P}$ by $\partial S/\partial \mathbf{r}$, and $\mathscr{H}$ by $-(\partial S/\partial t)$. Thus we get from (16.7)

$$\left(\nabla S - \frac{e}{c}\mathbf{A}\right)^2 - \frac{1}{c^2}\left(\frac{\partial S}{\partial t} + e\phi\right)^2 + m^2c^2 = 0. \tag{16.11}$$

§ 17. Equations of motion of a charge in a field

A charge located in a field not only is subjected to a force exerted by the field, but also in turn acts on the field, changing it. However, if the charge e is not large, the action of the charge on the field can be neglected. In this case, when considering the motion of the charge in a given field, we may assume that the field itself does not depend on the coordinates or the velocity of the charge. The precise conditions which the charge must fulfil in order to be

considered as small in the present sense, will be clarified later on (see § 75). In what follows we shall assume that this condition is fulfilled.

So we must find the equations of motion of a charge in a given electromagnetic field. These equations are obtained by varying the action, i.e. they are given by the Lagrange equations

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \mathbf{v}}\right) = \frac{\partial L}{\partial \mathbf{r}}, \tag{17.1}$$

where L is given by formula (16.4).

The derivative $\partial L/\partial \mathbf{v}$ is the generalized momentum of the particle (16.5). Further, we write

$$\frac{\partial L}{\partial \mathbf{r}} = \nabla L = \frac{e}{c}\operatorname{grad} \mathbf{A}\cdot\mathbf{v} - e\operatorname{grad}\phi.$$

But from a formula of vector analysis.

$$\operatorname{grad}(\mathbf{a}\cdot\mathbf{b}) = (\mathbf{a}\cdot\nabla)\mathbf{b} + (\mathbf{b}\cdot\nabla)\mathbf{a} + \mathbf{b}\times\operatorname{curl}\mathbf{a} + \mathbf{a}\times\operatorname{curl}\mathbf{b},$$

where $\mathbf{a}$ and $\mathbf{b}$ are two arbitrary vectors. Applying this formula to $\mathbf{A}\cdot\mathbf{v}$, and remembering that differentiation with respect to $\mathbf{r}$ is carried out for constant $\mathbf{v}$, we find

$$\frac{\partial L}{\partial \mathbf{r}} = \frac{e}{c}(\mathbf{v}\cdot\nabla)\mathbf{A} + \frac{e}{c}\mathbf{v}\times\operatorname{curl}\mathbf{A} - e\operatorname{grad}\phi.$$

So the Lagrange equation has the form:

$$\frac{d}{dt}\left(\mathbf{p} + \frac{e}{c}\mathbf{A}\right) = \frac{e}{c}(\mathbf{v}\cdot\nabla)\mathbf{A} + \frac{e}{c}\mathbf{v}\times\operatorname{curl}\mathbf{A} - e\operatorname{grad}\phi.$$

But the total differential $(d\mathbf{A}/dt)\,dt$ consists of two parts: the change $(\partial\mathbf{A}/\partial t)\,dt$ of the vector potential with time at a fixed point in space, and the change due to motion from one point in space to another at distance $d\mathbf{r}$. This second part is equal to $(d\mathbf{r}\cdot\nabla)\mathbf{A}$. Thus

$$\frac{d\mathbf{A}}{dt} = \frac{\partial \mathbf{A}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{A}.$$

Substituting this in the previous equation, we find

$$\frac{d\mathbf{p}}{dt} = -\frac{e}{c}\frac{\partial \mathbf{A}}{\partial t} - e\operatorname{grad}\phi + \frac{e}{c}\mathbf{v}\times\operatorname{curl}\mathbf{A}. \tag{17.2}$$

This is the equation of motion of a particle in an electromagnetic field. On the left side stands the derivative of the particle's momentum with respect to the time. Therefore the expression on the right of (17.2) is the force exerted on the charge in an electromagnetic field. We see that this force consists of two parts. The first part (first and second terms on the right side of 17.2) does not depend on the velocity of the particle. The second part (third term) depends on the velocity, being proportional to the velocity and perpendicular to it.

The force of the first type, per unit charge, is called the *electric field intensity;* we denote it by $\mathbf{E}$. So by definition,

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} - \operatorname{grad}\phi. \tag{17.3}$$

The factor of $\mathbf{v}/c$ in the force of the second type, per unit charge, is called the *magnetic field intensity*. We designate it by $\mathbf{H}$. So by definition,

$$\mathbf{H} = \operatorname{curl}\mathbf{A}. \tag{17.4}$$

If in an electromagnetic field, $\mathbf{E} \neq 0$ but $\mathbf{H} = 0$, then we speak of an electric field; if $\mathbf{E} = 0$ but $\mathbf{H} \neq 0$, then the field is said to be magnetic. In general, the electromagnetic field is a superposition of electric and magnetic fields.

We note that $\mathbf{E}$ is a polar vector while $\mathbf{H}$ is an axial vector.

The equation of motion of a charge in an electromagnetic field can now be written as

$$\frac{d\mathbf{p}}{dt} = e\mathbf{E} + \frac{e}{c}\mathbf{v} \times \mathbf{H}. \tag{17.5}$$

The expression on the right is called the *Lorentz force*. The first term (the force which the electric field exerts on the charge) does not depend on the velocity of the charge, and is along the direction of $\mathbf{E}$. The second part (the force exerted by the magnetic field on the charge) is proportional to the velocity of the charge and is directed perpendicular to the velocity and to the magnetic field $\mathbf{H}$.

For velocities small compared with the velocity of light, the momentum $\mathbf{p}$ is approximately equal to its classical expression $m\mathbf{v}$, and the equation of motion (17.5) becomes

$$m\frac{d\mathbf{v}}{dt} = e\mathbf{E} + \frac{e}{c}\mathbf{v} \times \mathbf{H}, \tag{17.6}$$

Next we derive the equation for the rate of change of the kinetic energy of the particle† with time, i.e. the derivative

$$\frac{d\mathscr{E}_{\text{kin}}}{dt} = \frac{d}{dt}\left(\frac{mc^2}{\sqrt{1 - \dfrac{v^2}{c^2}}}\right).$$

It is easy to check that

$$\frac{d\mathscr{E}_{\text{kin}}}{dt} = \mathbf{v} \cdot \frac{d\mathbf{p}}{dt}.$$

Substituting $d\mathbf{p}/dt$ from (17.5) and noting that $\mathbf{v} \times \mathbf{H} \cdot \mathbf{v} = 0$, we have

$$\frac{d\mathscr{E}_{\text{kin}}}{dt} = e\mathbf{E} \cdot \mathbf{v}. \tag{17.7}$$

The rate of change of the kinetic energy is the work done by the field on the particle per unit time. From (17.7) we see that this work is equal to the product of the velocity by the force which the electric field exerts on the charge. The work done by the field during a time dt, i.e. during a displacement of the charge by $d\mathbf{r}$, is clearly equal to $e\mathbf{E} \cdot d\mathbf{r}$.

† By "kinetic" we mean the energy (9.4), which includes the rest energy.

We emphasize the fact that work is done on the charge only by the electric field; the magnetic field does no work on a charge moving in it. This is connected with the fact that the force which the magnetic field exerts on a charge is always perpendicular to the velocity of the charge.

The equations of mechanics are invariant with respect to a change in sign of the time, that is, with respect to interchange of future and past. In other words, in mechanics the two time directions are equivalent. This means that if a certain motion is possible according to the equations of mechanics, then the reverse motion is also possible, in which the system passes through the same states in reverse order.

It is easy to see that this is also valid for the electromagnetic field in the theory of relativity. In this case, however, in addition to changing t into $-t$, we must reverse the sign of the magnetic field. In fact it is easy to see that the equations of motion (17.5) are not altered if we make the changes

$$t \to -t, \quad \mathbf{E} \to \mathbf{E}, \quad \mathbf{H} \to -\mathbf{H}. \tag{17.8}$$

According to (17.3) and (17.4), this does not change the scalar potential, while the vector potential changes sign:

$$\phi \to \phi, \quad \mathbf{A} \to -\mathbf{A}. \tag{17.9}$$

Thus, if a certain motion is possible in an electromagnetic field, then the reversed motion is possible in a field in which the direction of **H** is reversed.

PROBLEM

Express the acceleration of a particle in terms of its velocity and the electric and magnetic field intensities.

Solution: Substitute in the equation of motion (17.5) $\mathbf{p} = \mathbf{v}\,\mathcal{E}_{\text{kin}}/c^2$, and take the expression for $d\mathcal{E}_{\text{kin}}/dt$ from (17.7). As a result, we get

$$\dot{\mathbf{v}} = \frac{e}{m}\sqrt{1-\frac{v^2}{c^2}}\left\{\mathbf{E} + \frac{1}{c}\mathbf{v}\times\mathbf{H} - \frac{1}{c^2}\mathbf{v}(\mathbf{v}\cdot\mathbf{E})\right\}.$$

§ 18. Gauge invariance

Let us consider to what extent the potentials are uniquely determined. First of all we call attention to the fact that the field is characterized by the effect which it produces on the motion of a charge located in it. But in the equation of motion (17.5) there appear not the potentials, but the field intensities **E** and **H**. Therefore two fields are physically identical if they are characterized by the same vectors **E** and **H**.

If we are given potentials **A** and ϕ, then these uniquely determine (according to (17.3) and (17.4)) the fields **E** and **H**. However, to one and the same field there can correspond different potentials. To show this, let us add to each component of the potential the quantity $-\partial f/\partial x^k$, where f is an arbitrary function of the coordinates and the time. Then the potential A_k goes over into

$$A_k' = A_k - \frac{\partial f}{\partial x^k}. \tag{18.1}$$

As a result of this change there appears in the action integral (16.1) the additional term

$$\frac{e}{c}\frac{\partial f}{dx^k}dx^k = d\left(\frac{e}{c}f\right), \tag{18.2}$$

which is a total differential and has no effect on the equations of motion. (See *Mechanics*, § 2.)

If in place of the four-potential we introduce the scalar and vector potentials, and in place of x^i, the coordinates ct, x, y, z, then the four equations (18.1) can be written in the form

$$\mathbf{A}' = \mathbf{A} + \operatorname{grad} f, \quad \phi' = \phi - \frac{1}{c}\frac{\partial f}{\partial t}. \tag{18.3}$$

It is easy to check that electric and magnetic fields determined from equations (17.3) and (17.4) actually do not change upon replacement of $\mathbf{A}$ and ϕ by $\mathbf{A}'$ and ϕ', defined by (18.3). Thus the transformation of potentials (18.1) does not change the fields. The potentials are therefore not uniquely defined; the vector potential is determined to within the gradient of an arbitrary function, and the scalar potential to within the time derivative of the same function.

In particular, we see that we can add an arbitrary constant vector to the vector potential, and an arbitrary constant to the scalar potential. This is also clear directly from the fact that the definitions of $\mathbf{E}$ and $\mathbf{H}$ contain only derivatives of $\mathbf{A}$ and ϕ, and therefore the addition of constants to the latter does not affect the field intensities.

Only those quantities have physical meaning which are invariant with respect to the transformation (18.3) of the potentials; in particular all equations must be invariant under this transformation. This invariance is called *gauge invariance* (in German, *eichinvarianz*).†

This nonuniqueness of the potentials gives us the possibility of choosing them so that they fulfil one auxiliary condition chosen by us. We emphasize that we can set one condition, since we may choose the function f in (18.3) arbitrarily. In particular, it is always possible to choose the potentials so that the scalar potential ϕ is zero. If the vector potential is not zero, then it is not generally possible to make it zero, since the condition $\mathbf{A} = 0$ represents three auxiliary conditions (for the three components of $\mathbf{A}$).

§ 19. Constant electromagnetic field

By a constant electromagnetic field we mean a field which does not depend on the time. Clearly the potentials of a constant field can be chosen so that they are functions only of the coordinates and not of the time. A constant magnetic field is equal, as before, to $\mathbf{H} = \operatorname{curl} \mathbf{A}$. A constant electric field is equal to

$$\mathbf{E} = -\operatorname{grad} \phi. \tag{19.1}$$

Thus a constant electric field is determined only by the scalar potential and a constant magnetic field only by the vector potential.

We saw in the preceding section that the potentials are not uniquely determined. However, it is easy to convince oneself that if we describe the constant electromagnetic field in terms of potentials which do not depend on the time, then we can add to the scalar potential, without changing the fields, only an arbitrary constant (not depending on either the coordinates

† We emphasize that this is related to the assumed constancy of e in (18.2). Thus the gauge invariance of the equations of electrodynamics (see below) and the conservation of charge are closely related to one another.

or the time). Usually ϕ is subjected to the additional requirement that it has a definite value at some particular point in space; most frequently ϕ is chosen to be zero at infinity. Thus the arbitrary constant previously mentioned is determined, and the scalar potential of the constant field is thus determined uniquely.

On the other hand, just as before, the vector potential is not uniquely determined even for the constant electromagnetic field; namely, we can add to it the gradient of an arbitrary function of the coordinates.

We now determine the energy of a charge in a constant electromagnetic field. If the field is constant, then the Lagrangian for the charge also does not depend explicitly on the time. As we know, in this case the energy is conserved and coincides with the Hamiltonian.

According to (16.6), we have

$$\mathscr{E} = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} + e\phi. \tag{19.2}$$

Thus the presence of the field adds to the energy of the particle the term $e\phi$, the potential energy of the charge in the field. We note the important fact that the energy depends only on the scalar and not on the vector potential. This means that the magnetic field does not affect the energy of the charge. Only the electric field can change the energy of the particle. This is related to the fact that the magnetic field, unlike the electric field, does no work on the charge.

If the field intensities are the same at all points in space, then the field is said to be uniform. The scalar potential of a uniform electric field can be expressed in terms of the field intensity as

$$\phi = -\mathbf{E} \cdot \mathbf{r}. \tag{19.3}$$

In fact, since $\mathbf{E} = \text{const}$, $\nabla(\mathbf{E} \cdot \mathbf{r}) = (\mathbf{E} \cdot \nabla)\,\mathbf{r} = \mathbf{E}$.

The vector potential of a uniform magnetic field can be expressed in terms of its field intensity as

$$\mathbf{A} = \tfrac{1}{2}\mathbf{H} \times \mathbf{r}. \tag{19.4}$$

In fact, recalling that $\mathbf{H} = \text{const}$, we obtain with the aid of well-known formulas of vector analysis:

$$\text{curl}\,(\mathbf{H} \times \mathbf{r}) = \mathbf{H}\,\text{div}\,\mathbf{r} - (\mathbf{H} \cdot \nabla)\mathbf{r} = 2\mathbf{H}$$

(noting that $\text{div}\,\mathbf{r} = 3$).

The vector potential of a uniform magnetic field can also be chosen in the form

$$A_x = -Hy, \quad A_y = A_z = 0 \tag{19.5}$$

(the z axis is along the direction of $\mathbf{H}$). It is easily verified that with this choice for $\mathbf{A}$ we have $\mathbf{H} = \text{curl}\,\mathbf{A}$. In accordance with the transformation formulas (18.3), the potentials (19.4) and (19.5) differ from one another by the gradient of some function: formula (19.5) is obtained from (19.4) by adding ∇f, where $f = -xyH/2$.

PROBLEM

Give the variational principle for the trajectory of a particle (Maupertuis' principle) in a constant electromagnetic field in relativistic mechanics.

Solution: Maupertuis' principle consists in the statement that if the energy of a particle is conserved (motion in a constant field), then its trajectory can be determined from the variational equation

$$\delta \int \mathbf{P} \cdot d\mathbf{r} = 0,$$

where $\mathbf{P}$ is the generalized momentum of the particle, expressed in terms of the energy and the coordinate differentials, and the integral is taken along the trajectory of the particle.† Substituting $\mathbf{P} = \mathbf{p} + (e/c)\mathbf{A}$ and noting that the directions of $\mathbf{p}$ and $d\mathbf{r}$ coincide, we have

$$\delta \int \left(p dl + \frac{e}{c} \mathbf{A} \cdot d\mathbf{r} \right) = 0,$$

where $dl = \sqrt{d\mathbf{r}^2}$ is the element of arc. Determining p from

$$p^2 + m^2c^2 = \left(\frac{\mathscr{E} - e\phi}{c} \right)^2,$$

we obtain finally

$$\delta \int \left\{ \sqrt{\left(\frac{\mathscr{E} - e\phi}{c} \right)^2 - m^2c^2}\, dl + \frac{e}{c} \mathbf{A} \cdot d\mathbf{r} \right\} = 0.$$

§ 20. Motion in a constant uniform electric field

Let us consider the motion of a charge e in a uniform constant electric field $\mathbf{E}$. We take the direction of the field as the X axis. The motion will obviously proceed in a plane, which we choose as the XY plane. Then the equations of motion (17.5) become

$$\dot{p}_x = eE, \quad \dot{p}_y = 0$$

(where the dot denotes differentiation with respect to t), so that

$$p_x = eEt, \quad p_y = p_0. \tag{20.1}$$

The time reference point has been chosen at the moment when $p_x = 0$; p_0 is the momentum of the particle at that moment.

The kinetic energy of the particle (the energy omitting the potential energy in the field) is $\mathscr{E}_{\text{kin}} = c\sqrt{m^2c^2 + p^2}$. Substituting (20.1), we find in our case

$$\mathscr{E}_{\text{kin}} = \sqrt{m^2c^4 + c^2p_0^2 + (ceEt)^2} = \sqrt{\mathscr{E}_0^2 + (ceEt)^2}, \tag{20.2}$$

where $\mathscr{E}_0$ is the energy at $t = 0$.

According to (9.8) the velocity of the particle is $\mathbf{v} = \mathbf{p}c^2/\mathscr{E}_{\text{kin}}$. For the velocity $v_x = \dot{x}$ we have therefore

$$\frac{dx}{dt} = \frac{p_x c^2}{\mathscr{E}_{\text{kin}}} = \frac{c^2 eEt}{\sqrt{\mathscr{E}_0^2 + (ceEt)^2}}.$$

Integrating, we find

† See *Mechanics*, § 44.

$$x = \frac{1}{eE}\sqrt{\mathscr{E}_0^2 + (ceEt)^2}. \tag{20.3}$$

The constant of integration we set equal to zero.†

For determining y, we have

$$\frac{dy}{dt} = \frac{p_y c^2}{\mathscr{E}_{\text{kin}}} = \frac{p_0 c^2}{\sqrt{\mathscr{E}_0^2 + (ceEt)^2}},$$

from which

$$y = \frac{p_0 c}{eE}\sinh^{-1}\left(\frac{ceEt}{\mathscr{E}_0}\right). \tag{20.4}$$

We obtain the equation of the trajectory by expressing t in terms of y from (20.4) and substituting in (20.3). This gives:

$$x = \frac{\mathscr{E}_0}{eE}\cosh\frac{eE_y}{p_0 c}. \tag{20.5}$$

Thus in a uniform electric field a charge moves along a catenary curve.

If the velocity of the particle is $v << c$, then we can set $p_0 = mv_0$, $\mathscr{E}_0 = mc^2$, and expand (20.5) in series in powers of $1/c$. Then we get, to within terms of higher order,

$$x = \frac{eE}{2mv_0^2}y^2 + \text{const},$$

that is, the charge moves along a parabola, a result well known from classical mechanics.

§ 21. Motion in a constant uniform magnetic field

We now consider the motion of a charge e in a uniform magnetic field $\mathbf{H}$. We choose the direction of the field as the Z axis. We rewrite the equation of motion

$$\dot{\mathbf{p}} = \frac{e}{c}\mathbf{v} \times \mathbf{H}$$

in another form, by substituting for the momentum, from (9.8),

$$\mathbf{p} = \frac{\mathscr{E}\mathbf{v}}{c^2},$$

where $\mathscr{E}$ is the energy of the particle, which is constant in the magnetic field. The equation of motion then goes over into the form

$$\frac{\mathscr{E}}{c^2}\frac{d\mathbf{v}}{dt} = \frac{e}{c}\mathbf{v} \times \mathbf{H} \tag{21.1}$$

or, expressed in terms of components,

$$\dot{v}_x = \omega v_y, \quad \dot{v}_y = -\omega v_x, \quad \dot{v}_z = 0, \tag{21.2}$$

† This result (for $p_0 = 0$) coincides with the solution of the problem of relativistic motion with constant "proper acceleration" $w_0 = eE/m$ (see the problem in § 7). For the present case, the constancy of the acceleration is related to the fact that the electric field does not change for Lorentz transformations having velocities $\mathbf{V}$ along the direction of the field (see § 24).

where we have introduced the notation

$$\omega = \frac{ecH}{\mathscr{E}}. \tag{21.3}$$

We multiply the second equation of (21.2) by i, and add it to the first:

$$\frac{d}{dt}(v_x + iv_y) = -i\omega(v_x + iv_y),$$

so that

$$v_x + iv_y = ae^{-i\omega t},$$

where a is a complex constant. This can be written in the form $a = v_{0t}e^{-i\alpha}$ where v_{0t} and α are real. Then

$$v_x + iv_y = v_{0t}e^{-i(\omega t + \alpha)}$$

and, separating real and imaginary parts, we find

$$v_x = v_{0t}\cos(\omega t + \alpha), \quad v_y = -v_{0t}\sin(\omega t + \alpha). \tag{21.4}$$

The constants v_{0t}, and α are determined by the initial conditions; α is the initial phase, and as for v_{0t}, from (21.4) it is clear that

$$v_{0t} = \sqrt{v_x^2 + v_y^2},$$

that is, v_{0t}, is the velocity of the particle in the XY plane, and stays constant throughout the motion.

From (21.4) we find, integrating once more,

$$x = x_0 + r\sin(\omega t + \alpha), \quad y = y_0 + r\cos(\omega t + \alpha), \tag{21.5}$$

where

$$r = \frac{v_{0t}}{\omega} = \frac{v_{0t}\mathscr{E}}{ecH} = \frac{cp_t}{eH} \tag{21.6}$$

(p_t is the projection of the momentum on the XY plane). From the third equation of (21.2), we find $v_z = v_{0z}$ and

$$z = z_0 + v_{0z}t. \tag{21.7}$$

From (21.5) and (21.7), it is clear that the charge moves in a uniform magnetic field along a helix having its axis along the direction of the magnetic field and with a radius r given by (21.6). The velocity of the particle is constant. In the special case where $v_{0z} = 0$, that is, the charge has no velocity component along the field, it moves along a circle in the plane perpendicular to the field.

The quantity ω, as we see from the formulas, is the angular frequency of rotation of the particle in the plane perpendicular to the field.

If the velocity of the particle is low, then we can approximately set $\mathscr{E} = mc^2$. Then the frequency ω is changed to

$$\omega = \frac{eH}{mc}. \tag{21.8}$$

We shall now assume that the magnetic field remains uniform but varies slowly in magnitude and direction. Let us see how the motion of a charged particle changes in this case.

We know that when the conditions of the motion are changed slowly, certain quantities called adiabatic invariants remain constant. Since the motion in the plane perpendicular to the magnetic field is periodic, the adiabatic invariant is the integral

$$I = \frac{1}{2\pi} \oint \mathbf{P}_t \cdot d\mathbf{r},$$

taken over a complete period of the motion, i.e. over the circumference of a circle in the present case ($\mathbf{P}_t$ is the projection of the generalized momentum on the plane perpendicular to $\mathbf{H}$†). Substituting $\mathbf{P}_t = \mathbf{p}_t + (e/c)\,\mathbf{A}$, we have:

$$I = \frac{1}{2\pi} \oint \mathbf{P}_t \cdot d\mathbf{r} = \frac{1}{2\pi} \oint \mathbf{p}_t \cdot d\mathbf{r} + \frac{e}{2\pi c} \oint \mathbf{A} \cdot d\mathbf{r}.$$

In the first term we note that p_t is constant in magnitude and directed along $d\mathbf{r}$; we apply Stokes' theorem to the second term and write curl $\mathbf{A} = \mathbf{H}$:*

$$I = rp_t - \frac{e}{2c} Hr^2 = \frac{cp_t^2}{2eH}. \tag{21.9}$$

From this we see that, for slow variation of H, the tangential momentum p_t varies proportionally to $\sqrt{H}$.

This result can also be applied to another case, when the particle moves along a helical path in a magnetic field that is not strictly homogeneous (so that the field varies little over distances comparable with the radius and step of the helix). Such a motion can be considered as a motion in a circular orbit that shifts in the course of time, while relative to the orbit the field appears to change in time but remain uniform. One can then state that the component of the momentum transverse to the direction of the field varies according to the law: $p_t = \sqrt{CH}$, where C is a constant and H is a given function of the coordinates. On the other hand, just as for the motion in any constant magnetic field, the energy of the particle (and consequently the square of its momentum p^2) remains constant. Therefore the longitudinal component of the momentum varies according to the formula:

$$p_l^2 = p^2 - p_t^2 = p^2 - CH(x, y, z). \tag{21.10}$$

Since we should always have $p_l^2 \geq 0$, we see that penetration of the particle into regions of sufficiently high field ($CH > p^2$) is impossible. During motion in the direction of increasing field, the radius of the helical trajectory decreases proportionally to p_t/H (i.e. proportionally

† See *Mechanics*, § 49. In general the integrals $\oint p\,dq$, taken over a period of the particular coordinate q, are adiabatic invariants. In the present case the periods for the two coordinates in the plane perpendicular to $\mathbf{H}$ coincide, and the integral I which we have written is the sum of the two corresponding adiabatic invariants. However, each of these invariants individually has no special significance, since it depends on the (non-unique) choice of the vector potential of the field. The nonuniqueness of the adiabatic invariants which results from this is a reflection of the fact that, when we regard the magnetic field as uniform over all of space, we cannot in principle determine the electric field which results from changes in $\mathbf{H}$, since it will actually depend on the specific conditions at infinity.

*By inspecting the direction of motion of a charge along the orbit for a given direction of $\mathbf{H}$, we observe that it is counterclockwise if we look along $\mathbf{H}$. Hence the negative sign in the second term.

to $1/\sqrt{H}$), and the step proportionally to p_l. On reaching the boundary where p_l vanishes, the particle is reflected; while continuing to rotate in the same direction it begins to move opposite to the gradient of the field.

Inhomogeneity of the field also leads to another phenomenon—a slow transverse shift (*drift*) of the *guiding centre* of the helical trajectory of the particle (the name given to the centre of the circular orbit); problem 3 of the next section deals with this question.

PROBLEM

Determine the frequency of vibration of a charged spatial oscillator, placed in a constant, uniform magnetic field; the proper frequency of vibration of the oscillator (in the absence of the field) is ω_0.

Solution: The equations of forced vibration of the oscillator in a magnetic field (directed along the z axis) are:

$$\ddot{x} + \omega_0^2 x = \frac{eH}{me}\dot{y}, \quad \ddot{y} + \omega_0^2 y = \frac{eH}{mc}\dot{x}, \quad \ddot{z} + \omega_0^2 z = 0.$$

Multiplying the second equation by i and combining with the first, we find

$$\ddot{\zeta} + \omega_0^2 \zeta = -i\frac{eH}{mc}\dot{\zeta},$$

where $\zeta = x + iy$. From this we find that the frequency of vibration of the oscillator in a plane perpendicular to the field is

$$\omega = \sqrt{\omega_0^2 + \frac{1}{4}\left(\frac{cH}{mc}\right)^2} \pm \frac{eH}{2mc}.$$

If the field H is weak, this formula goes over into

$$\omega = \omega_0 \pm eH/2mc.$$

The vibration along the direction of the field remains unchanged.

§ 22. Motion of a charge in constant uniform electric and magnetic fields

Finally we consider the motion of a charge in the case where there are present both electric and magnetic fields, constant and uniform. We limit ourselves to the case where the velocity of the charge $v << c$, so that its momentum $\mathbf{p} = m\mathbf{v}$; as we shall see later, it is necessary for this that the electric field be small compared to the magnetic.

We choose the direction of $\mathbf{H}$ as the Z axis, and the plane passing through $\mathbf{H}$ and $\mathbf{E}$ as the YZ plane. Then the equation of motion

$$m\dot{\mathbf{v}} = e\mathbf{E} + \frac{e}{c}\mathbf{v} \times \mathbf{H}$$

can be written in the form

$$m\ddot{x} = \frac{e}{c}\dot{y}H, \quad m\ddot{y} = eE_y - \frac{e}{c}\dot{x}H, \quad m\ddot{z} = eE_z. \tag{22.1}$$

From the third equation we see that the charge moves with uniform acceleration in the Z direction, that is,

$$z = \frac{eE_z}{2m}t^2 + v_{0z}t. \tag{22.2}$$

Multiplying the second equation of (22.1) by i and combining with the first, we find

$$\frac{d}{dt}(\dot{x} + i\dot{y}) + i\omega(\dot{x} + i\dot{y}) = i\frac{e}{m}E_y$$

($\omega = eH/mc$). The integral of this equation, where $\dot{x} + i\dot{y}$ is considered as the unknown, is equal to the sum of the integral of the same equation without the right-hand term and a particular integral of the equation with the right-hand term. The first of these is $ae^{-i\omega t}$, the second is $eE_y/m\omega = cE_y/H$. Thus

$$\dot{x} + i\dot{y} = ae^{-i\omega t} + \frac{cE_y}{H}.$$

The constant a is in general complex. Writing it in the form $a = be^{i\alpha}$, with real b and α, we see that since a is multiplied by $e^{-i\omega t}$, we can, by a suitable choice of the time origin, give the phase α any arbitrary value. We choose this so that a is real. Then breaking up $\dot{x} + i\dot{y}$ into real and imaginary parts, we find

$$\dot{x} = a\cos\omega t + \frac{cE_y}{H}, \quad \dot{y} = -a\sin\omega t. \tag{22.3}$$

At $t = 0$ the velocity is along the X axis.

We see that the components of the velocity of the particle are periodic functions of the time. Their average values are:

$$\bar{\dot{x}} = \frac{cE_y}{H}, \quad \bar{y} = 0.$$

This average velocity of motion of a charge in crossed electric and magnetic fields is often called the electrical *drift* velocity. Its direction is perpendicular to both fields and independent of the sign of the charge. It can be written in vector form as:

$$\bar{\mathbf{v}} = \frac{c\mathbf{E} \times \mathbf{H}}{H^2}. \tag{22.4}$$

All the formulas of this section assume that the velocity of the particle is small compared with the velocity of light; we see that for this to be so, it is necessary in particular that the electric and magnetic fields satisfy the condition

$$\frac{E_y}{H} \ll 1, \tag{22.5}$$

while the absolute magnitudes of E_y and H can be arbitrary.

Integrating equation (22.3) again, and choosing the constant of integration so that at $t = 0$, $x = y = 0$, we obtain

$$x = \frac{a}{\omega}\sin\omega t + \frac{cE_y}{H}t; \quad y = \frac{a}{\omega}(\cos\omega t - 1). \tag{22.6}$$

Considered as parametric equations of a curve, these equations define a trochoid. Depending on whether a is larger or smaller in absolute value than the quantity cE_y/H, the projection of the trajectory on the plane XY has the forms shown in Figs. 6a and 6b, respectively.

If $a = -cE_y/H$, then (22.6) becomes

$$x = \frac{cE_y}{\omega H}(\omega t - \sin\omega t),$$

$$y = \frac{cE_y}{\omega H}(1 - \cos \omega t) \tag{22.7}$$

that is, the projection of the trajectory on the XY plane is a cycloid (Fig. 6c).

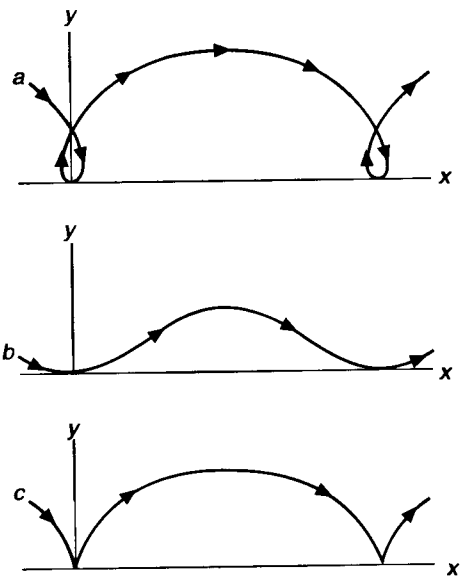


FIG. 6.

PROBLEMS

1. Determine the relativistic motion of a charge in parallel uniform electric and magnetic fields.

Solution: The magnetic field has no influence on the motion along the common direction of **E** and **H** (the z axis), which therefore occurs under the influence of the electric field alone; therefore according to § 20 we find:

$$z = \frac{\mathscr{E}_{\text{kin}}}{eE}, \quad \mathscr{E}_{\text{kin}} = \sqrt{\mathscr{E}_0^2 + (ceEt)^2}\ .$$

For the motion in the xy plane we have the equation

$$\dot{p}_x = \frac{e}{c} H v_y, \quad \dot{p}_y = -\frac{e}{c} H v_x,$$

or

$$\frac{d}{dt}(p_x + ip_y) = -i\frac{eH}{c}(v_x + iv_y) = -\frac{ieHc}{\mathscr{E}_{\text{kin}}}(p_x + ip_y).$$

Consequently

$$p_x + ip_y = p_t e^{-i\phi},$$

where p_t is the constant value of the projection of the momentum on the xy plane, and the auxiliary quantity ϕ is defined by the relation

$$d\phi = eHc\frac{dt}{\mathscr{E}_{\text{kin}}},$$

from which

$$ct = \frac{\mathscr{E}_0}{eE} \sinh \frac{E}{H}\phi. \tag{1}$$

Furthermore we have:

$$p_x + ip_y = p_t e^{-i\phi} = \frac{\mathscr{E}_{\text{kin}}}{c^2}(\dot{x} + i\dot{y}) = \frac{eH}{c}\frac{d(x+iy)}{d\phi},$$

so that

$$x = \frac{cp_t}{eH}\sin\phi, \qquad y = \frac{c\dot{p}_t}{eH}\cos\phi. \tag{2}$$

Formulas (1), (2) together with the formula

$$z = \frac{\mathscr{E}_0}{eE}\cosh\frac{E}{H}\phi, \tag{3}$$

determine the motion of the particle in parametric form. The trajectory is a helix with radius cp_t/eH and monotonically increasing step, along which the particle moves with decreasing angular velocity $\dot{\phi} = eHc/\mathscr{E}_{\text{kin}}$ and with a velocity along the z axis which tends toward the value c.

2. Determine the relativistic motion of a charge in electric and magnetic fields which are mutually perpendicular and equal in magnitude.†

Solution: Choosing the z axis along **H** and the y axis along **E** and setting $E = H$, we write the equations of motion:

$$\frac{dp_x}{dt} = \frac{e}{c}Ev_y, \qquad \frac{dp_y}{dt} = eE\left(1 - \frac{v_x}{c}\right), \qquad \frac{dp_z}{dt} = 0$$

and, as a consequence of them, formula (17.7),

$$\frac{d\mathscr{E}_{\text{kin}}}{dt} = eEv_y.$$

From these equations we have:

$$p_z = \text{const}, \quad \mathscr{E}_{\text{kin}} - cp_x = \text{const} \equiv \alpha.$$

Also using the equation

$$\mathscr{E}_{\text{kin}}^2 - c^2p_x^2 = (\mathscr{E}_{\text{kin}} + cp_x)(\mathscr{E}_{\text{kin}} - cp_x) = c^2p_y^2 + \varepsilon^2$$

(where $\varepsilon^2 = m^2c^4 + c^2p_z^2 = \text{const}$), we find:

$$\mathscr{E}_{\text{kin}} + cp_x = \frac{1}{\alpha}(c^2p_y^2 + \varepsilon^2),$$

and so

$$\mathscr{E}_{\text{kin}} = \frac{\alpha}{2} + \frac{c^2p_y^2 + \varepsilon^2}{2\alpha},$$

$$p_x = -\frac{\alpha}{2c} + \frac{c^2p_y^2 + \varepsilon^2}{2\alpha c}.$$

† The problem of motion in mutually perpendicular fields **E** and **H** which are not equal in magnitude can, by a suitable transformation of the reference system, be reduced to the problem of motion in a pure electric or a pure magnetic field (see § 25).

Furthermore, we write

$$\mathscr{E}_{\text{kin}} \frac{dp_y}{dt} = eE\left(\mathscr{E}_{\text{kin}} - \frac{\mathscr{E}_{\text{kin}} v_x}{c}\right) = eE(\mathscr{E}_{\text{kin}} - cp_x) = eE\alpha,$$

from which

$$2eEt = \left(1 + \frac{\varepsilon^2}{\alpha^2}\right)p_y + \frac{c^2}{3\alpha^2}p_y^3. \tag{1}$$

To determine the trajectory, we make a transformation of variables in the equations

$$\frac{dx}{dt} = \frac{c^2 p_x}{\mathscr{E}_{\text{kin}}}, \ldots$$

to the variable p_y by using the relation $dt = \mathscr{E}_{\text{kin}} dp_y / eE\alpha$, after which integration gives the formulas:

$$x = \frac{c}{2eE}\left(-1 + \frac{\varepsilon^2}{\alpha^2}\right)p_y + \frac{c^3}{6\alpha^2 eE}p_y^3, \tag{2}$$

$$y = \frac{c^2}{2\alpha eE}p_y^2, \qquad z = \frac{p_z c^2}{eE\alpha}p_y.$$

Formulas (1) and (2) completely determine the motion of the particle in parametric form (parameter p_y). We call attention to the fact that the velocity increases most rapidly in the direction perpendicular to **E** and **H** (the x axis).

3. Determine the velocity of drift of the guiding centre of the orbit of a nonrelativistic charged particle in a quasihomogeneous magnetic field (H. Alfven, 1940).

Solution: We assume first that the particle is moving in a circular orbit, i.e. its velocity has no longitudinal component (along the field). We write the equation of the trajectory in the form $\mathbf{r} = \mathbf{R}(t) + \zeta(t)$, where $\mathbf{R}(t)$ is the radius vector of the guiding centre (a slowly varying function of the time), while $\zeta(t)$ is a rapidly oscillating quantity describing the rotational motion about the guiding centre. We average the force $(e/c)\,\dot{\mathbf{r}} \times \mathbf{H}(\mathbf{r})$ acting on the particle over a period of the oscillatory (circular) motion (compare *Mechanics*, § 30). We expand the function $\mathbf{H}(\mathbf{r})$ in this expression in powers of ξ:

$$\mathbf{H}(\mathbf{r}) = \mathbf{H}(\mathbf{R}) + (\zeta \cdot \nabla)\mathbf{H}(\mathbf{R}).$$

On averaging, the terms of first order in $\zeta(t)$ vanish, while the second-degree terms give rise to an additional force

$$\mathbf{f} = \frac{e}{c}\dot{\zeta} \times (\zeta \cdot \nabla)\mathbf{H}.$$

For a circular orbit

$$\dot{\zeta} = \omega\zeta \times \mathbf{n}, \qquad \zeta = \frac{v_\perp}{\omega},$$

where $\mathbf{n}$ is a unit vector along $\mathbf{H}$; the frequency $\omega = eH/mc$; $v_\perp$ is the velocity of the particle in its circular motion. The average values of products of components of the vector ζ, rotating in a plane (the plane perpendicular to $\mathbf{n}$), are:

$$\overline{\zeta_\alpha \zeta_\beta} = \tfrac{1}{2}\zeta^2 \delta_{\alpha\beta},$$

where $\delta_{\alpha\beta}$ is the unit tensor in this plane. As a result we find:

$$\mathbf{f} = -\frac{mv_\perp^2}{2H}(\mathbf{n} \times \nabla) \times \mathbf{H}.$$

Because of the equations div $\mathbf{H} = 0$ and curl $\mathbf{H} = 0$ which the constant field $\mathbf{H}(\mathbf{R})$ satisfies, we have:

$$(\mathbf{n} \times \nabla) \times \mathbf{H} = -\mathbf{n} \operatorname{div} \mathbf{H} + (\mathbf{n} \cdot \nabla)\mathbf{H} + \mathbf{n} \times (\nabla \times \mathbf{H}) = (\mathbf{n} \cdot \nabla)\mathbf{H} = H(\mathbf{n} \cdot \nabla)\mathbf{n} + \mathbf{n}(\mathbf{n} \cdot \nabla H).$$

We are interested in the force transverse to $\mathbf{n}$, giving rise to a shift of the orbit; it is equal to

$$\mathbf{f} = -\frac{mv_\perp^2}{2}(\mathbf{n} \cdot \nabla)\,\mathbf{n} = \frac{mv_\perp^2}{2\rho}\,\boldsymbol{\nu},$$

where ρ is the radius of curvature of the force line of the field at the given point, and $\boldsymbol{\nu}$ is a unit vector directed from the centre of curvature to this point.

The case where the particle also has a longitudinal velocity $v_\parallel$ (along $\mathbf{n}$) reduces to the previous case if we go over to a reference frame which is rotating about the instantaneous centre of curvature of the force line (which is the trajectory of the guiding centre) with angular velocity $v_\parallel/\rho$. In this reference system the particle has no longitudinal velocity, but there is an additional transverse force, the centrifugal force $mv_\parallel^2/\rho$. Thus the total transverse force is

$$\mathbf{f}_\perp = \boldsymbol{\nu}\frac{m}{\rho}\left(v_\parallel^2 + \frac{v_\perp^2}{2}\right).$$

This force is equivalent to a constant electric field of strength $\mathbf{f}_\perp/e$. According to (22.4) it causes a drift of the guiding center of the orbit with a velocity

$$v_d = \frac{1}{\omega\rho}\left(v_\parallel^2 + \frac{v_\perp^2}{2}\right)\boldsymbol{\nu} \times \mathbf{n}.$$

The sign of this velocity depends on the sign of the charge.

§ 23. The electromagnetic field tensor

In § 17, we derived the equation of motion of a charge in a field, starting from the Lagrangian (16.4) written in three-dimensional form. We now derive the same equation directly from the action (16.1) written in four-dimensional notation.

The principle of least action states

$$\delta S = \delta \int_a^b \left(-mc\,ds - \frac{e}{c}A_i dx^i\right) = 0. \tag{23.1}$$

Noting that $ds = \sqrt{dx_i dx^i}$, we find (the limits of integration a and b are omitted for brevity):

$$\delta S = -\int \left(mc\frac{dx_i d\delta x^i}{ds} + \frac{e}{c}A_i d\delta x^i + \frac{e}{c}\delta A_i dx^i\right) = 0.$$

We integrate the first two terms in the integrand by parts. Also, in the first term we set $dx_i/ds = u_i$, where u_i are the components of the four-velocity. Then

$$\int \left(mc\,du_i\,\delta x^i + \frac{e}{c}\,\delta x^i\,dA_i - \frac{e}{c}\,\delta A_i\,dx^i\right) - \left[\left(mcu_i + \frac{e}{c}A_i\right)\delta x^i\right] = 0. \tag{23.2}$$

The second term in this equation is zero, since the integral is varied with fixed coordinate values at the limits. Furthermore:

$$\delta A_i = \frac{\partial A_i}{\partial x^k}\delta x^k, \quad dA_i = \frac{\partial A_i}{\partial x^k}dx^k,$$

and therefore

$$\int \left(mcdu_i \delta x^i + \frac{e}{c} \frac{\partial A_i}{\partial x^k} \delta x^i dx^k - \frac{e}{c} \frac{\partial A_i}{\partial x^k} dx^i \delta x^k \right) = 0.$$

In the first term we write $du_i = (du_i/ds)\, ds$, in the second and third, $dx^i = u^i ds$. In addition, in the third term we interchange the indices i and k (this changes nothing since the indices i and k are summed over). Then

$$\int \left[mc \frac{du_i}{ds} - \frac{e}{c} \left(\frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k} \right) u^k \right] \delta x^i ds = 0.$$

In view of the arbitrariness of δx^i, it follows that the integrand is zero, that is,

$$mc \frac{du_i}{ds} = \frac{e}{c} \left(\frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k} \right) u^k.$$

We now introduce the notion

$$F_{ik} = \frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k}. \tag{23.3}$$

The antisymmetric tensor F_{ik} is called the *electromagnetic field tensor.* The equation of motion then takes the form:

$$mc \frac{du^i}{ds} = \frac{e}{c} F^{ik} u_k. \tag{23.4}$$

These are the equations of motion of a charge in four-dimensional form.

The meaning of the individual components of the tensor F_{ik} is easily seen by substituting the values $A_i = (\phi, -\mathbf{A})$ in the definition (23.3). The result can be written as a matrix in which the index $i = 0, 1, 2, 3$ labels the rows, and the index k the columns:

$$F_{ik} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -H_z & H_y \\ -E_y & H_z & 0 & -H_x \\ -E_z & -H_y & H_x & 0 \end{pmatrix}, \quad F^{ik} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -H_z & H_y \\ E_y & H_z & 0 & -H_x \\ E_z & -H_y & H_x & 0 \end{pmatrix}. \tag{23.5}$$

More briefly, we can write (see § 6):

$$F_{ik} = (\mathbf{E}, \mathbf{H}), \quad F^{ik} = (-\mathbf{E}, \mathbf{H}).$$

Thus the components of the electric and magnetic field strengths are components of the same electromagnetic field four-tensor.

Changing to three-dimensional notation, it is easy to verify that the three space components ($i = 1, 2, 3$) of (23.4) are identical with the vector equation of motion (17.5), while the time component ($i = 0$) gives the work equation (17.7). The latter is a consequence of the equations of motion; the fact that only three of the four equations are independent can also easily be found directly by multiplying both sides of (23.4) by u^i. Then the left side of the equation vanishes because of the orthogonality of the four-vectors u^i and du_i/ds, while the right side vanishes because of the antisymmetry of F_{ik}.

If we admit only possible trajectories when we vary S, the first term in (23.2) vanishes identically. Then the second term, in which the upper limit is considered as variable, gives the differential of the action as a function of the coordinates. Thus

$$\delta S = -\left(mcu_i + \frac{e}{c}A_i\right)\delta x^i. \tag{23.6}$$

Then

$$-\frac{\partial S}{\partial x^i} = mcu_i + \frac{e}{c}A_i = p_i + \frac{e}{c}A_i. \tag{23.7}$$

The four-vector $-\delta S/\partial x^i$ is the four-vector P_i of the generalized momentum of the particle. Substituting the values of the components p_i and A_i, we find that

$$P^i = \left(\frac{\mathscr{E}_{\text{kin}} + e\phi}{c},\ \mathbf{p} + \frac{e}{c}\mathbf{A}\right). \tag{23.8}$$

As expected, the space components of the four-vector form the three-dimensional generalized momentum vector (16.5), while the time component is $\mathscr{E}/c$, where $\mathscr{E}$ is the total energy of the charge in the field.

§ 24. Lorentz transformation of the field

In this section we find the transformation formulas for fields, that is, formulas by means of which we can determine the field in one inertial system of reference, knowing the same field in another system.

The formulas for transformation of the potentials are obtained directly from the general formulas for transformation of four-vectors (6.1). Remembering that $A^i = (\phi, \mathbf{A})$, we get easily

$$\phi = \frac{\phi' + \frac{V}{c}A'_x}{\sqrt{1 - \frac{V^2}{c^2}}}, \quad A_x = \frac{A'_x + \frac{V}{c}\phi'}{\sqrt{1 - \frac{V^2}{c^2}}}, \quad A_y = A'_y, \quad A_z = A'_z. \tag{24.1}$$

The transformation formulas for an antisymmetric second-rank tensor (like F^{ik}) were found in problem 2 of § 6: the components F^{23} and F^{01} do not change, while the components F^{02}, F^{03}, and F^{12}, F^{13} transform like x^0 and x^1, respectively. Expressing the components of F^{ik} in terms of the components of the fields $\mathbf{E}$ and $\mathbf{H}$, according to (23.5), we then find the following formulas of transformation for the electric field:

$$E_x = E'_x, E_y = \frac{E'_y + \frac{V}{c}H'_z}{\sqrt{1 - \frac{V^2}{c^2}}}, \quad E_z = \frac{E'_z - \frac{V}{c}H'_y}{\sqrt{1 - \frac{V^2}{c}}}, \tag{24.2}$$

and for the magnetic field:

$$H_x = H'_x, H_y = \frac{H'_y - \frac{V}{c}E'_z}{\sqrt{1 - \frac{V^2}{c^2}}}, \quad H_z = \frac{H'_z - \frac{V}{c}E'_y}{\sqrt{1 - \frac{V^2}{c^2}}}, \tag{24.3}$$

Thus the electric and magnetic fields, like the majority of physical quantities, are relative; that is, their properties are different in different reference systems. In particular, the electric or the magnetic field can be equal to zero in one reference system and at the same time be present in another system.

The formulas (24.2), (24.3) simplify considerably for the case $V \ll c$. To terms of order V/c, we have:

$$E_x = E'_x, E_y = E'_y + \frac{V}{c} H'_z, E_z = E'_z - \frac{V}{c} H'_y;$$

$$H_x = H'_x, H_y = H'_y - \frac{V}{c} E'_z, H_z = H'_z + \frac{V}{c} E'_y.$$

These formulas can be written in vector form

$$\mathbf{E} = \mathbf{E}' + \frac{1}{c} \mathbf{H}' \times \mathbf{V}, \mathbf{H} = \mathbf{H}' - \frac{1}{c} \mathbf{E}' \times \mathbf{V}. \tag{24.4}$$

The formulas for the inverse transformation from K' to K are obtained directly from (24.2)–(24.4) by changing the sign of V and shifting the prime.

If the magnetic field $\mathbf{H}' = 0$ in the K' system, then, as we easily verify on the basis of (24.2) and (24.3), the following relation exists between the electric and magnetic fields in the K system:

$$\mathbf{H} = \frac{1}{c} \mathbf{V} \times \mathbf{E}. \tag{24.5}$$

If in the K' system, $\mathbf{E}' = 0$, then in the K system

$$\mathbf{E} = -\frac{1}{c} \mathbf{V} \times \mathbf{H}. \tag{24.6}$$

Consequently, in both cases, in the K system the magnetic and electric fields are mutually perpendicular.

These formulas also have a significance when used in the reverse direction: if the fields $\mathbf{E}$ and $\mathbf{H}$ are mutually perpendicular (but not equal in magnitude) in some reference system K, then there exists a reference system K' in which the field is pure electric or pure magnetic. The velocity $\mathbf{V}$ of this system (relative to K) is perpendicular to $\mathbf{E}$ and $\mathbf{H}$ and equal in magnitude to cH/E in the first case (where we must have $H < E$) and to cE/H in the second case (where $E < H$).

§ 25. Invariants of the field

From the electric and magnetic field intensities we can form invariant quantities, which remain unchanged in the transition from one inertial reference system to another.

The form of these invariants is easily found starting from the four-dimensional representation of the field using the antisymmetric four-tensor F^{ik}. It is obvious that we can form the following invariant quantities from the components of this tensor:

$$F_{ik}F^{ik} = \text{inv}, \tag{25.1}$$

$$e^{iklm}F_{ik}F_{lm} = \text{inv}, \tag{25.2}$$

where e^{iklm} is the completely antisymmetric unit tensor of the fourth rank (cf. § 6). The first

quantity is a scalar, while the second is a pseudoscalar (the product of the tensor F^{ik} with its dual tensor.†

Expressing F^{ik} in terms of the components of **E** and **H** using (23.5), it is easily shown that, in three-dimensional form, these invariants have the form:

$$H^2 - E^2 = \text{inv}, \tag{25.3}$$

$$\mathbf{E} \cdot \mathbf{H} = \text{inv}. \tag{25.4}$$

The pseudoscalar character of the second of these is here apparent from the fact that it is the product of the polar vector **E** with the axial vector **H** (whereas its square $(\mathbf{E} \cdot \mathbf{H})^2$ is a true scalar).

From the invariance of the two expressions presented, we get the following theorems. If the electric and magnetic fields are mutually perpendicular in any reference system, that is, $\mathbf{E} \cdot \mathbf{H} = 0$, then they are also perpendicular in every other inertial reference system. If the absolute values of **E** and **H** are equal to each other in any reference system, then they are the same in any other system.

The following inequalities are also clearly valid. If in any reference system $E > H$ (or $H > E$), then in every other system we will have $E > H$ (or $H > E$). If in any system of reference the vectors **E** and **H** make an acute (or obtuse) angle, then they will make an acute (or obtuse) angle in every other reference system.

By means of a Lorentz transformation we can always give **E** and **H** any arbitrary values, subject only to the condition that $E^2 - H^2$ and $\mathbf{E} \cdot \mathbf{H}$ have fixed values. In particular, we can always find an inertial system in which the electric and magnetic fields are parallel to each other at a given point. In this system $\mathbf{E} \cdot \mathbf{H} = EH$, and from the two equations

$$E^2 - H^2 = E_0^2 - H_0^2, \qquad EH = \mathbf{E}_0 \cdot \mathbf{H}_0 .$$

we can find the values of **E** and **H** in this system of reference ($\mathbf{E}_0$ and $\mathbf{H}_0$ are the electric and magnetic fields in the original system of reference).

The case where both invariants are zero is excluded. In this case, **E** and **H** are equal and mutually perpendicular in all reference systems.

If $\mathbf{E} \cdot \mathbf{H} = 0$, then we can always find a reference system in which $\mathbf{E} = 0$ or $\mathbf{H} = 0$ (according as $E^2 - H^2 <$ or > 0), that is, the field is purely magnetic or purely electric. Conversely, if in any reference system $\mathbf{E} = 0$ or $\mathbf{H} = 0$, then they are mutually perpendicular in every other system, in accordance with the statement at the end of the preceding section.

We shall give still another approach to the problem of finding the invariants of an antisymmetric four-tensor. From this method we shall, in particular, see that (25.3)–(25.4) are actually the only two independent invariants and at the same time we will explain some instructive mathematical properties of the Lorentz transformations when applied to such a four-tensor.

Let us consider the complex vector

$$\mathbf{F} = \mathbf{E} + i\mathbf{H}. \tag{25.5}$$

† We also note that the pseudoscalar (25.2) can also be expressed as a four-divergence:

$$e^{iklm} F_{ik} F_{lm} = 4 \frac{\partial}{\partial x^i} \left(e^{iklm} A_k \frac{\partial}{\partial x^l} A_m \right),$$

as can be easily verified by using the antisymmetry of e^{iklm}.

Using formulas (24.2)–(24.3), it is easy to see that a Lorentz transformation (along the x axis) for this vector has the form

$$F_x = F'_x, F_y = F'_y \cosh \phi - iF'_z \sinh \phi = F'_y \cos i\phi - F'_z \sin i\phi.$$

$$F_z = F'_z \cos i\phi + F'_y \sin i\phi, \tanh \phi = \frac{V}{c}. \qquad (25.6)$$

We see that a rotation in the x, t plane in four-space (which is what this Lorentz transformation is) for the vector $\mathbf{F}$ is equivalent to a rotation in the y, z plane through an imaginary angle in three-dimensional space. The set of all possible rotations in four-space (including also the simple rotations around the x, y, and z axes) is equivalent to the set of all possible rotations, through complex angles in three-dimensional space (where the six angles of rotation in four-space correspond to the three complex angles of rotation of the three-dimensional system).

The only invariant of a vector with respect to rotation is its square: $\mathbf{F}^2 = E^2 - H^2 + 2i\,\mathbf{E} \cdot \mathbf{H}$; thus the real quantities $E^2 - H^2$ and $\mathbf{E} \cdot \mathbf{H}$ are the only two independent invariants of the tensor F_{ik}.

If $\mathbf{F}^2 \neq 0$, the vector $\mathbf{F}$ can be written as $\mathbf{F} = a\mathbf{n}$, where $\mathbf{n}$ is a complex unit vector ($\mathbf{n}^2 = 1$). By a suitable complex rotation we can point $\mathbf{n}$ along one of the coordinate axes; it is clear that then $\mathbf{n}$ becomes real and determines the directions of the two vectors $\mathbf{E}$ and $\mathbf{H}$: $\mathbf{F} = (E + iH)\mathbf{n}$; in other words we get the result that $\mathbf{E}$ and $\mathbf{H}$ become parallel to one another.

PROBLEM

Determine the velocity of the system of reference in which the electric and magnetic fields are parallel.

Solution: Systems of reference K', satisfying the required condition, exist in infinite numbers. If we have found one such, then the same property will be had by any other system moving relative to the first with its velocity directed along the common direction of $\mathbf{E}$ and $\mathbf{H}$. Therefore it is sufficient to find one of these systems which has a velocity perpendicular to both fields. Choosing the direction of the velocity as the x axis, and making use of the fact that in K': $E'_z = H'_z = 0$, $E'_y H'_z - E'_z H'_y = 0$, we obtain with the aid of formulas (24.2) and (24.3) for the velocity $\mathbf{V}$ of the K' system relative to the original system the following equation:

$$\frac{\frac{\mathbf{V}}{c}}{1 + \frac{\mathbf{V}^2}{c^2}} = \frac{\mathbf{E} \times \mathbf{H}}{E^2 + H^2}$$

(we must choose that root of the quadratic equation for which $V < c$).

CHAPTER 4

THE ELECTROMAGNETIC FIELD EQUATIONS

§ 26. The first pair of Maxwell's equations

From the expressions

$$\mathbf{H} = \operatorname{curl} \mathbf{A}, \qquad \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} - \operatorname{grad} \phi$$

it is easy to obtain equations containing only **E** and **H**. To do this we find curl **E**:

$$\operatorname{curl} \mathbf{E} = -\frac{1}{c}\frac{\partial}{\partial t} \operatorname{curl} \mathbf{A} - \operatorname{curl} \operatorname{grad} \phi.$$

But the curl of any gradient is zero. Consequently,

$$\operatorname{curl} \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t}. \tag{26.1}$$

Taking the divergence of both sides of the equation curl **A** = **H**, and recalling that div curl = 0, we find

$$\operatorname{div} \mathbf{H} = 0. \tag{26.2}$$

The equations (26.1) and (26.2) are called the first pair of Maxwell's equations.† We note that these two equations still do not completely determine the properties of the fields. This is clear from the fact that they determine the change of the magnetic field with time (the derivative $\partial \mathbf{H}/\partial t$), but do not determine the derivative $\partial \mathbf{E}/\partial t$.

Equations (26.1) and (26.2) can be written in integral form. According to Gauss'theorem

$$\int \operatorname{div} \mathbf{H} dV = \oint \mathbf{H} \cdot d\mathbf{f},$$

where the integral on the right goes over the entire closed surface surrounding the volume over which the integral on the left is extended. On the basis of (26.2), we have

$$\oint \mathbf{H} \cdot d\mathbf{f} = 0. \tag{26.3}$$

† Maxwell's equations (the fundamental equations of electrodynamics) were first formulated by him in the 1860's.

The integral of a vector over a surface is called the *flux of the vector* through the surface. Thus the flux of the magnetic field through every closed surface is zero.

According to Stokes' theorem,

$$\int \operatorname{curl} \mathbf{E} \cdot d\mathbf{f} = \oint \mathbf{E} \cdot d\mathbf{l},$$

where the integral on the right is taken over the closed contour bounding the surface over which the left side is integrated. From (26.1) we find, integrating both sides for any surface,

$$\oint \mathbf{E} \cdot d\mathbf{l} = \frac{1}{c} \frac{\partial}{\partial t} \int \mathbf{H} \cdot d\mathbf{f}. \tag{26.4}$$

The integral of a vector over a closed contour is called the *circulation* of the vector around the contour. The circulation of the electric field is also called the *electromotive force* in the given contour. Thus the electromotive force in any contour is equal to minus the time derivative of the magnetic flux through a surface bounded by this contour.

The Maxwell equations (26.1) and (26.2) can be expressed in four-dimensional notation. Using the definition of the electromagnetic field tensor

$$F_{ik} = \partial A_k / \partial x^i - \partial A_i / \partial x^k,$$

it is easy to verify that

$$\frac{\partial F_{ik}}{\partial x^l} + \frac{\partial F_{kl}}{\partial x^i} + \frac{\partial F_{li}}{\partial x^k} = 0. \tag{26.5}$$

The expression on the left is a tensor of third rank, which is antisymmetric in all three indices. The only components which are not identically zero are those with $i \neq k \neq l$. Thus there are altogether four different equations which we can easily show [by substituting from (23.5)] coincide with equations (26.1) and (26.2).

We can construct the four-vector which is dual to this antisymmetric four-tensor of rank three by multiplying the tensor by e^{iklm} and contracting on three pairs of indices (see § 6). Thus (26.5) can be written in the form

$$e^{iklm} \frac{\partial F_{lm}}{\partial x^k} = 0, \tag{26.6}$$

which shows explicitly that there are only four independent equations.

§ 27. The action function of the electromagnetic field

The action function S for the whole system, consisting of an electromagnetic field as well as the particles located in it, must consist of three parts:

$$S = S_f + S_m + S_{mf}, \tag{27.1}$$

where S_m is that part of the action which depends only on the properties of the particles, that is, just the action for free particles. For a single free particle, it is given by (8.1). If there are several particles, then their total action is the sum of the actions for each of the individual particles. Thus,

$$S_m = -\Sigma mc \int ds. \tag{27.2}$$

The quantity S_{mf} is that part of the action which depends on the interaction between the particles and the field. According to § 16, we have for a system of particles:

$$S_{mf} = -\sum \frac{e}{c} \int A_k dx^k . \tag{27.3}$$

In each term of this sum, A_k is the potential of the field at that point of spacetime at which the corresponding particle is located. The sum $S_m + S_{mf}$ is already familiar to us as the action (16.1) for charges in a field.

Finally S_f is that part of the action which depends only on the properties of the field itself, that is, S_f is the action for a field in the absence of charges. Up to now, because we were interested only in the motion of charges in a *given* electromagnetic field, the quantity S_f, which does not depend on the particles, did not concern us, since this term cannot affect the motion of the particles. Nevertheless this term is necessary when we want to find equations determining the field itself. This corresponds to the fact that from the parts $S_m + S_{mf}$ of the action we found only two equations for the field, (26.1) and (26.2), which are not yet sufficient for complete determination of the field.

To establish the form of the action S_f for the field, we start from the following very important property of electromagnetic fields. As experiment shows, the electromagnetic field satisfies the so-called *principle of superposition.* This principle consists in the statement that the field produced by a system of charges is the result of a simple composition of the fields produced by each of the particles individually. This means that the resultant field intensity at each point is equal to the vector sum of the individual field intensities at that point.

Every solution of the field equations gives a field that can exist in nature. According to the principle of superposition, the sum of any such fields must be a field that can exist in nature, that is, must satisfy the field equations.

As is well known, linear differential equations have just this property, that the sum of any solutions is also a solution. Consequently the field equations must be linear differential equations.

From the discussion, it follows that under the integral sign for the action S_f there must stand an expression quadratic in the field. Only in this case will the field equations be linear; the field equations are obtained by varying the action, and in the variation the degree of the expression under the integral sign decreases by unity.

The potentials cannot enter into the expression for the action S_f, since they are not uniquely determined (in S_{mf} this lack of uniqueness was not important). Therefore S_f must be the integral of some function of the electromagnetic field tensor F_{ik}. But the action must be a scalar and must therefore be the integral of some scalar. The only such quantity is the product $F_{ik}F^{ik}$.†

† The function in the integrand of S_f must not include derivatives of F_{ik}, since the Lagrangian can contain, aside from the coordinates, only their first time derivatives. The role of "coordinates" (i.e., parameters to be varied in the principle of least action) is in this case played by the field potential A_k; this is analogous to the situation in mechanics where the Lagrangian of a mechanical system contains only the coordinates of the particles and their first time derivatives.

As for the quantity $e^{iklm}F_{ik}F_{lm}$ (§ 25), as pointed out in the footnote on p. 68, it is a complete four-divergence, so that adding it to the integrand in S_f would have no effect on the "equations of motion". It is interesting that this quantity is already excluded from the action for a reason independent of the fact that it is a pseudoscalar and not a true scalar.

Thus S_f must have the form:

$$S_f = a \iint F_{ik} F^{ik} dV dt, \quad dV = dx\, dy\, dz,$$

where the integral extends over all of space and the time between two given moments; a is some constant. Under the integral stands $F_{ik}F^{ik} = 2(H^2 - E^2)$. The field $\mathbf{E}$ contains the derivative $\partial \mathbf{A}/\partial t$; but it is easy to see that $(\partial \mathbf{A}/\partial t)^2$ must appear in the action with the positive sign (and therefore E^2 must have a positive sign). For if $(\partial \mathbf{A}/\partial t)^2$ appeared in S_f with a minus sign, then sufficiently rapid change of the potential with time (in the time interval under consideration) could always make S_f a negative quantity with arbitrarily large absolute value. Consequently S_f could not have a minimum, as is required by the principle of least action. Thus, a must be negative.

The numerical value of a depends on the choice of units for measurement of the field. We note that after the choice of a definite value for a and for the units of measurement of field, the units for measurement of all other electromagnetic quantities are determined.

From now on we shall use the *Gaussian system of units;* in this system a is a dimensionless quantity, equal to $-(1/16\pi)$.†

Thus the action for the field has the form

$$S_f = -\frac{1}{16\pi c} \int F_{ik} F^{ik} d\Omega, \quad d\Omega = c\, dt\, dx\, dy\, dz. \tag{27.4}$$

In three-dimensional form:

$$S_f = \frac{1}{8\pi} \int (E^2 - H^2) dV dt. \tag{27.5}$$

In other words, the Lagrangian for the field is

$$L_f = \frac{1}{8\pi} \int (E^2 - H^2) dV. \tag{27.6}$$

The action for field plus particles has the form

$$S = -\Sigma \int mc ds - \Sigma \int \frac{e}{c} A_k dx^k - \frac{1}{16\pi c} \int f_{ik} F^{ik} d\Omega. \tag{27.7}$$

We emphasize that now the charges are not assumed to be small, as in the derivation of the equation of motion of a charge in a given field. Therefore A_k and F_{ik} refer to the actual field, that is, the external field plus the field produced by the particles themselves; A_k and F_{ik} now depend on the positions and velocities of the charges.

§ 28. The four-dimensional current vector

Instead of treating charges as points, for mathematical convenience we frequently consider them to be distributed continuously in space. Then we can introduce the "charge density" ϱ

† In addition to the Gaussian system, one also uses the Heaviside system, in which $a = -\frac{1}{4}$. In this system of units the field equations have a more convenient form (4π does not appear) but on the other hand, π appears in the Coulomb law. Conversely, in the Gaussian system the field equations contain 4π, but the Coulomb law has a simple form.

such that ϱdV is the charge contained in the volume dV. The density ϱ is in general a function of the coordinates and the time. The integral of ϱ over a certain volume is the charge contained in that volume.

Here we must remember that charges are actually pointlike, so that the density ϱ is zero everywhere except at points where the point charges are located, and the integral $\int \varrho dV$ must be equal to the sum of the charges contained in the given volume. Therefore ϱ can be expressed with the help of the δ-function in the following form†:

$$\varrho = \sum_a e_a \delta(\mathbf{r} - \mathbf{r}_a) \tag{28.1}$$

where the sum goes over all the charges and $\mathbf{r}_a$ is the radius vector of the charge e_a.

The charge on a particle is, from its very definition, an invariant quantity, that is, it does not depend on the choice of reference system. On the other hand, the density ϱ is not generally an invariant—only the product ϱdV is invariant.

Multiplying the equality $de = \varrho dV$ on both sides with dx^i:

$$de\, dx^i = \varrho dV dx^i = \varrho dV dt \frac{dx^i}{dt}.$$

† The δ-function $\delta(x)$ is defined as follows: $\delta(x) = 0$, for all nonzero values of x; for $x = 0$, $\delta(0) = \infty$, in such a way that the integral

$$\int_{-\infty}^{+\infty} \delta(x)dx = 1. \tag{I}$$

From this definition there result the following properties: if $f(x)$ is any continuous function, then

$$\int_{-\infty}^{+\infty} f(x)\delta(x-a)\,dx = f(a), \tag{II}$$

and in particular,

$$\int_{-\infty}^{+\infty} f(x)\delta(x)\,dx = \mathbf{f}(0). \tag{III}$$

(The limits of integration, it is understood, need not be $\pm\infty$; the range of integration can be arbitrary, provided it includes the point at which the δ-function does not vanish.)

The meaning of the following equalities is that the left and right sides give the same result when introduced as factors under an integral sign:

$$\delta(-x) = \delta(x), \qquad \delta(ax) = \frac{1}{|a|}\delta(x). \tag{IV}$$

The last equality is a special case of the more general relation

$$\delta[\phi(x)] = \sum_i \frac{1}{|\phi'(a_i)|}\delta(x - a_i), \tag{V}$$

where $\phi(x)$ is a single-valued function (whose inverse need not be single-valued) and the a_i are the roots of the equation $\phi(x) = 0$.

Just as $\delta(x)$ was defined for one variable x, we can introduce a three-dimensional δ-function, $\delta(\mathbf{r})$, equal to zero everywhere except at the origin of the three-dimensional coordinate system, and whose integral overall space is unity. As such a function we can clearly use the product $\delta(x)\,\delta(y)\,\delta(z)$.

On the left stands a four-vector (since de is a scalar and dx^i is a four-vector). This means that the right side must be a four-vector. But $dV\,dt$ is a scalar, and so $\varrho(dx^i/dt)$ is a four-vector. This vector (we denote it by j^i) is called the *current four-vector:*

$$j^i = \varrho\frac{dx^i}{dt}. \tag{28.2}$$

The space components of this vector form the *current density vector,*

$$\mathbf{j} = \varrho\mathbf{v}, \tag{28.3}$$

where $\mathbf{v}$ is the velocity of the charge at the given point. The time component of the four-vector (28.2) is $c\varrho$. Thus

$$j^i = (c\varrho, \mathbf{j}). \tag{28.4}$$

The total charge present in all of space is equal to the integral $\int \varrho dV$ over all space. We can write this integral in four-dimensional form:

$$\int \varrho dV = \frac{1}{c}\int j^0 dV = \frac{1}{c}\int j^i dS_i, \tag{28.5}$$

where the integral is taken over the entire four-dimensional hyperplane perpendicular to the x^0 axis (clearly this integration means integration over the whole three-dimensional space). Generally, the integral

$$\frac{1}{c}\int j^i dS_i$$

over an arbitrary hypersurface is the sum of the charges whose world lines pass through this surface.

Let us introduce the current four-vector into the expression (27.7) for the action and transform the second term in that expression. Introducing in place of the point charges e a continuous distribution of charge with density ϱ, we must write this term as

$$-\frac{1}{c}\int \varrho A_i dx^i dV,$$

replacing the sum over the charges by an integral over the whole volume. Rewriting in the form

$$-\frac{1}{c}\int \varrho\frac{dx^i}{dt}A_i dV dt,$$

we see that this term is equal to

$$-\frac{1}{c^2}\int A_i j^i d\Omega.$$

Thus the action S takes the form

$$S = -\Sigma\int mc\,ds - \frac{1}{c^2}\int A_i j^i d\Omega - \frac{1}{16\pi c}\int F_{ik}F^{ik}d\Omega. \tag{28.6}$$

§ 29. The equation of continuity

The change with time of the charge contained in a certain volume is determined by the derivative

$$\frac{\partial}{\partial t}\int \varrho dV.$$

On the other hand, the change in unit time, say, is determined by the quantity of charge which in unit time leaves the volume and goes to the outside or, conversely, passes to its interior. The quantity of charge which passes in unit time through the element $d\mathbf{f}$ of the surface bounding our volume is equal to $\varrho\mathbf{v}\cdot d\mathbf{f}$, where $\mathbf{v}$ is the velocity of the charge at the point in space where the element $d\mathbf{f}$ is located. The vector $d\mathbf{f}$ is directed, as always, along the external normal to the surface, that is, along the normal toward the outside of the volume under consideration. Therefore $\varrho\mathbf{v}\cdot d\mathbf{f}$ is positive if charge leaves the volume, and negative if charge enters the volume. The total amount of charge leaving the given volume per unit time is consequently $\oint \varrho\mathbf{v}\cdot d\mathbf{f}$, where the integral extends over the whole of the closed surface bounding the volume.

From the equality of these two expressions, we get

$$\frac{\partial}{\partial t}\int \varrho dV = -\oint \varrho\mathbf{v}\cdot d\mathbf{f}. \tag{29.1}$$

The minus sign appears on the right, since the left side is positive if the total charge in the given volume increases. The equation (29.1) is the so-called *equation of continuity,* expressing the conservation of charge in integral form. Noting that $\varrho\mathbf{v}$ is the current density, we can rewrite (29.1) in the form

$$\frac{\partial}{\partial t}\int \varrho dV = -\oint \mathbf{j}\cdot d\mathbf{f}. \tag{29.2}$$

We also write this equation in differential form. To do this we apply Gauss' theorem to (29.2):

$$\oint \mathbf{j}\cdot d\mathbf{f} = \int \operatorname{div}\mathbf{j}\, dV.$$

and we find

$$\int\left(\operatorname{div}\mathbf{j} + \frac{\partial\varrho}{\partial t}\right)dV = 0.$$

Since this must hold for integration over an arbitrary volume, the integrand must be zero:

$$\operatorname{div}\mathbf{j} + \frac{\partial\varrho}{\partial t} = 0. \tag{29.3}$$

This is the equation of continuity in differential form.

It is easy to check that the expression (28.1) for ϱ in δ-function form automatically satisfies the equation (29.3). For simplicity we assume that we have altogether only one charge, so that

$$\varrho = e\delta(\mathbf{r} - \mathbf{r}_0).$$

The current $\mathbf{j}$ is then

$$\mathbf{j} = e\mathbf{v}\ \delta(\mathbf{r} - \mathbf{r}_0),$$

where $\mathbf{v}$ is the velocity of the charge. We determine the derivative $\partial\varrho/\partial t$. During the motion of the charge its coordinates change, that is, the vector $\mathbf{r}_0$ changes. Therefore

$$\frac{\partial\varrho}{\partial t} = \frac{\partial\varrho}{\partial\mathbf{r}_0} \cdot \frac{\partial\mathbf{r}_0}{\partial t}.$$

But $\partial\mathbf{r}_0/\partial t$ is just the velocity $\mathbf{v}$ of the charge. Furthermore, since ϱ is a function of $\mathbf{r} - \mathbf{r}_0$,

$$\frac{\partial\varrho}{\partial\mathbf{r}_0} = -\frac{\partial\varrho}{\partial\mathbf{r}}.$$

Consequently

$$\frac{\partial\varrho}{\partial t} = -\mathbf{v} \cdot \operatorname{grad}\varrho = -\operatorname{div}(\varrho\mathbf{v})$$

(the velocity $\mathbf{v}$ of the charge of course does not depend on $\mathbf{r}$). Thus we arrive at the equation (29.3).

It is easily verified that, in four-dimensional form, the continuity equation (29.3) is expressed by the statement that the four-divergence of the current four-vector is zero:

$$\frac{\partial j^i}{\partial x^i} = 0. \tag{29.4}$$

In the preceding section we saw that the total charge present in all of space can be written as

$$\frac{1}{c}\int j^i dS_i,$$

where the integration is extended over the hyperplane $x^0 = \text{const}$. At each moment of time, the total charge is given by such an integral taken over a different hyperplane perpendicular to the x^0 axis. It is easy to verify that the equation (29.4) actually leads to conservation of charge, that is, to the result that the integral $\int j^i dS_i$ is the same no matter what hyperplane $x^0 = \text{const}$ we integrate over. The difference between the integrals $\int j^i dS_i$ taken over two such hyperplanes can be written in the form $\oint j^i dS_i$, where the integral is taken over the whole closed hypersurface surrounding the four-volume between the two hyperplanes under consideration (this integral differs from the required integral because of the presence of the integral over the infinitely distant "sides" of the hypersurface which, however, drop out, since there are no charges at infinity). Using Gauss' theorem (6.15) we can transform this to an integral over the four-volume between the two hyperplanes and verify that

$$\oint j^i dS_i = \int \frac{\partial j^i}{\partial x^i} d\Omega = 0. \tag{29.5}$$

The proof presented clearly remains valid also for any two integrals $\int j^i dS_i$, in which the integration is extended over any two infinite hypersurfaces (and not just the hyperplanes $x^0 = \text{const}$) which each contain all of three-dimensional space. From this it follows that the integral

$$\frac{1}{c}\int j^i dS_i$$

is actually identical in value (and equal to the total charge in space) no matter over what such hypersurface the integration is taken.

We have already mentioned (see the footnote on p. 53) the close connection between the gauge invariance of the equations of electrodynamics and the law of conservation of charge. Let us show this once again using the expression for the action in the form (28.6). On replacing A_i by $A_i - (\partial f/\partial x^i)$, the integral

$$\frac{1}{c^2}\int j^i \frac{\partial f}{\partial x^i} d\Omega$$

is added to the second term in this expression. It is precisely the conservation of charge, as expressed in the continuity equation (29.4), that enables us to write the integrand as a four-divergence $\partial(fj^i)/\partial x^i$, after which, using Gauss' theorem, the integral over the four-volume is transformed into an integral over the bounding hypersurface; on varying the action, these integrals drop out and thus have no effect on the equations on motion.

§ 30. The second pair of Maxwell equations

In finding the field equations with the aid of the principle of least action we must assume the motion of the charges to be given and vary only the potentials (which serve as the "coordinates" of the system); on the other hand, to find the equations of motion we assumed the field to be given and varied the trajectory of the particle.

Therefore the variation of the first term in (28.6) is zero, and in the second we must not vary the current j^i. Thus,

$$\delta S = -\int \frac{1}{c}\left\{\frac{1}{c} j^i \delta A_i + \frac{1}{8\pi} F^{ik}\delta F_{ik}\right\} d\Omega = 0.$$

(where we have used the fact that $F^{ik}\delta F_{ik} \equiv F_{ik}\delta F^{ik}$). Substituting $F_{ik} = \partial A_k/\partial x^i - \partial A_i/\partial x^k$ we have

$$\delta S = -\int \frac{1}{c}\left\{\frac{1}{c} j^i \delta A_i + \frac{1}{8\pi} F^{ik}\frac{\partial}{\partial x^i}\delta A_k - \frac{1}{8\pi} F^{ik}\frac{\partial}{\partial x^k}\delta A_i\right\} d\Omega.$$

In the second term we interchange the indices i and k, over which the expressions are summed, and in addition replace F^{ik} by $-F^{ik}$. Then we obtain

$$\delta S = -\int \frac{1}{c}\left\{\frac{1}{c} j^i \delta A_i - \frac{1}{4\pi} F^{ik}\frac{\partial}{\partial x^k}\delta A_i\right\} d\Omega.$$

The second of these integrals we integrate by parts, that is, we apply Gauss' theorem:

$$\delta S = -\frac{1}{c}\int \left\{\frac{1}{c} j^i + \frac{1}{4\pi}\frac{\partial F^{ik}}{\partial x^k}\right\}\delta A_i d\Omega - \frac{1}{4\pi c}\int F^{ik}\delta A_i dS_k\Big|. \qquad (30.1)$$

In the second term we must insert the values at the limits of integration. The limits for the coordinates are at infinity, where the field is zero. At the limits of the time integration, that is, at the given initial and final time values, the variation of the potentials is zero, since in

accord with the principle of least action the potentials are given at these times. Thus the second term in (30.1) is zero, and we find

$$\int \left(\frac{1}{c} j^i + \frac{1}{4\pi} \frac{\partial F^{ik}}{\partial x^k} \right) \delta A_i d\Omega = 0.$$

Since according to the principle of least action, the variations δA_i are arbitrary, the coefficients of the δA_i must be set equal to zero:

$$\frac{\partial F^{ik}}{\partial x^k} = -\frac{4\pi}{c} j^i. \tag{30.2}$$

Let us express these four ($i = 0, 1, 2, 3$) equations in three-dimensional form. For $i = 1$:

$$\frac{\partial F^{11}}{\partial x} + \frac{\partial F^{12}}{\partial y} + \frac{\partial F^{13}}{\partial z} + \frac{1}{c} \frac{\partial F^{10}}{\partial t} = -\frac{4\pi}{c} j^1.$$

Substituting the values for the components of F^{ik}, we find

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} - \frac{1}{c} \frac{\partial E_x}{\partial t} = \frac{4\pi}{c} j_x.$$

This together with the two succeeding equations ($i = 2, 3$) can be written as one vector equation:

$$\operatorname{curl} \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \mathbf{j}. \tag{30.3}$$

Finally, the fourth equation ($i = 0$) gives

$$\operatorname{div} \mathbf{E} = 4\pi\varrho. \tag{30.4}$$

Equations (30.3) and (30.4) are the second pair of Maxwell equations.† Together with the first pair of Maxwell equations they completely determine the electromagnetic field, and are the fundamental equations of the theory of such fields, i.e. of *electrodynamics*.

Let us write these equations in integral form. Integrating (30.4) over a volume and applying Gauss' theorem

$$\int \operatorname{div} \mathbf{E}\, dV = \oint \mathbf{E} \cdot d\mathbf{f},$$

we get

$$\oint \mathbf{E} \cdot d\mathbf{f} = 4\pi \int \varrho dV. \tag{30.5}$$

Thus the flux of the electric field through a closed surface is equal to 4π times the total charge contained in the volume bounded by the surface.

Integrating (30.3) over an open surface and applying Stokes' theorem

† The Maxwell equations in a form applicable to point charges in the electromagnetic field in vacuum were formulated by H. A. Lorentz.

$$\int \operatorname{curl} \mathbf{H} \cdot d\mathbf{f} = \oint \mathbf{H} \cdot d\mathbf{l},$$

we find

$$\oint \mathbf{H} \cdot d\mathbf{l} = \frac{1}{c} \frac{\partial}{\partial t} \int \mathbf{E} \cdot d\mathbf{f} + \frac{4\pi}{c} \int \mathbf{j} \cdot d\mathbf{f}. \tag{30.6}$$

The quantity

$$\frac{1}{4\pi} \frac{\partial \mathbf{E}}{\partial t} \tag{30.7}$$

is called the "*displacement current*". From (30.6) written in the form

$$\oint \mathbf{H} \cdot d\mathbf{l} = \frac{4\pi}{c} \int \left(\mathbf{j} + \frac{1}{4\pi} \frac{\partial \mathbf{E}}{\partial t} \right) \cdot d\mathbf{f}, \tag{30.8}$$

we see that the circulation of the magnetic field around any contour is equal to $4\pi/c$ times the sum of the true current and displacement current passing through a surface bounded by this contour.

From the Maxwell equations we can obtain the already familiar continuity equation (29.3). Taking the divergence of both sides of (30.3), we find

$$\operatorname{div} \operatorname{curl} \mathbf{H} = \frac{1}{c} \frac{\partial}{\partial t} \operatorname{div} \mathbf{E} + \frac{4\pi}{c} \operatorname{div} \mathbf{j}.$$

But div curl $\mathbf{H} = 0$ and div $\mathbf{E} = 4\pi\varrho$, according to (30.4). Thus we arrive once more at equation (29.3). In four-dimensional form, from (30.2), we have:

$$\frac{\partial^2 F^{ik}}{\partial x^i \partial x^k} = -\frac{4\pi}{c} \frac{\partial j^i}{\partial x^i}.$$

But when the operator $\partial^2/\partial x^i \partial x^k$, which is symmetric in the indices i and k, is applied to the antisymmetric tensor F^{ik}, it gives zero identically and we arrive at the continuity equation (29.4) expressed in four-dimensional form.

§ 31. Energy density and energy flux

Let us multiply both sides of (30.3) by $\mathbf{E}$ and both sides of (26.1) by $\mathbf{H}$ and combine the resultant equations. Then we get

$$\frac{1}{c} \mathbf{E} \cdot \frac{\partial \mathbf{E}}{\partial t} + \frac{1}{c} \mathbf{H} \cdot \frac{\partial \mathbf{H}}{\partial t} = -\frac{4\pi}{c} \mathbf{j} \cdot \mathbf{E} - (\mathbf{H} \cdot \operatorname{curl} \mathbf{E} - \mathbf{E} \cdot \operatorname{curl} \mathbf{H}).$$

Using the well-known formula of vector analysis,

$$\operatorname{div} (\mathbf{a} \times \mathbf{b}) = \mathbf{b} \cdot \operatorname{curl} \mathbf{a} - \mathbf{a} \cdot \operatorname{curl} \mathbf{b},$$

we rewrite this relation in the form

$$\frac{1}{2c} \frac{\partial}{\partial t} (E^2 + H^2) = -\frac{4\pi}{c} \mathbf{j} \cdot \mathbf{E} - \operatorname{div} (\mathbf{E} \times \mathbf{H})$$

or

$$\frac{\partial}{\partial t}\left(\frac{E^2 + H^2}{8\pi}\right) = -\mathbf{j}\cdot\mathbf{E} - \operatorname{div}\mathbf{S}. \tag{31.1}$$

The vector

$$\mathbf{S} = \frac{c}{4\pi}\mathbf{E}\times\mathbf{H} \tag{31.2}$$

is called the *Poynting vector.*

We integrate (31.1) over a volume and apply Gauss' theorem to the second term on the right. Then we obtain

$$\frac{\partial}{\partial t}\int \frac{E^2 + H^2}{8\pi}\,dV = -\int \mathbf{j}\cdot\mathbf{E}\,dV - \oint \mathbf{S}\cdot d\mathbf{f}. \tag{31.3}$$

If the integral extends over all space, then the surface integral vanishes (the field is zero at infinity). Furthermore, we can express the integral $\int \mathbf{j}\cdot\mathbf{E}dV$ as a sum $\sum e\mathbf{v}\cdot\mathbf{E}$ over all the charges, and substitute from (17.7):

$$e\mathbf{v}\cdot\mathbf{E} = \frac{d}{dt}\mathscr{E}_{\text{kin}}.$$

Then (31.3) becomes

$$\frac{d}{dt}\left\{\int \frac{E^2 + H^2}{8\pi}\,dV + \sum \mathscr{E}_{\text{kin}}\right\} = 0. \tag{31.4}$$

Thus for the closed system consisting of the electromagnetic field and particles present in it, the quantity in brackets in this equation is conserved. The second term in this expression is the kinetic energy (including the rest energy of all the particles; see the footnote on p. 51), the first term is consequently the energy of the field itself. We can therefore call the quantity

$$W = \frac{E^2 + H^2}{8\pi} \tag{31.5}$$

the *energy density* of the electromagnetic field; it is the energy per unit volume of the field.

If we integrate over any finite volume, then the surface integral in (31.3) generally does not vanish, so that we can write the equation in the form

$$\frac{\partial}{\partial t}\left\{\int \frac{E^2 + H^2}{8\pi}\,dV + \sum \mathscr{E}_{\text{kin}}\right\} = -\oint \mathbf{S}\cdot d\mathbf{f}, \tag{31.6}$$

where now the second term in the brackets is summed only over the particles present in the volume under consideration. On the left stands the change in the total energy of field and particles per unit time. Therefore the integral $\oint \mathbf{S}\cdot d\mathbf{f}$ must be interpreted as the flux of field energy across the surface bounding the given volume, so that the Poynting vector $\mathbf{S}$ is this flux density—the amount of field energy passing through unit area of the surface in unit time.†

† We assume that at the given moment there are no charges on the surface itself. If this were not the case, then on the right we would have to include the energy flux transported by particles passing through the surface.

§ 32. The energy-momentum tensor

In the preceding section we derived an expression for the energy of the electromagnetic field. Now we derive this expression, together with one for the field momentum, in four-dimensional form. In doing this we shall for simplicity consider for the present an electromagnetic field without charges. Having in mind later applications (to the gravitational field), and also to simplify the calculation, we present the derivation in a general form, not specializing the nature of the system. So we consider any system whose action integral has the form

$$S = \int \Lambda\left(q, \frac{\partial q}{\partial x^i}\right) dV\, dt = \frac{1}{c}\int \Lambda d\Omega, \tag{32.1}$$

where Λ is some function of the quantities q, describing the state of the system, and of their first derivatives with respect to coordinates and time (for the electromagnetic field the components of the four-potential are the quantities q); for brevity we write here only one of the q's. We note that the space integral $\int \Lambda\, dV$ is the Lagrangian of the system, so that Λ can be considered as the *Lagrangian "density"*. The mathematical expression of the fact that the system is closed is the absence of any explicit dependence of Λ on the x^i, similarly to the situation for a closed system in mechanics, where the Lagrangian does not depend explicitly on the time.

The "equations of motion" (i.e. the field equations, if we are dealing with some field) are obtained in accordance with the principle of least action by varying S. We have (for brevity we write $q_{,i} \equiv \partial q/\partial x^i$),

$$\delta S = \frac{1}{c}\int \left(\frac{\partial \Lambda}{\partial q}\delta q + \frac{\partial \Lambda}{\partial q_{,i}}\delta q_{,i}\right) d\Omega$$

$$= \frac{1}{c}\int \left[\frac{\partial \Lambda}{\partial q}\delta q + \frac{\partial}{\partial x^i}\left(\frac{\partial \Lambda}{\partial q_{,i}}\delta q\right) - \delta q \frac{\partial}{\partial x^i}\frac{\partial \Lambda}{\partial q_{,i}}\right] d\Omega = 0.$$

The second term in the integrand, after transformation by Gauss' theorem, vanishes upon integration over all space, and we then find the following "equations of motion":

$$\frac{\partial}{\partial x^i}\frac{\partial \Lambda}{\partial q_{,i}} - \frac{\partial \Lambda}{\partial q} = 0 \tag{32.2}$$

(it is, of course, understood that we sum over any repeated index).

The remainder of the derivation is similar to the procedure in mechanics for deriving the conservation of energy. Namely, we write:

$$\frac{\partial \Lambda}{\partial x^i} = \frac{\partial \Lambda}{\partial q}\frac{\partial q}{\partial x^i} + \frac{\partial \Lambda}{\partial q_{,k}}\frac{\partial q_{,k}}{\partial x^i}.$$

Substituting (32.2) and noting that $q_{,k,i} = q_{,i,k}$, we find

$$\frac{\partial \Lambda}{\partial x^i} = \frac{\partial}{\partial x^k}\left(\frac{\partial \Lambda}{\partial q_{,k}}\right) q_{,i} + \frac{\partial \Lambda}{\partial q_{,k}}\frac{\partial q_{,i}}{\partial x^k} = \frac{\partial}{\partial x^k}\left(q_{,i}\frac{\partial \Lambda}{\partial q_{,k}}\right).$$

On the other hand, we can write

$$\frac{\partial \Lambda}{\partial x^i} = \delta_i^k \frac{\partial \Lambda}{\partial x^k},$$

so that, introducing the notation

$$T_i^k = q_{,i} \frac{\partial \Lambda}{\partial q_{,k}} - \delta_i^k \Lambda, \tag{32.3}$$

we can express the relation in the form

$$\frac{\partial T_i^k}{\partial x^k} = 0. \tag{32.4}$$

We note that if there is not one but several quantities $q^{(l)}$, then in place of (32.3) we must write

$$T_i^k = \sum_l q_{,i}^{(l)} \frac{\partial \Lambda}{\partial q_{,k}^{(l)}} - \delta_i^k \Lambda. \tag{32.5}$$

But in § 29 we saw that an equation of the form $\partial A^k/\partial x^k = 0$, i.e. the vanishing of the four-divergence of a vector, is equivalent to the statement that the integral $\int A^k dS_k$ of the vector over a hypersurface which contains all of three-dimensional space is conserved. It is clear that an analogous result holds for the divergence of a tensor; the equation (32.4) asserts that the vector $P^i = \text{const} \int T^{ik} dS_k$ is conserved.

This vector must be identified with the four-vector of momentum of the system. We choose the constant factor in front of the integral so that, in accord with our previous definition, the time component P^0 is equal to the energy of the system multiplied by $1/c$. To do this we note that

$$P^0 = \text{const} \int T^{0k} dS_k = \text{const} \int T^{00} dV$$

if the integration is extended over the hyperplane $x^0 = \text{const}$. On the other hand, according to (32.3),

$$T^{00} = \dot{q} \frac{\partial \Lambda}{\partial \dot{q}} - \Lambda. \qquad \left(\dot{q} \equiv \frac{\partial q}{\partial t}\right)$$

Comparing with the usual formulas relating the energy and the Lagrangian, we see that this quantity must be considered as the energy density of the system, and therefore $\int T^{00} dV$ is the total energy of the system. Thus we must set const = $1/c$, and we get finally for the four-momentum of the system the expression

$$P^i = \frac{1}{c} \int T^{ik} dS_k. \tag{32.6}$$

The tensor T^{ik} is called the *energy-momentum tensor* of the system.

It is necessary to point out that the definition of the tensor T^{ik} is not unique. In fact, if T^{ik} is defined by (32.3), then any other tensor of the form

$$T^{ik} + \frac{\partial}{\partial x^l} \psi^{ikl}, \quad \psi^{ikl} = -\psi^{ilk} \tag{32.7}$$

will also satisfy equation (32.4), since we have identically $\partial^2\psi^{ikl}/\partial x^k\partial x^l = 0$. The total four-momentum of the system does not change, since according to (6.17) we can write

$$\int \frac{\partial\psi^{ikl}}{\partial x^l}dS_k = \frac{1}{2}\int\left(dS_k\frac{\partial\psi^{ikl}}{\partial x^l} - dS_l\frac{\partial\psi^{ikl}}{\partial x^k}\right) = \frac{1}{2}\int\psi^{ikl}df^*_{kl},$$

where the integration on the right side of the equation is extended over the (ordinary) surface which "bounds" the hypersurface over which the integration on the left is taken. This surface is clearly located at infinity in the three-dimensional space, and since neither field nor particles are present at infinity this integral is zero. Thus the four-momentum of the system is, as it must be, a uniquely determined quantity. To define the tensor T^{ik} uniquely we can use the requirement that the four-tensor of angular momentum (see § 14) of the system be expressed in terms of the four-momentum by

$$M^{ik} = \int (x^i dP^k - x^k dP^i) = \frac{1}{c}\int (x^i T^{kl} - x^k T^{il})\,dS_l, \tag{32.8}$$

that is its "density" is expressed in terms of the "density" of momentum by the usual formula.

It is easy to determine what conditions the energy-momentum tensor must satisfy in order that this be valid. We note that the law of conservation of angular momentum can be expressed, as we already know, by setting equal to zero the divergence of the expression under the integral sign in M^{ik}. Thus

$$\frac{\partial}{\partial x^l}(x^i T^{kl} - x^k T^{il}) = 0. \tag{32.9}$$

Noting that $\partial x^i/\partial x^l = \delta^i_l$ and that $\partial T^{kl}/\partial x^l = 0$, we find from this

$$\delta^i_l T^{kl} - \delta^k_l T^{il} = T^{ki} - T^{ik} = 0$$

or

$$T^{ik} = T^{ki}, \tag{32.10}$$

that is, the energy-momentum tensor must be symmetric.

We note that T^{ik}, defined by formula (32.5), is generally speaking not symmetric, but can be made so by transformation (32.7) with suitable ψ^{ikl}. Later on (§ 94) we shall see that there is a direct method for obtaining a symmetric tensor T^{ik}.

As we mentioned above, if we carry out the integration in (32.6) over the hyperplane x^0 = const., then $\mathbf{P}^i$ takes on the form

$$P^i = \frac{1}{c}\int T^{i0}dV, \tag{32.11}$$

where the integration extends over the whole (three-dimensional) space. The space components of P^i form the three-dimensional momentum vector of the system and the time component is its energy multiplied by $1/c$. Thus the vector with components

$$\frac{1}{c}T^{10},\quad \frac{1}{c}T^{20},\quad \frac{1}{c}T^{30}$$

may be called the "*momentum density*", and the quantity

$$W = T^{00}$$

the "*energy density*".

To clarify the meaning of the remaining components of T^{ik}, we separate the conservation equation (32.4) into space and time parts:

$$\frac{1}{c}\frac{\partial T^{00}}{\partial t} + \frac{\partial T^{0\alpha}}{\partial x^\alpha} = 0, \quad \frac{1}{c}\frac{\partial T^{\alpha 0}}{\partial t} + \frac{\partial T^{\alpha\beta}}{\partial x^\beta} = 0. \tag{32.12}$$

We integrate these equations over a volume V in space. From the first equation

$$\frac{1}{c}\frac{\partial}{\partial t}\int T^{00}dV + \int \frac{\partial T^{0\alpha}}{\partial x^\alpha}dV = 0$$

or, transforming the second integral by Gauss' theorem,

$$\frac{\partial}{\partial t}\int T^{00}dV = -c\oint T^{0\alpha}df_\alpha, \tag{32.13}$$

where the integral on the right is taken over the surface surrounding the volume V (df_x, df_y, df_z are the components of the three-vector of the surface element $d\mathbf{f}$). The expression on the left is the rate of change of the energy contained in the volume V; from this it is clear that the expression on the right is the amount of energy transferred across the boundary of the volume V, and the vector $\mathbf{S}$ with components

$$cT^{01}, cT^{02}, cT^{03}$$

is its flux density—the amount of energy passing through unit surface in unit time. Thus we arrive at the important conclusion that the requirements of relativistic invariance, as expressed by the tensor character of the quantities T^{ik}, automatically lead to a definite connection between the energy flux and the momentum density: the energy flux density is equal to the momentum density multiplied by c^2.

From the second equation in (32.12) we find similarly:

$$\frac{\partial}{\partial t}\int \frac{1}{c}T^{\alpha 0}dV = -\oint T^{\alpha\beta}df_\beta. \tag{32.14}$$

On the left is the change of the momentum of the system in volume V per unit time, therefore $\oint T^{\alpha\beta}df_\beta$ is the momentum emerging from the volume V per unit time. Thus the components $T^{\alpha\beta}$ of the energy-momentum tensor constitute the three-dimensional tensor of momentum flux density; we denote it by $-\sigma_{\alpha\beta}$, where $\sigma_{\alpha\beta}$ is the stress tensor. The energy flux density is a vector; the density of flux of momentum, which is itself a vector, must obviously be a tensor (the component $T_{\alpha\beta}$ of this tensor is the amount of the α-component of the momentum passing per unit time through unit surface perpendicular to the x^β axis).

We give a table indicating the meanings of the individual components of the energy-momentum tensor:

$$T_d^{ik} = \begin{bmatrix} W & S_x/c & S_y/c & S_z/c \\ S_x/c & -\sigma_{xx} & -\sigma_{xy} & -\sigma_{xz} \\ S_y/c & -\sigma_{yx} & -\sigma_{yy} & -\sigma_{yz} \\ S_z/c & -\sigma_{zx} & -\sigma_{zy} & -\sigma_{zz} \end{bmatrix} \tag{32.15}$$

§ 33. Energy-momentum tensor of the electromagnetic field

We now apply the general relations obtained in the previous section to the electromagnetic field. For the electromagnetic field, the quantity standing under the integral sign in (32.1) is equal, according to (27.4), to

$$\Lambda = -\frac{1}{16\pi}F_{kl}F^{kl}.$$

The quantities q are the components of the four-potential of the field, A_k, so that the definition (32.5) of the tensor T_i^k becomes

$$T_i^k = \frac{\partial A_l}{\partial x^i}\frac{\partial \Lambda}{\partial\left(\frac{\partial A_l}{\partial x^k}\right)} - \delta_i^k \Lambda.$$

To calculate the derivatives of Λ which appear here, we find the variation $\delta\Lambda$. We have

$$\delta\Lambda = -\frac{1}{8\pi}F^{kl}\delta F_{kl} = -\frac{1}{8\pi}F^{kl}\left(\delta\frac{\partial A_l}{\partial x^k} - \delta\frac{\partial A_k}{\partial x^l}\right)$$

or, interchanging indices and making use of the fact that $F_{kl} = -F_{lk}$,

$$\delta\Lambda = -\frac{1}{4\pi}F^{kl}\delta\frac{\partial A_l}{\partial x^k}.$$

From this we see that

$$\frac{\partial\Lambda}{\partial\left(\frac{\partial A_l}{\partial x^k}\right)} = -\frac{1}{4\pi}F^{kl},$$

and therefore

$$T_i^k = -\frac{1}{4\pi}\frac{\partial A_l}{\partial x^i}F^{kl} + \frac{1}{16\pi}\delta_i^k F_{lm}F^{lm},$$

or, for the contravariant components:

$$T^{ik} = -\frac{1}{4\pi}\frac{\partial A^l}{\partial x_i}F_l^k + \frac{1}{16\pi}g^{ik}F_{lm}F^{lm}.$$

But this tensor is not symmetric. To symmetrize it we add the quantity

$$\frac{1}{4\pi}\frac{\partial A^i}{\partial x_l}F_l^k.$$

According to the field equation (30.2) in the absence of charges, $\partial F_l^k/\partial x_l = 0$, and therefore

$$\frac{1}{4\pi}\frac{\partial A^i}{\partial x_l}F_l^k = \frac{1}{4\pi}\frac{\partial}{\partial x^l}(A^i F^{kl}),$$

so that the change made in T^{ik} is of the form (32.7) and is admissible. Since $\partial A^l/\partial x_i$ –

$\partial A^i/\partial x_l = F^{il}$, we get finally the following expression for the energy-momentum tensor of the electromagnetic field:

$$T^{ik} = \frac{1}{4\pi}\left(-F^{il}F_l^k + \frac{1}{4}g^{ik}F_{lm}F^{lm}\right). \tag{33.1}$$

This tensor is obviously symmetric. In addition it has the property that

$$T_i^i = 0, \tag{33 2}$$

i.e. the sum of its diagonal terms is zero.

Let us express the components of the tensor T^{ik} in terms of the electric and magnetic field intensities. By using the values (23.5) for the components F^{ik}, we easily verify that the quantity T^{00} coincides with the energy density (31.5), while the components $cT^{0\alpha}$ are the same as the components of the Poynting vector (31.2). The space components $T^{\alpha\beta}$ form a three-dimensional tensor with components

$$-\sigma_{xx} = \frac{1}{8\pi}(E_y^2 + E_z^2 - E_x^2 + H_y^2 + H_z^2 - H_x^2),$$

$$-\sigma_{xy} = -\frac{1}{4\pi}(E_xE_y + H_xH_y),$$

etc., or

$$\sigma_{\alpha\beta} = \frac{1}{4\pi}\left\{+E_\alpha E_\beta + H_\alpha H_\beta - \frac{1}{2}\delta_{\alpha\beta}(E^2 + H^2)\right\}. \tag{33.3}$$

This tensor is called the *Maxwell strees tensor.*

To bring the tensor T_{ik} to diagonal form, we must transform to a reference system in which the vectors **E** and **H** (at the given point in space and moment in time) are parallel to one another or where one of them is equal to zero; as we know (§ 25), such a transformation is always possible except when **E** and **H** are mutually perpendicular and equal in magnitude. It is easy to see that after the transformation the only non-zero components of T^{ik} will be

$$T^{00} = -T^{11} = T^{22} = T^{33} = W$$

(the x axis has been taken along the direction of the field).

But if the vectors **E** and **H** are mutually perpendicular and equal in magnitude, the tensor T^{ik} cannot be brought to diagonal form.† The non-zero components in this case are

$$T^{00} = T^{33} = T^{30} = W$$

(where the x axis is taken along the direction of **E** and the y axis along **H**).

Up to now we have considered fields in the absence of charges. When charged particles are present, the energy-momentum tensor of the whole system is the sum of the energy-momentum tensors for the electromagnetic field and for the particles, where in the latter the particles are assumed not to interact with one another.

To determine the form of the energy-momentum tensor of the particles we must describe their mass distribution in space by using a "mass density" in the same way as we describe

† The fact that the reduction of the symmetric tensor T^{ik} to principal axes may be impossible is related to the fact that the four-space is pseudo-euclidean. (See also the problem in § 94.)

a distribution of point charges in terms of their density. Analogously to formula (28.1) for the charge density, we can write the mass density in the form

$$\mu = \sum_a m_a \delta(\mathbf{r} - \mathbf{r}_a), \tag{33.4}$$

where $\mathbf{r}_a$ are the radius-vectors of the particles, and the summation extends over all the particles of the system.

The "four-momentum density" of the particles is given by $\mu c u_i$. We know that this density is the component $T^{0\alpha}/c$ of the energy-momentum tensor, i.e. $T^{0x} = \mu c^2 u^\alpha (\alpha = 1, 2, 3)$. But the mass density is the time component of the four-vector $\mu/c(dx^k/dt)$ (in analogy to the charge density; see § 28). Therefore the energy-momentum tensor of the system of non-interacting particles is

$$T^{ik} = \mu c \frac{dx^i}{ds}\frac{dx^k}{dt} = \mu c\, u^i u^k \frac{ds}{dt}. \tag{33.5}$$

As expected, this tensor is symmetric.

We verify by a direct computation that the energy and momentum of the system, defined as the sum of the energies and momenta of field and particles, are actually conserved. In other words we shall verify the equation

$$\frac{\partial}{\partial x_k}(T^{(f)k}_i + T^{(p)k}_i) = 0, \tag{33.6}$$

which expresses these conservation laws.

Differentiating (33.1), we write

$$\frac{\partial T^{(f)k}_i}{\partial x^k} = \frac{1}{4\pi}\left(\frac{1}{2}F^{lm}\frac{\partial F_{lm}}{\partial x^i} - \frac{\partial F_{il}}{\partial x^k}F^{kl} - \frac{\partial F^{kl}}{\partial x^k}F_{il}\right).$$

Substituting from the Maxwell equations (26.5) and (30.2),

$$\frac{\partial F^{kl}}{\partial x^k} = \frac{4\pi}{c}j^l, \quad \frac{\partial F_{lm}}{\partial x^i} = -\frac{\partial F_{mi}}{\partial x^l} - \frac{\partial F_{il}}{\partial x^m},$$

we have:

$$\frac{\partial T^{(f)k}_i}{\partial x^k} = \frac{1}{4\pi}\left(-\frac{1}{2}\frac{\partial F_{mi}}{\partial x^l}F^{lm} - \frac{1}{2}\frac{\partial F_{il}}{\partial x^m}F^{lm} - \frac{\partial F_{il}}{\partial x^k}F^{kl} - \frac{4\pi}{c}F_{il}j^l\right).$$

By permuting the indices, we easily show that the first three terms on the right cancel one another, and we arrive at the result:

$$\frac{\partial T^{(f)k}_i}{\partial x_k} = -\frac{1}{c}F_{ik}j^k. \tag{33.7}$$

Differentiating the expression (33.5) for the energy-momentum tensor of the particles gives

$$\frac{\partial T^{(p)k}_i}{\partial x^k} = cu_i\frac{\partial}{\partial x^k}\left(\mu\frac{dx^k}{dt}\right) + \mu c\frac{dx^k}{dt}\frac{\partial u_i}{\partial x^k}.$$

The first term in this expression is zero because of the conservation of mass for non-interacting particles. In fact, the quantities $\mu(dx^k/dt)$ constitute the "mass current" four-vector, analogous to the charge current four-vector (28.2); the conservation of mass is expressed by equating to zero the divergence of this four-vector:

$$\frac{\partial}{\partial x^k}\left(\mu\frac{dx^k}{dt}\right)=0, \tag{33.8}$$

just as the conservation of charge is expressed by equation (29.4).

Thus we have:

$$\frac{\partial T^{(p)}{}_i^k}{\partial x^k}=\mu c\frac{dx^k}{dt}\frac{\partial u_i}{\partial x^k}=\mu c\frac{du_i}{dt}.$$

Next we use the equation of motion of the charges in the field, expressed in the four-dimensional form (23.4).

$$mc\frac{du_i}{ds}=\frac{e}{c}F_{ik}u^k.$$

Changing to continuous distributions of charge and mass, we have, from the definitions of the densities μ and ϱ: $\mu/m=\varrho/e$. We can therefore write the equation of motion in the form

$$\mu c\frac{du_i}{ds}=\frac{\varrho}{c}F_{ik}u^k$$

or

$$\mu c\frac{du_i}{dt}=\frac{1}{c}F_{ik}\varrho\,u^k\frac{ds}{dt}=\frac{1}{c}F_{ik}j^k.$$

Thus,

$$\frac{\partial T^{(p)}{}_i^k}{\partial x^k}=\frac{1}{c}F_{ik}j^k. \tag{33.9}$$

Combining this with (33.7), we find that we actually get zero, i.e. we arrive at equation (33.6).

PROBLEM

Find the law of transformation of the energy density, the energy flux density, and the components of the stress tensor under a Lorentz transformation.

Solution: Suppose that the K′ coordinate system moves relative to the K system along the x axis with velocity V. Applying the formulas of problem 1, § 6 to the symmetric tensor T^{ik}, we find:

$$W=\frac{1}{1-\frac{V^2}{c^2}}\left(W'+\frac{V}{c^2}S'_x-\frac{V^2}{c^2}\sigma'_{xx}\right),$$

$$S_x=\frac{1}{1-\frac{V^2}{c^2}}\left[\left(1+\frac{V^2}{c^2}\right)S'_x-VW'-V\sigma'_{xx}\right],$$

$$S_y = \frac{1}{\sqrt{1-\frac{V^2}{c^2}}}(S'_y - V\sigma'_{xy}),$$

$$\sigma_{xx} = \frac{1}{1-\frac{V^2}{c^2}}\left(\sigma'_{xx} - 2\frac{V}{c^2}S'_x - \frac{V^2}{c^2}W'\right),$$

$$\sigma_{yy} = \sigma'_{yy},\ \sigma_{zz} = \sigma'_{zz},\ \sigma_{yz} = \sigma'_{yz},$$

$$\sigma_{xy} = \frac{1}{\sqrt{1-\frac{V^2}{c^2}}}\left(\sigma'_{xy} - \frac{V}{c^2}S'_y\right)$$

and similar formulas for S_z and σ_{xz}.

§ 34. The virial theorem

Since the sum of the diagonal terms of the energy-momentum tensor of the electromagnetic field is equal to zero, the sum T_i^i for any system of interacting particles reduces to the trace of the energy-momentum tensor for the particles alone. Using (33.5), we therefore have:

$$T_i^i = T^{(p)i}_i = \mu c u_i u^i \frac{ds}{dt} = \mu c \frac{ds}{dt} = \mu c^2 \sqrt{1-\frac{v^2}{c^2}}.$$

Let us rewrite this result, shifting to a summation over the particles, i.e. writing μ as the sum (33.4). We then get finally:

$$T_i^i = \sum_a m_a c^2 \sqrt{1-\frac{v_a^2}{c^2}}\,\delta(\mathbf{r}-\mathbf{r}_a). \tag{34.1}$$

We note that, according to this formula, we have for every system:

$$T_i^i \geq 0, \tag{34.2}$$

where the equality sign holds only for the electromagnetic field without charges.

Let us consider a closed system of charged particles carrying out a finite motion, in which all the quantities (coordinates, momenta) characterizing the system vary over finite ranges.†

We average the equation

$$\frac{1}{c}\frac{\partial T^{\alpha 0}}{\partial t} + \frac{\partial T^{\alpha\beta}}{\partial x^\beta} = 0$$

[see (32.11)] with respect to the time. The average of the derivative $\partial T^{\alpha 0}/\partial t$, like the average of the derivative of any bounded quantity, is zero.‡ Therefore we get

† Here we also assume that the electromagnetic field of the system goes to zero sufficiently rapidly at infinity. In specific cases this condition may require the neglect of radiation of electromagnetic waves by the system.

‡ Let $f(t)$ be such a quantity. Then the average value of the derivative df/dt over a certain time interval T is

$$\overline{\frac{df}{dt}} = \frac{1}{T}\int_0^T \frac{df}{dt}dt = \frac{f(T)-f(0)}{T}.$$

Since $f(t)$ varies only within finite limits, then as T increases without limit, the average value of df/dt clearly goes to zero.

$$\frac{\partial}{\partial x^\beta}\overline{T_\alpha^\beta} = 0.$$

We multiply this equation by x^α and integrate over all space. We transform the integral by Gauss' theorem, keeping in mind that at infinity $T_\alpha^\beta = 0$, and so the surface integral vanishes:

$$\int x^\alpha \frac{\overline{\partial T_\alpha^\beta}}{\partial x^\beta} dV = -\int \frac{\partial x^\alpha}{\partial x^\beta}\overline{T_\alpha^\beta}\, dV = -\int \delta_\beta^\alpha \overline{T_\alpha^\beta}\, dV = 0,$$

or finally,

$$\int \overline{T_\alpha^\alpha}\, dV = 0. \tag{34.3}$$

On the basis of this equality we can write for the integral of $\overline{T_i^i} = \overline{T_\alpha^\alpha} + \overline{T_0^0}$:

$$\int \overline{T_i^i}\, dV = \int \overline{T_0^0}\, dV = \mathscr{E},$$

where $\mathscr{E}$ is the total energy of the system.

Finally, substituting (34.1) we get:

$$\mathscr{E} = \sum_a m_a c^2 \overline{\sqrt{1 - \frac{v_a^2}{c^2}}}. \tag{34.4}$$

This relation is the relativistic generalization of the *virial theorem* of classical mechanics. (See *Mechanics*, § 10.) For low velocities, it becomes

$$\mathscr{E} - \sum_a m_a c^2 = -\sum_a \overline{\frac{m_a v_a^2}{2}},$$

that is, the total energy (minus the rest energy) is equal to the negative of the average value of the kinetic energy—in agreement with the result given by the classical virial theorem for a system of charged particles (interacting according to the Coulomb law).

We must point out that our formulas have a quite formal character and need to be made more precise. The point is that the electromagnetic field energy contains terms that give an infinite contribution to the electromagnetic self-energy of point charges (see § 37). To give meaning to the corresponding expressions we should omit these terms, considering that the intrinsic electromagnetic energy is already included in the kinetic energy of the particle (9.4). This means that we should "renormalize" the energy making the replacement

$$\mathscr{E} \to \mathscr{E} - \sum_a \int \frac{\mathbf{E}_a^2 + \mathbf{H}_a^2}{8\pi} dV$$

in (34.4), where $\mathbf{E}_a$ and $\mathbf{H}_a$ are the fields produced by the a'th particle. Similarly in (34.3) we should make the replacement†

$$\int T_a^a\, dV \to \int T_a^a\, dV + \sum_a \frac{\mathbf{E}_a^2 + \mathbf{H}_a^2}{8\pi} dV.$$

† Note that without this change the expression $-\int T_a^a\, dV = \int \frac{\mathbf{E}_a^2 + \mathbf{H}_a^2}{8\pi} dV + \sum_a \frac{m_a v_a^2}{\sqrt{1 - v_a^2/c^2}}$ is essentially positive and cannot vanish.

§ 35. The energy-momentum tensor for macroscopic bodies

In addition to the energy-momentum tensor for a system of point particles (33.5), we shall also need the expression for this tensor for macroscopic bodies which are treated as being continuous.

The flux of momentum through the element $d\mathbf{f}$ of the surface of the body is just the force acting on this surface element. Therefore $-\sigma_{\alpha\beta}\,df_\beta$ is the α-component of the force acting on the element. Now we introduce a reference system in which a given element of volume of the body is at rest. In such a reference system, Pascal's law is valid, that is, the pressure p applied to a given portion of the body is transmitted equally in all directions and is everywhere perpendicular to the surface on which it acts.† Therefore we can write $\sigma_{\alpha\beta}\,df_\beta = -p\,df_\alpha$, so that the stress tensor is $\sigma_{\alpha\beta} = -p\delta_{\alpha\beta}$. As for the components $T^{\alpha 0}$, which represent the momentum density, they are equal to zero for the given volume element in the reference system we are using. The component T^{00} is as always the energy density of the body, which we denote by ε; ε/c^2 is then the mass density of the body, i.e. the mass per unit volume. We emphasize that we are talking here about the unit "proper" volume, that is, the volume in the reference system in which the given portion of the body is at rest.

Thus, in the reference system under consideration, the energy-momentum tensor (for the given portion of the body) has the form:

$$T^{ik} = \begin{pmatrix} \varepsilon & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}. \tag{35.1}$$

Now it is easy to find the expression for the energy-momentum tensor in an arbitrary reference system. To do this we introduce the four-velocity u^i for the macroscopic motion of an element of volume of the body. In the rest frame of the particular element, $u^i = (1, 0)$. The expression for T^{ik} must be chosen so that in this reference system it takes on the form (35.1). It is easy to verify that this is

$$T^{ik} = (p + \varepsilon)u^i u^k - pg^{ik}, \tag{35.2}$$

or, for the mixed components,

$$T_i^k = (p + \varepsilon)u_i u^k - p\delta_i^k.$$

This expression gives the energy-momentum tensor for a macroscopic body. The expressions for the energy density W, energy flow vector $\mathbf{S}$ and stress tensor $\sigma_{\alpha\beta}$ are:

$$W = \frac{\varepsilon + p\dfrac{v^2}{c^2}}{1 - \dfrac{v^2}{c^2}}, \qquad \mathbf{S} = \frac{(p + \varepsilon)\mathbf{v}}{1 - \dfrac{v^2}{c^2}}, \tag{35.3}$$

† Strictly speaking, Pascal's law is valid for liquids and gases. However, for solid bodies the maximum possible difference in the stress in different directions is negligible in comparison with the stresses which can play a role in the theory of relativity, so that its consideration is of no interest.

$$\sigma_{\alpha\beta} = -\frac{(p+\varepsilon)v_\alpha v_\beta}{c^2\left(1-\frac{v^2}{c^2}\right)} - p\delta_{\alpha\beta}.$$

If the velocity v of the macroscopic motion is small compared with the velocity of light, then we have approximately:

$$\mathbf{S} = (p+\varepsilon)\mathbf{v}.$$

Since S/c^2 is the momentum density, we see that in this case the sum $(p+\varepsilon)/c^2$ plays the role of the mass density of the body.

The expression for T^{ik} simplifies in the case where the velocities of all the particles making up the body are small compared with the velocity of light (the velocity of the macroscopic motion itself can be arbitrary). In this case we can neglect, in the energy density ε, all terms small compared with the rest energy, that is, we can write $\mu_0 c^2$ in place of ε, where μ_0 is the sum of the masses of the particles present in unit (proper) volume of the body (we emphasize that in the general case, μ_0 must differ from the actual mass density ε/c^2 of the body, which includes also the mass corresponding to the energy of microscopic motion of the particles in the body and the energy of their interactions). As for the pressure determined by the energy of microscopic motion of the molecules, in the case under consideration it is also clearly small compared with the rest energy $\mu_0 c^2$. Thus we find

$$T^{ik} = \mu_0 c^2 u^i u^k. \tag{35.4}$$

From the expression (35.2), we get

$$T_i^i = \varepsilon - 3p. \tag{35.5}$$

The general property (34.2) of the energy-momentum tensor of an arbitrary system now shows that the following inequality is always valid for the pressure and density of a macroscopic body:

$$p < \frac{\varepsilon}{3}. \tag{35.6}$$

Let us compare the relation (35.5) with the general formula (34.1) which we saw was valid for an arbitrary system. Since we are at present considering a macroscopic body, the expression (34.1) must be averaged over all the values of $\mathbf{r}$ in unit volume. We obtain the result

$$\varepsilon - 3p = \sum_a m_a c^2 \sqrt{1-\frac{v_a^2}{c^2}} \tag{35.7}$$

(the summation extends over all particles in unit volume).

The right side of this equation tends to zero in the ultrarelativistic limit, so in this limit the equation of state of matter is: †

$$p = \frac{\varepsilon}{3}. \tag{35.8}$$

† This limiting equation of state is obtained here assuming an electromagnetic interaction between the particles. We shall assume (when this is needed in Chapter 14) that it remains valid for the other possible interactions between particles, though there is at present no proof of this assumption.

We apply our formula to an ideal gas, which we assume to consist of identical particles. Since the particles of an ideal gas do not interact with one another, we can use formula (33.5) after averaging it. Thus for an ideal gas,

$$T^{ik} = nmc\overline{\frac{dx^i}{dt} \cdot \frac{dx^k}{ds}},$$

where n is the number of particles in unit volume and the dash means an average over all the particles. If there is no macroscopic motion in the gas then we can use for T^{ik} the expression (35.1). Comparing the two formulas, we arrive at the equations:

$$\varepsilon = nm\overline{\left(\frac{c^2}{\sqrt{1-\frac{v^2}{c^2}}}\right)}, \quad p = \frac{nm}{3}\overline{\left(\frac{v^2}{\sqrt{1-\frac{v^2}{c^2}}}\right)}. \tag{35.9}$$

These equations determine the density and pressure of a relativistic ideal gas in terms of the velocity of its particles; the second of these replaces the well-known formula $p = nm\overline{v^2}/3$ of the nonrelativistic kinetic theory of gases.

CHAPTER 5

CONSTANT ELECTROMAGNETIC FIELDS

§ 36. Coulomb's law

For a constant electric, or as it is usually called, *electrostatic* field, the Maxwell equations have the form:

$$\operatorname{div} \mathbf{E} = 4\pi\rho, \tag{36.1}$$

$$\operatorname{curl} \mathbf{E} = 0. \tag{36.2}$$

The electric field **E** is expressed in terms of the scalar potential alone by the relation

$$\mathbf{E} = -\operatorname{grad} \phi. \tag{36.3}$$

Substituting (36.3) in (36.1), we get the equation which is satisfied by the potential of a constant electric field:

$$\Delta\phi = -4\pi\rho. \tag{36.4}$$

This equation is called the *Poisson equation.* In particular, in vacuum, i.e., for $\varrho = 0$, the potential satisfies the *Laplace equation*

$$\Delta\phi = 0. \tag{36.5}$$

From the last equation it follows, in particular, that the potential of the electric field can nowhere have a maximum or a minimum. For in order that ϕ have an extreme value, it would be necessary that the first derivatives of ϕ with respect to the coordinates be zero, and that the second derivatives $\partial^2\phi/\partial x^2$, $\partial^2\phi/\partial y^2$, $\partial^2\phi/\partial z^2$ all have the same sign. The last is impossible, since in that case (36.5) could not be satisfied.

We now determine the field produced by a point charge. From symmetry considerations, it is clear that it is directed along the radius-vector from the point at which the charge e is located. From the same consideration it is clear that the value E of the field depends only on the distance R from the charge. To find this absolute value, we apply equation (36.1) in the integral form (30.5). The flux of the electric field through a spherical surface of radius R circumscribed around the charge e is equal to $4\pi R^2 E$; this flux must equal $4\pi e$. From this we get

$$E = \frac{e}{R^2}.$$

In vector notation:

$$\mathbf{E} = \frac{e\mathbf{R}}{R^3}. \tag{36.6}$$

Thus the field produced by a point charge is inversely proportional to the square of the distance from the charge. This is the *Coulomb law.* The potential of this field is, clearly,

$$\phi = \frac{e}{R}. \tag{36.7}$$

If we have a system of charges, then the field produced by this system is equal, according to the principle of superposition, to the sum of the fields produced by each of the particles individually. In particular, the potential of such a field is

$$\phi = \sum_a \frac{e_a}{R_a},$$

where R_a is the distance from the charge e_a to the point at which we are determining the potential. If we introduce the charge density ϱ, this formula takes on the form

$$\phi = \int \frac{\varrho}{R} dV, \tag{36.8}$$

where R is the distance from the volume element dV to the given point of the field.

We note a mathematical relation which is obtained from (36.4) by substituting the values of ϱ and ϕ for a point charge, i.e. $\varrho = e\delta(\mathbf{R})$ and $\phi = e/R$. We then find

$$\Delta\left(\frac{1}{R}\right) = -4\pi\delta(\mathbf{R}). \tag{36.9}$$

§ 37. Electrostatic energy of charges

We determine the energy of a system of charges. We start from the enegy of the field, that is, from the expression (31.5) for the energy density. Namely, the energy of the system of charges must be equal to

$$U = \frac{1}{8\pi}\int E^2 dV,$$

where $\mathbf{E}$ is the field produced by these charges, and the integral goes over all space. Substituting $\mathbf{E} = -$ grad ϕ, U can be changed to the following form:

$$U = -\frac{1}{8\pi}\int \mathbf{E} \cdot \text{grad}\, \phi\, dV = -\frac{1}{8\pi}\int \text{div}\,(\mathbf{E}\phi)\, dV + \frac{1}{8\pi}\int \phi\, \text{div}\, \mathbf{E}\, dV.$$

According to Gauss' theorem, the first integral is equal to the integral of $\mathbf{E}\phi$ over the surface bounding the volume of integration, but since the integral is taken over all space and since the field is zero at infinity, this integral vanishes. Substituting in the second integral, div $\mathbf{E}$ $= 4\pi\varrho$, we find the following expression for the energy of a system of charges:

$$U = \frac{1}{2}\int \varrho\phi\, dV. \tag{37.1}$$

For a system of point charges, e_a, we can write in place of the integral a sum over the charges

$$U = \frac{1}{2}\sum_a e_a\phi_a, \tag{37.2}$$

where ϕ_a is the potential of the field produced by all the charges, at the point where the charge e_a is located.

If we apply our formula to a single elementary charged particle (say, an electron), and the field which the charge itself produces, we arrive at the result that the charge must have a certain "self"-potential energy equal to $e\phi/2$, where ϕ is the potential of the field produced by the charge at the point where it is located. But we know that in the theory of relativity every elementary particle must be considered as pointlike. The potential $\phi = e/R$ of its field becomes infinite at the point $R = 0$. Thus according to electrodynamics, the electron would have to have an infinite "self-energy", and consequently also an infinite mass. The physical absurdity of this result shows that the basic principles of electrodynamics itself lead to the result that its application must be restricted to definite limits.

We note that in view of the infinity obtained from electrodynamics for the self-energy and mass, it is impossible within the framework of classical electrodynamics itself to pose the question whether the total mass of the electron is electrodynamic (that is, associated with the electromagnetic self-energy of the particle).†

Since the occurrence of the physically meaningless infinite self-energy of the elementary particle is related to the fact that such a particle must be considered as pointlike, we can conclude that electrodynamics as a logically closed physical theory presents internal contradictions when we go to sufficiently small distances. We can pose the question as to the order of magnitude of such distances. We can answer this question by noting that for the electromagnetic self-energy of the electron we should obtain a value of the order of the rest energy mc^2. If, on the other hand, we consider an electron as possessing a certain radius R_0, then its self-potential energy would be of order e^2/R_0. From the requirement that these two quantities be of the same order, $e^2/R_0 \sim mc^2$, we find

$$R_0 \sim \frac{e^2}{mc^2}. \tag{37.3}$$

This dimension (the "radius" of the electron) determines the limit of applicability of electrodynamics to the electron, and follows already from its fundamental principles. We must, however, keep in mind that actually the limits of applicability of the classical electrodynamics which is presented here lie must higher, because of the occurrence of quantum phenomena.‡

We now turn again to formula (37.2). The potentials ϕ_a which appear there are equal, from Coulomb's law, to

$$\phi_a = \sum \frac{e_b}{R_{ab}}, \tag{37.4}$$

where R_{ab} is the distance between the charges e_a, e_b. The expression for the energy (37.2) consists of two parts. First, it contains an infinite constant, the self-energy of the charges, not depending on their mutual separations. The second part is the energy of interaction of the charges, depending on their separations. Only this part has physical interest. It is equal to

$$U' = \tfrac{1}{2}\sum e_a \phi_a', \tag{37.5}$$

† From the purely formal point of view, the finiteness of the electron mass can be handled by introducing an infinite negative mass of nonelectromagnetic origin which compensates the infinity of the electromagnetic mass (mass "renormalization"). However, we shall see later (§ 75) that this does not eliminate all the internal contradictions of classical electrodynamics.

‡ Quantum effects become important for distances of the order of $\hbar/mc$, where $\hbar$ is Planck's constant. The ratio of these distances to R_0 is of order $\hbar c/e^2 \sim 137$.

where

$$\phi_a' = \sum_{b \neq a} \frac{e_b}{R_{ab}} \tag{37.6}$$

is the potential at the point of location of e_a, produced by all the charges other than e_a. In other words, we can write

$$U' = \frac{1}{2} \sum_{a \neq b} \frac{e_a e_b}{R_{ab}}. \tag{37.7}$$

In particular, the energy of interaction of two charges is

$$U' = \frac{e_1 e_2}{R_{12}}. \tag{37.8}$$

§ 38. The field of a uniformly moving charge

We determine the field produced by a charge e, moving uniformly with velocity V. We call the laboratory frame the system K; the system of reference moving with the charge is the K' system. Let the charge be located at the origin of coordinates of the K' system. The system K' moves relative to K along the X axis; the axes Y and Z are parallel to Y' and Z'. At the time $t = 0$ the origins of the two systems coincide. The coordinates of the charge in the K system are consequently $x = Vt$, $y = z = 0$. In the K' system, we have a constant electric field with vector potential $\mathbf{A}' = 0$, and scalar potential equal to $\phi' = e/R'$, where $R'^2 = x'^2 + y'^2 + z'^2$. In the K system, according to (24.1) for $\mathbf{A}' = 0$,

$$\phi = \frac{\phi'}{\sqrt{1 - \frac{V^2}{c^2}}} = \frac{e}{R'\sqrt{1 - \frac{V^2}{c^2}}}. \tag{38.1}$$

We must now express R' in terms of the coordinates x, y, z, in the K system. According to the formulas for the Lorentz transformation

$$x' = \frac{x - Vt}{\sqrt{1 - \frac{V^2}{c^2}}}, \qquad y' = y, \qquad z' = z,$$

from which

$$R'^2 = \frac{(x - Vt)^2 + \left(1 - \frac{V^2}{c^2}\right)(y^2 + z^2)}{1 - \frac{V^2}{c^2}}. \tag{38.2}$$

Substituting this in (38.1) we find

$$\phi = \frac{e}{R^*} \tag{38.3}$$

where we have introduced the notation

$$R^{*2} = (x - Vt)^2 + \left(1 - \frac{V^2}{c^2}\right)(y^2 + z^2). \tag{38.4}$$

The vector potential in the K system is equal to

$$\mathbf{A} = \phi\frac{\mathbf{V}}{c} = \frac{e\mathbf{V}}{cR^*}. \tag{38.5}$$

In the K' system the magnetic field $\mathbf{H}'$ is absent and the electric field is

$$\mathbf{E}' = \frac{e\mathbf{R}'}{R'^3}.$$

From formula (24.2), we find

$$E_x = E'_x = \frac{ex'}{R'^3}, \quad E_y = \frac{E'_y}{\sqrt{1 - \frac{V^2}{c^2}}} = \frac{ey'}{R'^3\sqrt{1 - \frac{V^2}{c^2}}},$$

$$E_z = \frac{ez'}{R'^3\sqrt{1 - \frac{V^2}{c^2}}}.$$

Substituting for R', x', y', z', their expressions in terms of x, y, z, we obtain

$$\mathbf{E} = \left(1 - \frac{V^2}{c^2}\right)\frac{e\mathbf{R}}{R^{*3}}, \tag{38.6}$$

where $\mathbf{R}$ is the radius vector from the charge e to the field point with coordinates x, y, z (its components are $x - Vt$, y, z).

This expression for $\mathbf{E}$ can be written in another form by introducing the angle θ between the direction of motion and the radius vector $\mathbf{R}$. It is clear that $y^2 + z^2 = R^2 \sin^2\theta$, and therefore R^{*2} can be written in the form:

$$R^{*2} = R^2\left(1 - \frac{V^2}{c^2}\sin^2\theta\right). \tag{38.7}$$

Then we have for $\mathbf{E}$,

$$\mathbf{E} = \frac{e\mathbf{R}}{R^3}\frac{1 - \frac{V^2}{c^2}}{\left(1 - \frac{V^2}{c^2}\sin^2\theta\right)^{3/2}}. \tag{38.8}$$

For a fixed distance R from the charge, the value of the field E increases as θ increases from 0 to $\pi/2$ (or as θ decreases from π to $\pi/2$). The field along the direction of motion ($\theta = 0$, π) has the smallest value; it is equal to

$$E_{\parallel} = \frac{e}{R^2}\left(1 - \frac{V^2}{c^2}\right).$$

The largest field is that perpendicular to the velocity ($\theta = \pi/2$), equal to

$$E_\perp = \frac{e}{R^2}\frac{1}{\sqrt{1-\frac{V^2}{c^2}}}.$$

We note that as the velocity increases, the field $E_\parallel$ decreases, while $E_\perp$ increases. We can describe this pictorially by saying that the electric field of a moving charge is "contracted" in the direction of motion. For velocities V close to the velocity of light, the denominator in formula (38.8) is close to zero in a narrow interval of values θ around the value $\theta = \pi/2$. The "width" of this interval is, in order of magnitude,

$$\Delta\theta \sim \sqrt{1-\frac{V^2}{c^2}}.$$

Thus the electric field of a rapidly moving charge at a given distance from it is large only in a narrow range of angles in the neighbourhood of the equatorial plane, and the width of this interval decreases with increasing V like $\sqrt{1-(V^2/c^2)}$.

The magnetic field in the K system is

$$\mathbf{H} = \frac{1}{c}\mathbf{V}\times\mathbf{E} \tag{38.9}$$

[see (24.5)]. In particular, for $V \ll c$ the electric field is given approximately by the usual formula for the Coulomb law, $\mathbf{E} = e\mathbf{R}/R^3$, and the magnetic field is

$$\mathbf{H} = \frac{e}{c}\frac{\mathbf{V}\times\mathbf{R}}{R^3}, \tag{38.10}$$

PROBLEM

Determine the force (in the K system) between two charges moving with the same velocity $\mathbf{V}$.

Solution: We shall determine the force $\mathbf{F}$ by computing the force acting on one of the charges (e_1) in the field produced by the other (e_2). Using (38.9), we have

$$\mathbf{F} = e_1\mathbf{E}_2 + \frac{e_1}{c}\mathbf{V}\times\mathbf{H}_2 = e_1\left(1-\frac{V^2}{c^2}\right)\mathbf{E}_2 + \frac{e_1}{c^2}\mathbf{V}(\mathbf{V}\cdot\mathbf{E}_2).$$

Substituting for $\mathbf{E}_2$ from (38.8), we get for the components of the force in the direction of motion (F_x) and perpendicular to it (F_y):

$$F_x = \frac{e_1e_2}{R^2}\frac{\left(1-\frac{V^2}{c^2}\right)\cos\theta}{\left(1-\frac{V^2}{c^2}\sin^2\theta\right)^{3/2}}, \qquad F_y = \frac{e_1e_2}{R^2}\frac{\left(1-\frac{V^2}{c^2}\right)^2\sin\theta}{\left(1-\frac{V^2}{c^2}\sin^2\theta\right)^{3/2}},$$

where $\mathbf{R}$ is the radius vector from e_2 to e_1, and θ is the angle between $\mathbf{R}$ and $\mathbf{V}$.

§ 39. Motion in the Coulomb field

We consider the motion of a particle with mass m and charge e in the field produced by

a second charge e'; we assume that the mass of this second charge is so large that it can be considered as fixed. Then our problem becomes the study of the motion of a charge e in a centrally symmetric electric field with potential $\phi = e'/r$.

The total energy $\mathscr{E}$ of the particle is equal to

$$\mathscr{E} = c\sqrt{p^2 + m^2c^2} + \frac{\alpha}{r},$$

where $\alpha = ee'$. If we use polar coordinates in the plane of motion of the particle, then as we know from mechanics,

$$p^2 = (M^2/r^2) + p_r^2,$$

where p_r is the radial component of the momentum, and M is the constant angular momentum of the particle. Then

$$\mathscr{E} = c\sqrt{p_r^2 + \frac{M^2}{r^2} + m^2c^2} + \frac{\alpha}{r}. \tag{39.1}$$

We discuss the question whether the particle during its motion can approach arbitrarily close to the centre. First of all, it is clear that this is never possible if the charges e and e' repel each other, that is, if e and e' have the same sign. Furthermore, in the case of attraction (e and e' of opposite sign), arbitrarily close approach to the centre is not possible if $Mc > |\alpha|$, for in this case the first term in (39.1) is always large than the second, and for $r \to 0$ the right side of the equation would approach infinity. On the other hand, if $Mc < |\alpha|$, then as $r \to 0$, this expression can remain finite (here it is understood that p_r approaches infinity). Thus, if

$$cM < |\alpha|, \tag{39.2}$$

the particle during its motion "falls in" toward the charge attracting it, in contrast to non-relativistic mechanics, where for the Coulomb field such a collapse is generally impossible (with the exception of the one case $M = 0$, where the particle e moves on a line toward the particle e').

A complete determination of the motion of a charge in a Coulomb field starts most conveniently from the Hamilton-Jacobi equation. We choose polar coordinates r, ϕ, in the plane of the motion. The Hamilton–Jacobi equation (16.11) has the form

$$-\frac{1}{c^2}\left(\frac{\partial S}{\partial t} + \frac{\alpha}{r}\right)^2 + \left(\frac{\partial S}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial S}{\partial \phi}\right)^2 + m^2c^2 = 0.$$

We seek an S of the form

$$S = -\mathscr{E}t + M\phi + f(r),$$

where $\mathscr{E}$ and M are the constant energy and angular momentum of the moving particle. The result is

$$S = -\mathscr{E}t + M\phi + \int \sqrt{\frac{1}{c^2}\left(\mathscr{E} - \frac{\alpha}{r}\right)^2 - \frac{M^2}{r^2} - m^2c^2}\, dr. \tag{39.3}$$

The trajectory is determined by the equation $\partial S/\partial M = \text{const}$. Integration of (39.3) leads to the following results for the trajectory:

(a) If $Mc > |\alpha|$,

$$(c^2M^2 - \alpha^2)\frac{1}{r} = c\sqrt{(M\mathscr{E})^2 - m^2c^2(M^2c^2 - \alpha^2)}\cos\left(\phi\sqrt{1 - \frac{\alpha^2}{c^2M^2}}\right) - \mathscr{E}\alpha. \quad (39.4)$$

(b) If $Mc < |\alpha|$,

$$(\alpha^2 - M^2c^2)\frac{1}{r} = \pm c\sqrt{(M\mathscr{E})^2 + m^2c^2(\alpha^2 - M^2c^2)}\cosh\left(\phi\sqrt{\frac{\alpha^2}{c^2M^2} - 1}\right) + \mathscr{E}\alpha. \quad (39.5)$$

(c) If $Mc = |\alpha|$,

$$\frac{2\mathscr{E}\alpha}{r} = \mathscr{E}^2 - m^2c^4 - \phi^2\left(\frac{\mathscr{E}\alpha}{cM}\right)^2. \quad (39.6)$$

The integration constant is contained in the arbitrary choice of the reference line for measurement of the angle ϕ.

In (39.4) the ambiguity of sign in front of the square root is unimportant, since it already contains the arbitrary reference origin of the angle ϕ under the cos. In the case of attraction ($\alpha < 0$) the trajectory corresponding to this equation lies entirely at finite values of r (finite motion), if $\mathscr{E} < mc^2$. If $\mathscr{E} > mc^2$, then r can go to infinity (infinite motion). The finite motion corresponds to motion in a closed orbit (ellipse) in nonrelativistic mechanics. From (39.4) it is clear that in relativistic mechanics the trajectory can never be closed; when the angle ϕ changes by 2π, the distance r from the centre does not return to its initial value. In place of ellipses we here get orbits in the form of open "rosettes". Thus, whereas in nonrelativistic mechanics the finite motion in a Coulomb field leads to a closed orbit, in relativistic mechanics the Coulomb field loses this property.

In (39.5) we must choose the positive sign for the root in case $\alpha < 0$, and the negative sign if $\alpha > 0$ [the opposite choice of sign would correspond to a reversal of the sign of the root in (39.1)].

For $\alpha < 0$ the trajectories (39.5) and (39.6) are spirals in which the distance r approaches 0 as $\phi \to \infty$. The time required for the "falling in" of the charge to the coordinate origin is finite. This can be verified by noting that the dependence of the coordinate r on the time is determined by the equation $\partial S/\partial\mathscr{E} = \text{const}$; substituting (39.3), we see that the time is determined by an integral which converges for $r \to 0$.

PROBLEMS

1. Determine the angle of deflection of a charge passing through a repulsive Coulomb field ($a > 0$).

Solution: The angle of deflection χ equals $\chi = \pi - 2\phi_0$, where $2\phi_0$ is the angle between the two asymptotes of the trajectory (39.4). We find

$$\chi = \pi - \frac{2cM}{\sqrt{c^2M^2 - a^2}}\tan^{-1}\left(\frac{v\sqrt{c^2M^2 - a^2}}{ca}\right),$$

where v is the velocity of the charge at infinity.

2. Determine the effective scattering cross section at small angles for the scattering of particles in a Coulomb field.

Solution: The effective cross section $d\sigma$ is the ratio of the number of particles scattered per second into a given element do of solid angle to the flux density of impinging particles (i.e., to the number of particles crossing one square centimetre, per second, of a surface perpendicular to the beam of particles).

Since the angle of deflection χ of the particle during its passage through the field is determined by the *impact parameter* ϱ (i.e. the distance from the centre to the line along which the particle would move in the absence of the field),

$$d\sigma = 2\pi\varrho d\varrho = 2\pi\varrho \frac{d\varrho}{d\chi} d\chi = \varrho \frac{d\varrho}{d\chi} \frac{do}{\sin\chi},$$

where $do = 2\pi \sin\chi d\chi$.† The angle of deflection (for small angles) can be taken equal to the ratio of the change in momentum to its initial value. The change in momentum is equal to the time integral of the force acting on the charge, in the direction perpendicular to the direction of motion; it is approximately $(a/r^2) \cdot (\varrho/r)$. Thus we have

$$\chi = \frac{1}{p} \int_{-\infty}^{+\infty} \frac{a\varrho\, dt}{(\varrho^2 + v^2t^2)^{3/2}} = \frac{2a}{p\varrho v}$$

(v is the velocity of the particles). From this we find the effective cross section for small χ:

$$d\sigma = 4\left(\frac{a}{pv}\right)^2 \frac{do}{\chi^4}.$$

In the nonrelativistic case, $p \equiv mv$, and the expression coincides with the one obtained from the Rutherford formula‡ for small χ.

§ 40. The dipole moment

We consider the field produced by a system of charges at large distances, that is, at distances large compared with the dimensions of the system.

We introduce a coordinate system with origin anywhere within the system of charges. Let the radius vectors of the various charges be $\mathbf{r}_a$. The potential of the field produced by all the charges at the point having the radius vector $\mathbf{R}_0$ is

$$\phi = \sum_a \frac{e_a}{|\mathbf{R}_0 - \mathbf{r}_a|} \tag{40.1}$$

(the summation goes over all charges); here $\mathbf{R}_0 - \mathbf{r}_a$ are the radius vectors from the charges e_a to the point where we are finding the potential.

We must investigate this expansion for large $\mathbf{R}_0$ ($\mathbf{R}_0 >> \mathbf{r}_a$). To do this, we expand it in powers or $\mathbf{r}_a/R_0$, using the formula

$$f(\mathbf{R}_0 - \mathbf{r}) = f(\mathbf{R}_0) - \mathbf{r} \cdot \operatorname{grad} f(\mathbf{R}_0)$$

(in the grad, the differentiation applies to the coordinates of the vector $\mathbf{R}_0$). To terms of first order,

$$\phi = \frac{\sum e_a}{R_0} - \sum e_a \mathbf{r}_a \cdot \operatorname{grad} \frac{1}{R_0}. \tag{40.2}$$

† See *Mechanics*, § 18.

‡ See *Mechanics*, § 19.

The sum

$$\mathbf{d} = \Sigma\, e_a \mathbf{r}_a \tag{40.3}$$

is called the *dipole moment* of the system of charges. It is important to note that if the sum of all the charges, Σe_a, is zero, then the dipole moment does not depend on the choice of the origin of coordinates, for the radius vectors $\mathbf{r}_a$ and $\mathbf{r}'_a$ of one and the same charge in two different coordinate systems are related by

$$\mathbf{r}'_a = \mathbf{r}_a + \mathbf{a},$$

where $\mathbf{a}$ is some constant vector. Therefore if $\Sigma e_a = 0$, the dipole moment is the same in both systems:

$$\mathbf{d}' = \Sigma\, e_a \mathbf{r}'_a = \Sigma\, e_a \mathbf{r}_a + \mathbf{a}\, \Sigma\, e_a = \mathbf{d}.$$

If we denote by e_a^+, $\mathbf{r}_a^+$ and e_a^-, $\mathbf{r}_a^-$ the positive and negative charges of the system and their radius vectors, then we can write the dipole moment in the form

$$\mathbf{d} = \Sigma\, e_a^+ \mathbf{r}_a^+ - \Sigma\, e_a^- \mathbf{r}_a^- = \mathbf{R}_a^+\, \Sigma\, e_a^+ - \mathbf{R}_a^-\, \Sigma\, e_a^- \tag{40.4}$$

where

$$\mathbf{R}^+ \frac{\Sigma\, e_a^+ \mathbf{r}_a^+}{\Sigma\, e_a^+}, \qquad \mathbf{R}^- = \frac{\Sigma\, e_a^- \mathbf{r}_a^-}{\Sigma\, e_a^-} \tag{40.5}$$

are the radius vectors of the "charge centres" for the positive and negative charges. If $\Sigma\, e_a^+ = \Sigma\, e_a^- = e$, then

$$\mathbf{d} = e\mathbf{R}_{+-}, \tag{40.6}$$

where $\mathbf{R}_{+-} = \mathbf{R}^+ - \mathbf{R}^-$ is the radius vector from the centre of negative to the centre of positive charge. In particular, if we have altogether two charges, then $\mathbf{R}_{+-}$ is the radius vector between them.

If the total charge of the system is zero, then the potential of the field of this system at large distances is

$$\phi = -\mathbf{d} \cdot \nabla \frac{1}{R_0} = \frac{\mathbf{d} \cdot \mathbf{R}_0}{R_0^3}. \tag{40.7}$$

The field intensity is:

$$\mathbf{E} = -\operatorname{grad} \frac{\mathbf{d} \cdot \mathbf{R}_0}{R_0^3} = -\frac{1}{R_0^3} \operatorname{grad}(\mathbf{d} \cdot \mathbf{R}_0) - (\mathbf{d} \cdot \mathbf{R}_0) \operatorname{grad} \frac{1}{R_0^3},$$

or finally,

$$\mathbf{E} = \frac{3(\mathbf{n} \cdot \mathbf{d})\mathbf{n} - \mathbf{d}}{R_0^3}, \tag{40.8}$$

where $\mathbf{n}$ is a unit vector along $\mathbf{R}_0$. Another useful expression for the field is

$$\mathbf{E} = (\mathbf{d} \cdot \nabla)\nabla \frac{1}{R_0}, \tag{40.9}$$

Thus the potential of the field at large distances produced by a system of charges with total

charge equal to zero is inversely proportional to the square of the distance, and the field intensity is inversely proportional to the cube of the distance. This field has axial symmetry around the direction of **d**. In a plane passing through this direction (which we choose as the z axis), the components of the vector **E** are:

$$E_z = d\frac{3\cos^2\theta - 1}{R_0^3}, \quad E_x = d\frac{3\sin\theta\cos\theta}{R_0^3}. \tag{40.10}$$

The radial and tangential components in this plane are

$$E_R = d\frac{2\cos\theta}{R_0^3}, \quad E_\theta = -d\frac{\sin\theta}{R_0^3}. \tag{40.11}$$

§ 41. Multipole moments

In the expansion of the potential in powers of $1/R_0$,

$$\phi = \phi^{(0)} + \phi^{(1)} + \phi^{(2)} + \dots, \tag{41.1}$$

the term $\phi^{(n)}$ is proportional to $1/R_0^{n+1}$. We saw that the first term, $\phi^{(0)}$, is determined by the sum of all the charges; the second term, $\phi^{(1)}$, sometimes called the dipole potential of the system, is determined by the dipole moment of the system.

The third term in the expansion is

$$\phi^{(2)} = \frac{1}{2}\sum e x_\alpha x_\beta \frac{\partial^2}{\partial X_\alpha \partial X_\beta}\left(\frac{1}{R_0}\right), \tag{41.2}$$

where the sum goes over all charges; we here drop the index numbering the charges; x_α are the components of the vector **r**, and X_α those of the vector $\mathbf{R}_0$. This part of the potential is usually called the *quadrupole potential*. If the sum of the charges and the dipole moment of the system are both equal to zero, the expansion begins with $\phi^{(2)}$.

In the expression (41.2) there enter the six quantities $\sum e x_\alpha x_\beta$. However, it is easy to see that the field depends not on six independent quantities, but only on five. This follows from the fact that the function $1/R_0$ satisfies the Laplace equation, that is,

$$\Delta\left(\frac{1}{R_0}\right) = \delta_{\alpha\beta}\frac{\partial^2}{\partial X_\alpha \partial X_\beta}\left(\frac{1}{R_0}\right) = 0.$$

We can therefore write $\phi^{(2)}$ in the form

$$\phi^{(2)} = \frac{1}{2}\sum e\left(x_\alpha x_\beta - \frac{1}{3}r^2\delta_{\alpha\beta}\right)\frac{\partial^2}{\partial X_\alpha \partial X_\beta}\left(\frac{1}{R_0}\right).$$

The tensor

$$D_{\alpha\beta} = \sum e(3x_\alpha x_\beta - r^2\delta_{\alpha\beta}) \tag{41.3}$$

is called the *quadrupole moment* of the system. From the definition of $D_{\alpha\beta}$ it is clear that the sum of its diagonal elements is zero:

$$D_{\alpha\alpha} = 0. \tag{41.4}$$

Therefore the symmetric tensor $D_{\alpha\beta}$ has altogether five independent components. With the

aid of $D_{\alpha\beta}$, we can write

$$\phi^{(2)} = \frac{D_{\alpha\beta}}{6} \frac{\partial^2}{\partial X_\alpha \partial X_\beta} \left(\frac{1}{R_0} \right), \tag{41.5}$$

or, performing the differentiation,

$$\frac{\partial^2}{\partial X_\alpha \partial X_\beta} \frac{1}{R_0} = \frac{3X_\alpha X_\beta}{R_0^5} - \frac{\delta_{\alpha\beta}}{R_0^3},$$

and using the fact that $\delta_{\alpha\beta} D_{\alpha\beta} = D_{\alpha\alpha} = 0$,

$$\phi^{(2)} = \frac{D_{\alpha\beta} n_\alpha n_\beta}{2R_0^3}. \tag{41.6}$$

Like every symmetric three-dimensional tensor, the tensor $D_{\alpha\beta}$ can be brought to principal axes. Because of (41.4), in general only two of the three principal values will be independent. If it happens that the system of charges is symmetric around some axis (the z axis)† then this axis must be one of the principal axes of the tensor $D_{\alpha\beta}$, the location of the other two axes in the x, y plane is arbitrary, and the three principal values are related to one another:

$$D_{xx} = D_{yy} = -\frac{1}{2} D_{zz}. \tag{41.7}$$

Denoting the component D_{zz} by D (in this case it is simply called the quadrupole moment), we get for the potential

$$\phi^{(2)} = \frac{D}{4R_0^3} (3\cos^2\theta - 1) = \frac{D}{2R_0^3} P_2(\cos\theta), \tag{41.8}$$

where θ is the angle between $\mathbf{R}_0$ and the z axis, and P_2 is a Legendre polynomial.

Just as we did for the dipole moment in the preceding section, we can easily show that the quadrupole moment of a system does not depend on the choice of the coordinate origin, if both the total charge and the dipole moment of the system are equal to zero.

In similar fashion we could also write the succeeding terms of the expansion (41.1). The l'th term of the expansion defines a tensor (which is called the tensor of the 2^l-pole moment) of rank l, symmetric in all its indices and vanishing when contracted on any pair of indices; it can be shown that such a tensor has $2l + 1$ independent components.‡

We shall express the general term in the expansion of the potential in another form, by using the well-known formula of the theory of spherical harmonics

$$\frac{1}{|\mathbf{R}_0 - \mathbf{r}|} = \frac{1}{\sqrt{R_0^2 + r^2 - 2rR_0\cos\chi}} = \sum_{l=0}^{\infty} \frac{r^l}{R_0^{l+1}} P_l(\cos\chi), \tag{41.9}$$

where χ is the angle between $\mathbf{R}_0$ and $\mathbf{r}$. We introduce the spherical angles Θ, Φ and θ, ϕ, formed by the vectors $\mathbf{R}_0$ and $\mathbf{r}$, respectively, with the fixed coordinate axes, and use the addition theorem for the spherical harmonics:

† We are assuming a symmetry axis of any order higher than the second.

‡ Such a tensor is said to be *irreducible*. The vanishing on contraction means that no tensor of lower rank can be formed from the components.

$$P_l(\cos\chi) = \sum_{m=-l}^{l} \frac{(l-|m|)!}{(l+|m|)!} P_l^{|m|}(\cos\Theta) P_l^{|m|}(\cos\theta) e^{-im(\Phi-\phi)}, \tag{41.10}$$

where the P_l^m are the associated Legendre polynomials.

We also introduce the spherical functions†

$$Y_{lm}(\theta,\phi) = (-1)^{m} i^{l} \sqrt{\frac{2l+1(l-m)!}{4\pi(l+m)!}} P_l^m(\cos\theta) e^{im\phi},\ m \geq 0,$$

$$Y_{l,-|m|}(\theta,\phi) = (-1)^{l-m} Y^*_{l,|m|}. \tag{41.11}$$

Then the expansion (41.9) takes the form:

$$\frac{1}{|\mathbf{R}_0 - \mathbf{r}|} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{r^l}{R_0^{l+1}} \frac{4\pi}{2l+1} Y^*_{lm}(\Theta,\Phi) Y_{lm}(\theta,\Phi).$$

Carrying out this expansion in each term of (40.1), we finally get the following expression for the l'th term of the expansion of the potential:

$$\phi^{(l)} = \frac{1}{R_0^{l+1}} \sum_{m=-l}^{l} \sqrt{\frac{4\pi}{2l+1}}\, Q_m^{(l)} Y^*_{lm}(\Theta,\Phi), \tag{41.12}$$

where

$$Q_m^{(l)} = \sum_a e_a r_a^l \sqrt{\frac{4\pi}{2l+1}}\, Y_{lm}(\theta_a,\phi_a). \tag{41.13}$$

The set of $2l+1$ quantities $Q_m^{(l)}$ form the 2^l-pole moment of the system of charges.

The quantities $Q_m^{(1)}$ defined in this way are related to the components of the dipole moment vector $\mathbf{d}$ by the formulas

$$Q_0^{(1)} = id_z, \quad Q_{\pm1}^{(1)} = \mp \frac{i}{\sqrt{2}}(d_x \pm id_y). \tag{41.14}$$

The quantities $Q_m^{(2)}$ are related to the tensor components $D_{\alpha\beta}$ by the relations

$$\begin{aligned} Q_0^{(2)} &= -\frac{1}{2} D_{zz}, \quad Q_{\pm1}^{(2)} = \pm \frac{1}{\sqrt{6}}(D_{xz} \pm iD_{yz}), \\ Q_{\pm2}^{(2)} &= -\frac{1}{2\sqrt{6}}(D_{xx} - D_{yy} \pm 2iD_{xy}). \end{aligned} \tag{41.15}$$

PROBLEM

Determine the quadrupole moment of a uniformly charged ellipsoid with respect to its centre.

Solution: Replacing the summation in (41.3) by an integration over the volume of the ellipsoid, we have:

$$D_{xx} = \rho \iiint (2x^2 - y^2 - z^2)\, dx\, dy\, dz, \text{ etc.}$$

† In accordance with the definition used to quantum mechanics.

Let us choose the coordinate axes along the axes of the ellipsoid with the origin at its centre; from symmetry considerations it is obvious that these axes are the principal axes of the tensor $D_{\alpha\beta}$. By means of the transformation

$$x = x'a, \quad y = y'b, \quad z = z'c$$

the integration over the volume of the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

is reduced to integration over the volume of the unit sphere

$$x'^2 + y'^2 + z'^2 = 1.$$

As a result we obtain:

$$D_{xx} = \frac{e}{5}(2a^2 - b^2 - c^2), \quad D_{yy} = \frac{e}{5}(2b^2 - a^2 - c^2),$$

$$D_{zz} = \frac{e}{5}(2c^2 - a^2 - b^2),$$

where $e = (4\pi/3)abc\varrho$ is the total charge of the ellipsoid.

§ 42. System of charges in an external field

We now consider a system of charges located in an external electric field. We designate the potential of this external field by $\phi(\mathbf{r})$. The potential energy of each of the charges is $e_a\phi(\mathbf{r}_a)$, and the total potential energy of the system is

$$U = \sum_a e_a \phi(\mathbf{r}_a). \tag{42.1}$$

We introduce another coordinate system with its origin anywhere within the system of charges; $\mathbf{r}_a$ is the radius vector of the charge e_a in these coordinates.

Let us assume that the external field changes slowly over the region of the system of charges, i.e. is quasiuniform with respect to the system. Then we can expand the energy U in powers of $\mathbf{r}_a$:

$$U = U^{(0)} + U^{(1)} + U^{(2)} + \ldots; \tag{42.2}$$

in this expansion the first term is

$$U^{(0)} = \phi_0 \sum e_a, \tag{42.3}$$

where ϕ_0 is the value of the potential at the origin. In this approximation, the energy of the system is the same as it would be if all the charges were located at one point (the origin).

The second term in the expansion is

$$U^{(1)} = (\text{grad } \phi)_0 \cdot \sum e_a \mathbf{r}_a$$

Introducing the field intensity $\mathbf{E}_0$ at the origin and the dipole moment $\mathbf{d}$ of the system, we have

$$U^{(1)} = -\mathbf{d} \cdot \mathbf{E}_0. \tag{42.4}$$

The total force acting on the system in the external quasiuniform field is, to the order we are considering,

$$\mathbf{F} = \mathbf{E}_0 \sum e_a + [\nabla(\mathbf{d} \cdot \mathbf{E})]_0.$$

If the total charge is zero, the first term vanishes, and

$$\mathbf{F} = (\mathbf{d} \cdot \nabla)\mathbf{E}, \tag{42.5}$$

i.e. the force is determined by the derivatives of the field intensity (taken at the origin). The total moment of the forces acting on the system is

$$\mathbf{K} = \sum (\mathbf{r}_a \times e_a \mathbf{E}_0) = \mathbf{d} \times \mathbf{E}_0, \tag{42.6}$$

i.e. to lowest order it is determined by the field intensity itself.

Let us assume that there are two systems, each having total charge zero, and with dipole moments $\mathbf{d}_1$ and $\mathbf{d}_2$, respectively. Their mutual distance is assumed to be large in comparison with their internal dimensions. Let us determine their potential energy of interaction, U. To do this we regard one of the systems as being in the field of the other. Then

$$U = -\mathbf{d}_2 \cdot \mathbf{E}_1.$$

where $\mathbf{E}_1$ is the field of the first system. Substituting (40.8) for $\mathbf{E}_1$, we find:

$$U = \frac{(\mathbf{d}_1 \cdot \mathbf{d}_2)R^2 - 3(\mathbf{d}_1 \cdot \mathbf{R})(\mathbf{d}_2 \cdot \mathbf{R})}{R^5}, \tag{42.7}$$

where $\mathbf{R}$ is the vector separation between the two systems.

For the case where one of the systems has a total charge different from zero (and equal to e), we obtain similarly

$$U = e\frac{\mathbf{d} \cdot \mathbf{R}}{R^3}, \tag{42.8}$$

where $\mathbf{R}$ is the vector directed from the dipole to the charge.

The next term in the expansion (42.1) is

$$U^{(2)} = \frac{1}{2}\sum e x_\alpha x_\beta \frac{\partial^2 \phi_0}{\partial x_\alpha \partial x_\beta}.$$

Here, as in § 41, we omit the index numbering the charge; the value of the second derivative of the potential is taken at the origin; but the potential ϕ satisfies Laplace's equation,

$$\frac{\partial^2 \phi}{\partial x_\alpha^2} = \delta_{\alpha\beta}\frac{\partial^2 \phi}{\partial x_\alpha \partial x_\beta} = 0.$$

Therefore we can write

$$U^{(2)} = \frac{1}{2}\frac{\partial^2 \phi_0}{\partial x_\alpha \partial x_\beta}\sum e\left(x_\alpha x_\beta - \frac{1}{3}\delta_{\alpha\beta} r^2\right)$$

or, finally,

$$U^{(2)} = \frac{D_{\alpha\beta}}{6}\frac{\partial^2 \phi_0}{\partial x_\alpha \partial x_\beta}. \tag{42.9}$$

The general term in the series (42.2) can be expressed in terms of the 2^l-pole moments $D_m^{(l)}$ defined in the preceding section. To do this, we first expand the potential $\phi(\mathbf{r})$ in spherical harmonics; the general form of this expansion is

$$\phi(\mathbf{r}) = \sum_{l=0}^{\infty} r^l \sum_{m=-l}^{l} \sqrt{\frac{4\pi}{2l+1}}\, a_{lm} Y_{lm}(\theta, \phi), \tag{42.10}$$

where r, θ, ϕ are the spherical coordinates of a point and the a_{lm} are constants. Forming the sum (42.1) and using the definition (41.13), we obtain:

$$U^{(l)} = \sum_{m=-l}^{l} a_{lm} Q_m^{(l)}. \tag{42.11}$$

§ 43. Constant magnetic field

Let us consider the magnetic field produced by charges which perform a finite motion, in which the particles are always within a finite region of space and the momenta also always remain finite. Such a motion has a "stationary" character, and it is of interest to consider the time average magnetic field $\overline{\mathbf{H}}$, produced by the charges; this field will now be a function only of the coordinates and not of the time, that is, it will be constant.

In order to find equations for the average magnetic field $\overline{\mathbf{H}}$, we take the time average of the Maxwell equations

$$\operatorname{div} \mathbf{H} = 0, \quad \operatorname{curl} \mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c}\mathbf{j}.$$

The first of these gives simply

$$\operatorname{div} \overline{\mathbf{H}} = 0. \tag{43.1}$$

In the second equation the average value of the derivative $\partial \mathbf{E}/\partial t$, like the derivative of any quantity which varies over a finite range, is zero (cf. the footnote on p. 90). Therefore the second Maxwell equation becomes

$$\operatorname{curl} \overline{\mathbf{H}} = \frac{4\pi}{c}\overline{\mathbf{j}}. \tag{43.2}$$

These two equations determine the constant field $\overline{\mathbf{H}}$.

We introduce the average vector potential $\overline{\mathbf{A}}$ in accordance with

$$\operatorname{curl} \overline{\mathbf{A}} = \overline{\mathbf{H}}.$$

We substitute this in equation (43.2). We find

$$\operatorname{grad} \operatorname{div} \overline{\mathbf{A}} - \Delta \overline{\mathbf{A}} = \frac{4\pi}{c}\overline{\mathbf{j}}.$$

But we know that the vector potential of a field is not uniquely defined, and we can impose an arbitrary auxiliary condition on it. On this basis, we choose the potential $\overline{\mathbf{A}}$ so that

$$\operatorname{div} \overline{\mathbf{A}} = 0. \tag{43.3}$$

Then the equation defining the vector potential of the constant magnetic field becomes

$$\Delta \overline{\mathbf{A}} = -\frac{4\pi}{c}\overline{\mathbf{j}}. \tag{43.4}$$

It is easy to find the solution of this equation by noting that (43.4) is completely analogous to the Poisson equation (36.4) for the scalar potential of a constant electric field, where in place of the charge density ϱ we here have the current density $\bar{\mathbf{j}}/c$. By analogy with the solution (36.8) of the Poisson equation, we can write

$$\bar{\mathbf{A}} = \frac{1}{c}\int \frac{\bar{\mathbf{j}}}{R}\,dV, \tag{43.5}$$

where R is the distance from the field point to the volume element dV.

In formula (43.5) we can go over from the integral to a sum over the charges, by substituting in place of $\mathbf{j}$ the product $\varrho\mathbf{v}$, and recalling that all the charges are pointlike. In this we must keep in mind that in the integral (43.5), R is simply an integration variable, and is therefore not subject to the averaging process. If we write in place of the integral

$$\int \frac{\mathbf{j}}{R}\,dV, \quad \text{the sum} \quad \Sigma \frac{e_a \mathbf{v}_a}{R_a},$$

then R_a here are the radius vectors of the various particles, which change during the motion of the charges. Therefore we must write

$$\bar{\mathbf{A}} = \frac{1}{c}\Sigma \overline{\frac{e_a \mathbf{v}_a}{R_a}}, \tag{43.6}$$

where we average the whole expression under the summation sign.

Knowing $\bar{\mathbf{A}}$, we can also find the magnetic field,

$$\bar{\mathbf{H}} = \operatorname{curl} \bar{\mathbf{A}} = \operatorname{curl} \frac{1}{c}\int \frac{\bar{\mathbf{j}}}{R}\,dV.$$

The curl operator refers to the coordinates of the field point. Therefore the curl can be brought under the integral sign and $\bar{\mathbf{j}}$ can be treated as constant in the differentiation. Applying the well-known formula

$$\operatorname{curl} f\mathbf{a} = f \operatorname{curl} \mathbf{a} + \operatorname{grad} f \times \mathbf{a},$$

where f and $\mathbf{a}$ are an arbitrary scalar and vector, to the product $\bar{\mathbf{j}}\cdot 1/R$, we get

$$\operatorname{curl} \frac{\bar{\mathbf{j}}}{R} = \operatorname{grad} \frac{1}{R} \times \bar{\mathbf{j}} = \frac{\bar{\mathbf{j}} \times \mathbf{R}}{R^3},$$

and consequently,

$$\bar{\mathbf{H}} = \frac{1}{c}\int \frac{\bar{\mathbf{j}} \times \mathbf{R}}{R^3}\,dV \tag{43.7}$$

(the radius vector $\mathbf{R}$ is directed from dV to the field point). This is the *law* of *Biot and Savart.*

§ 44. Magnetic moments

Let us consider the average magnetic field produced by a system of charges in stationary motion, at large distances from the system.

We introduce a coordinate system with its origin anywhere within the system of charges, just as we did in § 40. Again we denote the radius vectors of the various charges by $\mathbf{r}_a$, and

the radius vector of the point at which we calculate the field by $\mathbf{R}_0$. Then $\mathbf{R}_0 - \mathbf{r}_a$ is the radius vector from the charge e_a to the field point. According to (43.6), we have for the vector potential:

$$\overline{\mathbf{A}} = \frac{1}{c} \Sigma \overline{\frac{e_a \mathbf{v}_a}{|\mathbf{R}_0 - \mathbf{r}_a|}}. \tag{44.1}$$

As in § 40, we expand this expression in powers of $\mathbf{r}_a$. To terms of first order (we omit the index a), we have

$$\overline{\mathbf{A}} = \frac{1}{cR_0} \Sigma e\overline{\mathbf{v}} - \frac{1}{c} \Sigma \overline{e\mathbf{v}\left(\mathbf{r} \cdot \nabla \frac{1}{R_0}\right)}.$$

In the first term we can write

$$\Sigma e\overline{\mathbf{v}} = \overline{\frac{d}{dt} \Sigma e\mathbf{r}}.$$

But the average value of the derivative of a quantity changing within a finite interval (like $\Sigma e\mathbf{r}$) is zero. Thus there remains for $\overline{\mathbf{A}}$ the expression

$$\overline{\mathbf{A}} = -\frac{1}{c} \Sigma \overline{e\mathbf{v}\left(\mathbf{r} \cdot \nabla \frac{1}{R_0}\right)} = \frac{1}{cR_0^3} \Sigma \overline{e\mathbf{v}(\mathbf{r} \cdot \mathbf{R}_0)}.$$

We transform this expression as follows. Noting that $\mathbf{v} = \dot{\mathbf{r}}$, we can write (remembering that $\mathbf{R}_0$ is a constant vector)

$$\Sigma e(\mathbf{R}_0 \cdot \mathbf{r})\mathbf{v} = \frac{1}{2}\frac{d}{dt} \Sigma e\mathbf{r}(\mathbf{r} \cdot \mathbf{R}_0) + \frac{1}{2} \Sigma e[\mathbf{v}(\mathbf{r} \cdot \mathbf{R}_0) - \mathbf{r}(\mathbf{v} \cdot \mathbf{R}_0)].$$

Upon substitution of this expression in $\overline{\mathbf{A}}$, the average of the first term (containing the time derivative) again goes to zero, and we get

$$\overline{\mathbf{A}} = \frac{1}{2cR_0^3} \Sigma \overline{e[\mathbf{v}(\mathbf{r} \cdot \mathbf{R}_0) - \mathbf{r}(\mathbf{v} \cdot \mathbf{R}_0)]}.$$

We introduce the vector

$$\mathcal{m} = \frac{1}{2c} \Sigma e\mathbf{r} \times \mathbf{v}, \tag{44.2}$$

which is called the *magnetic moment* of the system. Then we get for $\overline{\mathbf{A}}$:

$$\overline{\mathbf{A}} = \frac{\overline{\mathcal{m}} \times \mathbf{R}_0}{R_0^3} = \nabla \frac{1}{R_0} \times \overline{\mathcal{m}} \tag{44.3}$$

Knowing the vector potential, it is easy to find the magnetic field. With the aid of the formula

$$\text{curl } (\mathbf{a} \times \mathbf{b}) = (\mathbf{b} \cdot \nabla)\mathbf{a} - (\mathbf{a} \cdot \nabla)\,\mathbf{b} + \mathbf{a} \text{ div } \mathbf{b} - \mathbf{b} \text{ div } \mathbf{a},$$

we find

$$\overline{\mathbf{H}} = \text{curl } \overline{\mathbf{A}} = \text{curl}\left(\frac{\overline{\mathcal{m}} \times \mathbf{R}_0}{R_0^3}\right) = \overline{\mathcal{m}} \text{ div } \frac{\mathbf{R}_0}{R_0^3} - (\overline{\mathcal{m}} \cdot \nabla)\frac{\mathbf{R}_0}{R_0^3}.$$

Furthermore,

$$\operatorname{div} \frac{\mathbf{R}_0}{R_0^3} = \mathbf{R}_0 \cdot \operatorname{grad} \frac{1}{R_0^3} + \frac{1}{R_0^3} \operatorname{div} \mathbf{R}_0 = 0$$

and

$$(\bar{m} \cdot \nabla) \frac{\mathbf{R}_0}{R_0^3} = \frac{1}{R_0^3} (\bar{m} \cdot \nabla)\mathbf{R}_0 + \mathbf{R}_0(\bar{m} \cdot \nabla) \frac{1}{R_0^3} = \frac{\bar{m}}{R_0^3} - \frac{3\mathbf{R}_0(\bar{m} \cdot \mathbf{R}_0)}{R_0^5}.$$

Thus,

$$\bar{\mathbf{H}} = \frac{3\mathbf{n}(\bar{m} \cdot \mathbf{n}) - \bar{m}}{R_0^3}, \tag{44.4}$$

where $\mathbf{n}$ is again the unit vector along $\mathbf{R}_0$. We see that the magnetic field is expressed in terms of the magnetic moment by the same formula by which the electric field was expressed in terms of the dipole moment [see (40.8)].

If all the charges of the system have the same ratio of charge to mass, then we can write

$$\bar{m} = \frac{1}{2c} \cdot \Sigma e\mathbf{r} \times \mathbf{v} = \frac{e}{2mc} \Sigma \bar{m}\mathbf{r} \times \mathbf{v}.$$

If the velocities of all the charges $v << c$ then $m\mathbf{v}$ is the momentum $\mathbf{p}$ of the charge and we get

$$\bar{m} = \frac{e}{2mc} \Sigma \mathbf{r} \times \mathbf{p} = \frac{e}{2mc} \mathbf{M}, \tag{44.5}$$

where $\mathbf{M} = \Sigma \mathbf{r} \times \mathbf{p}$ is the mechanical angular momentum of the system. Thus in this case, the ratio of magnetic moment to the angular momentum is constant and equal to $e/2mc$.

PROBLEM

Find the ratio of the magnetic moment to the angular momentum for a system of two charges (velocities $v << c$).

Solution: Choosing the origin of coordinates as the centre of mass of the two particles we have $m_1\mathbf{r}_1 + m_2\mathbf{r}_2 = 0$ and $\mathbf{p}_1 = -\mathbf{p}_2 = \mathbf{p}$, where $\mathbf{p}$ is the momentum of the relative motion. With the aid of these relations, we find

$$\bar{m} = \frac{1}{2c} \left(\frac{e_1}{m_1^2} + \frac{e_2}{m_2^2} \right) \frac{m_1 m_2}{m_1 + m_2} \mathbf{M}.$$

§ 45. Larmor's theorem

Let us consider a system of charges in an external constant uniform magnetic field. The time average of the force acting on the system,

$$\bar{\mathbf{F}} = \Sigma \frac{e}{c} \overline{\mathbf{v} \times \mathbf{H}} = \overline{\frac{d}{dt} \Sigma \frac{e}{c} \mathbf{r} \times \mathbf{H}},$$

is zero, as is the time average of the time derivative of any quantity which varies over a finite range. The average value of the moment of the forces is

$$\overline{\mathbf{K}} = \Sigma \frac{e}{c} \overline{(\mathbf{r} \times (\mathbf{v} \times \mathbf{H}))}$$

and is different from zero. It can be expressed in terms of the magnetic moment of the system, by expanding the vector triple product:

$$\mathbf{K} = \Sigma \frac{e}{c} \{\mathbf{v}(\mathbf{r} \cdot \mathbf{H}) - \mathbf{H}(\mathbf{v} \cdot \mathbf{r})\} = \Sigma \frac{e}{c} \left\{ \mathbf{v}(\mathbf{r} \cdot \mathbf{H}) - \frac{1}{2} \mathbf{H} \frac{d}{dt} \mathbf{r}^2 \right\}.$$

The second term gives zero after averaging, so that

$$\overline{\mathbf{K}} = \Sigma \frac{e}{c} \overline{\mathbf{v}(\mathbf{r} \cdot \mathbf{H})} = \frac{1}{2c} \Sigma e \{\overline{\mathbf{v}(\mathbf{r} \cdot \mathbf{H})} - \overline{\mathbf{r}(\mathbf{v} \cdot \mathbf{H})}\}$$

[the last transformation is analogous to the one used in deriving (44.3)], or finally

$$\overline{\mathbf{K}} = \overline{\boldsymbol{m}} \times \mathbf{H}. \tag{45.1}$$

We call attention to the analogy with formula (42.6) for the electrical case.

The Lagrangian for a system of charges in an external constant uniform magnetic field contains (compared with the Lagrangian for a closed system) the additional term

$$L_H = \Sigma \frac{e}{c} \mathbf{A} \cdot \mathbf{v} = \Sigma \frac{e}{2c} (\mathbf{H} \times \mathbf{r}) \cdot \mathbf{v} = \Sigma \frac{e}{2c} (\mathbf{r} \times \mathbf{v}) \cdot \mathbf{H} \tag{45.2}$$

[where we have used the expression (19.4) for the vector potential of a uniform field]. Introducing the magnetic moment of the system, we have:

$$L_H = \boldsymbol{m} \cdot \mathbf{H}. \tag{45.3}$$

We call attention to the analogy with the electric field; in a uniform electric field, the Lagrangian of a system of charges with total charge zero contains the term

$$L_E = \mathbf{d} \cdot \mathbf{E},$$

which in that case is the negative of the potential energy of the charge system (see § 42).

We now consider a system of charges performing a finite motion (with velocities $v \ll c$) in the centrally symmetric electric field produced by a certain fixed charge. We transform from the laboratory coordinate system to a system rotating uniformly around an axis passing through the fixed particle. From the well-known formula, the velocity $\mathbf{v}$ of the particle in the new coordinate system is related to its velocity $\mathbf{v}'$ in the old system by the relation

$$\mathbf{v}' = \mathbf{v} + \boldsymbol{\Omega} \times \mathbf{r},$$

where $\mathbf{r}$ is the radius vector of the particle and Ω is the angular velocity of the rotating coordinate system. In the fixed system the Lagrangian of the system of charges is

$$L = \Sigma \frac{mv'^2}{2} - U,$$

where U is the potential energy of the charges in the external field plus the energy of their mutual interactions. The quantity U is a function of the distances of the charges from the fixed particle and of their mutual separations; when transformed to the rotating system it obviously remains unchanged. Therefore in the new system the Lagrangian is

$$L = \Sigma \frac{m}{2} (\mathbf{v} + \boldsymbol{\Omega} \times \mathbf{r})^2 - U.$$

Let us assume that all the charges have the same charge-to-mass ratio e/m, and set

$$\boldsymbol{\Omega} = \frac{e}{2mc}\mathbf{H}. \tag{45.4}$$

Then for sufficiently small H (when we can neglect terms in H^2) the Lagrangian becomes:

$$L = \Sigma \frac{mv^2}{2} + \frac{1}{2c}\Sigma e\mathbf{H} \times \mathbf{r} \cdot \mathbf{v} - U.$$

We see that it coincides with the Lagrangian which would have described the motion of the charges in the laboratory system of coordinates in the presence of a constant magnetic field (see (45.2)).

Thus we arrive at the result that, in the nonrelativistic case, the behaviour of a system of charges all having the same e/m, performing a finite motion in a centrally symmetric electric field and in a weak uniform magnetic field $\mathbf{H}$, is equivalent to the behaviour of the same system of charges in the same electric field in a coordinate system rotating uniformly with the angular velocity (45.4). This assertion is the content of the *Larmor theorem*, and the angular velocity $\Omega = eH/2mc$ is called the *Larmor frequency*.

We can approach this same problem from a different point of view. If the magnetic field $\mathbf{H}$ is sufficiently weak, the Larmor frequency will be small compared to the frequencies of the finite motion of the system of charges. Then we may consider the averages, over times small compared to the period $2\pi/\Omega$, of quantities describing the system. These new quantities will vary slowly in time (with frequency Ω).

Let us consider the change in the average angular momentum $\mathbf{M}$ of the system. According to a well-known equation of mechanics, the derivative of $\mathbf{M}$ is equal to the moment $\mathbf{K}$ of the forces acting on the system. We therefore have, using (45.1):

$$\frac{d\overline{\mathbf{M}}}{dt} = \overline{\mathbf{K}} = \overline{m} \times \mathbf{H}.$$

If the e/m ratio is the same for all particles of the system, the angular momentum and magnetic moment are proportional to one another, and we find by using formulas (44.5) and (45.4):

$$\frac{d\overline{\mathbf{M}}}{dt} = -\Omega \times \overline{\mathbf{M}}. \tag{45.5}$$

This equation states that the vector $\overline{\mathbf{M}}$ (and with it the megnetic moment $\overline{m}$) rotates with angular velocity $-\Omega$ around the direction of the field, while its absolute magnitude and the angle which it makes with this direction remain fixed. (This motion is called the *Larmor precession*.)

CHAPTER 6

ELECTROMAGNETIC WAVES

§ 46. The wave equation

The electromagnetic field in vacuum is determined by the Maxwell equations in which we must set $\rho = 0$, $\mathbf{j} = 0$. We write them once more:

$$\operatorname{curl} \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t}, \qquad \operatorname{div} \mathbf{H} = 0, \tag{46.1}$$

$$\operatorname{curl} \mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t}, \qquad \operatorname{div} \mathbf{E} = 0. \tag{46.2}$$

These equations possess nonzero solutions. This means that an electromagnetic field can exist even in the absence of any charges.

Electromagnetic fields occurring in vacuum in the absence of charges are called *electromagnetic waves*. We now take up the study of the properties of such waves.

First of all we note that such fields must necessarily be time-varying. In fact, in the contrary case, $\partial \mathbf{H}/\partial t = \partial \mathbf{E}/\partial t = 0$ and the equations (46.1) and (46.2) go over into the equations (36.1), (36.2) and (43.1), (43.2) of a constant field in which, however, we now have $\rho = 0$, $\mathbf{j} = 0$. But the solution of these equations which is given by formulas (36.8) and (43.5) becomes zero for $\rho = 0$, $\mathbf{j} = 0$.

We derive the equations determining the potentials of electromagnetic waves.

As we already know, because of the ambiguity in the potentials we can always subject them to an auxiliary condition. For this reason, we choose the potentials of the electromagnetic wave so that the scalar potential is zero:

$$\phi = 0. \tag{46.3}$$

Then

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}, \qquad \mathbf{H} = \operatorname{curl} \mathbf{A}. \tag{46.4}$$

Substituting these two expressions in the first of equations (46.2), we get

$$\operatorname{curl} \operatorname{curl} \mathbf{A} = -\Delta \mathbf{A} + \operatorname{grad} \operatorname{div} \mathbf{A} = -\frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2}. \tag{46.5}$$

Despite the fact that we have already imposed one auxiliary condition on the potentials, the potential $\mathbf{A}$ is still not completely unique. Namely, we can add to it the gradient of an

arbitrary function which does not depend on the time (meantime leaving ϕ unchanged). In particular, we can choose the potentials of the electromagnetic wave so that

$$\operatorname{div} \mathbf{A} = 0. \tag{46.6}$$

In fact, substituting for $\mathbf{E}$ from (46.4) in $\operatorname{div} \mathbf{E} \doteq 0$, we have

$$\operatorname{div} \frac{\partial \mathbf{A}}{\partial t} = \frac{\partial}{\partial t} \operatorname{div} \mathbf{A} = 0,$$

that is, div $\mathbf{A}$ is a function only of the coordinates. This function can always be made zero by adding to $\mathbf{A}$ the gradient of a suitable time-independent function.

The equation (46.5) now becomes

$$\Delta \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = 0. \tag{46.7}$$

This is the equation which determines the potentials of electromagnetic waves. It is called the *d'Alembert equation,* or the *wave equation.*†

Applying to (46.7) the operators curl and $\partial/\partial t$, we can verify that the electric and magnetic fields $\mathbf{E}$ and $\mathbf{H}$ satisfy the same wave equation.

We repeat the derivation of the wave equation in four-dimensional form. We write the second pair of Maxwell equations for the field in the absence of charges in the form

$$\frac{\partial F^{ik}}{\partial x^k} = 0$$

(This is equation (30.2) with $j^i = 0$.) Substituting F^{ik}, expressed in terms of the potentials,

$$F^{ik} = \frac{\partial A^k}{\partial x_i} - \frac{\partial A^i}{\partial x_k},$$

we get

$$\frac{\partial^2 A^k}{\partial x_i \partial x^k} - \frac{\partial^2 A^i}{\partial x_k \partial x^k} = 0. \tag{46.8}$$

We impose on the potentials the auxiliary condition:

$$\frac{\partial A^k}{\partial x^k} = 0. \tag{46.9}$$

(This condition is called the *Lorentz condition,* and potentials that satisfy it are said to be in the *Lorentz gauge.*) Then the first term in (46.8) drops out and there remains

$$\frac{\partial^2 A^i}{\partial x_k \partial x^k} \equiv g^{kl} \frac{\partial^2 A^i}{\partial x^k \partial x^l} = 0. \tag{46.10}$$

† The wave equation is sometimes written in the form $\Box \mathbf{A} = 0$, where

$$\Box = -\frac{\partial^2}{\partial x_t \partial x^t} = \Delta - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}$$

is called the *d'Alembertian operator.*

This is the wave equation expressed in four-dimensional form.†

In three-dimensional form, the condition (46.9) is:

$$\frac{1}{c}\frac{\partial \phi}{\partial t} + \operatorname{div} \mathbf{A} = 0. \tag{46.11}$$

It is more general than the conditions $\phi = 0$ and $\operatorname{div} \mathbf{A} = 0$ that were used earlier; potentials that satisfy those conditions also satisfy (46.11). But unlike them the Lorentz condition has a relativistically invariant character: potentials satisfying it in one frame satisfy it in any other frame (whereas condition (46.6) is generally violated if the frame is changed).

§ 47. Plane waves

We consider the special case of electromagnetic waves in which the field depends only on one coordinate, say x (and on the time). Such waves are said to be *plane*. In this case the equation for the field becomes

$$\frac{\partial^2 f}{\partial t^2} - c^2 \frac{\partial^2 f}{\partial x^2} = 0, \tag{47.1}$$

where by f is understood any component of the vectors **E** or **H**.

To solve this equation, we rewrite it in the form

$$\left(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\right) f = 0,$$

and introduce new variables

$$\xi = t - \frac{x}{c}, \qquad \eta = t + \frac{x}{c}$$

so that $t = \frac{1}{2}(\eta + \xi), x = \frac{c}{2}(\eta - \xi)$. Then

$$\frac{\partial}{\partial \xi} = \frac{1}{2}\left(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\right), \qquad \frac{\partial}{\partial \eta} = \frac{1}{2}\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\right),$$

so that the equation for f becomes

$$\frac{\partial^2 f}{\partial \xi \partial \eta} = 0.$$

The solution obviously has the form $f = f_1(\xi) + f_2(\eta)$, where f_1 and f_2 are arbitrary functions. Thus

$$f = f_1\left(t - \frac{x}{c}\right) + f_2\left(t + \frac{x}{c}\right). \tag{47.2}$$

† It should be mentioned that the condition (46.9) still does not determine the choice of the potentials uniquely. We can add to **A** a term grad f, and subtract a term $1/c\ (\partial f/\partial t)$ from ϕ, where the function f is not arbitrary but must satisfy the wave equation $\Box f = 0$.

Suppose, for example, $f_2 = 0$, so that

$$f = f_1\left(t - \frac{x}{c}\right).$$

Let us clarify the meaning of this solution. In each plane $x = \text{const}$, the field changes with the time; at each given moment the field is different for different x. It is clear that the field has the same values for coordinates x and times t which satisfy the relation $t - (x/c) = \text{const}$, that is,

$$x = \text{const} + ct.$$

This means that if, at some time $t = 0$, the field at a certain point x in space had some definite value, then after an interval of time t the field has that same value at a distance ct along the X axis from the original place. We can say that all the values of the electromagnetic field are propagated in space along the X axis with a velocity equal to the velocity of light, c.

Thus,

$$f_1\left(t - \frac{x}{c}\right)$$

represents a plane wave moving in the positive direction along the X axis. It is easy to show that

$$f_2\left(t + \frac{x}{c}\right)$$

represents a wave moving in the opposite, negative, direction along the X axis.

In § 46 we showed that the potentials of the electromagnetic wave can be chosen so that $\phi = 0$, and $\operatorname{div} \mathbf{A} = 0$. We choose the potentials of the plane wave which we are now considering in this same way. The condition $\operatorname{div} \mathbf{A} = 0$ gives in this case

$$\frac{\partial A_x}{\partial x} = 0,$$

since all quantities are independent of y and z. According to (47.1) we then have also $\partial^2 A_x/\partial t^2 = 0$, that is, $\partial A_x/\partial t = \text{const}$. But the derivative $\partial \mathbf{A}/\partial t$ determines the electric field, and we see that the nonzero component A_x represents in this case the presence of a constant longitudinal electric field. Since such a field has no relation to the electromagnetic wave, we can set $A_x = 0$.

Thus the vector potential of the plane wave can always be chosen perpendicular to the X axis, i.e. to the direction of propagation of that wave.

We consider a plane wave moving in the positive direction of the X axis; in this wave, all quantities, in particular also $\mathbf{A}$, are functions only of $t - (x/c)$. From the formulas

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}, \qquad \mathbf{H} = \operatorname{curl} \mathbf{A},$$

we therefore obtain

$$\mathbf{E} = -\frac{1}{c}\mathbf{A}', \quad \mathbf{H} = \nabla \times \mathbf{A} = \nabla\left(t - \frac{x}{c}\right) \times \mathbf{A}' = -\frac{1}{c}\mathbf{n} \times \mathbf{A}', \tag{47.3}$$

where the prime denotes differentiation with respect to $t - (x/c)$ and $\mathbf{n}$ is a unit vector along the direction of propagation of the wave. Substituting the first equation in the second, we obtain

$$\mathbf{H} = \mathbf{n} \times \mathbf{E}. \tag{47.4}$$

We see that the electric and magnetic fields $\mathbf{E}$ and $\mathbf{H}$ of a plane wave are directed perpendicular to the direction of propagation of the wave. For this reason, electromagnetic waves are said to be *transverse*. From (47.4) it is clear also that the electric and magnetic fields of the plane wave are perpendicular to each other and equal to each other in absolute value.

The energy flux in the plane wave, i.e. its Poynting vector is

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

and since $\mathbf{E} \cdot \mathbf{n} = 0$,

$$\mathbf{S} = \frac{c}{4\pi} E^2 \mathbf{n} = \frac{c}{4\pi} H^2 \mathbf{n}.$$

Thus the energy flux is directed along the direction of propagation of the wave. Since

$$W = \frac{1}{8\pi} (E^2 + H^2) = \frac{E^2}{4\pi}$$

is the energy density of the wave, we can write

$$\mathbf{S} = cW\mathbf{n}, \tag{47.5}$$

in accordance with the fact that the field propagates with the velocity of light.

The momentum per unit volume of the electromagnetic field is $\mathbf{S}/c^2$. For a plane wave this gives $(W/c)\mathbf{n}$. We call attention to the fact that the relation between energy W and momentum W/c for the electromagnetic wave is the same as for a particle moving with the velocity of light [see (9.9)].

The flux of momentum of the field is determined by the components $\sigma_{\alpha\beta}$ of the Maxwell stress tensor (33.3). Choosing the direction of propagation of the wave as the X axis, we find that the only nonzero component of $T^{\alpha\beta}$ is

$$T^{xx} = -\sigma_{xx} = W. \tag{47.6}$$

As it must be, the flux of momentum is along the direction of propagation of the wave, and is equal in magnitude to the energy density.

Let us find the law of transformation of the energy density of a plane electromagnetic wave when we change from one inertial reference system to another. To do this we start from the formula

$$W = \frac{1}{1 - \dfrac{V^2}{c^2}} \left(W' + 2 \frac{V}{c^2} S'_x + \frac{V^2}{c^2} \sigma'_{xx} \right)$$

(see the problem in § 33) and must substitute

$$S'_x = cW' \cos \alpha', \; \sigma'_{xx} = -W' \cos^2 \alpha',$$

where α' is the angle (in the K' system) between the X' axis (along which the velocity $\mathbf{V}$ is directed) and the direction of propagation of the wave. We find:

$$W = W' \frac{\left(1 + \frac{V}{c} \cos \alpha'\right)^2}{1 - \frac{V^2}{c^2}}. \tag{47.7}$$

Since $W = E^2/4\pi = H^2/4\pi$, the absolute values of the field intensities in the wave transform like $\sqrt{W}$.

PROBLEMS

1. Determine the force exerted on a wall from which an incident plane electromagnetic wave is reflected (with reflection coefficient R).

Solution: The force **f** acting on unit area of the wall is given by the flux of momentum through this area, i.e., it is the vector with components

$$f_\alpha = -\sigma_{\alpha\beta} N_\beta - \sigma'_{\alpha\beta} N_\beta,$$

where **N** is the vector normal to the surface of the wall, and $\sigma_{\alpha\beta}$ and $\sigma'_{\alpha\beta}$ are the components of the energy-momentum tensors for the incident and reflected waves. Using (47.6), we obtain:

$$\mathbf{f} = W\mathbf{n}(\mathbf{N} \cdot \mathbf{n}) + W'\mathbf{n}'(\mathbf{N} \cdot \mathbf{n}').$$

From the definition of the reflection coefficient, we have: $W' = RW$. Also introducing the angle of incidence θ (which is equal to the reflection angle) and writing out components, we find the normal force ("light pressure")

$$f_N = W(1 + R) \cos^2 \theta$$

and the tangential force

$$f_t = W(1 - R) \sin \theta \cos \theta.$$

2. Use the Hamilton–Jacobi method to find the motion of a charge in the field of a plane electromagnetic wave with vector potential $\mathbf{A}[t - (x/c)]$.

Solution: We write the Hamilton–Jacobi equation in four-dimensional form:

$$g^{ik}\left(\frac{\partial S}{\partial x^i} + \frac{e}{c} A_i\right)\left(\frac{\partial S}{\partial x^k} + \frac{e}{c} A_k\right) = m^2 c^2. \tag{1}$$

The fact that the field is a plane wave means that the A^i are functions of one independent variable, which can be written in the form $\xi = k_i x^i$, where k^i is a constant four-vector with its square equal to zero, $k_i k^i = 0$ (see the following section). We subject the potentials to the Lorentz condition

$$\frac{\partial A^i}{\partial x^i} = \frac{dA^i}{d\xi} k_i = 0;$$

for the variables field this is equivalent to the condition $A^i k_i = 0$.

We seek a solution of equation (1) in the form

$$S = -f_i x^i + F(\xi),$$

where $f^i = (f^0, \mathbf{f})$ is a constant vector satisfying the condition $f_i f^i = m^2 c^2$ ($S = -f_i x^i$ is the solution of the Hamilton–Jacobi equation for a free particle with four-momentum $p^i = f^i$). Substitution in (1) gives the equation

$$\frac{e^2}{c^2}A_iA^i - 2\gamma\frac{dF}{d\xi} - \frac{2e}{c}f_iA^i = 0,$$

where the constant $\gamma = k_if^i$. Having determined F from this equation, we get

$$S = -f_ix^i - \frac{e}{c\gamma}\int f_iA^i d\xi + \frac{e^2}{2\gamma c^2}\int A_iA^i d\xi. \tag{2}$$

Changing to three-dimensional notation with a fixed reference frame, we choose the direction of propagation of the wave as the x axis. Then $\xi = ct - x$, while the constant $\gamma = f^0 - f^1$. Denoting the two-dimensional vector f_y, f_z by $\boldsymbol{\kappa}$, we find from the condition $f_if^i = (f^0)^2 - (f^1)^2 - \boldsymbol{\kappa}^2 = m^2c^2$,

$$f^0 + f^1 = \frac{m^2c^2 + \boldsymbol{\kappa}^2}{\gamma}.$$

We choose the potentials in the gauge in which $\phi = 0$, while $\mathbf{A}(\xi)$ lies in the yz plane. Then equation (2) takes the form:

$$S = \boldsymbol{\kappa}\cdot\mathbf{r} - \frac{\gamma}{2}(ct + x) - \frac{m^2c^2 + \boldsymbol{\kappa}^2}{2\gamma}\xi + \frac{e}{c\gamma}\int \boldsymbol{\kappa}\cdot\mathbf{A}d\xi - \frac{e^2}{2\gamma c^2}\int \mathbf{A}^2 d\xi.$$

According to the general rules (*Mechanics*, § 47), to determine the motion we must equate the derivatives $\partial S/\partial\kappa$, $\partial S/\partial\gamma$ to certain new constants, which can be made to vanish by a suitable choice of the coordinate and time origins. We thus obtain the parametric equations in ξ:

$$y = \frac{1}{\gamma}\kappa_y\xi - \frac{e}{c\gamma}\int A_y d\xi, \quad z = \frac{1}{\gamma}\kappa_z\xi - \frac{e}{c\gamma}\int A_z d\xi,$$

$$x = \frac{1}{2}\left(\frac{m^2c^2 + \kappa^2}{\gamma^2} - 1\right)\xi - \frac{e}{c\gamma^2}\int \boldsymbol{\kappa}\cdot\mathbf{A}d\xi + \frac{e^2}{2\gamma^2c^2}\int \mathbf{A}^2 d\xi, \quad ct = \xi + x.$$

The generalized momentum $\mathbf{P} = \mathbf{p} + (e/c)\mathbf{A}$ and the energy $\mathscr{E}$ are found by differentiating the action with respect to the coordinates and the time; this gives:

$$p_y = \kappa_y - \frac{e}{c}A_y, \quad p_z = \kappa_z - \frac{e}{c}A_z,$$

$$p_x = -\frac{\gamma}{2} + \frac{m^2c^2 + \kappa^2}{2\gamma} - \frac{e}{c\gamma}\boldsymbol{\kappa}\cdot\mathbf{A} + \frac{e^2}{2\gamma c^2}\mathbf{A}^2,$$

$$\mathscr{E} = (\gamma + p_x)c.$$

If we average these over the time, the terms of first degree in the periodic function $\mathbf{A}(\xi)$ vanish. We assume that the reference system has been chosen so that the particle is at rest in it on the average, i.e. so that its averaged momentum is zero. Then

$$\boldsymbol{\kappa} = 0, \quad \gamma^2 = m^2c^2 + \frac{e^2}{c^2}\overline{A^2}.$$

The final formulas for determining the motion have the form:

$$x = \frac{e^2}{2\gamma^2c^2}\int (\mathbf{A}^2 - \overline{\mathbf{A}^2})d\xi, \quad y = -\frac{e}{c\gamma}\int A_y d\xi, \quad z = -\frac{e}{c\gamma}\int A_z\, d\xi,$$

$$ct = \xi + \frac{e^2}{2\gamma^2 c^2}\int (\mathbf{A}^2 - \overline{\mathbf{A}^2})\, d\xi; \tag{3}$$

$$p_x = \frac{e^2}{2\gamma c^2}(\mathbf{A}^2 - \overline{\mathbf{A}^2}), \quad p_y = -\frac{e}{c}A_y, \quad p_z = -\frac{e}{c}A_z,$$

$$\mathscr{E} = c\gamma + \frac{e^2}{2\gamma c}(\mathbf{A}^2 - \overline{\mathbf{A}^2}). \tag{4}$$

§ 48. Monochromatic plane waves

A very important special case of electromagnetic waves is a wave in which the field is a simply periodic function of the time. Such a wave is said to be *monochromatic*. All quantities (potentials, field components) in a monochromatic wave depend on the time through a factor of the form $\cos(\omega t + \alpha)$. The quantity ω is called the *cyclic frequency* of the wave (we shall simply call it the *frequency*).

In the wave equation, the second derivative of the field with respect to the time is now $\partial^2 f/\partial t^2 = -\omega^2 f$, so that the distribution of the field in space is determined for a monochromatic wave by the equation

$$\Delta f + \frac{\omega^2}{c^2} f = 0. \tag{48.1}$$

In a plane wave (propagating along the x axis), the field is a function only of $t - (x/c)$. Therefore, if the plane wave is monochromatic, its field is a simply periodic function of $t - (x/c)$. The vector potential of such a wave is most conveniently written as the real part of a complex expression:

$$\mathbf{A} = \mathrm{Re}\,\{\mathbf{A}_0 e^{-i\omega(t-\frac{x}{c})}\} \tag{48.2}$$

Here $\mathbf{A}_0$ is a certain constant complex vector. Obviously, the fields $\mathbf{E}$ and $\mathbf{H}$ of such a wave have analogous forms with the same frequency ω. The quantity

$$\lambda = \frac{2\pi c}{\omega} \tag{48.3}$$

is called the *wavelength;* it is the period of variation of the field with the coordinate x at a fixed time t.

The vector

$$\mathbf{k} = \frac{\omega}{c}\mathbf{n} \tag{48.4}$$

(where $\mathbf{n}$ is a unit vector along the direction of propagation of the wave) is called the *wave vector.* In terms of it we can write (48.2) in the form

$$\mathbf{A} = \mathrm{Re}\,\{\mathbf{A}_0 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t}\}, \tag{48.5}$$

which is independent of the choice of coordinate axes. The quantity which appears multiplied by i in the exponent is called the *phase* of the wave.

So long as we perform only linear operations, we can omit the sign Re for taking the real part, and operate with complex quantities as such.† Thus, substituting

$$\mathbf{A} = \mathbf{A}_0 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$$

in (47.3), we find the relation between the intensities and the vector potential of a plane monochromatic wave in the form

$$\mathbf{E} = ik\mathbf{A}, \qquad \mathbf{H} = i\mathbf{k}\times\mathbf{A}. \tag{48.6}$$

We now treat in more detail the direction of the field of a monochromatic wave. To be specific, we shall talk of the electric field

$$\mathbf{E} = \mathrm{Re}\,\{\mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}\}$$

(everything stated below applies equally well, of course, to the magnetic field). The quantity $\mathbf{E}_0$ is a certain complex vector. Its square $\mathbf{E}_0^2$ is (in general) a complex number. If the argument of this number is -2α (i.e. $\mathbf{E}_0^2 = |\,\mathbf{E}_0^2\,| e^{-2i\alpha}$), the vector $\mathbf{b}$ defined by

$$\mathbf{E}_0 = \mathbf{b}e^{-i\alpha} \tag{48.7}$$

will have its square real, $\mathbf{b}^2 = |\,\mathbf{E}_0\,|^2$. With this definition, we write:

$$\mathbf{E} = \mathrm{Re}\,\{\mathbf{b}e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t-\alpha)}\}. \tag{48.8}$$

We write $\mathbf{b}$ in the form

$$\mathbf{b} = \mathbf{b}_1 + i\mathbf{b}_2,$$

where $\mathbf{b}_1$ and $\mathbf{b}_2$ are real vectors. Since $\mathbf{b}^2 = \mathbf{b}_1^2 - \mathbf{b}_2^2 + 2i\mathbf{b}_1\cdot\mathbf{b}_2$ must be a real quantity, $\mathbf{b}_1\cdot\mathbf{b}_2 = 0$, i.e. the vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ are mutually perpendicular. We choose the direction of $\mathbf{b}_1$ as the y axis (and the x axis along the direction of propagation of the wave). We then have from (48.8):

$$\begin{aligned} E_y &= b_1 \cos(\omega t - \mathbf{k}\cdot\mathbf{r} + \alpha), \\ E_z &= \pm\, b_2 \sin(\omega t - \mathbf{k}\cdot\mathbf{r} + \alpha), \end{aligned} \tag{48.9}$$

where we use the plus (minus) sign if $\mathbf{b}_2$ is along the positive (negative) z axis. From (48.9) it follows that

† If two quantities $\mathbf{A}(t)$ and $\mathbf{B}(t)$ are written in complex form

$$\mathbf{A}(t) = \mathbf{A}_0 e^{-i\omega t}, \qquad \mathbf{B}(t) = \mathbf{B}_0 e^{-i\omega t},$$

then in forming their product we must first, of course, separate out the real part. But if, as it frequently happens, we are interested only in the time average of this product, it can be computed as

$$\tfrac{1}{2}\,\mathrm{Re}\,\{\mathbf{A}\cdot\mathbf{B}^*\}.$$

In fact, we have:

$$\mathrm{Re}\,\mathbf{A}\cdot\mathrm{Re}\,\mathbf{B} = \tfrac{1}{4}(\mathbf{A}_0 e^{-i\omega t} + \mathbf{A}_0^* e^{i\omega t})\cdot(\mathbf{B}_0 e^{-i\omega t} + \mathbf{B}_0^* e^{i\omega t}).$$

When we average, the terms containing factors $e^{\pm 2i\omega t}$ vanish, so that we are left with

$$\overline{\mathrm{Re}\,\mathbf{A}\cdot\mathrm{Re}\,\mathbf{B}} = \tfrac{1}{4}(\mathbf{A}_0\cdot\mathbf{B}_0^* + \mathbf{A}_0^*\cdot\mathbf{B}_0) = \tfrac{1}{2}\,\mathrm{Re}\,(\mathbf{A}\cdot\mathbf{B}^*).$$

$$\frac{E_y^2}{b_1^2} + \frac{E_z^2}{b_2^2} = 1. \tag{48.10}$$

Thus we see that, at each point in space, the electric field vector rotates in a plane perpendicular to the direction of propagation of the wave, while its endpoint describes the ellipse (48.10). Such a wave is said to be *elliptically polarized.* The rotation occurs in the direction of (opposite to) a right-hand screw rotating along the x axis, if we have the plus (minus) sign in (48.9).

If $b_1 = b_2$, the ellipse (48.10) reduces to a circle, i.e. the vector $\mathbf{E}$ rotates while remaining constant in magnitude. In this case we say that the wave is *circularly polarized.* The choice of the directions of the y and z axes is now obviously arbitrary. We note that in such a wave the ratio of the y and z components of the complex amplitude $\mathbf{E}_0$ is

$$\frac{E_{0z}}{E_{0y}} = \pm i \tag{48.11}$$

for rotation in the same (opposite) direction as that of a right-hand screw *right* and *left* polarizations).†

Finally, if b_1 or b_2 equals zero, the field of the wave is everywhere and always parallel (or antiparallel) to one and the same direction. In this case the wave is said to be *linearly polarized,* or plane polarized. An elliptically polarized wave can clearly be treated as the superposition of two plane polarized waves.

Now let us turn to the definition of the wave vector and introduce the four-dimensional wave vector with components

$$k^i = \left(\frac{\omega}{c}, \mathbf{k}\right). \tag{48.12}$$

That these quantities actually form a four-vector is obvious from the fact that we get a scalar the phase of the wave) when we nultiply by x^i:

$$k_i x^i = \omega t - \mathbf{k} \cdot \mathbf{r}. \tag{48.13}$$

From the definitions (48.4) and (48.12) we see that the square of the wave four-vector is zero:

$$k^i k_i = 0. \tag{48.14}$$

This relation also follows directly from the fact that the expression

$$\mathbf{A} = \mathbf{A}_0 e^{-ik_i x^i}$$

must be a solution of the wave equation (46.10).

As is the case for every plane wave, in a monochromatic wave propagating along the x axis only the following components of the energy-momentum tensor are different from zero (see § 47):

$$T^{00} = T^{01} = T^{11} = W.$$

By means of the wave four-vector, these equations can be written in tensor form as

$$T^{ik} = \frac{Wc^2}{\omega^2} k^i k^k. \tag{48.15}$$

† We assume that the coordinate axes form a right-handed system.

Finally, by using the law of transformation of the wave four-vector we can easily treat the so-called *Doppler effect*—the change in frequency ω of the wave emitted by a source moving with respect to the observer, as compared to the "true" frequency ω_0 of the same source in the reference system (K_0) in which it is at rest.

Let V be the velocity of the source, i.e. the velocity of the K_0 system relative to K. According to the general formula for transformation of four-vectors, we have:

$$k^{(0)0} = \frac{k^0 - \frac{V}{c}k^1}{\sqrt{1 - \frac{V^2}{c^2}}}$$

(the velocity of the K system relative to K_0 is $-V$). Substituting $k^0 = \omega/c$, $k^1 = k\cos\alpha = \omega/c\cos\alpha$, where α is the angle (in the K system) between the direction of emission of the wave and the direction of motion of the source, and expressing ω in terms of ω_0, we obtain:

$$\omega = \omega_0 \frac{\sqrt{1 - \frac{V^2}{c^2}}}{1 - \frac{V}{c}\cos\alpha} \tag{48.16}$$

This is the required formula. For $V \ll c$, and if the angle α is not too close to $\pi/2$, it gives:

$$\omega \cong \omega_0\left(1 + \frac{V}{c}\cos\alpha\right). \tag{48.17}$$

For $\alpha = \pi/2$, we have:

$$\omega = \omega_0\sqrt{1 - \frac{V^2}{c^2}} \cong \omega_0\left(1 - \frac{V^2}{2c^2}\right); \tag{48.18}$$

in this case the relative change in frequency is proportional to the square of the ratio V/c.

PROBLEMS

1. Determine the direction and magnitude of the axes of the polarization ellipse in terms of the complex amplitude $\mathbf{E}_0$.

Solution: The problem consists in determining the vector $\mathbf{b} = \mathbf{b}_1 + i\mathbf{b}_2$, whose square is real. We have from (48.7):

$$\mathbf{E}_0 \cdot \mathbf{E}_0^* = b_1^2 + b_2^2, \quad \mathbf{E}_0 \times \mathbf{E}_0^* = -2i\mathbf{b}_1 \times \mathbf{b}_2, \tag{1}$$

or

$$b_1^2 \cdot b_2^2 = A^2 + B^2, \quad b_1 b_2 = AB\sin\delta,$$

where we have introduced the notation

$$|E_{0y}| = A, \quad |E_{0z}| = B, \quad \frac{E_{0z}}{B} = \frac{E_{0y}}{A}e^{i\delta}$$

for the absolute values of E_{0y} and E_{0z} and for the phase difference δ between them. Then

$$2b_{1,2} = \sqrt{A^2 + B^2 + 2AB\sin\delta} \pm \sqrt{A^2 + B^2 - 2AB\sin\delta}, \tag{2}$$

from which we get the magnitudes of the semiaxes of the polarization ellipse.

To determine their directions (relative to the arbitrary initial axes y and z) we start from the equality

$$\mathrm{Re}\,\{(\mathbf{E}_0 \cdot \mathbf{b}_1)(\mathbf{E}_0^* \cdot \mathbf{b}_2)\} = 0,$$

which is easily verified by substituting $\mathbf{E}_0 = (\mathbf{b}_1 + i\mathbf{b}_2)\, e^{-ia}$. Writing out this equality in the y, z coordinates, we get for the angle θ between the direction of $\mathbf{b}_1$ and the y axis:

$$\tan 2\theta = \frac{2AB\cos\delta}{A^2 - B^2}. \tag{3}$$

The direction of rotation of the field is determined by the sign of the x component of the vector $\mathbf{b}_1 \times \mathbf{b}_2$. Taking its expression from (1)

$$2i(\mathbf{b}_1 \times \mathbf{b}_2)_x = E_{0z}E_{0y}^* - E_{0z}^* E_{0y} = |E_{0y}|^2 \left\{ \left(\frac{E_{0z}}{E_{0y}}\right) - \left(\frac{E_{0z}}{E_{0y}}\right)^* \right\},$$

we see that the direction of $\mathbf{b}_1 \times \mathbf{b}_2$ (whether it is along or opposite to the positive direction of the x axis), and the sign of the rotation (whether in the same direction, or opposite to the direction of a right-hand screw along the x axis) are given by the sign of the imaginary part of the ratio E_{0z}/E_{0y} (plus for the first case and minus for the second). This is a generalization of the rule (48.11) for the case of circular polarization.

2. Determine the motion of a charge in the field of a plane monochromatic linearly polarized wave.

Solution: Choosing the direction of the field $\mathbf{E}$ of the wave as the y axis, we write:

$$E_y = E = E_0 \cos\omega\xi, \quad A_y = A = -\frac{cE_0}{\omega}\sin\omega\xi$$

($\xi = t - x/c$). From formulas (3) and (4) of problem 2, § 47, we find (in the reference system in which the particle is at rest on the average) the following representation of the motion in terms of the parameter $\eta = \omega\xi$):

$$x = -\frac{e^2E_0^2c}{8\gamma^2\omega^3}\sin 2\eta, \quad y = -\frac{eE_0c}{\gamma\omega^2}\cos\eta, \quad z = 0,$$

$$t = \frac{\eta}{\omega} - \frac{e^2E_0^2}{8\gamma^2\omega^3}\sin 2\eta, \quad \gamma^2 = m^2c^2 + \frac{e^2E_0^2}{2\omega^2};$$

$$p_x = -\frac{e^2E_0^2}{4\gamma\omega^2}\cos 2\eta, \quad p_y = \frac{eE_0}{\omega}\sin\eta, \quad p_z = 0.$$

The charge moves in the x, y plane in a symmetric figure-8 curve with its longitudinal axis along the y axis. During a period of the motion, η varies from 0 to 2π.

3. Determine the motion of a charge in the field of a circularly polarized wave.

Solution: For the field of the wave we have:

$$E_y = E_0\cos\omega\xi, \quad E_z = E_0\sin\omega\xi,$$

$$A_y = -\frac{cE_0}{\omega}\sin\omega\xi, \quad A_z = \frac{cE_0}{\omega}\cos\omega\xi.$$

The motion is given by the formulas:

$$x = 0, \quad y = -\frac{ecE_0}{\gamma\omega^2}\cos\omega t, \quad z = -\frac{ecE_0}{\gamma\omega^2}\sin\omega t,$$

$$p_z = 0, \quad p_y = \frac{eE_0}{\omega}\sin\omega t, \quad p_z = -\frac{eE_0}{\omega}\cos\omega t,$$

$$\gamma^2 = m^2c^2 + \frac{c^2E_0^2}{\omega^2}.$$

Thus the charge moves in the y, z plane along a circle of radius $ecE_0/\gamma\omega^2$ with a momentum having the constant magnitude $p = eE_0/\omega$; at each instant the direction of the momentum $\mathbf{p}$ is opposite to the direction of the magnetic field $\mathbf{H}$ of the wave.

§ 49. Spectral resolution

Every wave can be subjected to the process of spectral resolution, i.e. can be represented as a superposition of monochromatic waves with various frequencies. The character of this expansion varies according to the character of the time dependence of the field.

One category consists of those cases where the expansion contains frequencies forming a discrete sequence of values. The simplest case of this type arises in the resolution of a purely periodic (though not monochromatic) field. This is the usual expansion in Fourier series; it contains the frequencies which are integral multiples of the "fundamental" frequency $\omega_0 = 2\pi/T$, where T is the period of the field. We write it in the form

$$f = \sum_{n=-\infty}^{\infty} f_n e^{-in\omega_0 t} \tag{49.1}$$

(where f is any of the quantities describing the field). The quantities f_n are defined in terms of the function f by the integrals

$$f_n = \frac{1}{T}\int_{-T/2}^{T/2} f(t)e^{in\omega_0 t}dt. \tag{49.2}$$

Because $f(t)$ must be real,

$$f_{-n} = f_n^*. \tag{49.3}$$

In more complicated cases, the expansion may contain integral multiples (and sums of integral multiples) of several different incommensurable fundamental frequencies.

When the sum (49.1) is squared and averaged over the time, the products of terms with different frequencies give zero because they contain oscillating factors. Only terms of the form $f_nf_{-n} = |f_n|^2$ remain. Thus the average of the square of the field, i.e. the average intensity of the wave, is the sum of the intensities of its monochromatic components:

$$\overline{f^2} = \sum_{n=-\infty}^{\infty} |f_n|^2 = 2\sum_{n=1}^{\infty} |f_n|^2. \tag{49.4}$$

(where it is assumed that the average of the function f over a period is zero, i.e. $f_0 = \bar{f} = 0$).

Another category consists of fields which are expandable in a Fourier integral containing a continuous distribution of different frequencies. For this to be possible, the function $f(t)$ must satisfy certain definite conditions; usually we consider functions which vanish for $t \to \pm\infty$. Such an expansion has the form

$$f(t) = \int_{-\infty}^{\infty} f_\omega e^{-i\omega t}\frac{d\omega}{2\pi}, \tag{49.5}$$

where the Fourier components are given in terms of the function $f(t)$ by the integrals

$$f_\omega = \int_{-\infty}^{\infty} f(t)e^{i\omega t}dt. \tag{49.6}$$

Analogously to (49.3),

$$f_{-\omega} = f_\omega^*. \tag{49.7}$$

Let us express the total intensity of the wave, i.e. the integral of f^2 over all time, in terms of the intensity of the Fourier components. Using (49.5) and (49.6), we have:

$$\int_{-\infty}^{\infty} f^2 dt = \int_{-\infty}^{\infty} \left\{ f \int_{-\infty}^{\infty} f_\omega e^{-i\omega t} \frac{d\omega}{2\pi} \right\} dt = \int_{-\infty}^{\infty} \left\{ f_\omega \int_{-\infty}^{\infty} f e^{-i\omega t} dt \right\} \frac{d\omega}{2\pi} = \int_{-\infty}^{\infty} f_\omega f_{-\omega} \frac{d\omega}{2\pi},$$

or, using (49.7),

$$\int_{-\infty}^{\infty} f^2 dt = \int_{-\infty}^{\infty} |f_\omega|^2 \frac{d\omega}{2\pi} = 2\int_0^{\infty} |f_\omega|^2 \frac{d\omega}{2\pi}. \tag{49.8}$$

§ 50. Partially polarized light

Every monochromatic wave is, by definition, necessarily polarized. However we usually have to deal with waves which are only approximately monochromatic, and which contain frequencies in a small interval $\Delta\omega$. We consider such a wave, and let ω be some average frequency for it. Then its field (to be specific we shall consider the electric field **E**) at a fixed point in space can be writen in the form

$$\mathbf{E}_0(t)e^{-i\omega t},$$

where the complex amplitude $\mathbf{E}_0(t)$ is some slowly varying function of the time (for a strictly monochromatic wave $\mathbf{E}_0$ would be constant). Since $\mathbf{E}_0$ determines the polarization of the wave, this means that at each point of the wave, its polarization changes with time, such a wave is said to be *partially polarized.*

The polarization properties of electromagnetic waves, and of light in particular, are observed experimentally by passing the light to be investigated through various bodies† and then observing the intensity of the transmitted light. From the mathematical point of view this means that we draw conclusions concerning the polarization properties of the light from the values of certain quadratic functions of its field. Here of course we are considering the time averages of such functions.

Quadratic functions of the field are made up of terms proportional to the products $E_\alpha E_\beta$, $E_\alpha^* E_\beta^*$ or $E_\alpha E_\beta^*$. Products of the form

$$E_\alpha E_\beta = E_{0\alpha}E_{0\beta}e^{-2i\omega\tau}, \quad E_\alpha^* E_\beta^* = E_{0\alpha}^* E_{0\beta}^* e^{2i\omega t},$$

which contain the rapidly oscillating factors $e^{\pm 2i\omega t}$ give zero when the time average is taken. The products $E_\alpha E_\beta^* = E_{0\alpha}E_{0\beta}^*$ do not contain such factors, and so their averages are not

† For example, through a Nicol prism.

zero. Thus we see that the polarization properties of the light are completely characterized by the tensor

$$J_{\alpha\beta} = \overline{E_{0\alpha} E^*_{0\beta}}\,. \tag{50.1}$$

Since the vector $\mathbf{E}_0$ always lies in a plane perpendicular to the direction of the wave, the tensor $J_{\alpha\beta}$ has altogether four components (in this section the indices α, β are understood to take on only two values: $\alpha, \beta = 1, 2$, corresponding to the y and z axes; the x axis is along the direction of propagation of the wave).

The sum of the diagonal elements of the tensor $J_{\alpha\beta}$ (we denote it by J) is a real quantity—the average value of the square modulus of the vector $\mathbf{E}_0$ (or $\mathbf{E}$):

$$J \equiv J_{\alpha\alpha} = \overline{\mathbf{E}_0 \cdot \mathbf{E}^*_0}\,. \tag{50.2}$$

This quantity determines the intensity of the wave, as measured by the energy flux density. To eliminate this quantity which is not directly related to the polarization properties, we introduce in place of $J_{\alpha\beta}$ the tensor

$$\rho_{\alpha\beta} = \frac{J_{\alpha\beta}}{J}, \tag{50.3}$$

for which $\rho_{\alpha\alpha} = 1$; we call it the *polarization tensor.*

From the definition (50.1) we see that the components of the tensor $J_{\alpha\beta}$, and consequently also $\rho_{\alpha\beta}$, are related by

$$\rho_{\alpha\beta} = \rho^*_{\beta\alpha} \tag{50.4}$$

(i.e. the tensor is hermitian). Consequently the diagonal components ρ_{11} and ρ_{22} are real (with $\rho_{11} + \rho_{22} = 1$) while $\rho_{21} = \rho^*_{12}$. Thus the polarization is characterized by three real parameters.

Let us study the conditions that the tensor $\rho_{\alpha\beta}$ must satisfy for completely polarized light. In this case $\mathbf{E}_0 = \text{const}$, and so we have simply

$$J_{\alpha\beta} = J\rho_{\alpha\beta} = E_{0\alpha} E^*_{0\beta} \tag{50.5}$$

(without averaging), i.e. the components of the tensor can be written as products of components of some constant vector. The necessary and sufficient condition for this is that the determinant vanish:

$$|\rho_{\alpha\beta}| = \rho_{11}\rho_{22} - \rho_{12}\rho_{21} = 0. \tag{50.6}$$

The opposite case is that of unpolarized or *natural* light. Complete absence of polarization means that all directions (in the yz plane) are equivalent. In other words the polarization tensor must have the form:

$$\rho_{\alpha\beta} = \tfrac{1}{2}\delta_{\alpha\beta}\,. \tag{50.7}$$

The determinant is $|\rho_{\alpha\beta}| = \frac{1}{4}$.

In the general case of arbitrary polarization the determinant has values between 0 and $\frac{1}{4}$.†

† The fact that the determinant is positive for any tensor of the form (50.1) is easily seen by considering the averaging, for simplicity, as a summation over discrete values, and using the well-known algebraic inequality

$$\left|\sum_{a,b} x_a y_b\right|^2 \leq \sum_a |x_a|^2 \sum_b |y_b|^2.$$

By the *degree of polarization* we mean the positive quantity P, defined from

$$|\rho_{\alpha\beta}| = \tfrac{1}{4}(1 - P^2). \tag{50.8}$$

It runs from the value 0 for unpolarized to 1 for polarized light.

An arbitrary tensor $\rho_{\alpha\beta}$ can be split into two parts—a symmetric and an antisymmetric part. Of these, the first

$$S_{\alpha\beta} = \tfrac{1}{2}(\rho_{\alpha\beta} + \rho_{\beta\alpha})$$

is real because of the hermiticity of $\rho_{\alpha\beta}$. The antisymmetric part is pure imaginary. Like any antisymmetric tensor of rank equal to the number of dimensions, it reduces to a pseudoscalar (see the footnote on p. 18):

$$\tfrac{1}{2}(\rho_{\alpha\beta} - \rho_{\beta\alpha}) = -\frac{i}{2} e_{\alpha\beta} A,$$

where A is a real pseudoscalar, $e_{\alpha\beta}$ is the unit antisymmetric tensor (with components $e_{12} = -e_{21} = 1$). Thus the polarization tensor has the form:

$$\rho_{\alpha\beta} = S_{\alpha\beta} - \frac{i}{2} e_{\alpha\beta} A, \quad S_{\alpha\beta} = S_{\beta\alpha}, \tag{50.9}$$

i.e. it reduces to one real symmetric tensor and one pseudoscalar.

For a circularly polarized wave, the vector $\mathbf{E}_0 = \text{const}$, where

$$E_{02} = \pm i E_{01}.$$

It is easy to see that then $S_{\alpha\beta} = \frac{1}{2}\delta_{\alpha\beta}$, while $A = \pm 1$. On the other hand, for a linearly polarized wave the constant vector $\mathbf{E}_0$ can be chosen to be real, so that $A = 0$. In the general case the quantity A may be called the degree of circular polarization; it runs through values from +1 to −1, where the limiting values correspond to right- and left-circularly polarized waves, respectively.

The real symmetric tensor $S_{\alpha\beta}$, like any symmetric tensor, can be brought to principal axes, with different principal values which we denote by λ_1 and λ_2. The directions of the principal axes are mutually perpendicular. Denoting the unit vectors along these directions by $\mathbf{n}^{(1)}$ and $\mathbf{n}^{(2)}$, we can write $S_{\alpha\beta}$ in the form:

$$S_{\alpha\beta} = \lambda_1 n_\alpha^{(1)} n_\beta^{(1)} + \lambda_2 n_\alpha^{(2)} n_\beta^{(2)}, \quad \lambda_1 + \lambda_2 = 1. \tag{50.10}$$

The quantities λ_1 and λ_2 are positive and take on values from 0 to 1.

Suppose that $A = 0$, so that $\rho_{\alpha\beta} = S_{\alpha\beta}$. Each of the two terms in (50.10) has the form of a product of two components of a constant vector ($\sqrt{\lambda_1}\,\mathbf{n}^{(1)}$ or $\sqrt{\lambda_2}\,\mathbf{n}^{(2)}$). In other words, each of the terms corresponds to linearly polarized light. Furthermore, we see that there is no term in (50.10) containing products of components of the two waves. This means that the two parts can be regarded as physically independent of one another, or, as one says, they are *incoherent*. In fact, if two waves are independent, the average value of the product $E_\alpha^{(1)} E_\beta^{(2)}$ is equal to the product of the averages of each of the factors, and since each of them is zero,

$$\overline{E_\alpha^{(1)} E_\beta^{(2)}} = 0.$$

Thus we arrive at the result that in this case ($A = 0$) the partially polarized light can be represented as a superposition of two incoherent waves (with intensities proportional to λ_1 and λ_2), linearly polarized along mutually perpendicular directions.† (In the general case of a complex tensor $\rho_{\alpha\beta}$ one can show that the light can be represented as a superposition of two incoherent elliptically polarized waves, whose polarization ellipses are similar and mutually perpendicular (see problem 2).)

Let ϕ be the angle between the axis 1 (the y axis) and the unit vector $\mathbf{n}^{(1)}$; then

$$\mathbf{n}^{(1)} = (\cos\phi, \sin\phi), \quad \mathbf{n}^{(2)} = (-\sin\phi, \cos\phi).$$

Introducing the quantity $l = \lambda_1 - \lambda_2$ (assume $\lambda_1 > \lambda_2$), we write the components of the tensor (50.10) in the following form:

$$S_{\alpha\beta} = \frac{1}{2}\begin{pmatrix} 1 + l\cos 2\phi & l\sin 2\phi \\ l\sin 2\phi & 1 - l\cos 2\phi \end{pmatrix}. \tag{50.11}$$

Thus, for an arbitrary choice of the axes y and z, the polarization properties of the wave can be characterized by the following three real parameters: A—the degree of circular polarization, l—the degree of maximum linear polarization, and ϕ—the angle between the direction $\mathbf{n}^{(1)}$ of maximum polarization and the y axis.

In place of these parameters one can choose another set of three parameters:

$$\xi_1 = l\sin 2\phi, \quad \xi_2 = A, \quad \xi_3 = l\cos 2\phi \tag{50.12}$$

(the *Stokes parameters*). The polarization tensor is expressed in terms of them as

$$\rho_{\alpha\beta} = \frac{1}{2}\begin{pmatrix} 1 + \xi_3 & \xi_1 - i\xi_2 \\ \xi_1 + i\xi_2 & 1 - \xi_3 \end{pmatrix}. \tag{50.13}$$

All three parameters run through values from -1 to $+1$. The parameter ξ_3 characterizes the linear polarization along the y and z axes: the value $\xi_3 = 1$ corresponds to complete linear polarization along the y axis, and $\xi_3 = -1$ to complete polarization along the z axis. The parameter ξ_1 characterizes the linear polarization along directions making an angle of 45° with the y axis: the value $\xi_1 = 1$ means complete polarization at an angle $\phi = \pi/4$, while $\xi_1 = -1$ means complete polarization at $\phi = -\pi/4$.‡

The determinant of (50.13) is equal to

$$|\rho_{\alpha\beta}| = \tfrac{1}{4}(1 - \xi_1^2 - \xi_2^2 - \xi_3^2). \tag{50.14}$$

Comparing with (50.8), we see that

$$P = \sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2}\,. \tag{50.15}$$

† The determinant $|S_{\alpha\beta}| = \lambda_1\lambda_2$; suppose that $\lambda_1 > \lambda_2$; then the degree of polarization, as defined in (50.8), is $P = 1 - 2\lambda_2$. In the present case ($A = 0$) one frequently characterizes the degree of polarization by using the *depolarization coefficient*, defined as the ratio λ_2/λ_1.

‡ For a completely elliptically polarized wave with axes of the ellipse $\mathbf{b}_1$ and $\mathbf{b}_2$ (see § 48), the Stokes parameters are:

$$\xi_1 = 0, \quad \xi_2 = \pm 2b_1b_2, \quad \xi_3 = b_1^2 - b_2^2.$$

Here the y axis is along $\mathbf{b}_1$, while the two signs in ξ_2 correspond to directions of $\mathbf{b}_2$ along and opposite to the direction on the z axis.

Thus, for a given overall degree of polarization P, different types of polarization are possible, characterized by the values of the three quantities ξ_2, ξ_2, ξ_3, the sum of whose squares is fixed; they form a sort of vector of fixed length.

We note that the quantities $\xi_2 = A$ and $\sqrt{\xi_1^2 + \xi_3^2} = l$ are invariant under Lorentz transformations. This remark is already almost obvious from the very meaning of these quantities as degrees of circular and linear polarization.†

PROBLEMS

1. Resolve an arbitrary partially polarized light wave into its "natural" and "polarized" parts.

Solution: This resolution means the representation of the tensor $J_{\alpha\beta}$ in the form

$$J_{\alpha\beta} = \tfrac{1}{2} J^{(n)} \delta_{\alpha\beta} + E_{0\alpha}^{(p)} E_{0\beta}^{(p)*}.$$

The first term corresponds to the natural, and the second to the polarized parts of the light. To determine the intensities of the parts we note that the determinant

$$|J_{\alpha\beta} - \tfrac{1}{2} J^{(n)} \delta_{\alpha\beta}| = |E_{0\alpha}^{(p)} E_{0\beta}^{(p)*}| = 0.$$

Writing $J_{\alpha\beta} = J_{\rho\alpha\beta}$ in the form (50.13) and solving the equation, we get

$$J^{(n)} = J(1 - P).$$

The intensity of the polarized part is $J^{(p)} = |\mathbf{E}_0^{(p)}|^2 = J - J^{(n)} = JP$.

The polarized part of the light is in general an elliptically polarized wave, where the directions of the axes of the ellipse coincide with the principal axes of the tensor $S_{\alpha\beta}$. The lengths b_1 and b_2 of the axes of the ellipse and the angle ϕ formed by the axis $\mathbf{b}_1$ and the y axis are given by the equations:

$$b_1^2 + b_2^2 = JP, \quad 2b_1 b_2 = JP\xi_2, \quad \tan 2\phi = \frac{\xi_1}{\xi_3}.$$

2. Represent an arbitrary partially polarized wave as a superposition of two incoherent elliptically polarized waves.

Solution: For the hermitian tensor $\rho_{\alpha\beta}$ the "principal axes" are determined by two unit complex vectors $\mathbf{n}(\mathbf{n} \cdot \mathbf{n}^* = 1)$, satisfying the equations

$$\rho_{\alpha\beta} n_\beta = \lambda n_\alpha. \tag{1}$$

The principal values λ_1 and λ_2 are the roots of the equation

$$|\rho_{\alpha\beta} - \lambda \delta_{\alpha\beta}| = 0.$$

Multiplying (1) on both sides by n_α^*, we have:

$$\lambda = \rho_{\alpha\beta} n_\alpha^* n_\beta = \frac{1}{J} \overline{|E_{0\alpha} n_\alpha^*|^2},$$

† For a direct proof, we note that since the field of the wave is transverse in any reference frame, it is clear from the start that the tensor $\rho_{\alpha\beta}$ remains two-dimensional in any new frame. The transformation of $\rho_{\alpha\beta}$ into $\rho'_{\alpha\beta}$ leaves unchanged the sum of absolute squares $\rho_{\alpha\beta}\rho_{\alpha\beta}^*$ (in fact, the form of the transformation does not depend on the specific polarization properties of the light, while for a completely polarized wave this sum is 1 in any reference system). Because this transformation is real, the real and imaginary parts of the tensor $\rho_{\alpha\beta}$ (50.9) transform independently, so that the sums of the squares of the components of each separately remain constant, and are expressed in terms of l and A.

from which we see that λ_1, λ_2 are real and positive. Multiplying the equations

$$\rho_{\alpha\beta} n_\beta^{(1)} = \lambda_1 n_\alpha^{(1)}, \quad \rho_{\alpha\beta}^* n_\beta^{(2)*} = \lambda_2 n_\alpha^{(2)*}$$

for the first by $n_\alpha^{(2)*}$ and for the second by $n_\alpha^{(1)}$, taking the difference of the results and using the hermiticity of $\rho_{\alpha\beta}$, we get:

$$(\lambda_1 - \lambda_2) n_\alpha^{(1)} n_\alpha^{(2)*} = 0.$$

It then follows that $\mathbf{n}^{(1)} \cdot \mathbf{n}^{(2)*} = 0$, i.e. the unit vectors $\mathbf{n}^{(1)}$ and $\mathbf{n}^{(2)}$ are mutually orthogonal.

The expansion of the wave is provided by the formula

$$\rho_{\alpha\beta} = \lambda_1 n_\alpha^{(1)} n_\beta^{(1)*} + \lambda_2 n_\alpha^{(2)} n_\beta^{(2)*}.$$

One can always choose the complex amplitude so that, of the two mutually perpendicular components, one is real and the other imaginary (compare § 48). Setting

$$n_1^{(1)} = b_1, \quad n_2^{(1)} = ib_2$$

(where now b_1 and b_2 are understood to be normalized by the condition $b_1^2 + b_2^2 = 1$), we get from the equation $\mathbf{n}^{(1)} \cdot \mathbf{n}^{(2)*} = 0$:

$$n_1^{(2)} = ib_2, \quad n_2^{(2)} = b_1.$$

We then see that the ellipses of the two elliptically polarized vibrations are similar (have equal axis ratio), and one of them is turned through 90° relative to the other.

3. Find the law of transformation of the Stokes parameters for a rotation of they y, z axes through and angle ϕ.

Solution: The law is determined by the connection of the Stokes parameters to the components of the two-dimensional tensor in the yz plane, and is given by the formulas

$$\xi_1' = \xi_1 \cos 2\phi - \xi_3 \sin 2\phi, \qquad \xi_3' = \xi_1 \sin 2\phi + \xi_3 \cos 2\phi, \qquad \xi_2' = \xi_2.$$

§ 51. The Fourier resolution of the electrostatic field

The field produced by charges can also be formally expanded in plane waves (in a Fourier integral). This expansion, however, is essentially different from the expansion of electromagnetic waves in vacuum, for the field produced by charges does not satisfy the homogeneous wave equation, and therefore each term of this expansion does not satisfy the equation. From this it follows that for the plane waves into which the field of charges can be expanded, the relation $k^2 = \omega^2/c^2$, which holds for plane monochromatic electromagnetic waves, is not fulfilled.

In particular, if we formally represent the electrostatic field as a superposition of plane waves, then the "frequency" of these waves is clearly zero, since the field under consideration does not depend on the time. The wave vectors themselves are, of course, different from zero.

We consider the field produced by a point charge e, located at the origin of coordinates. The potential ϕ of this field is determined by the equation (see § 36)

$$\Delta\phi = -4\pi e\delta(\mathbf{r}). \tag{51.1}$$

We expand ϕ in a Fourier integral, i.e. we represent it in the form

$$\phi = \int_{-\infty}^{+\infty} e^{i\mathbf{k}\cdot\mathbf{r}}\phi_{\mathbf{k}}\frac{d^3k}{(2\pi)^3} \tag{51.2}$$

where d^3k denotes $dk_x dk_y dk_z$. In this formula $\phi_{\mathbf{k}} = \int \phi(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}dV$. Applying the Laplace operator to both sides of (51.2), we obtain

$$\Delta\phi = -\int_{-\infty}^{+\infty} k^2 e^{i\mathbf{k}\cdot\mathbf{r}}\phi_{\mathbf{k}}\frac{d^3k}{(2\pi)^3},$$

so that the Fourier component of the expression $\Delta\phi$ is

$$(\Delta\phi)_{\mathbf{k}} = -k^2\phi_{\mathbf{k}}.$$

On the other hand, we can find $(\Delta\phi)_{\mathbf{k}}$ by taking Fourier components of both sides of equation (51.1),

$$(\Delta\phi)_{\mathbf{k}} = -\int 4\pi e\delta(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}dV = -4\pi e.$$

Equating the two expressions obtained for $(\Delta\phi)_{\mathbf{k}}$, we find

$$\phi_{\mathbf{k}} = \frac{4\pi e}{k^2}. \tag{51.3}$$

This formula solves our problem.

Just as for the potential ϕ, we can expand the field

$$\mathbf{E} = \int_{-\infty}^{+\infty} \mathbf{E}_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}\frac{d^3k}{(2\pi)^3}. \tag{51.4}$$

With the aid of (51.2), we have

$$\mathbf{E} = -\operatorname{grad}\int_{-\infty}^{+\infty} \phi_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}\frac{d^3k}{(2\pi)^3} = -\int i\mathbf{k}\phi_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}}\frac{d^3k}{(2\pi)^3}.$$

Comparing with (51.4), we obtain

$$\mathbf{E}_{\mathbf{k}} = -i\mathbf{k}\phi_{\mathbf{k}} = -i\frac{4\pi e\mathbf{k}}{k^2}. \tag{51.5}$$

From this we see that the field of the waves, into which we have resolved the Coulomb field, is directed along the wave vector. Therefore these waves can be said to be *longitudinal.*

§ 52. Characteristic vibrations of the field

We consider an electromagnetic field (in the absence of charges) in some finite volume of space. To simplify further calculations we assume that this volume has the form of a rectangular parallelepiped with sides A, B, C, respectively. Then we can expand all quantities characterizing the field in this parallelepiped in a triple Fourier series (for the three coordinates). This expansion can be written (e.g. for the vector potential) in the form:

$$\mathbf{A} = \sum_{\mathbf{k}} \mathbf{A}_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{52.1}$$

explicitly indicating that $\mathbf{A}$ is real. The summation extends here over all possible values of the vector $\mathbf{k}$ whose components run through the values

$$k_x = \frac{2\pi n_x}{A}, \quad k_y = \frac{2\pi n_y}{B}, \quad k_z = \frac{2\pi n_z}{C}, \tag{52.2}$$

where n_x, n_y, n_z are positive or negative integers. Since $\mathbf{A}$ is real, the coefficients in the expansion (52.1) are related by the equations $\mathbf{A}_{-\mathbf{k}} = \mathbf{A}_{\mathbf{k}}^*$. From the equation div $\mathbf{A} = 0$ it follows that for each $\mathbf{k}$,

$$\mathbf{k} \cdot \mathbf{A}_{\mathbf{k}} = 0, \tag{52.3}$$

i.e., the complex vectors $\mathbf{A}_{\mathbf{k}}$ are "perpendicular" to the corresponding wave vectors $\mathbf{k}$. The vectors $\mathbf{A}_{\mathbf{k}}$ are, of course, functions of the time; from the wave equation (46.7), they satisfy the equation

$$\ddot{\mathbf{A}}_{\mathbf{k}} + c^2 k^2 \mathbf{A}_{\mathbf{k}} = 0. \tag{52.4}$$

If the dimensions A, B, C of the volume are sufficiently large, then neighbouring values of k_x, k_y, k_z (for which n_x, n_y, n_z differ by unity) are very close to one another. In this case we may speak of the number of possible values of k_x, k_y, k_z in the small intervals Δk_x, Δk_y, Δk_z.

Since to neighbouring values of, say, k_x, there correspond values of n_x differing by unity, the number Δn_x of possible values of k_x in the interval Δk_x is equal simply to the number of values of n_x in the corresponding interval. Thus, we obtain

$$\Delta n_x = \frac{A}{2\pi}\Delta k_x, \quad \Delta n_y = \frac{B}{2\pi}\Delta k_y, \quad \Delta n_z = \frac{C}{2\pi}\Delta k_z.$$

The total number Δn of possible values of the vector $\mathbf{k}$ with components in the intervals Δk_x, Δk_y, Δk_z is equal to the product $\Delta n_x\, \Delta n_y\, \Delta n_z$, that is,

$$\Delta n = \frac{V}{(2\pi)^3}\Delta k_x \Delta k_y \Delta k_z, \tag{52.5}$$

where $V = ABC$ is the volume of the field. It is easy to determine from this the number of possible values of the wave vector having absolute values in the interval Δk, and directed into the element of solid angle Δo. To get this we need only transform to polar coordinates in the "k space" and write in place of $\Delta k_x \Delta k_y \Delta k_z$ the element of volume in these coordinates. Thus

$$\Delta n = \frac{V}{(2\pi)^3} k^2 \Delta k \Delta o. \tag{52.6}$$

Replacing Δo by 4π, we find the number of possible values of $\mathbf{k}$ with absolute value in the interval Δk and pointing in all directions: $\Delta n = (V/2\pi^2)k^2\Delta k$.

We calculate the total energy

$$\mathscr{E} = \frac{1}{8\pi}\int (\mathbf{E}^2 + \mathbf{H}^2)\, dV$$

of the field, expressing it in terms of the quantities $\mathbf{A}_{\mathbf{k}}$. For the electric and magnetic fields we have

$$\mathbf{E} = -\frac{1}{c}\dot{\mathbf{A}} = -\frac{1}{c}\sum_{\mathbf{k}} \dot{\mathbf{A}}_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}},$$

$$\mathbf{H} = \operatorname{curl} \mathbf{A} = i \sum_{\mathbf{k}} (\mathbf{k} \times \mathbf{A}_{\mathbf{k}}) e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{52.7}$$

When calculating the squares of these sums, we must keep in mind that all products of terms with wave vectors $\mathbf{k}$ and $\mathbf{k}'$ such that $\mathbf{k} \neq \mathbf{k}'$ give zero on integration over the whole volume. In fact, such terms contain factors of the form $e^{i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{r}}$, and the integral, e.g. of

$$\int_0^A e^{i\frac{2\pi}{A}n_x x}\, dx,$$

with integer n_x different from zero, gives zero. In those terms with $\mathbf{k}' = -\mathbf{k}$, the exponentials drop out and integration over dV gives just the volume V.

As a result, we obtain

$$\mathscr{E} = \frac{V}{8\pi} \sum_{\mathbf{k}} \left\{ \frac{1}{c^2} \dot{\mathbf{A}}_{\mathbf{k}} \cdot \dot{\mathbf{A}}^*_{\mathbf{k}} + (\mathbf{k} \times \mathbf{A}_{\mathbf{k}}) \cdot (\mathbf{k} \times \mathbf{A}^*_{\mathbf{k}}) \right\}.$$

From (52.3), we have

$$(\mathbf{k} \times \mathbf{A}_{\mathbf{k}}) \cdot (\mathbf{k} \times \mathbf{A}^*_{\mathbf{k}}) = k^2 \mathbf{A}_{\mathbf{k}} \cdot \mathbf{A}^*_{\mathbf{k}},$$

so that

$$\mathscr{E} = \frac{V}{8\pi c^2} \sum_{\mathbf{k}} \{ \dot{\mathbf{A}}_{\mathbf{k}} \cdot \dot{\mathbf{A}}^*_{\mathbf{k}} + k^2 c^2 \mathbf{A}_{\mathbf{k}} \cdot \mathbf{A}^*_{\mathbf{k}} \}. \tag{52.8}$$

Each term of this sum corresponds to one of the terms of the expansion (52.1).

Because of (52.4), the vectors $\mathbf{A}_{\mathbf{k}}$ are harmonic functions of the time with frequencies $\omega_{\mathbf{k}} = ck$, depending only on the absolute value of the wave vector. Depending on the choice of these functions, the terms in the expansion (52.1) can represent standing or running plane waves. We shall write the expansion so that its terms describe running waves. To do this we write it in the form

$$\mathbf{A} = \sum_{\mathbf{k}} (\mathbf{a}_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} + \mathbf{a}^*_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}}) \tag{52.9}$$

which explicitly exhibits that $\mathbf{A}$ is real, and each of the vectors $\mathbf{a}_{\mathbf{k}}$ depends on the time according to the law

$$\mathbf{a}_{\mathbf{k}} \sim e^{-i\omega_k t}, \quad \omega_{\mathbf{k}} = ck. \tag{52.10}$$

Then each individual term in the sum (52.9) will be a function only of the difference $\mathbf{k} \cdot \mathbf{r} - \omega_{\mathbf{k}} t$, which corresponds to a wave propagating in the $\mathbf{k}$ direction.

Comparing the expansions (52.9) and (52.1), we find that their coefficients are related by the formulas

$$\mathbf{A}_{\mathbf{k}} = \mathbf{a}_{\mathbf{k}} + \mathbf{a}^*_{-\mathbf{k}},$$

and from (52.10) the time derivatives are related by

$$\dot{\mathbf{A}}_{\mathbf{k}} = -ick(\mathbf{a}_{\mathbf{k}} - \mathbf{a}^*_{-\mathbf{k}}).$$

Substituting in (52.8), we express the field energy in terms of the coefficients of the expansion (52.9). Terms with products of the form $\mathbf{a}_{\mathbf{k}} \cdot \mathbf{a}_{-\mathbf{k}}$ or $\mathbf{a}^*_{\mathbf{k}} \cdot \mathbf{a}^*_{-\mathbf{k}}$ cancel one another; also noting

that the sums $\Sigma \mathbf{a}_\mathbf{k} \cdot \mathbf{a}_\mathbf{k}^*$ and $\Sigma \mathbf{a}_{-\mathbf{k}} \mathbf{a}_{-\mathbf{k}}^*$ differ only in the labelling of the summation index, and therefore coincide, we finally obtain:

$$\mathscr{E} = \sum_\mathbf{k} \mathscr{E}_\mathbf{k}, \quad \mathscr{E}_\mathbf{k} = \frac{k^2 V}{2\pi} \mathbf{a}_\mathbf{k} \cdot \mathbf{a}_\mathbf{k}^*. \tag{52.11}$$

Thus the total energy of the field is expressed as a sum of the energies $\mathscr{E}_\mathbf{k}$, associated with each of the plane waves individually.

In a completely analogous fashion, we can calculate the total momentum of the field,

$$\frac{1}{c^2} \int \mathbf{S} dV = \frac{1}{4\pi c} \int \mathbf{E} \times \mathbf{H} dV,$$

for which we obtain

$$\sum_\mathbf{k} \frac{\mathbf{k}}{k} \frac{\mathscr{E}_\mathbf{k}}{c}. \tag{52.12}$$

This result could have been anticipated in view of the relation between the energy and momentum of a plane wave (see § 47).

The expansion (52.9) succeeds in expressing the field in terms of a series of discrete parameters (the vectors $\mathbf{a}_\mathbf{k}$), in place of the description in terms of a continuous series of parameters, which is essentially what is done when we give the potential $\mathbf{A}(x, y, z, t)$ at all points of space. We now make a transformation of the variables $\mathbf{a}_\mathbf{k}$, which has the result that the equations of the field take on a form similar to the canonical equations (Hamilton equations) of mechanics.

We introduce the real "canonical variables" $\mathbf{Q}_\mathbf{k}$ and $\mathbf{P}_\mathbf{k}$ according to the relations

$$\mathbf{Q}_\mathbf{k} = \sqrt{\frac{V}{4\pi c^2}} (\mathbf{a}_\mathbf{k} + \mathbf{a}_\mathbf{k}^*), \tag{52.13}$$

$$\mathbf{P}_\mathbf{k} = -i\omega_k \sqrt{\frac{V}{4\pi c^2}} (\mathbf{a}_\mathbf{k} - \mathbf{a}_\mathbf{k}^*) = \dot{\mathbf{Q}}_\mathbf{k}.$$

The Hamiltonian of the field is obtained by substituting these expressions in the energy (52.11):

$$\mathscr{H} = \sum_\mathbf{k} \mathscr{H}_\mathbf{k} = \sum_\mathbf{k} \tfrac{1}{2} (\mathbf{P}_\mathbf{k}^2 + \omega_k^2 \mathbf{Q}_\mathbf{k}^2). \tag{52.14}$$

Then the Hamilton equation $\partial \mathscr{H} / \partial \mathbf{P}_\mathbf{k} = \dot{\mathbf{Q}}_\mathbf{k}$ coincide with $\mathbf{P}_\mathbf{k} = \dot{\mathbf{Q}}_\mathbf{k}$, which is thus a consequence of the equations of motion. (This was achieved by an appropriate choice of the coefficient in (52.13).) The equations of motion, $\partial \mathscr{H} / \partial \mathbf{Q}_\mathbf{k} = -\dot{\mathbf{P}}_\mathbf{k}$, become the equations

$$\ddot{\mathbf{Q}}_\mathbf{k} + \omega_k^2 \mathbf{Q}_\mathbf{k} = 0, \tag{52.15}$$

that is, they are identical with the equations of the field.

Each of the vectors $\mathbf{Q}_\mathbf{k}$ and $\mathbf{P}_\mathbf{k}$ is perpendicular to the wave vector $\mathbf{k}$, i.e. has two independent components. The direction of these vectors determines the direction of polarization of the corresponding travelling wave. Denoting the two components of the vector $\mathbf{Q}_\mathbf{k}$ (in the plane perpendicular to $\mathbf{k}$) by $Q_{\mathbf{k}j}$, $j = 1, 2$, we have

$$\mathbf{Q}_\mathbf{k}^2 = \sum_j Q_{\mathbf{k}j}^2,$$

and similarly for $\mathbf{P_k}$. Then

$$\mathscr{H} = \sum_{\mathbf{k}j} \mathscr{H}_{\mathbf{k}j}, \quad \mathscr{H}_{\mathbf{k}j} = \tfrac{1}{2}(P_{\mathbf{k}j}^2 + \omega_k^2 Q_{\mathbf{k}j}^2). \tag{52.16}$$

We see that the Hamiltonian splits into a sum of independent terms $\mathscr{H}_{\mathbf{k}j}$, each of which contains only one pair of the quantities $Q_{\mathbf{k}j}$, $P_{\mathbf{k}j}$. Each such term corresponds to a travelling wave with a definite wave vector and polarization. The quantity $\mathscr{H}_{\mathbf{k}j}$ has the form of the Hamiltonian of a one-dimensional "oscillator", performing a simple harmonic vibration. For this reason, one sometimes refers to this result as the expansion of the field in terms of oscillators.

We give the formulas which express the field explicitly in terms of the variables $\mathbf{P_k}$, $\mathbf{Q_k}$. From (52.13), we have

$$\mathbf{a_k} = \frac{i}{k}\sqrt{\frac{\pi}{V}}(\mathbf{P_k} - i\omega_k \mathbf{Q_k}), \qquad \mathbf{a_k^*} = -\frac{i}{k}\sqrt{\frac{\pi}{V}}(\mathbf{P_k} + i\omega_k \mathbf{Q_k}). \tag{52.17}$$

Substituting these expressions in (52.1), we obtain for the vector potential of the field:

$$\mathbf{A} = 2\sqrt{\frac{\pi}{V}} \sum_{\mathbf{k}} \frac{1}{k}(ck\mathbf{Q_k} \cos \mathbf{k}\cdot\mathbf{r} - \mathbf{P_k} \sin \mathbf{k}\cdot\mathbf{r}). \tag{52.18}$$

For the electric and magnetic fields, we find

$$\mathbf{E} = -2\sqrt{\frac{\pi}{V}} \sum_{\mathbf{k}} (ck\mathbf{Q_k} \sin \mathbf{k}\cdot\mathbf{r} + \mathbf{P_k} \cos \mathbf{k}\cdot\mathbf{r}),$$

$$\mathbf{H} = -2\sqrt{\frac{\pi}{V}} \sum_{\mathbf{k}} \frac{1}{k}\{ck(\mathbf{k}\times\mathbf{Q_k}) \sin \mathbf{k}\cdot\mathbf{r} + (\mathbf{k}\times\mathbf{P_k}) \cos \mathbf{k}\cdot\mathbf{r}\}. \tag{52.19}$$

CHAPTER 7

THE PROPAGATION OF LIGHT

§ 53. Geometrical optics

A plane wave is characterized by the property that its direction of propagation and amplitude are the same everywhere. Arbitrary electromagnetic waves, of course, do not have this property. Nevertheless, a great many electromagnetic waves, which are not plane, have the property that within each small region of space they can be considered to be plane. For this, it is clearly necessary that the amplitude and direction of the wave remain practically constant over distances of the order of the wavelength. If this condition is satisfied, we can introduce the so-called *wave surface,* i.e. a surface at all of whose points the phase of the wave is the same (at a given time). (The wave surfaces of a plane wave are obviously planes perpendicular to the direction of propagation of the wave.) In each small region of space we can speak of a direction of propagation of the wave, normal to the wave surface. In this way we can introduce the concept of *rays*—curves whose tangents at each point coincide with the direction of propagation of the wave.

The study of the laws of propagation of waves in this case constitutes the domain of *geometrical optics.* Consequently, geometrical optics considers the propagation of waves, in particular of light, as the propagation of rays, completely divorced from their wave properties. In other words, geometrical optics corresponds to the limiting case of small wavelength, $\lambda \to 0$.

We now take up the derivation of the fundamental equation of geometrical optics—the equation determining the direction of the rays. Let f be any quantity describing the field of the wave (any component of $\mathbf{E}$ or $\mathbf{H}$). For a plane monochromatic wave, f has the form

$$f = ae^{i(\mathbf{k}\cdot\mathbf{r}-\omega t+\alpha)} = ae^{i(-k_i x^i+\alpha)} \tag{53.1}$$

(we omit the Re; it is understood that we take the real part of all expressions).

We write the expression for the field in the form

$$f = ae^{i\psi}. \tag{53.2}$$

In case the wave is not plane, but geometrical optics is applicable, the amplitude a is, generally speaking, a function of the coordinates and time, and the phase ψ, which is called the *eikonal,* does not have a simple form, as in (53.1). It is essential, however, that ψ be a large quantity. This is clear immediately from the fact that it changes by 2π when we move through one wavelength, and geometrical optics corresponds to the limit $\lambda \to 0$.

Over small space regions and time intervals the eikonal ψ can be expanded in series; to terms of first order, we have

$$\psi = \psi_0 + \mathbf{r} \cdot \frac{\partial \psi}{\partial \mathbf{r}} + t \frac{\partial \psi}{\partial t}$$

(the origin for coordinates and time has been chosen within the space region and time interval under consideration; the derivatives are evaluated at the origin). Comparing this expression with (53.1), we can write

$$\mathbf{k} = \frac{\partial \psi}{\partial \mathbf{r}} \equiv \operatorname{grad} \psi, \quad \omega = -\frac{\partial \psi}{\partial t}, \tag{53.3}$$

which corresponds to the fact that in each small region of space (and each small interval of time) the wave can be considered as plane. In four-dimensional form, the relation (53.3) is expressed as

$$k_i = -\frac{\partial \psi}{\partial x^i}, \tag{53.4}$$

where k_i is the wave four-vector.

We saw in § 48 that the components of the four-vector k^i are related by $k_i k^i = 0$. Substituting (53.4), we obtain the equation

$$\frac{\partial \psi}{\partial x_i} \frac{\partial \psi}{\partial x^i} = 0. \tag{53.5}$$

This equation, the *eikonal equation,* is the fundamental equation of geometrical optics.

The eikonal equation can also be derived by direct transition to the limit $\lambda \to 0$ in the wave equation. The field f satisfies the wave equation

$$\frac{\partial^2 f}{\partial x_i \partial x^i} = 0.$$

Substituting $f = ae^{i\psi}$, we obtain

$$\frac{\partial^2 a}{\partial x_i \partial x^i} e^{i\psi} + 2i \frac{\partial a}{\partial x_i} \frac{\partial \psi}{\partial x^i} e^{i\psi} + if \frac{\partial^2 \psi}{\partial x_i \partial x^i} - \frac{\partial \psi}{\partial x_i} \cdot \frac{\partial \psi}{\partial x^i} f = 0. \tag{53.6}$$

But the eikonal ψ, as we pointed out above, is a large quantity; therefore we can neglect the first three terms compared with the fourth, and we arrive once more at equation (53.5).

We shall give certain relations which, in their application to the propagation of light in vacuum, lead only to completely obvious results. Nevertheless, they are important because, in their general form, these derivations apply also to the propagation of light in material media.

From the form of the eikonal equation there results a remarkable analogy between geometrical optics and the mechanics of material particles. The motion of a material particle is determined by the Hamilton–Jacobi equation (16.11). This equation, like the eikonal equation, is an equation in the first partial derivatives and is of second degree. As we know, the action S is related to the momentum $\mathbf{p}$ and the Hamiltonian $\mathcal{H}$ of the particle by the relations

$$\mathbf{p} = \frac{\partial S}{\partial \mathbf{r}}, \quad \mathcal{H} = -\frac{\partial S}{\partial t}.$$

Comparing these formulas with the formulas (53.3), we see that the wave vector plays the same role in geometrical optics as the momentum of the particle in mechanics, while the frequency plays the role of the Hamiltonian, i.e., the energy of the particle. The absolute magnitude k of the wave vector is related to the frequency by the formula $k = \omega/c$. This relation is analogous to the relation $p = \mathscr{E}/c$ between the momentum and energy of a particle with zero mass and velocity equal to the velocity of light.

For a particle, we have the Hamilton equations

$$\dot{\mathbf{p}} = -\frac{\partial \mathscr{H}}{\partial \mathbf{r}}, \quad \mathbf{v} = \dot{\mathbf{r}} = \frac{\partial \mathscr{H}}{\partial \mathbf{p}}.$$

In view of the analogy we have pointed out, we can immediately write the corresponding equations for rays:

$$\dot{\mathbf{k}} = -\frac{\partial \omega}{\partial \mathbf{r}}, \quad \dot{\mathbf{r}} = \frac{\partial \omega}{\partial \mathbf{k}}. \tag{53.7}$$

In vacuum, $\omega = ck$, so that $\dot{\mathbf{k}} = 0$, $\mathbf{v} = c\mathbf{n}$ ($\mathbf{n}$ is a unit vector along the direction of propagation); in other words, as it must be, in vacuum the rays are straight lines, along which the light travels with velocity c.

The analogy between the wave vector of a wave and the momentum of a particle is made especially clear by the following consideration. Let us consider a wave which is a superposition of monochromatic waves with frequencies in a certain small interval and occupying some finite region in space (this is called a *wave packet*). We calculate the four-momentum of the field of this wave, using formula (32.6) with the energy-momentum tensor (48.15) (for each monochromatic component). Replacing k^i in this formula by some average value, we obtain an expression of the form

$$P^i = Ak^i, \tag{53.8}$$

where the coefficient of proportionality A between the two four-vectors P^i and k^i is some scalar. In three-dimensional form this relation gives:

$$\mathbf{P} = A\mathbf{k}, \quad \mathscr{E} = A\omega. \tag{53.9}$$

Thus we see that the momentum and energy of a wave packet transform, when we go from one reference system to another, like the wave vector and the frequency.

Pursuing the analogy, we can establish for geometrical optics a principle analogous to the principle of least action in mechanics. However, it cannot be written in Hamiltonian form as $\delta \int L\, dt = 0$, since it turns out to be impossible to introduce, for rays, a function analogous to the Lagrangian of a particle. Since the Lagrangian of a particle is related to the Hamiltonian $\mathscr{H}$ by the equation $L = \mathbf{p} \cdot \partial \mathscr{H}/\partial \mathbf{p} - \mathscr{H}$, replacing the Hamiltonian $\mathscr{H}$ by the frequency ω and the momentum by the wave vector $\mathbf{k}$, we should have to write for the Lagrangian in optics $\mathbf{k} \cdot \partial\omega/\partial\mathbf{k} - \omega$. But this expression is equal to zero, since $\omega = ck$. The impossibility of introducing a Lagrangian for rays is also clear directly from the consideration mentioned earlier that the propagation of rays is analogous to the motion of particles with zero mass.

If the wave has a definite constant frequency ω, then the time dependence of its field is given by a factor of the form $e^{-i\omega t}$. Therefore for the eikonal of such a wave we can write

$$\psi = -\omega t + \psi_0(x, y, z), \tag{53.10}$$

where ψ_0 is a function only of the coordinates. The eikonal equation (53.5) now takes the form

$$(\text{grad}\,\psi_0)^2 = \frac{\omega^2}{c^2}. \tag{53.11}$$

The wave surfaces are the surfaces of constant eikonal, i.e. the family of surfaces of the form $\psi_0(x, y, z) = \text{const}$. The rays themselves are at each point normal to the corresponding wave surface; their direction is determined by the gradient $\nabla\psi_0$.

As is well known, in the case where the energy is constant, the principle of least action for particles can also be written in the form of the so-called *principle of Maupertuis:*

$$\delta S = \delta \int \mathbf{p} \cdot d\mathbf{l} = 0,$$

where the integration extends over the trajectory of the particle between two of its points. In this expression the momentum is assumed to be a function of the energy and the coordinates. The analogous principle for rays is called *Fermat's principle.* In this case, we can write by analogy:

$$\delta\psi = \delta \int \mathbf{k} \cdot d\mathbf{l} = 0. \tag{53.12}$$

In vacuum, $\mathbf{k} = (\omega/c)\mathbf{n}$, and we obtain ($d\mathbf{l} \cdot \mathbf{n} = dl$):

$$\delta \int dl = 0, \tag{53.13}$$

which corresponds to rectilinear propagation of the rays.

§ 54. Intensity

In geometrical optics, the light wave can be considered as a bundle of rays. The rays themselves, however, determine only the direction of propagation of the light at each point; there remains the question of the distribution of the light intensity in space.

On some wave surface of the bundle of rays under consideration, we isolate an infinitesimal surface element. From differential geometry it is known that every surface has, at each of its points, two (generally different) principal radii of curvature. Let *ac* and *bd* (Fig. 7) be elements of the principal circles of curvature, constructed at a given element of the wave surface. Then the rays passing through *a* and *c* meet at the corresponding centre of curvature O_1, while the rays passing through *b* and *d* meet at the other centre of curvature O_2.

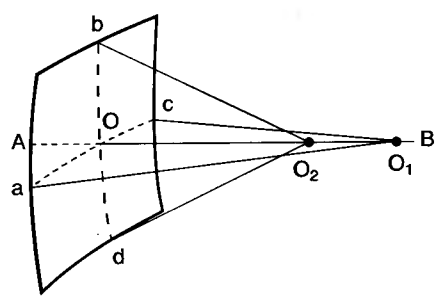


FIG. 7.

For fixed angular openings of the beams starting from O_1 and O_2, the lengths of the arcs ac and bd are, clearly, proportional to the corresponding radii of curvature R_1 and R_2 (i.e. to the lengths O_1O and O_2O). The area of the surface element is proportional to the product of the lengths ac and bd, i.e., proportional to R_1R_2. In other words, if we consider the element of the wave surface bounded by a definite set of rays, then as we move along them the area of the element will change proportionally to R_1R_2.

On the other hand, the intensity, i.e. the energy flux density, is inversely proportional to the surface area through which a given amount of light energy passes. Thus we arrive at the result that the intensity is

$$I = \frac{\text{const}}{R_1 R_2}. \tag{54.1}$$

This formula must be understood as follows. On each ray (AB in Fig. 7) there are definite points O_1 and O_2, which are the centres of curvature of all the wave surfaces intersecting the given ray. The distances OO_1 and OO_2 from the point O where the wave surface intersects the ray, to the points O_1 and O_2, are the radii of curvature R_1 and R_2 of the wave surface at the point O. Thus formula (54.1) determines the change in intensity of the light along a given ray as a function of the distances from definite points on this ray. We emphasize that this formula cannot be used to compare intensities at different points on a single wave surface.

Since the intensity is determined by the square modulus of the field, we can write for the change of the field itself along the ray

$$f = \frac{\text{const}}{\sqrt{R_1 R_2}} e^{ikR}, \tag{54.2}$$

where in the phase factor e^{ikR} we can write either e^{ikR_1} or e^{ikR_2}. The quantities e^{ikR_1} and e^{ikR_2} (for a given ray) differ from each other only by a constant factor, since the difference $R_1 - R_2$, the distance between the two centres of curvature, is a constant.

If the two radii of curvature of the wave surface coincide, then (54.1) and (54.2) have the form:

$$I = \frac{\text{const}}{R^2}, \quad f = \frac{\text{const}}{R} e^{ikR}. \tag{54.3}$$

This happens always when the light is emitted from a point source (the wave surfaces are then concentric spheres and R is the distance from the light source).

From (54.1) we see that the intensity becomes infinite at the points $R_1 = 0$, $R_2 = 0$, i.e. at the centres of curvature of the wave surface. Applying this to all the rays in a bundle, we find that the intensity of the light in the given bundle becomes infinite, generally, on two surfaces—the geometrical loci of all the centres of curvature of the wave surfaces. These surfaces are called *caustics*. In the special case of a beam of rays with spherical wave surfaces, the two caustics fuse into a single point (*focus*).

We note from well-known results of differential geometry concerning the properties of the loci of centres of curvature of a family of surfaces, that the rays are tangent to the caustic.

It is necessary to keep in mind that (for convex wave surfaces) the centres of curvature of the wave surfaces can turn out to lie not on the rays themselves, but on their extensions beyond the optical system from which they emerge. In such cases we speak of *imaginary caustics* (or foci). In this case the intensity of the light does not become infinite anywhere.

As for the increase of intensity to infinity, in actuality we must understand that the intensity does become large at points on the caustic, but it remains finite (see the problem

in § 59). The formal increase to infinity means that the approximation of geometrical optics is never applicable in the neighbourhood of the caustic. To this is related the fact that the change in phase along the ray can be determined from formula (54.2) only over sections of the ray which do not include its point of tangency to the caustic. Later (in § 59), we shall show that actually in passing through the caustic the phase of the field decreases by $\pi/2$. This means that if, on the section of the ray before its first intersection with the caustic, the field is proportional to the factor e^{ikx} (x is the coordinate along the ray), then after passage through the caustic the field will be proportional to $e^{i(kx-(\pi/2))}$. The same thing occurs in the neighbourhood of the point of tangency to the second caustic, and beyond that point the field is proportional to $e^{i(kx-\pi)}$.†

§ 55. The angular eikonal

A light ray travelling in vacuum and impinging on a transparent body will, on its emergence from this body, generally have a direction different from its initial direction. This change in direction will, of course, depend on the specific properties of the body and on its form. However, it turns out that one can derive general laws relating to the change in direction of a light ray on passage through an arbitrary material body. In this it is assumed only that geometrical optics is applicable to rays propagating in the interior of the body under consideration. As is customary, we shall call such transparent bodies, through which rays of light propagate, *optical systems.*

Because of the analogy mentioned in § 53, between the propagation of rays and the motion of particles, the same general laws are valid for the change in direction of motion of a particle, initially moving in a straight line in vacuum, then passing through some electromagnetic field, and once more emerging into vacuum. For definiteness, we shall, however, always speak later of the propagation of light rays.

We saw in a previous section that the eikonal equation, describing the propagation of the rays, can be written in the form (53.11) (for light of a definite frequency). From now on we shall, for convenience, designate by ψ the eikonal ψ_0 divided by the constant ω/c. Then the basic equation of geometrical optics has the form:

$$(\nabla\psi)^2 = 1. \tag{55.1}$$

Each solution of this equation describes a definite beam of rays, in which the direction of the rays passing through a given point in space is determined by the gradient of ψ at that point. However, for our purposes this description is insufficient, since we are seeking general relations determining the passage through an optical system not of a single definite bundle of rays, but of arbitrary rays. Therefore we must use an eikonal expressed in such a form that it describes all the generally possible rays of light, i.e. rays passing through any pair of points in space. In its usual form the eikonal $\psi(\mathbf{r})$ is the phase of the rays in a certain bundle passing through the point $\mathbf{r}$. Now we must introduce the eikonal as a function $\psi(\mathbf{r}, \mathbf{r}')$ of the coordinates of two points ($\mathbf{r}$, $\mathbf{r}'$ are the radius vectors of the initial and end points of the ray). A ray can pass through each pair of points $\mathbf{r}$, $\mathbf{r}'$, and $\psi(\mathbf{r}, \mathbf{r}')$ is the phase difference (or, as it is called, the *optical path length*) of this ray between the points $\mathbf{r}$ and $\mathbf{r}'$. From now on we shall always understand by $\mathbf{r}$ and $\mathbf{r}'$ the radius vectors to points on the ray before and after its passage through the optical system.

† Although formula (54.2) itself is not valid near the caustic, the change in phase of the field corresponds formally to a change in sign (i.e. multiplication by $e^{i\pi}$) of R_1 or R_2 in this formula.

If in $\psi(\mathbf{r}, \mathbf{r}')$ one of the radius vectors, say $\mathbf{r}'$, is fixed, then ψ as a function of $\mathbf{r}$ describes a definite bundle of rays, namely, the bundle of rays passing through the point $\mathbf{r}'$. Then ψ must satisfy equation (55.1), where the differentiations are applied to the components of $\mathbf{r}$. Similarly, if $\mathbf{r}$ is assumed fixed, we again obtain an equation for $\psi(\mathbf{r}, \mathbf{r}')$, so that

$$(\nabla_{\mathbf{r}}\psi)^2 = 1, \quad (\nabla_{\mathbf{r}'}\psi)^2 = 1. \tag{55.2}$$

The direction of the ray is determined by the gradient of its phase. Since $\psi(\mathbf{r}, \mathbf{r}')$ is the difference in phase at the points $\mathbf{r}$ and $\mathbf{r}'$, the direction of the ray at the point $\mathbf{r}'$ is given by the vector $\mathbf{n}' = \partial\psi/\partial\mathbf{r}'$, and at the point $\mathbf{r}$ by the vector $\mathbf{n} = -\partial\psi/\partial\mathbf{r}$. From (55.2) it is clear that $\mathbf{n}$ and $\mathbf{n}'$ are unit vectors:

$$\mathbf{n}^2 = \mathbf{n}'^2 = 1. \tag{55.3}$$

The four vectors $\mathbf{r}$, $\mathbf{r}'$, $\mathbf{n}$, $\mathbf{n}'$ are interrelated, since two of them ($\mathbf{n}$, $\mathbf{n}'$) are derivatives of a certain function ψ with respect to the other two ($\mathbf{r}$, $\mathbf{r}'$). The function ψ itself satisfies the auxiliary conditions (55.2).

To obtain the relation between $\mathbf{n}$, $\mathbf{n}'$, $\mathbf{r}$, $\mathbf{r}'$, it is convenient to introduce, in place of ψ, another quantity, on which no auxiliary condition is imposed (i.e., is not required to satisfy any differential equations). This can be done as follows. In the function ψ the independent variables are $\mathbf{r}$ and $\mathbf{r}'$, so that for the differential $d\psi$ we have

$$d\psi = \frac{\partial\psi}{\partial\mathbf{r}} \cdot d\mathbf{r} + \frac{\partial\psi}{\partial\mathbf{r}'} \cdot d\mathbf{r}' = -\mathbf{n} \cdot d\mathbf{r} + \mathbf{n}' \cdot d\mathbf{r}'.$$

We now make a Legendre transformation from $\mathbf{r}$, $\mathbf{r}'$ to the new independent variables $\mathbf{n}$, $\mathbf{n}'$, that is, we write

$$d\psi = -d(\mathbf{n} \cdot \mathbf{r}) + \mathbf{r} \cdot d\mathbf{n} + d(\mathbf{n}' \cdot \mathbf{r}'(- \mathbf{r}' \cdot d\mathbf{n}'.$$

from which, introducing the function

$$\chi = \mathbf{n}' \cdot \mathbf{r}' - \mathbf{n} \cdot \mathbf{r} - \psi, \tag{55.4}$$

we have

$$d\chi = -\mathbf{r} \cdot d\mathbf{n} + \mathbf{r}' \cdot d\mathbf{n}'. \tag{55.5}$$

The function χ is called the *angular eikonal*; as we see from (55.5), the independent variables in it are $\mathbf{n}$ and $\mathbf{n}'$. No auxiliary conditions are imposed on χ. In fact, equation (55.3) now states only a condition referring to the independent variables: of the three components n_x, n_y, n_z, of the vector $\mathbf{n}$ (and similarly for $\mathbf{n}'$), only two are independent. As independent variables we shall use n_y, n_z, n'_y, n'_z; then

$$n_x = \sqrt{1 - n_y^2 - n_z^2}, \quad n'_x = \sqrt{1 - n_y'^2 - n_z'^2}.$$

Substituting these expressions in

$$d\chi = -x\,dn_x - y\,dn_y - z\,dn_z + x'dn'_x + y'\,dn'_y + z'\,dn'_z,$$

we obtain for the differential $d\chi$:

$$d\chi = -\left(y - \frac{n_y}{n_x}x\right)dn_y - \left(z - \frac{n_z}{n_x}x\right)dn_z + \left(y' - \frac{n'_y}{n'_x}x'\right)dn'_y + \left(z' - \frac{n'_z}{n'_x}x'\right)dn'_z.$$

From this we obtain, finally, the following equations:

$$y - \frac{n_y}{n_x}x = -\frac{\partial\chi}{\partial n_y}, \quad z - \frac{n_z}{n_x}x = -\frac{\partial\chi}{\partial n_z},$$
$$y' - \frac{n'_y}{n'_x}x' = \frac{\partial\chi}{\partial n'_y}, \quad z' - \frac{n'_z}{n'_x}x' = \frac{\partial\chi}{\partial n'_z}, \tag{55.6}$$

which is the relation sought between $\mathbf{n}$, $\mathbf{n}'$, $\mathbf{r}$, $\mathbf{r}'$. The function χ characterizes the special properties of the body through which the rays pass (or the properties of the field, in the case of the motion of a charged particle).

For fixed values of $\mathbf{n}$, $\mathbf{n}'$, each of the two pairs of equations (55.6) represent a straight line. These lines are precisely the rays before and after passage through the optical system. Thus the equation (55.6) directly determines the path of the ray on the two sides of the optical system.

§ 56. Narrow bundles of rays

In studying the passage of beams of rays through optical systems, special interest attaches to bundles whose rays all pass through one point (such bundles are said to be *homocentric*).

After passage through an optical system, homocentric bundles in general cease to be homocentric, i.e. after passing through a body the rays no longer come together in any one point. Only in exceptional cases will the rays starting from a luminous point come together after passage through an optical system and all meet at one point (the image of the luminous point).†

One can show (see § 57) that the only case for which all homocentric bundles remain strictly homocentric after passage through the optical system is the case of identical imaging, i.e. the case where the image differs from the object only in its position or orientation, or is mirror inverted.

Thus no optical system can give a completely sharp image of an object (having finite dimensions) except in the trivial case of identical imaging.‡ Only approximate, but not completely sharp images can be produced of an extended body, in any case other than for identical imaging.

The most important case where there is approximate transition of homocentric bundles into homocentric bundles is that of sufficiently narrow beams (i.e. beams with a small opening angle) passing close to a particular line (for a given optical system). This line is called the *optic axis* of the system.

Nevertheless, we must note that even infinitely narrow bundles of rays (in the three-dimensional case) are in general not homocentric; we have seen (Fig. 7) that even in such a bundle different rays intersect at different points (this phenomenon is called *astigmatism*). Exceptions are those points of the wave surface at which the two principal radii of curvature are equal—a small region of the surface in the neighbourhood of such points can be considered as spherical, and the corresponding narrow bundle of rays is homocentric.

† The point of intersection can lie either on the rays themselves or on their continuations; depending on this, the image is said to be *real* or *virtual*.

‡ Such imaging can be produced with a plane mirror.

We consider an optical system having axial symmetry.† The axis of symmetry of the system is also its optical axis. The wave surface of a bundle of rays travelling along this axis also has axial symmetry; as we know, surfaces of rotation have equal radii of curvature at their points of intersection with the symmetry axis. Therefore a narrow bundle moving in this direction remains homocentric.

To obtain general quantitative relations, determining image formation with the aid of narrow bundles, passing through an axially-symmetric optical system, we use the general equations (55.6) after determining first of all the form of the function χ in the case under consideration.

Since the bundles of rays are narrow and move in the neighbourhood of the optical axis, the vectors $\mathbf{n}$, $\mathbf{n}'$ for each bundle are directed almost along this axis. If we choose the optical axis as the X axis, then the components, n_y, n_z, n'_y, n'_z will be small compared with unity. As for the components n_x, n'_x; $n_x \approx 1$ and n'_x can be approximately equal to either $+1$ or -1. In the first case the rays continue to travel almost in their original direction, emerging into the space on the other side of the optical system, which in this case is called a *lens*. In the second the rays change their direction to almost the reverse; such an optical system is called a *mirror*.

Making use of the smallness of n_y, n_z, n'_y, n'_z, we expand the angular eikonal $\chi(n_y, n_z, n'_y, n'_z)$ in series and stop at the first terms. Because of the axial symmetry of the whole system, χ must be invariant with respect to rotations of the coordinate system around the optical axis. From this it is clear that in the expansion of χ there can be no terms of first order, proportional to the first powers of the y- and z-components of the vectors $\mathbf{n}$ and $\mathbf{n}'$; such terms would not have the required invariance. The terms of second order which have the required property are the squares $\mathbf{n}^2$ and $\mathbf{n}'^2$ and the scalar product $\mathbf{n} \cdot \mathbf{n}'$. Thus, to terms of second order, the angular eikonal of an axially-symmetric optical system has the form

$$\chi = \text{const} + \frac{g}{2}(n_y^2 + n_z^2) + f(n_y n'_y + n_z n'_z) + \frac{h}{2}(n'^2_y + n'^2_z), \tag{56.1}$$

where f, g, h are constants.

For definiteness, we now consider a lens, so that we set $n'_x \approx 1$; for a mirror, as we shall show later, all the formulas have a similar appearance. Now substituting the expression (56.1) in the general equations (55.6), we obtain:

$$\begin{aligned} n_y(x-g) - fn'_y &= y, \quad fn_y + n'_y(x'+h) = y', \\ n_z(x-g) - fn'_z &= z, \quad fn_z + n'_z(x'+h) = z'. \end{aligned} \tag{56.2}$$

We consider a homocentric bundle emanating from the point x, y, z; let the point x', y', z' be the point in which all the rays of the bundle intersect after passing through the lens. If the first and second pairs of equations (56.2) were independent, then these four equations, for given x, y, z, x', y', z', would determine one definite set of values n_y, n_z, n'_y, n'_z, that is, there would be just *one* ray starting from the point x, y, z, which would pass through the point x', y', z'. In order that all rays starting from x, y, z shall pass through x', y', z', it is consequently necessary that the equations (56.2) not be independent, that is, one pair of these equations must be a consequence of the other. The necessary condition for this dependence is that the

† It can be shown that the problem of image formation with the aid of narrow bundles, moving in the neighbourhood of the optical axis in a nonaxially-symmetric system, can be reduced to image formation in an axially-symmetric system plus a subsequent rotation of the image thus obtained, relative to the object.

coefficients in the one pair of equations be proportional to the coefficients of the other pair. Thus we must have

$$\frac{x-g}{f} = -\frac{f}{x'+h} = \frac{y}{y'} = \frac{z}{z'}. \tag{56.3}$$

In particular,

$$(x-g)(x'+h) = -f^2. \tag{56.4}$$

The equations we have obtained give the required connection between the coordinates of the image and object for image formation using narrow bundles.

The points $x = g$ and $x' = -h$ on the optical axis are called the *principal foci* of the optical system. Let us consider bundles of rays parallel to the optical axis. The source point of such rays is, clearly, located at infinity on the optical axis, that is, $x = \infty$. From (56.3) we see that in this case, $x' = -h$. Thus a parallel bundle of rays, after passage through the optical system, intersects at the principal focus. Conversely, a bundle of rays emerging from the principal focus becomes parallel after passage through the system.

In the equation (56.3) the coordinates x and x' are measured from the same origin of coordinates, lying on the optical axis. It is, however, more convenient to measure the coordinates of object and image from different origins, choosing them at the corresponding principal foci. As positive direction of the coordinates we choose the direction from the corresponding focus toward the side to which the light travels. Designating the new coordinates of object and image by capital letters, we have

$$X = x - g, \quad X' = x' + h, \quad Y = y, \quad Y' = y', \quad Z = z, \quad Z' = z'.$$

The equations of image formation (56.3) and (56.4) in the new coordinates take the form

$$XX' = -f^2, \tag{56.5}$$

$$\frac{Y'}{Y} = \frac{Z'}{Z} = \frac{f}{X} = -\frac{X'}{f}. \tag{56.6}$$

The quantity f is called the *principal focal length* of the system.

The ratio Y'/Y is called the *lateral magnification.* As for the *longitudinal magnification,* since the coordinates are not simply proportional to each other, it must be written in differential form, comparing the length of an element of the object (along the direction of the axis) with the length of the corresponding element in the image. From (56.5) we get for the "longitudinal magnification"

$$\left|\frac{dX'}{dX}\right| = \frac{f^2}{X^2} = \left(\frac{Y'}{Y}\right)^2. \tag{56.7}$$

We see from this that even for an infinitely small object, it is impossible to obtain a geometrically similar image. The longitudinal magnification is never equal to the transverse (except in the trivial case of identical imaging).

A bundle passing through the point $X = f$ on the optical axis intersects once more at the point $X' = -f$ on the axis; these two points are called *principal points.* From equation (56.2) ($n_y X - f n'_y = Y, n_z X - f n'_z = Z$) it is clear that in this case ($X = f$, $Y = Z = 0$), we have the equations $n_y = n'_y, n_z = n'_z$. Thus every ray starting from a principal point crosses the optical axis again at the other principal point in a direction parallel to its original direction.

If the coordinates of object and image are measured from the principal points (and not from the principal foci), then for these coordinates ξ and ξ', we have

$$\xi' = X' + f, \quad \xi = X - f.$$

Substituting in (56.5) it is easy to obtain the equations of image formation in the form

$$\frac{1}{\xi} - \frac{1}{\xi'} = -\frac{1}{f}. \tag{56.8}$$

One can show that for an optical system with small thickness (for example, a mirror or a thin lens), the two principal points almost coincide. In this case the equation (56.8) is particularly convenient, since in it ξ and ξ' are then measured practically from one and the same point.

If the focal distance is positive, then objects located in front of the focus ($X > 0$) are imaged erect ($Y'/Y > 0$); such optical systems are said to be *converging*. If $f < 0$, then for $X > 0$ we have $Y'/Y < 0$, that is, the object is imaged in inverted form; such systems are said to be *diverging*.

There is one limiting case of image formation which is not contained in the formulas (56.8); this is the case where all three coefficients f, g, h are infinite (i.e. the optical system has an infinite focal distance and its principal foci are located at infinity). Going to the limit of infinite f, g, h in (56.4) we obtain

$$x' = \frac{h}{g}x + \frac{f^2 - gh}{g}.$$

Since we are interested only in the case where the object and its image are located at finite distances from the optical system, f, g, h must approach infinity in such fashion that the ratios h/g, $(f^2 - gh)/g$ are finite. Denoting them, respectively, by α^2 and β, we have

$$x' = \alpha^2 x + \beta.$$

For the other two coordinates we now have from the general equation (56.7):

$$\frac{y'}{y} = \frac{z'}{z} = \pm\alpha.$$

Finally, again measuring the coordinates x and x' from different origins, namely from some arbitrary point on the axis and from the image of this point, respectively, we finally obtain the equations of image formation in the simple form

$$X' = \alpha^2 X, \quad Y' = \pm\alpha Y, \quad Z' = \pm\alpha Z. \tag{56.9}$$

Thus the longitudinal and transverse magnifications are constants (but not equal to each other). This case of image formation is called *telescopic*.

All the equations (56.5) through (56.9), derived by us for lenses, apply equally to mirrors, and even to an optical system without axial symmetry, if only the image formation occurs by means of narrow bundles of rays travelling near the optical axis. In this, the reference points for the x coordinates of object and image must always be chosen along the optical axis from corresponding points (principal foci or principal points) in the direction of propagation of the ray. In doing this, we must keep in mind that for an optical system not possessing axial symmetry, the directions of the optical axis in front of and beyond the system do not lie in the same plane.

PROBLEMS

1. Find the focal distance for image formation with the aid of two axially-symmetric optical systems whose optical axes, coincide.

Solution: Let f_1 and f_2 be the focal lengths of the two systems. For each system separately, we have

$$X_1X_1' = -f_1^2, \quad X_2X_2' = -f_2^2.$$

Since the image produced by the first system acts as the object for the second, then denoting by l the distance between the rear principal focus of the first system and the front focus of the second, we have $X_2 = X_1' - l$; expressing X_2' in terms of X_1, we obtain

$$X_2' = \frac{X_1 f_2^2}{f_1^2 + lX_1}$$

or

$$\left(X_1 + \frac{f_1^2}{l}\right)\left(X_2' - \frac{f_2^2}{l}\right) = -\left(\frac{f_1 f_2}{l}\right)^2,$$

from which it is clear that the principal foci of the composite system are located at the points $X_1 = -f_1^2/l$, $X_2' = f_2^2/l$ and the focal length is

$$f = -\frac{f_1 f_2}{l}$$

(to choose the sign of this expression, we must write the corresponding equation for the transverse magnification).

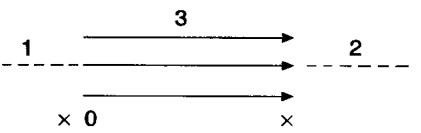


FIG. 8.

In case $l = 0$, the focal length $f = \infty$, that is, the composite system gives telescopic image formation. In this case we have $X_2' = X_1(f_2/f_1)^2$, that is, the parameter α in the general formula (56.9) is $\alpha = f_2/f_1$.

2. Find the focal length for charged particles of a "magnetic lens" in the form of a longitudinal homogeneous field in the section of length l (Fig. 8).†

Solution: The kinetic energy of the particle is conserved during its motion in a magnetic field; therefore the Hamilton–Jacobi equation for the reduced action $S_0(\mathbf{r})$ (where the total action is $S = -\mathscr{E}t + S_0$) is

$$\left(\nabla S_0 - \frac{e}{c}\mathbf{A}\right)^2 = p^2,$$

where

$$p^2 = \frac{\mathscr{E}^2}{c^2} - m^2c^2 = \text{const.}$$

Using formula (19.4) for the vector potential of the homogeneous magnetic field, choosing the x axis along the field direction and considering this axis as the optical axis of an axially-symmetric optical system, we get the Hamilton–Jacobi equation in the form:

† This might be the field inside a long solenoid, when we neglect the disturbance of the homogeneity of the field near the ends of the solenoid.

$$\left(\frac{\partial S_0}{\partial x}\right)^2 + \left(\frac{\partial S_0}{\partial r}\right)^2 + \frac{e^2}{4c^2} H^2 r^2 = p^2, \tag{1}$$

where r is the distance from the x axis, and S_0 is a function of x and r.

For narrow beams of particles propagating close to the optical axis, the coordinate r is small, so that accordingly we try to find S_0 as a power series in r. The first two terms of this series are

$$S_0 = px + \tfrac{1}{2}\sigma(x) r^2, \tag{2}$$

where $\sigma(x)$ satisfies the equation

$$p\sigma'(x) + \sigma^2 - \frac{e^2}{4c^2} H^2 = 0. \tag{3}$$

In region 1 in front of the lens, we have:

$$\sigma^{(1)} = \frac{p}{x - x_1}$$

where $x_1 < 0$ is a constant. This solution corresponds to a free beam of particles, emerging along straight line rays from the point $x = x_1$ on the optical axis in region 1. In fact, the action function for the free motion of a particle with a momentum p in a direction out from the point $x = x_1$ is

$$S_0 = p\sqrt{r^2 + (x - x_1)^2} \cong p(x - x_1) + \frac{pr^2}{2(x - x_1)}.$$

Similarly, in region 2 behind the lens we write:

$$\sigma^{(2)} = \frac{p}{x - x_2},$$

where the constant x_2 is the coordinate of the image of the point x_1.

In region 3 inside the lens, the solution of equation (3) is obtained by separation of variables, and gives:

$$\sigma^{(3)} = \frac{eH}{2c} \cot\left(\frac{eH}{2cp} x + C\right),$$

where C is an arbitrary constant.

The constant C and x_2 (for given x_1) are determined by the requirements of continuity of $\sigma(x)$ for $x = 0$ and $x = l$:

$$-\frac{p}{x_1} = \frac{eH}{2c} \cot C, \quad \frac{p}{l - x_2} = \frac{eH}{2c} \cot\left(\frac{eH}{2cp} l + C\right).$$

Eliminating the constant C from these equations, we find:

$$(x_1 - g)(x_2 + h) = -f^2,$$

where†

$$g = -\frac{2cp}{eH} \cot \frac{eHl}{2cp}, \quad h = g + l,$$

$$f = \frac{2cp}{eH \sin \dfrac{eHl}{2cp}}.$$

† The value of f is given with the correct sign. However, to show this requires additional investigation.

§ 57. Image formation with broad bundles of rays

The formation of images with the aid of narrow bundles of rays, which was considered in the previous section, is approximate; it is the more exact (i.e. the sharper) the narrower the bundles. We now go over to the question of image formation with bundles of rays of arbitrary breadth.

In contrast to the formation of an image of an object by narrow beams, which can be achieved for any optical system having axial symmetry, image formation with broad beams is possible only for specially constituted optical systems. Even with this limitation, as already pointed out in § 56, image formation is not possible for all points in space.

The later derivations are based on the following essential remark. Suppose that all rays, starting from a certain point O and travelling through the optical system, intersect again at some other point O'. It is easy to see that the optical path length ψ is the same for all these rays. In the neighbourhood of each of the points O, O', the wave surfaces for the rays intersecting in them are spheres with centres at O and O', respectively, and, in the limit as we approach O and O', degenerate to these points. But the wave surfaces are the surfaces of constant phase, and therefore the change in phase along different rays, between their points of intersection with two given wave surfaces, is the same. From what has been said, it follows that the total change in phase between the points O and O' is the same (for the different rays).

Let us consider the conditions which must be fulfilled in order to have formation of an image of a small line segment using broad beams; the image is then also a small line segment. We choose the directions of these segments as the directions of the ξ and ξ' axes, with origins at any two corresponding points O and O' of the object and image. Let ψ be the optical path length for the rays starting from O and reaching O'. For the rays starting from a point infinitely near to O with coordinate $d\xi$, and arriving at a point of the image with coordinate $d\xi'$, the optical path length is $\psi + d\psi$, where

$$d\psi = \frac{\partial \psi}{\partial \xi}\, d\xi + \frac{\partial \psi}{\partial \xi'} d\xi'.$$

We introduce the "magnification"

$$\alpha_\xi = \frac{d\xi'}{d\xi}$$

as the ratio of the length $d\xi'$ of the element of the image to the length $d\xi$ of the imaged element. Because of the smallness of the line segment which is being imaged, the quantity α can be considered constant along the line segment. Writing, as usual, $\partial\psi/\partial\xi = -n_\xi$, $\partial\psi/\partial\xi' = n'_\xi$ (n_ξ, n'_ξ are the cosines of the angles between the directions of the ray and the corresponding axes ξ and ξ'), we obtain

$$d\psi = (\alpha_\xi n'_\xi - n_\xi)\, d\xi.$$

As for every pair of corresponding points of object and image, the optical path length $\psi + d\psi$ must be the same for all rays starting from the point $d\xi$ and arriving at the point $d\xi'$. From this we obtain the condition:

$$\alpha_\xi n'_\xi - n_\xi = \text{const.} \tag{57.1}$$

This is the condition we have been seeking, which the paths of the rays in the optical system

must satisfy in order to have image formation for a small line segment using broad beams. The relation (57.1) must be fulfilled for all rays starting from the point O.

Let us apply this condition to image formation by means of an axially-symmetric optical system. We start with the image of a line segment coinciding with the optical axis (x axis); clearly the image also coincides with the axis. A ray moving along the optical axis ($n_x = 1$), because of the axial symmetry of the system, does not change its direction after passing through it, that is, n'_x is also 1. From this it follows that const in (57.1) is equal in this case to $\alpha_x - 1$, and we can rewrite (57.1) in the form

$$\frac{1 - n_x}{1 - n'_x} = \alpha_x .$$

Denoting by θ and θ' the angles subtended by the rays with the optical axis at points of the object and image, we have

$$1 - n_x = 1 - \cos\theta = 2\sin^2\frac{\theta}{2}, \quad 1 - n'_x = 1 - \cos\theta' = 2\sin^2\frac{\theta'}{2}.$$

Thus we obtain the condition for image formation in the form

$$\frac{\sin\frac{\theta}{2}}{\sin\frac{\theta'}{2}} = \text{const} = \sqrt{\alpha_x} . \tag{57.2}$$

Next, let us consider the imaging of a small portion of a plane perpendicular to the optical axis of an axially symmetric system; the image will obviously also be perpendicular to this axis. Applying (57.1) to an arbitrary segment lying in the plane which is to be imaged, we get:

$$\alpha_r \sin\theta' - \sin\theta = \text{const},$$

where θ and θ' are again the angles made by the beam with the optical axis. For rays emerging from the point of intersection of the object plane with the optical axis, and directed along this axis ($\theta = 0$), we must have $\theta' = 0$, because of symmetry. Therefore const is zero, and we obtain the condition for imaging in the form

$$\frac{\sin\theta}{\sin\theta'} = \text{const} = \alpha_r . \tag{57.3}$$

As for the formation of an image of a three-dimensional object using broad beams, it is easy to see that this is impossible even for a small volume, since the conditions (57.2) and (57.3) are incompatible.

§ 58. The limits of geometrical optics

From the definition of a monochromatic plane wave, its amplitude is the same everywhere and at all times. Such a wave is infinite in extent in all directions in space, and exists over the whole range of time from $-\infty$ to $+\infty$. Any wave whose amplitude is not constant everywhere at all times can only be more or less monochromatic. We now take up the question of the "degree of non-monochromaticity" of a wave.

Let us consider an electromagnetic wave whose amplitude at each point is a function of the time. Let ω_0 be some average frequency of the wave. Then the field of the wave, for

example the electric field, at a given point has the form $\mathbf{E}_0(t)e^{-i\omega_0 t}$. This field, although it is of course not monochromatic, can be expanded in monochromatic waves, that is, in a Fourier integral. The amplitude of the component in the expansion, with frequency ω, is proportional to the integral

$$\int_{-\infty}^{+\infty} \mathbf{E}_0(t)e^{i(\omega-\omega_0)t}dt.$$

The factor $e^{i(\omega-\omega_0)t}$ is a periodic function whose average value is zero. If $\mathbf{E}_0$ were exactly constant, then the integral would be exactly zero, for $\omega \neq \omega_0$. If, however, $\mathbf{E}_0(t)$ is variable, but hardly changes over a time interval of order $1/|\omega - \omega_0|$, then the integral is almost equal to zero, the more exactly the slower the variation of $\mathbf{E}_0$. In order for the integral to be significantly different from zero, it is necessary that $\mathbf{E}_0(t)$ vary significantly over a time interval of the order of $1/|\omega - \omega_0|$.

We denote by Δt the order of magnitude of the time interval during which the amplitude of the wave at a given point in space changes significantly. From these considerations, it now follows that the frequencies deviating most from ω_0, which appear with reasonable intensity in the spectral resolution of this wave, are determined by the condition $1/|\omega - \omega_0| \sim \Delta t$. If we denote by $\Delta\omega$ the frequency interval (around the average frequency ω_0) which enters in the spectral resolution of the wave, then we have the relation

$$\Delta\omega\Delta t \sim 1. \tag{58.1}$$

We see that a wave is the more monochromatic (i.e. the smaller $\Delta\omega$) the larger Δt, i.e. the slower the variation of the amplitude at a given point in space.

Relations similar to (58.1) are easily derived for the wave vector. Let Δx, Δy, Δz be the orders of magnitude of distances along the X, Y, Z axes, in which the wave amplitude changes significantly. At a given time, the field of the wave as a function of the coordinates has the form

$$\mathbf{E}_0(\mathbf{r})e^{i\mathbf{k}_0\cdot\mathbf{r}},$$

where $\mathbf{k}_0$ is some average value of the wave vector. By a completely analogous derivation to that for (58.1) we can obtain the interval $\Delta\mathbf{k}$ of values contained in the expansion of the wave into a Fourier integral:

$$\Delta k_x \Delta x \sim 1, \quad \Delta k_y \Delta y \sim 1, \quad \Delta k_z \Delta z \sim 1. \tag{58.2}$$

Let us consider, in particular, a wave which is radiated during a finite time interval. We denote by Δt the order of magnitude of this interval. The amplitude at a given point in space changes significantly during the time Δt in the course of which the wave travels completely past the point. Because of the relations (58.1) we can now say that the "lack of monochromaticity" of such a wave, $\Delta\omega$, cannot be smaller than $1/\Delta t$ (it can of course be larger):

$$\Delta\omega \gtrsim \frac{1}{\Delta t}. \tag{58.3}$$

Similarly, if Δx, Δy, Δz are the orders of magnitude of the extension of the wave in space, then for the spread in the values of components of the wave vector, entering in the resolution of the wave, we obtain

$$\Delta k_x \gtrsim \frac{1}{\Delta x}, \quad \Delta k_y \gtrsim \frac{1}{\Delta y}, \quad \Delta k_z \lesssim \frac{1}{\Delta z}. \tag{58.4}$$

From these formulas it follows that if we have a beam of light of finite width, then the direction of propagation of the light in such a beam cannot be strictly constant. Taking the X axis along the (average) direction of light in the beam, we obtain

$$\theta_y \gtrsim \frac{1}{k\Delta y} \sim \frac{\lambda}{\Delta y}, \tag{58.5}$$

where θ_y is the order of magnitude of the deviation of the beam from its average direction in the XY plane and λ is the wavelength.

On the other hand, the formula (58.5) answers the question of the limit of sharpness of optical image formation. A beam of light whose rays, according to geometrical optics, would all intersect in a point, actually gives an image not in the form of a point but in the form of a spot. For the width Δ of this spot, we obtain, according to (58.5),

$$\Delta \sim \frac{1}{k\theta} \sim \frac{\lambda}{\theta}, \tag{58.6}$$

where θ is the opening angle of the beam. This formula can be applied not only to the image but also to the object. Namely, we can state that in observing a beam of light emerging from a luminous point, this point cannot be distinguished from a body of dimensions λ/θ. In this way formula (58.6) determines the limiting *resolving power* of a microscope. The minimum value of Δ, which is reached for $\theta \sim 1$, is λ, in complete agreement with the fact that the limit of geometrical optics is determined by the wavelength of the light.

PROBLEM

Determine the order of magnitude of the smallest width of a light beam produced from a parallel beam at a distance l from a diaphragm.

Solution: Denoting the size of the aperture in the diaphragm by d, we have from (58.5) for the angle of deflection of the beam (the "diffraction angle"), λ/d, so that the width of the beam is of order $d + (\lambda/d)l$. The smallest value of this quantity $\sim \sqrt{\lambda l}$.

§ 59. Diffraction

The laws of geometrical optics are strictly correct only in the ideal case when the wavelength can be considered to be infinitely small. The more poorly this condition is fulfilled, the greater are the deviations from geometrical optics. Phenomenon which are the consequence of such deviations are called *diffraction phenomena.*

Diffraction phenomena can be observed, for example, if along the path of propagation of the light† there is an obstacle—an opaque body (we call it a *screen*) of arbitrary form or, for example, if the light passes through holes in opaque screens. If the laws of geometrical optics were strictly satisfied, there would be beyond the screen regions of "shadow" sharply delineated from regions where light falls. The diffraction has the consequence that, instead of a sharp boundary between light and shadow, there is a quite complex distribution of the

† In what follows, in discussing diffraction we shall talk of the diffraction of light; all these same considerations also apply, of course, to any electromagnetic wave.

intensity of the light. These diffraction phenomena appear the more strongly the smaller the dimensions of the screens and the apertures in them, or the greater the wavelength.

The problems of the theory of diffraction consists in determining, for given positions and shapes of the objects (and locations of the light sources), the distribution of the light, that is, the electromagnetic field over all space. The exact solution of this problem is possible only through solution of the wave equation with suitable boundary conditions at the surface of the body, these conditions being determined also by the optical properties of the material. Such a solution usually presents great mathematical difficulties.

However, there is an approximate method which for many cases is a satisfactory solution of the problem of the distribution of light near the boundary between light and shadow. This method is applicable to cases of small deviation from geometrical optics, i.e. when firstly, the dimensions of all bodies are large compared with the wavelength (this requirement applies both to the dimensions of screens and apertures and also to the distances from the bodies to the points of emission and observation of the light); and secondly when there are only small deviations of the light from the directions of the rays given by geometrical optics.

Let us consider a screen with an aperture through which the light passes from given sources. Figure 9 shows the screen in profile (the heavy line); the light travels from left to right. We denote by u some one of the components of **E** or **H**. Here we shall understand u to mean a function only of the coordinates, i.e. without the factor $e^{-i\omega t}$ determining the time dependence. Our problem is to determine the light intensity, that is, the field u, at any point of observation P beyond the screen. For an approximate solution of this problem in cases where the deviations from geometrical optics are small, we may assume that at the points of the aperture the field is the same as it would have been in the absence of the screen. In other words, the values of the field here are those which follow directly from geometrical optics. At all points immediately behind the screen, the field can be set equal to zero. In this the properties of the screen (i.e. of the screen material) obviously play no part. It is also obvious that in the cases we are considering, what is important for the diffraction is only the shape of the edge of the aperture, while the shape of the opaque screen is unimportant.

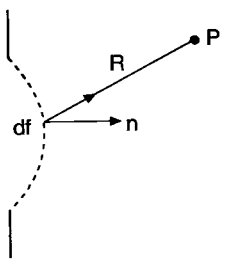


FIG. 9.

We introduce some surface which covers the aperture in the screen and is bounded by its edges (a profile of such a surface is shown in Fig. 9 as a dashed line). We break up this surface into sections with area df, whose dimensions are small compared with the size of the aperture, but large compared with the wavelength of the light. We can then consider each of these sections through which the light passes as if it were itself a source of light waves spreading out on all sides from this section. We shall consider the field at the point P to be the result of superposition of the fields produced by all the sections df of the surface covering the aperture. (This is called *Huygens' principle.*)

The field produced at the point P by the section df is obviously proportional to the value u of the field at the section df itself (we recall that the field at df is assumed to be the same as it would have been in the absence of the screen). In addition, it is proportional to the projection df_n of the area df on the plane perpendicular to the direction n of the ray coming from the light source to df. This follows from the fact that no matter what shape the element df has, the same rays will pass through it provided its projection df_n remain fixed, and therefore its effect on the field at P will be the same.

Thus the field produced at the point P by the section df is proportional to $u\,df_n$. Furthermore, we must still take into account the change in the amplitude and phase of the wave during its propagation from df to P. The law of this change is determined by formula (54.3). Therefore $u\,df_n$ must be multiplied by $(1/R)e^{ikR}$ (where R is the distance from df to P, and k is the absolute value of the wave vector of the light), and we find that the required field is

$$au\frac{e^{ikR}}{R}df_n,$$

where a is an as yet unknown constant. The field at the point P, being the result of the addition of the fields produced by all the elements df, is consequently equal to

$$u_P = a\int u\frac{e^{ikR}}{R}df_n, \tag{59.1}$$

where the integral extends over the surface bounded by the edge of the aperture. In the approximation we are considering, this integral cannot, of course, depend on the form of this surface. Formula (59.1) is, obviously, applicable not only to diffraction by an aperture in a screen, but also to diffraction by a screen around which the light passes freely. In that case the surface of integration in (59.1) extends on all sides from the edge of the screen.

To determine the constant a, we consider a plane wave propagating along the X axis; the wave surfaces are parallel to the plane YZ. Let u be the value of the field in the YZ plane. Then at the point P, which we choose on the X axis, the field is equal to $u_p = ue^{ikx}$. On the other hand, the field at the point P can be determined starting from formula (59.1), choosing as surface of integration, for example, the YZ plane. In doing this, because of the smallness of the angle of diffraction, only those points of the YZ plane are important in the integral which lie close to the origin, i.e. the points for which $y, z \ll x$ (x is the coordinate of the point P). Then

$$R = \sqrt{x^2 + y^2 + z^2} \approx x + \frac{y^2 + z^2}{2x},$$

and (59.1) gives

$$u_p = au\frac{e^{ikx}}{x}\int_{-\infty}^{+\infty} e^{i\frac{ky^2}{2x}}dy\int_{-\infty}^{+\infty} e^{i\frac{kz^2}{2x}}dz,$$

where u is a constant (the field in the YZ plane); in the factor $1/R$, we can put $R \simeq x = \text{const}$. By the substitution $y = \xi\sqrt{2x/k}$ these two integrals can be transformed to the integral

$$\int_{-\infty}^{+\infty} e^{i\xi^2}d\xi = \int_{-\infty}^{+\infty}\cos\xi^2 d\xi + i\int_{-\infty}^{+\infty}\sin\xi^2 d\xi = \sqrt{\frac{\pi}{2}}(1+i),$$

and we get

$$u_p = aue^{ikx}\frac{2i\pi}{k}.$$

On the other hand, $u_p = ue^{ikx}$, and consequently

$$a = \frac{k}{2\pi i}.$$

Substituting in (59.1), we obtain the solution to our problem in the form

$$u_p = \int \frac{ku}{2\pi iR} e^{ikR} df_n. \tag{59.2}$$

In deriving formula (59.2), the light source was assumed to be essentially a point, and the light was assumed to be strictly monochromatic. The case of a real, extended source, which emits non-monochromatic light, does not, however, require special treatment. Because of the complete independence (incoherence) of the light emitted by different points of the source, and the incoherence of the different spectral components of the emitted light, the total diffraction pattern is simply the sum of the intensity distributions obtained from the diffraction of the independent components of the light.

Let us apply formula (59.2) to the solution of the problem of the change in phase of a ray on passing through its point of tangency to the caustic (see the end of § 54). We choose as our surface of integration in (59.2) any wave surface, and determine the field u_p at a point P, lying on some given ray at a distance x from its point of intersection with the wave surface we have chosen (we choose this point as coordinate origin O, and as YZ plane the plane tangent to the wave surface at the point O). In the integration of (59.2) only a small area of the wave surface in the neighbourhood of O is important. If the XY and XZ planes are chosen to coincide with the principal planes of curvature of the wave surface at the point O, then near this point the equation of the surface is

$$X = \frac{y^2}{2R_1} + \frac{z^2}{2R_2},$$

where R_1 and R_2 are the radii of curvature. The distance R from the point on the wave surface with coordinates X, y, z, to the point P with coordinates x, 0, 0, is

$$R = \sqrt{(x-X)^2 + y^2 + z^2} \simeq x + \frac{y^2}{2}\left(\frac{1}{x} - \frac{1}{R_1}\right) + \frac{z^2}{2}\left(\frac{1}{x} - \frac{1}{R_2}\right).$$

On the wave surface, the field u can be considered constant; the same applies to the factor $1/R$. Since we are interested only in changes in the place of the wave, we drop coefficients and write simply

$$u_p \sim \frac{1}{i}\int e^{ikR} df_n \simeq \frac{e^{ikx}}{i} \int_{-\infty}^{+\infty} dy e^{ik\frac{y^2}{2}\left(\frac{1}{x}-\frac{1}{R_1}\right)} \int_{-\infty}^{+\infty} dz e^{ik\frac{z^2}{2}\left(\frac{1}{x}-\frac{1}{R_2}\right)}. \tag{59.3}$$

The centres of curvature of the wave surface lie on the ray we are considering, at the points $x = R_1$ and $x = R_2$; these are the points where the ray is tangent to the caustic. Suppose $R_2 < R_1$. For $x < R_2$, the coefficients of i in the exponentials appearing in the two integrands

are positive, and each of these integrals is proportional to $(1 + i)$. Therefore on the part of the ray before its first tangency to the caustic, we have $u_p \sim e^{ikx}$. For $R_2 < x < R_1$, that is, on the segment of the ray between its two points of tangency, the integral over y is proportional to $1 + i$, but the integral over z is proportional to $1 - i$, so that their product does not contain i. Thus we have here $u_p \sim -ie^{ikx} = e^{i(kx-(\pi/2))}$, that is, as the ray passes in the neighbourhood of the first caustic, its phase undergoes an additional change of $-\pi/2$. Finally, for $x > R_1$, we have $u_p \sim -e^{ikx} = e^{i(kx-\pi)}$, that is, on passing in the neighbourhood of the second caustic, the phase once more changes by $-\pi/2$.

PROBLEM

Determine the distribution of the light intensity in the neighbourhood of the point where the ray is tangent to the caustic.

Solution: To solve the problem, we use formula (59.2), taking the integral in it over any wave surface which is sufficiently far from the point of tangency of the ray to the caustic. In Fig. 10, ab is a section of this wave surface, and $a'b'$ is a section of the caustic; $\alpha'b'$ is the evolute of the curve ab. We are interested in the intensity distribution in the neighbourhood of the point O where the ray QO is tangent to the caustic; we assume the length D of the segment QO of the ray to be large. We denote by x the distance from the point O along the normal to the caustic, and assume positive values x for points on the normal in the direction of the centre of curvature.

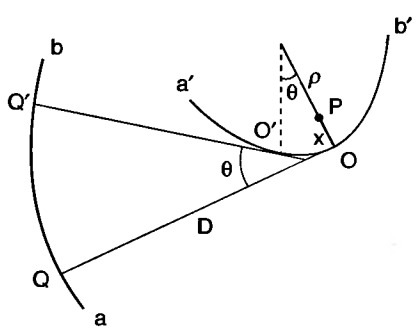


FIG. 10.

The integrand in (59.2) is a function of the distance R from the arbitrary point Q' on the wave surface to the point P. From a well-known property of the evolute, the sum of the length of the segment $Q'O'$ of the tangent at the point O' and the length of the arc OO' is equal to the length QO of the tangent at the point O. For points O and O' which are near to each other we have $OO' = \theta\varrho$ (ϱ is the radius of curvature of the caustic at the point O). Therefore the length $Q'O' = D - \theta\varrho$. The distance $Q'O$ (along a straight line) is approximately (the angle θ is assumed to be small)

$$Q'O \cong Q'O' + \varrho \sin\theta = D - \theta\varrho + \varrho \sin\theta \cong D - \varrho\frac{\theta^3}{6}.$$

Finally, the distance $R = Q'P$ is equal to $R \cong Q'O - x \sin\theta \cong Q'O - x\theta$, that is,

$$R \cong D - x\theta - \tfrac{1}{6}\varrho\theta^3.$$

Substituting this expression in (59.2), we obtain

$$u_p \sim \int_{-\infty}^{+\infty} e^{-ikr\theta - i\frac{k\varrho}{6}\theta^3}\, d\theta = 2\int_0^{\infty} \cos\left(kx\theta + \frac{k\varrho}{6}\theta^3\right) d\theta$$

(the slowly varying factor $1/D$ in the integrand is unimportant compared with the exponential factor, so we assume it constant). Introducing the new integration variable $\xi = (k\varrho/2)^{1/3}\,\theta$, we get

$$u_p \sim \Phi\left(x\left(\frac{2k^2}{\varrho}\right)^{1/3}\right),$$

where $\Phi(t)$ is the Airy function.†

For the intensity $I \sim |u_p|^2$, we write:

$$I = 2A\left(\frac{2k^2}{\varrho}\right)^{1/6}\Phi^2\left(x\left(\frac{2k^2}{\varrho}\right)^{1/3}\right)$$

(concerning the choice of the constant factor, cf, below).

For large positive values of x, we have from this the asymptotic formula

$$I \approx \frac{A}{2\sqrt{x}}\exp\left\{-\frac{4x^{3/2}}{3}\sqrt{\frac{2k^2}{\varrho}}\right\},$$

that is, the intensity drops exponentially (shadow region). For large negative values of x, we have

$$I \approx \frac{2A}{\sqrt{-x}}\sin^2\left\{\frac{2(-x)^{3/2}}{3}\sqrt{\frac{2k^2}{\rho}} + \frac{\pi}{4}\right\},$$

that is, the intensity oscillates rapidly; its average value over these oscillations is

$$\bar{I} = \frac{A}{\sqrt{-x}}.$$

From this meaning of the constant A is clear—it is the intensity far from the caustic which would be obtained from geometrical optics neglecting diffraction effects.

† The Airy function $\Phi(t)$ is defined as

$$\Phi(t) = \frac{1}{\sqrt{\pi}}\int_0^\infty \cos\left(\frac{\xi^3}{3} + \xi t\right)d\xi. \tag{1}$$

(see *Quantum Mechanics,* Mathematical Appendices, § b). For large positive values of the argument, the asymptotic expression for $\Phi(t)$ is

$$\Phi(t) \approx \frac{1}{2t^{1/4}}\exp\left(-\frac{2}{3}t^{3/2}\right), \tag{2}$$

that is, $\Phi(t)$ goes exponentially to zero. For large negative values of t, the function $\Phi(t)$ oscillates with decreasing amplitude according to the law:

$$\Phi(t) \approx \frac{1}{(-t)^{1/4}}\sin\left(\frac{2}{3}(-t)^{3/2} + \frac{\pi}{4}\right). \tag{3}$$

The Airy function is related to the MacDonald function (modified Hankel function) of order 1/3:

$$\Phi(t) = \sqrt{t/3\pi}\,K_{1/3}(\tfrac{2}{3}t^{3/2}). \tag{4}$$

Formula (2) corresponds to the asymptotic expansion of $K_\nu(t)$:

$$Kv(t) \approx \sqrt{\frac{\pi}{2t}}e^{-t}\,.$$

The function $\Phi(t)$ attains its largest value, 0.949, for $t = -1.02$; correspondingly, the maximum intensity is reached at $x(2k^2/\varrho)^{1/3} = -1.02$, where

$$I = 2.03\, Ak^{1/3}\varrho^{-1/6}.$$

At the point where the ray is tangent to the caustic ($x = 0$), we have $I = 0.89\, Ak^{1/3}\varrho^{-1/6}$ [since $\Phi(0) = 0.629$].

Thus near the caustic the intensity is proportional to $k^{1/3}$, that is, to $\lambda^{-1/3}$ (λ is the wavelength). For $\lambda \to 0$, the intensity goes to infinity, as it should (see § 54).

§ 60. Fresnel diffraction

If the light source and the point P at which we determine the intensity of the light are located at finite distances from the screen, then in determining the intensity at the point P, only those points are important which lie in a small region of the wave surface over which we integrate in (59.2)—the region which lies near the line joining the source and the point P. In fact, since the deviations from geometrical optics are small, the intensity of the light arriving at P from various points of the wave surface decreases very rapidly as we move away from this line. Diffraction phenomena in which only a small portion of the wave surface plays a role are called *Fresnel diffraction* phenomena.

Let us consider the Fresnel diffraction by a screen. From what we have just said, for a given point P only a small region at the edge of the screen is important for this diffraction. But over sufficiently small regions, the edge of the screen can always be considered to be a straight line. We shall therefore, from now on, understand the edge of the screen to mean just such a small straight line segment.

We choose as the XY plane a plane passing through the light source Q (Fig. 11) and through the line of the edge of the screen. Perpendicular to this, we choose the plane XZ so that it passes through the point Q and the point of observation P, at which we try to determine the light intensity. Finally, we choose the origin of coordinates O on the line of the edge of the screen, after which the positions of all three axes are completely determined.

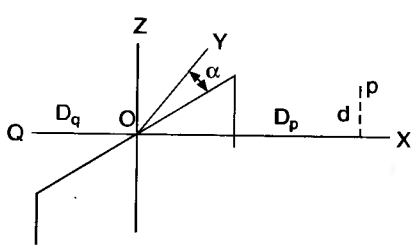


FIG. 11.

Let the distance from the light source Q to the origin be D_q. We denote the x-coordinate of the point of observation P by D_p, and its z-coordinate, i.e. its distance from the XY plane, by d. According to geometrical optics, the light should pass only through points lying above the XY plane; the region below the XY plane is the region which according to geometrical optics should be in shadow (region of geometrical shadow).

We now determine the distribution of light intensity on the screen near the edge of the geometrical shadow, i.e. for values of d small compared with D_p and D_q. A negative d means that the point P is located within the geometrical shadow.

As the surface of integration in (59.2) we choose the half-plane passing through the line of the edge of the screen and perpendicular to the XY plane. The coordinates x and y of points

on this surface are related by the equation $x = y \tan \alpha$ (α is the angle between the line of the edge of the screen and the Y axis), and the z-coordinate is positive. The field of the wave produced by the source Q, at the distance R_q from it, is proportional to the factor e^{ikR_q}. Therefore the field u on the surface of integration is proportional to

$$u \sim \exp\{ik\sqrt{y^2 + z^2 + (D_q + y \tan \alpha)^2}\}.$$

In the integral (59.2) we must now substitute for R,

$$R = y^2 + (z - d)^2 + (D_p - y \tan \alpha)^2.$$

The slowly varying factors in the integrand are unimportant compared with the exponential. Therefore we may consider $1/R$ constant, and write $dy\, dz$ in place of df_n. We then find that the field at the point P is

$$u_p \sim \int_{-\infty}^{+\infty} \int_{0}^{\infty} \exp\{ik\sqrt{(D_q + y \tan \alpha)^2 + y^2 + z^2}$$

$$+ \sqrt{(D_p - y \tan \alpha)^2 + (z - d)^2 + y^2}\}\, dy\, dz. \qquad (60.1)$$

As we have already said, the light passing through the point P comes mainly from points of the plane of integration which are in the neighbourhood of O. Therefore in the integral (60.1) only values of y and z which are small (compared with D_q and D_p) are important. For this reason we can write

$$\sqrt{(D_q + y \tan \alpha)^2 + y^2 + z^2} \simeq D_q + \frac{y^2 + z^2}{2D_q} + y \tan \alpha,$$

$$\sqrt{(D_p - y \tan \alpha)^2 + (z - d)^2 + y^2} \simeq D_p + \frac{(z - d)^2 + y^2}{2D_p} - y \tan \alpha.$$

We substitute this in (60.1). Since we are interested only in the field as a function of the distance d, the constant factor $\exp\{ik(D_p + D_q)\}$ can be omitted; the integral over y also gives an expression not containing d, so we omit it also. We then find

$$u_p \sim \int_{0}^{\infty} \exp\left\{ik\left(\frac{1}{2D_q} z^2 + \frac{1}{2D_p}(z - d)^2\right)\right\} dz.$$

This expression can also be written in the form

$$u_p \sim \exp\left\{ik \frac{d^2}{2(D_p + D_q)}\right\} \int_{0}^{\infty} \exp\left\{ik \frac{\frac{1}{2}\left[\left(\frac{1}{D_p} + \frac{1}{D_q}\right)z - \frac{d}{D_p}\right]^2}{\frac{1}{D_p} + \frac{1}{D_q}}\right\} dz. \qquad (60.2)$$

The light intensity is determined by the square of the field, that is, by the square modulus $|u_p|^2$. Therefore, when calculating the intensity, the factor standing in front of the integral is

irrelevant, since when multiplied by the complex conjugate expression it gives unity. An obvious substitution reduces the integral to

$$u_p \sim \int_{-w}^{\infty} e^{i\eta^2}\, d\eta, \tag{60.3}$$

where

$$w = d\sqrt{\frac{kD_q}{2D_p(D_q + D_p)}}. \tag{60.4}$$

Thus, the intensity I at the point P is :

$$I = \frac{I_0}{2}\left|\sqrt{\frac{2}{\pi}}\int_{-w}^{\infty} e^{i\eta^2}\, d\eta\right|^2 = \frac{I_0}{2}\left\{\left(C(w^2) + \frac{1}{2}\right)^2 + \left(S(w^2) + \frac{1}{2}\right)^2\right\}, \tag{60.5}$$

where

$$C(z) = \sqrt{\frac{2}{\pi}}\int_0^{\sqrt{z}} \cos\eta^2\, d\eta, \quad S(z) = \sqrt{\frac{2}{\pi}}\int_0^{\sqrt{z}} \sin\eta^2\, d\eta$$

are called the *Fresnel integrals.* Formula (60.5) solves our problem of determining the light intensity as a function of d. The quantity I_0 is the intensity in the illuminated region at points not too near the edge of the shadow; more precisely, at those points with $w >> 1$ ($C(\infty) = S(\infty) = \frac{1}{2}$ in the limit $w \to \infty$).

The region of geometrical shadow corresponds to negative w. It is easy to find the asymptotic form of the function $I(w)$ for large negative values of w. To do this we proceed as follows. Integrating by parts, we have

$$\int_{|w|}^{\infty} e^{i\eta^2}\, d\eta = -\frac{1}{2i|w|}\, e^{iw^2} + \frac{1}{2i}\int_{|w|}^{\infty} e^{i\eta^2}\, \frac{d\eta}{\eta^2}.$$

Integrating by parts once more on the right side of the equation and repeating this process, we obtain an expansion in powers of $1/|w|$:

$$\int_{|w|}^{\infty} e^{i\eta^2}\, d\eta = e^{iw^2}\left[-\frac{1}{2i|w|} + \frac{1}{4|w|^3} - \ldots\right]. \tag{60.6}$$

Although an infinite series of this type does not converge, nevertheless, because the sucessive terms decrease very rapidly for large values of $|w|$, the first term already gives a good representation of the function on the left for sufficiently large $|w|$ (such a series is said to be *asymptotic*). Thus, for the intensity $I(w)$, (60.5), we obtain the following asymptotic formula, valid for large negative values of w:

$$I = \frac{I_0}{4\pi w^2}. \tag{60.7}$$

We see that in the region of geometric shadow, far from its edge, the intensity goes to zero as the inverse square of the distance from the edge of the shadow.

We now consider positive values of w, that is, the region above the XY plane. We write

$$\int_{-w}^{\infty} e^{i\eta^2}\,d\eta = \int_{-\infty}^{+\infty} e^{i\eta^2}\,d\eta - \int_{-\infty}^{-w} e^{i\eta^2}\,d\eta = (1+i)\sqrt{\frac{\pi}{2}} - \int_{w}^{\infty} e^{i\eta^2}\,d\eta.$$

For sufficiently large w, we can use an asymptotic representation for the integral standing on the right side of the equation, and we have

$$\int_{-w}^{\infty} e^{i\eta^2}\,d\eta \cong (1+i)\sqrt{\frac{\pi}{2}} + \frac{1}{2iw}e^{iw^2}. \tag{60.8}$$

Substituting this expression in (60.5), we obtain

$$I = I_0\left(1 + \sqrt{\frac{1}{\pi}}\,\frac{\sin\left(w^2 - \frac{\pi}{4}\right)}{w}\right). \tag{60.9}$$

Thus in the illuminated region, far from the edge of the shadow, the intensity has an infinite sequence of maxima and minima, so that the ratio I/I_0 oscillates on both sides of unity. With increasing w, the amplitude of these oscillations decreases inversely with the distance from the edge of the geometric shadow, and the positions of the maxima and minima steadily approach one another.

For small w, the function $I(w)$ has qualitatively this same character (Fig. 12). In the region of the geometric shadow, the intensity decreases monotonically as we move away from the boundary of the shadow. (On the boundary itself, $I/I_0 = \frac{1}{4}$.) For positive w, the intensity has alternating maxima and minima. At the first (largest) maximum, $I/I_0 = 1.37$.

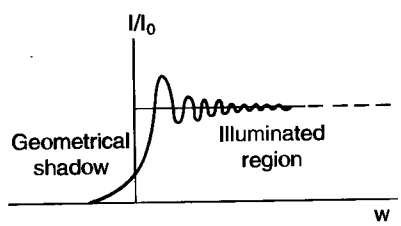


FIG. 12.

§ 61. Fraunhofer diffraction

Of special interest for physical applications are those diffraction phenomena which occur when a plane parallel bundle of rays is incident on a screen. As a result of the diffraction, the beam ceases to be parallel, and there is light propagation along directions other than the initial one. Let us consider the problem of determining the distribution over direction of the intensity of the diffracted light at large distances beyond the screen (this formulation of the problem corresponds to *Fraunhofer diffraction*). Here we shall again restrict ourselves to

the case of small deviations from geometrical optics, i.e. we shall assume that the angles of deviation of the rays from the initial direction (the diffraction angles) are small.

This problem can be solved by starting from the general formula (59.2) and passing to the limit where the light source and the point of observation are at infinite distances from the screen. A characteristic feature of the case we are considering is that, in the integral which determines the intensity of the diffracted light, the whole wave surface over which the integral is taken is important (in contrast to the case of Fresnel diffraction, where only the portions of the wave surface near the edge of the screens are important).†

However, it is simpler to treat this problem anew, without recourse to the general formula (59.2).

Let us denote by u_0 the field which would exist beyond the screens if geometrical optics were rigorously valid. This field is a plane wave, but its cross-section has certain regions (corresponding to the "shadows" of opaque screens) in which the field is zero. We denote by S the part of the plane cross-section on which the field u_0 is different from zero; since each such plane is a wave surface of the plane wave, $u_0 = \text{const}$ over the whole surface S.

Actually, however, a wave with a limited cross-sectional area cannot be strictly plane (see § 58). In its spatial Fourier expansion there appear components with wave vectors having different directions, and this is precisely the origin of the diffraction.

Let us expand the field u_0 into a two-dimensional Fourier integral with respect to the coordinates y, z in the plane of the transverse cross-section of the wave. For the Fourier components, we have:

$$u_{\mathbf{q}} = \iint u_0 e^{-i\mathbf{q}\cdot\mathbf{r}}\, dy\, dz, \tag{61.1}$$

where the vectors $\mathbf{q}$ are constant vectors in the y, z plane; the integration actually extends only over that portion S of the y, z plane on which u_0 is different from zero. If $\mathbf{k}$ is the wave vector of the incident wave, the field component $u_{\mathbf{q}}e^{i\mathbf{q}\cdot\mathbf{r}}$ gives the wave vector $\mathbf{k}' = \mathbf{k} + \mathbf{q}$. Thus the vector $\mathbf{q} = \mathbf{k}' - \mathbf{k}$ determines the change in the wave vector of the light in the diffraction. Since the absolute values $k = k' = \omega/c$, the small diffraction angles θ_y, θ_z in the xy- and xz-planes are related to the components of the vector q by the equations

$$q_y = \frac{\omega}{c}\theta_y, \quad q_z = \frac{\omega}{c}\theta_z. \tag{61.2}$$

For small deviations from geometrical optics, the components in the expansion of the field u_0 can be assumed to be identical with the components of the actual field of the diffracted light, so that formula (61.1) solves our problem.

† The criteria for Fresnel and Fraunhofer diffraction are easily found by returning to formula (60.2) and applying it, for example, to a slit of width a (instead of to the edge of an isolated screen). The integration over z in (60.2) should then be taken between the limits from 0 to a. Fresnel diffraction corresponds to the case when the term containing z^2 in the exponent of the integrand is important, and the upper limit of the integral can be replaced by ∞. For this to be the case, we must have

$$ka^2\left(\frac{1}{D_p} + \frac{1}{D_q}\right) \gg 1.$$

On the other hand, if this inequality is reversed, the term in z^2 can be dropped; this corresponds to the case of Fraunhofer diffraction.

The intensity distribution of the diffracted light is given by the square $|u_{\mathbf{q}}|^2$ as a function of the vector $\mathbf{q}$. The quantitative connection with the intensity of the incident light is established by the formula

$$\iint u_0^2 \, dy \, dz = \iint |u_{\mathbf{q}}|^2 \frac{dq_y \, dq_z}{(2\pi)^2} \tag{61.3}$$

[compare (49.8)]. From this we see that the relative intensity diffracted into the solid angle $do = d\theta_y \, d\theta_z$ is given by

$$\frac{|u_{\mathbf{q}}|^2}{u_0^2} \frac{dq_y \, dq_z}{(2\pi)^2} = \left(\frac{\omega}{2\pi c}\right)^2 \left|\frac{u_{\mathbf{q}}}{u_0}\right|^2 do. \tag{61.4}$$

Let us consider the Fraunhofer diffraction from two screens which are "complementary": the first screen has holes where the second is opaque and conversely. We denote by $u^{(1)}$ and $u^{(2)}$ the field of the light diffracted by these screens (when the same light is incident in both cases). Since $u_{\mathbf{q}}^{(1)}$ and $u_{\mathbf{q}}^{(2)}$ are expressed by integrals (61.1) taken over the surfaces of the apertures in the screens, and since the apertures in the two screens complement one another to give the whole plane, the sum $u_{\mathbf{q}}^{(1)} + u_{\mathbf{q}}^{(2)}$ is the Fourier component of the field obtained in the absence of the screens, i.e. it is simply the incident light. But the incident light is a rigorously plane wave with definite direction of propagation, so that $u_{\mathbf{q}}^{(1)} + u_{\mathbf{q}}^{(2)} = 0$ for all nonzero values of $\mathbf{q}$. Thus we have $u_{\mathbf{q}}^{(1)} = -u_{\mathbf{q}}^{(2)}$, or for the corresponding intensities,

$$|u_{\mathbf{q}}^{(1)}|^2 = |u_{\mathbf{q}}^{(2)}|^2 \text{ for } \mathbf{q} \neq 0. \tag{61.5}$$

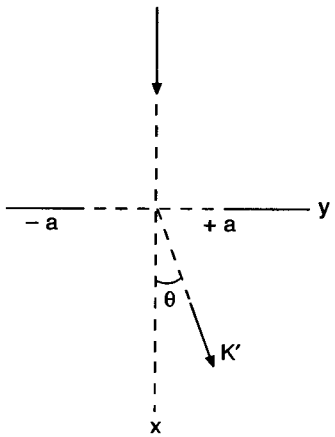
This means that complementary screens give the same distribution of intensity of the diffracted light (this is called *Babinet's principle*).

We call attention here to one interesting consequence of the Babinet principle. Let us consider a black body, i.e. one which absorbs completely all the light falling on it. According to geometrical optics, when such a body is illuminated, there is produced behind it a region of geometrical shadow, whose cross-sectional area is equal to the area of the body in the direction perpendicular to the direction of incidence of the light. However, the presence of diffraction causes the light passing by the body to be partially deflected from its initial direction. As a result, at large distances behind the body there will not be complete shadow but, in addition to the light propagating in the original direction, there will also be a certain amount of light propagating at small angles to the original direction. It is easy to determine the intensity of this scattered light. To do this, we point out that according to Babinet's principle, the amount of light deviated because of diffraction by the body under consideration is equal to the amount of light which would be deviated by diffraction from an aperture cut in an opaque screen, the shape and size of the aperture being the same as that of the transverse section of the body. But in Fraunhofer diffraction from an aperture *all* the light passing through the aperture is deflected. From this it follows that the total amount of light scattered by a black body is equal to the amount of light falling on its surface and absorbed by it.

PROBLEMS

1. Calculate the Fraunhofer diffraction of a plane wave normally incident on an infinite slit (of width $2a$) with parallel sides cut in an opaque screen.

Solution: We choose the plane of the slit as the yz plane, with the z axis along the slit (Fig. 13 shows a section of the screen). For normally incident light, the plane of the slit is one of the wave surfaces, and we choose it as the surface of integration in (61.1). Since the slit is infinitely long, the light is deflected only in the xy plane [since the integral (61.1)] becomes zero for $q_z \neq 0$].



Therefore the field should be expanded only in the y coordinate:

$$u_q = u_0 \int_{-a}^{a} e^{-iqy} dy = \frac{2u_0}{q} \sin qa.$$

The intensity of the diffracted light in the angular range $d\theta$ is

$$udI = \frac{I_0}{2a} \left| \frac{u_q}{u_0} \right|^2 \frac{dq}{2\pi} = \frac{I_0 \sin^2 ka\theta}{\pi ak \quad \theta^2} d\theta,$$

where $k = \omega/c$, and I_0 is the total intensity of the light incident on the slit.

$dI/d\theta$ as a function of diffraction angle has the form shown in Fig. 14. As θ increases toward either side from $\theta = 0$, the intensity goes through a series of maxima with rapidly decreasing height. The successive maxima are separated by minima at the points $\theta = n\pi/ka$ (where n is an integer); at the minima, the intensity falls to zero.

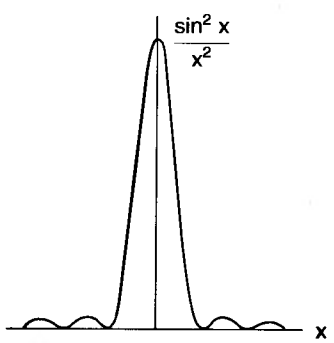


FIG. 14.

2. Calculate the Fraunhofer diffraction by a diffraction grating—a plane screen in which are cut a series of identical parallel slits (the width of the slits is $2a$, the width of opaque screen between neighbouring slits is $2b$, and the number of slits is N).

Solution: We choose the plane of the grating as the yz plane, with the z axis parallel to the slits. Diffraction occurs only in the xy plane, and integration of (61.1) gives:

$$u_q = u'_q \sum_{n=0}^{N-1} e^{-2inqd} = u'_q \frac{1 - e^{-2iNqd}}{1 - e^{-2iqd}},$$

where $d = a + b$, and u'_q is the result of the integration over a single slit. Using the results of problem 1, we get:

$$dI = \frac{I_0 a}{N\pi}\left(\frac{\sin Nqd}{\sin qd}\right)^2\left(\frac{\sin qa}{qa}\right)^2 dq = \frac{I_0}{N\pi ak}\left(\frac{\sin Nk\theta d}{\sin k\theta d}\right)^2 \frac{\sin^2 ka\theta}{\theta^2} d\theta$$

(I_0 is the total intensity of the light passing through all the slits).

For the case of a large number of slits ($N \to \infty$), this formula can be written in another form. For values $q = \pi n/d$, where n is an integer, dI/dq has a maximum; near such a maximum (i.e. for $qd = n\pi + \varepsilon$, with ε small)

$$dI = I_0 a\left(\frac{\sin qa}{qa}\right)^2 \frac{\sin^2 N\varepsilon}{\pi N\varepsilon^2} dq.$$

But for $N \to \infty$, we have the formula†

$$\lim_{N\to\infty} \frac{\sin^2 Nx}{\pi Nx^2} = \delta(x).$$

We therefore have, in the neighbourhood of each maximum:

$$dI = I_0 \frac{a}{d}\left(\frac{\sin qa}{qa}\right)^2 \delta(\varepsilon)\, d\varepsilon,$$

i.e., in the limit the widths of the maxima are infinitely narrow and the total light intensity in the n'th maximum is

$$I^{(n)} = I_0 \frac{d}{\pi^2 a} \frac{\sin^2(n\pi a/d)}{n^2}.$$

3. Find the distribution of intensity over direction for the diffraction of light which is incident normal to the plane of a circular aperture of radius a.

Solution: We introduce cylindrical coordinates z, r, ϕ with the z axis passing through the centre of the aperture and perpendicular to its plane. It is obvious that the diffraction is symmetric about the z axis, so that the vector $\mathbf{q}$ has only a radial component $q_r = q = k\theta$. Measuring the angle ϕ from the direction $\mathbf{q}$, and integrating in (61.1) over the plane of the aperture, we find:

† For $x \neq 0$ the function on the left side of the equation is zero, while according to a well-known formula of the theory of Fourier series,

$$\lim_{N\to\infty}\left(\frac{1}{\pi}\int_{-a}^{a} f(x)\frac{\sin^2 Nx}{Nx^2}dx\right) = f(0).$$

From this we see that the properties of this function actually coincide with those of the δ-function (see the footnote on p. 74).

$$u_q = u_0 \int_0^a \int_0^{2\pi} e^{-iqr\cos\phi}\, rd\phi dr = 2\pi u_0 \int_0^a J_0(qr) r dr,$$

where J_0 is the zero'th order Bessel function. Using the well-known formula

$$\int_0^a J_0(qr) r dr = \frac{a}{q} J_1(aq),$$

we then have

$$u_q = 2\pi \frac{u_0 a}{q} J_1(aq),$$

and according to (61.4) we obtain for the intensity of the light diffracted into the element of solid angle do:

$$dI = I_0 \frac{J_1^2(ak\theta)}{\pi\theta^2} do,$$

where I_0 is the total intensity of the light incident on the aperture.

CHAPTER 8

THE FIELD OF MOVING CHARGES

§ 62. The retarded potentials

In Chapter 5 we studied the constant field, produced by charges at rest, and in Chapter 6, the variable field in the absence of charges. Now we take up the study of varying fields in the presence of arbitrarily moving charges.

We derive equations determining the potentials for arbitrarily moving charges. This derivation is most conveniently done in four-dimensional form, repeating the derivation at the end of § 46, with the one change that we use the second pair of Maxwell equations in the form (30.2)

$$\frac{\partial F^{ik}}{\partial x^k} = -\frac{4\pi}{c} j^i.$$

The same right-hand side also appears in (46.8), and after imposing the Lorentz condition

$$\frac{\partial A^i}{\partial x^i} = 0, \quad \text{i.e.} \quad \frac{1}{c}\frac{\partial \phi}{\partial t} + \operatorname{div} \mathbf{A} = 0, \tag{62.1}$$

on the potentials, we get

$$\frac{\partial^2 A^i}{\partial x_k \partial x^k} = \frac{4\pi}{c} j^i. \tag{62.2}$$

This is the equation which determines the potentials of an arbitrary electromagnetic field. In three-dimensional form it is written as two equations, for $\mathbf{A}$ and for ϕ:

$$\Delta \mathbf{A} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = -\frac{4\pi}{c}\mathbf{j}, \tag{62.3}$$

$$\Delta \phi - \frac{1}{c^2}\frac{\partial^2 \phi}{\partial t^2} = -4\pi\varrho. \tag{62.4}$$

For constant fields, these reduce to the already familiar equations (36.4) and (43.4), and for variable fields without charges, to the homogeneous wave equation.

As we know, the solution of the inhomogeneous linear equations (62.3) and (62.4) can be represented as the sum of the solution of these equations without the right-hand side, and a particular integral of these equations with the right-hand side. To find the particular solution, we divide the whole space into infinitely small regions and determine the field produced by

the charges located in one of these volume elements. Because of the linearity of the field equations, the actual field will be the sum of the fields produced by all such elements.

The charge de in a given volume element is, generally speaking, a function of the time. If we choose the origin of coordinates in the volume element under consideration, then the charge density is $\varrho = de(t)\ \delta(\mathbf{R})$, where $\mathbf{R}$ is the distance from the origin. Thus we must solve the equation

$$\Delta\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = -4\pi\, de(t)\ \delta(\mathbf{R}). \tag{62.5}$$

Everywhere, except at the origin, $\delta(\mathbf{R}) = 0$, and we have the equation

$$\Delta\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = 0. \tag{62.6}$$

It is clear that in the case we are considering ϕ has central symmetry, i.e. ϕ is a function only of R. Therefore if we write the Laplace operator in spherical coordinates, (62.6) reduces to

$$\frac{1}{R^2}\frac{\partial}{\partial R}\left(R^2\frac{\partial\phi}{\partial R}\right) - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = 0.$$

To solve this equation, we make the substitution $\phi = \chi(R, t)/R$. Then, we find for χ

$$\frac{\partial^2\chi}{\partial R^2} - \frac{1}{c^2}\frac{\partial^2\chi}{\partial t^2} = 0.$$

But this is the equation of plane waves, whose solution has the form (see § 47):

$$\chi = f_1\left(t - \frac{R}{c}\right) + f_2\left(t + \frac{R}{c}\right).$$

Since we only want a particular solution of the equation, it is sufficient to choose only one of the functions f_1 and f_2. Usually it turns out to be convenient to take $f_2 = 0$ (concerning this, see below). Then, everywhere except at the origin, ϕ has the form

$$\phi = \frac{\chi\left(t - \frac{R}{c}\right)}{R}. \tag{62.7}$$

So far the function χ is arbitrary; we now choose it so that we also obtain the correct value for the potential at the origin. In other words, we must select χ so that at the origin equation (62.5) is satisfied. This is easily done noting that as $R \to 0$, the potential increases to infinity, and therefore its derivatives with respect to the coordinates increase more rapidly than its time derivative. Consequently as $R \to 0$, we can, in equation (62.5), neglect the term $(1/c^2)/(\partial^2\phi/\partial t^2)$ compared with $\Delta\phi$. Then (62.5) goes over into the familiar equation (36.9) leading to the Coulomb law. Thus, near the origin, (62.7) must go over into the Coulomb law, from which it follows that $\chi(t) = de(t)$, that is,

$$\phi = \frac{de\left(t - \frac{R}{c}\right)}{R}.$$

From this it is easy to get to the solution of equation (62.4) for an arbitrary distribution of charges $\varrho(x, y, z, t)$. To do this, it is sufficient to write $de = \varrho dV$ (dV is the volume element) and integrate over the whole space. To this solution of the inhomogeneous equation (62.4) we can still add the solution ϕ_0 of the same equation without the right-hand side. Thus, the general solution has the form:

$$\phi(\mathbf{r}, t) = \int \frac{1}{R} \varrho \left(\mathbf{r}', t - \frac{R}{c} \right) dV' + \phi_0, \tag{62.8}$$

$$\mathbf{R} = \mathbf{r} - \mathbf{r}', \qquad dV' = dx'\, dy'\, dz'$$

where

$$\mathbf{r} = (x, y, z), \qquad \mathbf{r}' = (x', y', z');$$

R is the distance from the volume element dV to the "field point" at which we determine the potential. We shall write this expression briefly as

$$\phi = \int \frac{\varrho_{t-(R/c)}}{R} dV + \phi_0, \tag{62.9}$$

where the subscript means that the quantity ϱ is to be taken at the time $t - (R/c)$, and the prime on dV has been omitted.

Similarly we have for the vector potential:

$$\mathbf{A} = \frac{1}{c} \int \frac{\mathbf{j}_{t-(R/c)}}{R} dV + \mathbf{A}_0, \tag{62.10}$$

where $\mathbf{A}_0$ is the solution of equation (62.3) without the right-hand term.

The potentials (62.9) and (62.10) (without ϕ_0 and $\mathbf{A}_0$) are called the retarded potentials.

In case the charges are at rest (i.e. density ρ independent of the time), formula (62.9) goes over into the well-known formula (36.8) for the electrostatic field; for the case of stationary motion of the charges, formula (62.10), after averaging, goes over into formula (43.5) for the vector potential of a constant magnetic field.

The quantities $\mathbf{A}_0$ and ϕ_0 in (62.9) and (62.10) are to be determined so that the conditions of the problem are fulfilled. To do this it is clearly sufficient to impose initial conditions, that is, to fix the values of the field at the initial time. However we do not usually have to deal with such initial conditions. Instead we are usually given conditions at large distances form the system of charges throughout all of time. Thus, we may be told that radiation is incident on the system from outside. Corresponding to this, the field which is developed as a result of the interaction of this radiation with the system can differ from the external field only by the radiation originating from the system. This radiation emitted by the system must, at large distances, have the form of waves spreading out from the system, that is, in the direction of increasing R. But precisely this condition is satisfied by the retarded potentials. Thus these solutions represent the field produced by the system, while ϕ_0 and $\mathbf{A}_0$ must be set equal to the external field acting on the system.

§ 63. The Lienard–Wiechert potentials

Let us determine the potentials for the field produced by a charge carrying out an assigned motion along a trajectory $\mathbf{r} = \mathbf{r}_0(t)$.

According to the formulas for the retarded potentials, the field at the point of observation $P(x, y, z)$ at time t is determined by the state of motion of the charge at the earlier time t', for which the time of propagation of the light signal from the point $\mathbf{r}_0(t')$, where the charge was located, to the field point P just coincides with the difference $t - t'$. Let $\mathbf{R}(t) = \mathbf{r} - \mathbf{r}_0(t)$ be the radius vector from the charge e to the point P; like $\mathbf{r}_0(t)$ it is a given function of the time. Then the time t' is determined by the equation

$$t' + \frac{R(t')}{c} = t. \tag{63.1}$$

For each value of t this equation has just one root t'.†

In the system of reference in which the particle is at rest at time t', the potential at the point of observation at time t is just the Coulomb potential,

$$\phi = \frac{e}{R(t')}, \qquad \mathbf{A} = 0. \tag{63.2}$$

The expressions for the potentials in an arbitrary reference system can be found directly by finding a four-vector which for $\mathbf{v} = 0$ coincides with the expressions just given for ϕ and $\mathbf{A}$. Noting that, according to (63.1), ϕ in (63.2) can also be written in the form

$$\phi = \frac{e}{c(t - t')},$$

we find that the required four-vector is:

$$A^i = e\frac{u^i}{R_k u^k}, \tag{63.3}$$

where u^k is the four-velocity of the charge, $R^k = [c(t - t'), \mathbf{r} - \mathbf{r}']$, where x', y', z', t' are related by the equation (63.1), which in four-dimensional form is

$$R_k R^k = 0. \tag{63.4}$$

Now once more transforming to three-dimensional notation, we obtain, for the potentials of the field produced by an arbitrarily moving point charge, the following expressions:

$$\phi = \frac{e}{\left(R - \frac{\mathbf{v} \cdot \mathbf{R}}{c}\right)}, \qquad \mathbf{A} = \frac{e\mathbf{v}}{c\left(R - \frac{\mathbf{v} \cdot \mathbf{R}}{c}\right)}, \tag{63.5}$$

where $\mathbf{R}$ is the radius vector, taken from the point where the charge is located to the point of observation P, and all the quantities on the right sides of the equations must be evaluated at the time t', determined from (63.1). The potentials of the field, in the form (63.5), are called the *Lienard–Wiechert potentials.*

† This point is obvious but it can be verified directly. To do this we choose the field point P and the time of observation t as the origin O of the four-dimensional coordinate system and construct the light cone (§ 2) with its vertex at O. The lower half of the cone, containing the absolute past (with respect to the event O), is the geometrical locus of world points such that signals sent from them reach O. The points in which this hypersurface intersects the world line of the charge are precisely the roots of (63.1). But since the velocity of a particle is always less than the velocity of light, the inclination of its world line relative to the time axis is everywhere less than the slope of the light cone. It then follows that the world line of the particle can intersect the lower half of the light cone in only one point.

To calculate the intensities of the electric and magnetic fields from the formulas

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} - \operatorname{grad}\phi, \qquad \mathbf{H} = \operatorname{curl}\mathbf{A},$$

we must differentiate ϕ and $\mathbf{A}$ with respect to the coordinates x, y, z of the point, and the time t of observation. But the formulas (63.5) express the potentials as functions of t', and only through the relation (63.1) as implicit functions of x, y, z, t. Therefore to calculate the required derivatives we must first calculate the derivatives of t'. Differentiating the relation $R(t') = c(t - t')$ with respect to t, we get

$$\frac{\partial R}{\partial t} = \frac{\partial R}{\partial t'}\frac{\partial t'}{\partial t} = -\frac{\mathbf{R}\cdot\mathbf{v}}{R}\frac{\partial t'}{\partial t} = c\left(1 - \frac{\partial t'}{\partial t}\right).$$

(The value of $\partial R/\partial t'$ is obtained by differentiating the identity $R^2 = \mathbf{R}^2$ and substituting $\partial\mathbf{R}(t')/\partial t' = -\mathbf{v}(t')$. The minus sign is present because $\mathbf{R}$ is the radius vector from the charge e to the point P, and not the reverse.)

Thus,

$$\frac{\partial t'}{\partial t} = \frac{1}{1 - \dfrac{\mathbf{v}\cdot\mathbf{R}}{Rc}}. \tag{63.6}$$

Similarly differentiating the same relation with respect to the coordinates, we find

$$\operatorname{grad} t' = -\frac{1}{c}\operatorname{grad} R(t') = -\frac{1}{c}\left(\frac{\partial R}{\partial t'}\operatorname{grad} t' + \frac{\mathbf{R}}{R}\right),$$

so that

$$\operatorname{grad} t' = -\frac{\mathbf{R}}{c\left(R - \dfrac{\mathbf{R}\cdot\mathbf{v}}{c}\right)}. \tag{63.7}$$

With the aid of these formulas, there is no difficulty in carrying out the calculation of the fields $\mathbf{E}$ and $\mathbf{H}$. Omitting the intermediate calculations, we give the final results:

$$\mathbf{E} = e\frac{1 - \dfrac{v^2}{c^2}}{\left(R - \dfrac{\mathbf{R}\cdot\mathbf{v}}{c}\right)^3}\left(\mathbf{R} - \frac{\mathbf{v}}{c}R\right) + \frac{e}{c^2\left(R - \dfrac{\mathbf{R}\cdot\mathbf{v}}{c}\right)^3}\mathbf{R}\times\left\{\left(\mathbf{R} - \frac{\mathbf{v}}{c}R\right)\times\dot{\mathbf{v}}\right\}, \tag{63.8}$$

$$\mathbf{H} = \frac{1}{R}\mathbf{R}\times\mathbf{E}. \tag{63.9}$$

Here, $\dot{\mathbf{v}} = \partial\mathbf{v}/\partial t'$; all quantities on the right sides of the equations refer to the time t'. It is interesting to note that the magnetic field turns out to be everywhere perpendicular to the electric.

The electric field (63.8) consists of two parts of different type. The first term depends only on the velocity of the particle (and not on its acceleration) and varies at large distances like $1/\mathbf{R}^2$. The second term depends on the acceleration, and for large R it varies like $1/R$. Later

(§ 66) we shall see that this latter term is related to the electromagnetic waves radiated by the particle.

As for the first term, since it is independent of the acceleration it must correspond to the field produced by a uniformly moving charge. In fact, for constant velocity the difference

$$\mathbf{R}_{t'} - \frac{\mathbf{v}}{c} R_{t'} = \mathbf{R}_{t'} - \mathbf{v}(t - t')$$

is the distance $\mathbf{R}_t$ from the charge to the point of observation at precisely the moment of observation. It is also easy to show directly that

$$R_{t'} - \frac{1}{c} \mathbf{R}_{t'} \cdot \mathbf{v} = \sqrt{R_t^2 - \frac{1}{c^2} (\mathbf{v} \times \mathbf{R}_t)^2} = R_t \sqrt{1 - \frac{v^2}{c^2} \sin^2 \theta_t},$$

where θ_t is the angle between $\mathbf{R}_t$ and $\mathbf{v}$. Consequently the first term in (63.8) is identical with the expression (38.8).

PROBLEM

Derive the Lienard–Wiechert potentials by integrating (62.9)–(62.10).

Solution: We write formula (62.8) in the form:

$$\phi(\mathbf{r}, t) = \iint \frac{\varrho(\mathbf{r}', \tau)}{|\mathbf{r} - \mathbf{r}'|} \delta\left(\tau - t + \frac{1}{c} |\mathbf{r} - \mathbf{r}'|\right) d\tau dV'$$

(and similarly for $\mathbf{A}(\mathbf{r}, t)$), introducing the additional delta function and thus eliminating the implicit arguments in the function ϱ. For a point charge, moving in a trajectory $\mathbf{r} = \mathbf{r}_0(t)$; we have:

$$\varrho(\mathbf{r}', \tau) = e\delta[\mathbf{r}' - \mathbf{r}_0(\tau)].$$

Substituting this expression and integrating over dV', we get:

$$\phi(\mathbf{r}, t) = e \int \frac{d\tau}{|\mathbf{r} - \mathbf{r}_0(\tau)|} \delta\left[\tau - t + \frac{1}{c} |\mathbf{r} - \mathbf{r}_0(\tau)|\right],$$

The τ integration is done using the formula

$$\delta[F(\tau)] = \frac{\delta(\tau - t')}{F'(t')}$$

[where t' is the root of $F(t') = 0$], and gives formula (63.5).

§ 64. Spectral resolution of the retarded potentials

The field produced by moving charges can be expanded into monochromatic waves. The potentials of the different monochromatic components of the field have the form $\phi_\omega e^{-i\omega t}$, $\mathbf{A}_\omega e^{-i\omega t}$. The charge and current densities of the system of charges producing the field can also be expanded in a Fourier series or integral. It is clear that each Fourier component of ϱ and $\mathbf{j}$ is responsible for the creation of the corresponding monochromatic component of the field.

In order to express the Fourier components of the field in terms of the Fourier components of the charge density and current, we substitute in (62.9) for ϕ and ϱ respectively, $\phi_\omega e^{-i\omega t}$, and $\varrho_\omega e^{-i\omega t}$. We then obtain

$$\phi_\omega e^{-i\omega t} = \int \varrho_\omega \frac{e^{-i\omega\left(t-\frac{R}{c}\right)}}{R} dV.$$

Factoring $e^{-i\omega t}$ and introducing the absolute value of the wave vector $k = \omega/c$, we have:

$$\phi_\omega = \int \varrho_\omega \frac{e^{ikR}}{R} dV. \tag{64.1}$$

Similarly, for $\mathbf{A}_\omega$ we get

$$\mathbf{A}_\omega = \int \mathbf{j}_\omega \frac{e^{ikR}}{cR} dV. \tag{64.2}$$

We note that formula (64.1) represents a generalization of the solution of the Poisson equation to a more general equation of the form

$$\Delta\phi_\omega + k^2\phi_\omega = -4\pi\varrho_\omega \tag{64.3}$$

(obtained from equations (62.4) for ϱ, ϕ depending on the time through the factor $e^{-i\omega t}$).

If we were dealing with expansion into a Fourier integral, then the Fourier components of the charge density would be

$$\varrho_\omega = \int_{-\infty}^{+\infty} \varrho e^{i\omega t} dt.$$

Substituting this expression in (64.1), we get

$$\phi_\omega = \iint_{-\infty}^{+\infty} \frac{\varrho}{R} e^{i(\omega t + kR)} dV\, dt. \tag{64.4}$$

We must still go over from the continuous distribution of charge density to the point charges whose motion we are actually considering. Thus, if there is just one point charge, we set

$$\varrho = e\delta[\mathbf{r} - \mathbf{r}_0(t)],$$

where $\mathbf{r}_0(t)$ is the radius vector of the charge, and is a given function of the time. Substituting this expression in (64.4) and carrying out the space integration [which reduces to replacing $\mathbf{r}$ by $\mathbf{r}_0(t)$], we get:

$$\phi_\omega = e \int_{-\infty}^{\infty} \frac{1}{R(t)} e^{i\omega[t+R(t)/c]} dt, \tag{64.5}$$

where now $R(t)$ is the distance from the moving particle to the point of observation. Similarly we find for the vector potential:

$$\mathbf{A}_\omega = \frac{e}{c} \int_{-\infty}^{\infty} \frac{\mathbf{v}(t)}{R(t)} e^{i\omega[t+R(t)/c]} dt, \tag{64.6}$$

where $\mathbf{v} = \dot{\mathbf{r}}_0(t)$ is the velocity of the particle.

Formulas analogous to (64.5), (64.6) can also be written for the case where the spectral resolution of the charge and current densities contains a discrete series of frequencies. Thus, for a periodic motion of a point charge (with period $T = 2\pi/\omega_0$) the spectral resolution of the field contains only frequencies of the form $n\omega_0$, and the corresponding components of the vector potential are

$$\mathbf{A}_n = \frac{e}{cT}\int_0^T \frac{\mathbf{v}(t)}{R(t)}\, e^{in\omega_0[t+R(t)/c]}\, dt \tag{64.7}$$

(and similarly for ϕ_n). In both (64.6) and (64.7) the Fourier components are defined in accordance with § 49.

PROBLEM

Find the expansion in plane waves of the field of a charge in uniform rectilinear motion.

Solution: We proceed in similar fashion to that used in § 51. We write the charge density in the form $\varrho = e\delta(\mathbf{r} - \mathbf{v}t)$, where $\mathbf{v}$ is the velocity of the particle. Taking Fourier components of the equation $\Box\phi = -4\pi e\,\delta(\mathbf{r} - \mathbf{v}t)$, we find $(\Box\phi)_{\mathbf{k}} = -4\pi e\, e^{-i(\mathbf{v}\cdot\mathbf{k})t}$.

On the other hand, from

$$\phi = \int e^{i\mathbf{k}\cdot\mathbf{r}}\phi_{\mathbf{k}}\,\frac{d^3k}{(2\pi)^3}$$

we have

$$(\Box\phi)_{\mathbf{k}} = -k^2\phi_{\mathbf{k}} - \frac{1}{c^2}\frac{\partial^2\phi_{\mathbf{k}}}{\partial t^2}.$$

Thus,

$$\frac{1}{c^2}\frac{\partial^2\phi_{\mathbf{k}}}{\partial t^2} + k^2\phi_{\mathbf{k}} = 4\pi e e^{-i(\mathbf{k}\cdot\mathbf{v})t},$$

from which, finally

$$\phi_{\mathbf{k}} = 4\pi e\,\frac{e^{-i(\mathbf{k}\cdot\mathbf{v})t}}{k^2 - \left(\frac{\mathbf{k}\cdot\mathbf{v}}{c}\right)^2}.$$

From this it follows that the wave with wave vector $\mathbf{k}$ has the frequency $\omega = \mathbf{k}\cdot\mathbf{v}$. Similarly, we obtain for the vector potential,

$$\mathbf{A}_{\mathbf{k}} = 4\pi e\,\frac{\mathbf{v}e^{-i(\mathbf{k}\cdot\mathbf{v})t}}{k^2 - \left(\frac{\mathbf{k}\cdot\mathbf{v}}{c}\right)^2}.$$

Finally, we have for the fields,

$$\mathbf{E}_{\mathbf{k}} = -i\mathbf{k}\phi_{\mathbf{k}} + i\frac{\mathbf{k}\cdot\mathbf{v}}{c}\mathbf{A}_{\mathbf{k}} = 4\pi e\, i\,\frac{-\mathbf{k} + \frac{(\mathbf{k}\cdot\mathbf{v})}{c^2}\mathbf{v}}{k^2 - \left(\frac{\mathbf{k}\cdot\mathbf{v}}{c}\right)^2}\, e^{-i(\mathbf{k}\cdot\mathbf{v})t},$$

$$\mathbf{H}_{\mathbf{k}} = i\mathbf{k} \times \mathbf{A}_{\mathbf{k}} = \frac{4\pi e}{c} i \frac{\mathbf{k} \times \mathbf{v}}{k^2 - \left(\frac{\mathbf{k} \cdot \mathbf{v}}{c}\right)^2} e^{-i(\mathbf{k} \cdot \mathbf{v})t}.$$

§ 65. The Lagrangian to terms of second order

In ordinary classical mechanics, we can describe a system of particles interacting with each other with the aid of a Lagrangian which depends only on the coordinates and velocities of these particles (at one and the same time). The possibility of doing this is, in the last analysis, dependent on the fact that in mechanics the velocity of propagation of interactions is assumed to be infinite.

We already know that because of the finite velocity of propagation, the field must be considered as an independent system with its own "degrees of freedom". From this it follows that if we have a system of interacting particles (charges), then to describe it we must consider the system consisting of these particles and the field. Therefore, when we take into account the finite velocity of propagation of interactions, it is impossible to describe the system of interacting particles rigorously with the aid of a Lagrangian, depending only on the coordinates and velocities of the particles and containing no quantities related to the internal "degrees of freedom" of the field.

However, if the velocity v of all the particles is small compared with the velocity of light, then the system can be described by a certain approximate Lagrangian. It turns out to be possible to introduce a Lagrangian describing the system, not only when all powers of v/c are neglected (classical Lagrangian), but also to terms of second order, v^2/c^2. This last remark is related to the fact that the radiation of electromagnetic waves by moving charges (and consequently, the appearance of a "self"-field) occurs only in the third approximation in v/c (see later, in § 67).†

As a preliminary, we note that in zero'th approximation, that is, when we completely neglect the retardation of the potentials, the Lagrangian for a system of charges has the form

$$L^{(0)} = \sum_a \frac{1}{2} m_a \mathbf{v}_a^2 - \sum_{a>b} \frac{e_a e_b}{R_{ab}} \tag{65.1}$$

(the summation extends over the charges which make up the system). The second term is the potential energy of interaction as it would be for charges at rest.

To get the next approximation, we proceed in the following fashion. The Lagrangian for a charge e_a in an external field is

$$L_a = -m_a c^2 \sqrt{1 - \frac{v_a^2}{c^2}} - e_a \phi + \frac{e_a}{c} \mathbf{A} \cdot \mathbf{v}_a. \tag{65.2}$$

Choosing any one of the charges of the system, we determine the potentials of the field produced by all the other charges at the position of the first, and express them in terms of the coordinates and velocities of the charges which produce this field (this can be done only approximately—for ϕ, to terms of order v^2/c^2, and for $\mathbf{A}$, to terms in v/c). Substituting the expressions for the potentials obtained in this way in (65.2), we get the Lagrangian for one

† For systems consisting of particles with the same charge-to-mass ratio, the appearance of radiation is put off to the fifth approximation in v/c; in such a case there is a Lagrangian to terms of fourth order in v/c. [See B.M. Barker and R.F. O'Connel, *Can. J. Phys.* **58**, 1659 (1980).]

of the charges of the system (for a given motion of the other charges). From this, one can then easily find the Lagrangian for the whole system.

We start from the expressions for the retarded potentials

$$\phi = \int \frac{\varrho_{t-R/c}}{R}\, dV, \qquad \mathbf{A} = \frac{1}{c}\int \frac{\mathbf{j}_{t-R/c}}{R}\, dV.$$

If the velocities of all the charges are small compared with the velocity of light, then the charge distribution does not change significantly during the time R/c. Therefore we can expand $\rho_{t-R/c}$ and $\mathbf{j}_{t-R/c}$ in series of powers of R/c. For the scalar potential we thus find, to terms of second order:

$$\phi = \int \frac{\varrho dV}{R} - \frac{1}{c}\frac{\partial}{\partial t}\int \varrho dV + \frac{1}{2c^2}\frac{\partial^2}{\partial t^2}\int R\varrho dV$$

(ϱ without indices is the value of ϱ at time t; the time differentiations can clearly be taken out from under the integral sign). But $\int \varrho dV$ is the constant total charge of the system. Therefore the second term in our expression is zero, so that

$$\phi = \int \frac{\varrho dV}{R} + \frac{1}{2c^2}\frac{\partial^2}{\partial t^2}\int R\varrho dV. \tag{65.3}$$

We can proceed similarly with $\mathbf{A}$. But the expression for the vector potential in terms of the current density already contains $1/c$, and when substituted in the Lagrangian is multiplied once more by $1/c$. Since we are looking for a Lagrangian which is correct only to terms of second order, we can limit ourselves to the first term in the expansion of $\mathbf{A}$, that is,

$$\mathbf{A} = \frac{1}{c}\int \frac{\varrho \mathbf{v}}{R}\, dV \tag{65.4}$$

(we have substituted $\mathbf{j} = \varrho\mathbf{v}$).

Let us first assume that there is only a single point charge e. Then we obtain from (65.3) and (65.4),

$$\phi = \frac{e}{R} + \frac{e}{2c^2}\frac{\partial^2 R}{\partial t^2}, \qquad \mathbf{A} = \frac{e\mathbf{v}}{cR}, \tag{65.5}$$

where R is the distance from the charge.

We choose in place of ϕ and $\mathbf{A}$ other potentials ϕ' and $\mathbf{A}'$, making the transformation (see § 18):

$$\phi' = \phi - \frac{1}{c}\frac{\partial f}{\partial t}, \qquad \mathbf{A}' = \mathbf{A} + \operatorname{grad} f,$$

in which we choose for f the function

$$f = \frac{e}{2c}\frac{\partial R}{\partial t}.$$

Then we get†

† These potentials no longer satisfy the Lorentz condition (62.1), nor the equations (62.3)–(62.4).

$$\phi' = \frac{e}{R}, \qquad \mathbf{A}' = \frac{e\mathbf{v}}{cR} + \frac{e}{2c}\nabla\frac{\partial R}{\partial t}.$$

To calculate $\mathbf{A}'$ we note first of all that $\nabla(\partial R/\partial t) = (\partial/\partial t)\nabla R$. The grad operator here means differentiation with respect to the coordinates of the field point at which we seek the value of $\mathbf{A}'$. Therefore ∇R is the unit vector $\mathbf{n}$ directed from the charge e to the field point, so that

$$\mathbf{A}' = \frac{e\mathbf{v}}{cR} + \frac{e}{2c}\dot{\mathbf{n}}.$$

We also write:

$$\dot{\mathbf{n}} = \frac{\partial}{\partial t}\left(\frac{\mathbf{R}}{R}\right) = \frac{\dot{\mathbf{R}}}{R} - \frac{\mathbf{R}\dot{R}}{R^2}.$$

But the derivative $-\dot{\mathbf{R}}$ for a given field point is the velocity $\mathbf{v}$ of the charge, and the derivative $\dot{R}$ is easily determined by differentiating $R^2 = \mathbf{R}^2$, that is, by writing

$$R\dot{R} = \mathbf{R}\cdot\dot{\mathbf{R}} = -\mathbf{R}\cdot\mathbf{v}.$$

Thus,

$$\dot{\mathbf{n}} = \frac{-\mathbf{v} + \mathbf{n}(\mathbf{n}\cdot\mathbf{v})}{R}.$$

Substituting this in the expression for $\mathbf{A}'$, we get finally:

$$\phi' = \frac{e}{R}, \qquad \mathbf{A}' = \frac{e[\mathbf{v} + (\mathbf{v}\cdot\mathbf{n})\mathbf{n}]}{2cR}. \tag{65.6}$$

If there are several charges then we must, clearly, sum these expressions over all the charges.

Substituting these expressions in (65.2), we obtain the Lagrangian L_a for the charge e_a (for a fixed motion of the other charges). In doing this we must also expand the first term in (65.2) in powers of v_a/c, retaining terms up to the second order. Thus we find:

$$L_a = \frac{m_a v_a^2}{2} + \frac{1}{8}\frac{m_a v_a^4}{c^2} - e_a \sum_b{}' \frac{e_b}{R_{ab}} + \frac{e_a}{2c^2}\sum_b{}' \frac{e_b}{R_{ab}}[\mathbf{v}_a\cdot\mathbf{v}_b + (\mathbf{v}_a\cdot\mathbf{n}_{ab})(\mathbf{v}_b\cdot\mathbf{n}_{ab})]$$

(the summation goes over all the charges except e_a; $\mathbf{n}_{ab}$ is the unit vector from e_b to e_a).

From this, it is no longer difficult to get the Lagrangian for the whole system. It is easy to convince oneself that this function is not the sum of the L_a for all the charges, but has the form

$$L = \sum_a \frac{m_a v_a^2}{2} + \sum_a \frac{m_a v_a^4}{8c^2} - \sum_{a>b}\frac{e_a e_b}{R_{ab}} + \sum_{a>b}\frac{e_a e_b}{2c^2 R_{ab}}[\mathbf{v}_a\cdot\mathbf{v}_b + (\mathbf{v}_a\cdot\mathbf{n}_{ab})(\mathbf{v}_b\cdot\mathbf{n}_{ab})]. \tag{65.7}$$

Actually, for each of the charges under a given motion of all the others, this function L goes over into L_a as given above. The expression (65.7) determines the Lagrangian of a system of charges correctly to terms of second order. (It was first obtained by C. G. Darwin, 1922.)

Finally we find the Hamiltonian of a system of charges in this same approximation. This could be done by the general rule for calculating $\mathcal{H}$ from L; however it is simpler to proceed as follows. The second and fourth terms in (65.7) are small corrections to $L^{(0)}$(65.1). On the

other hand, we know from mechanics that for small changes of L and $\mathscr{H}$, the additions to them are equal in magnitude and opposite in sign (here the variations of L are considered for constant coordinates and velocities, while the changes in $\mathscr{H}$ refer to constant coordinates and momenta).†

Therefore we can at once write $\mathscr{H}$, subtracting from

$$\mathscr{H}^{(0)} = \sum_a \frac{p_a^2}{2m_a} + \sum_{a>b} \frac{e_a e_b}{R_{ab}}$$

the second and fourth terms of (65.7), replacing the velocities in them by the first approximation $\mathbf{v}_a = \mathbf{p}_a/m_a$. Thus,

$$\mathscr{H} = \sum_a \frac{p_a^2}{2m_a} - \sum_a \frac{p_a^4}{8c^2 m_a^3} + \sum_{a>b} \frac{e_a e_b}{R_{ab}} -$$

$$- \sum_{a>b} \frac{e_a e_b}{2c^2 m_a m_b R_{ab}} [\mathbf{p}_a \cdot \mathbf{p}_b + (\mathbf{p}_a \cdot \mathbf{n}_{ab})(\mathbf{p}_b \cdot \mathbf{n}_{ab})]. \tag{65.8}$$

PROBLEMS

1. Determine (correctly to terms of second order) the centre of inertia of a system of interacting particles.

Solution: The problem is solved most simply by using the formula

$$\mathbf{R} = \frac{\sum_a \mathscr{E}_a \mathbf{r}_a + \int W\mathbf{r}\, dV}{\sum_a \mathscr{E}_a + \int W\, dV}$$

[see (14.6)], where $\mathscr{E}_a$ is the kinetic energy of the particle (including its rest energy), and W is the energy density of the field produced by the particles. Since the $\mathscr{E}_a$ contain the large quantities $m_a c^2$, it is sufficient, in obtaining the next approximation, to consider only those terms in $\mathscr{E}_a$ and W which do not contain c, i.e. we need consider only the nonrelativistic kinetic energy of the particles and the energy of the electrostatic field. We then have:

$$\int W\mathbf{r}\, dV = \frac{1}{8\pi} \int E^2 \mathbf{r}\, dV$$

$$= \frac{1}{8\pi} \int (\nabla\varphi)^2 \mathbf{r}\, dV$$

$$= \frac{1}{8\pi} \int \left(d\mathbf{f} \cdot \nabla \frac{\varphi^2}{2} \right) \mathbf{r} - \frac{1}{8\pi} \int \nabla \frac{\varphi^2}{2}\, dV - \frac{1}{8\pi} \int \varphi \Delta\varphi \cdot \mathbf{r}\, dV;$$

the integral over the infinitely distant surface vanishes; the second integral also is transformed into a surface integral and vanishes, while we substitute $\Delta\varphi = -4\pi\varrho$ in the third integral and obtain:

$$\int W\mathbf{r}\, dV = \frac{1}{2} \int \rho\varphi\mathbf{r}\, dV = \frac{1}{2} \sum_a e_a \varphi_a \mathbf{r}_a,$$

† See *Mechanics*, § 40.

where φ_a is the potential produced at the point $\mathbf{r}_a$ by all the charges other than e_a.†

Finally, we get:

$$\mathbf{R} = \frac{1}{\mathscr{E}} \sum_a \mathbf{r}_a \left(m_a c^2 + \frac{{p_a}^2}{2m_a} + \frac{e_a}{2} \sum_b{}' \frac{e_b}{R_{ab}} \right)$$

(with a summation over all b except $b = a$), where

$$\mathscr{E} = \sum_a \left(m_a c^2 + \frac{p_a^2}{2m_a} + \sum_{a>b} \frac{e_a e_b}{R_{ab}} \right)$$

is the total energy of the system. Thus in this approximation the coordinates of the centre of inertia can actually be expressed in terms of quantities referring only to the particles.

2. Write the Hamiltonian in second approximation for a system of two particles, omitting the motion of the system as a whole.

Solution: We choose a system of reference in which the total momentum of the two particles is zero. Expressing the momenta as derivatives of the action, we have

$$\mathbf{p}_1 + \mathbf{p}_2 = \partial S/\partial \mathbf{r}_1 + \partial S/\partial \mathbf{r}_2 = 0.$$

From this it is clear that in the reference system chosen the action is a function of $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$, the difference of the radius vectors of the two particles. Therefore we have $\mathbf{p}_2 = -\mathbf{p}_1 = \mathbf{p}$, where $\mathbf{p} = \partial S/\partial \mathbf{r}$ is the momentum of the relative motion of the particles. The Hamiltonian is

$$\mathscr{H} = \frac{1}{2}\left(\frac{1}{m_1} + \frac{1}{m_2}\right)p^2 - \frac{1}{8c^2}\left(\frac{1}{{m_1}^3} + \frac{1}{{m_2}^3}\right)p^4 + \frac{e_1 e_2}{r} + \frac{e_1 e_2}{2m_1 m_2 c^2 r}[p^2 + (\mathbf{p}\cdot\mathbf{n})^2].$$

† The elimination of the self-field of the particles corresponds to the mass "renormalization" mentioned in the footnote on p. 97).