The experimental basis of electromagnetism: the direct-current circuit

1946 Proc. Phys. Soc. 58 634

(http://iopscience.iop.org/0959-5309/58/6/302)

THE EXPERIMENTAL BASIS OF ELECTRO-MAGNETISM: THE DIRECT-CURRENT CIRCUIT

By NORMAN R. CAMPBELL AND L. HARTSHORN

MS. received 25 *March* 1946

ABSTRACT. The purpose of the enquiry, of which this paper forms the first part, is to show to what extent the working principles of electromagnetism can be soundly based on real experimental facts—that is to say, on experiments that have actually been performed, as distinct from the imaginary experiments, which are common to most expositions of the subject, but which are either quite impracticable or incapable of being performed with an accuracy that would be regarded as significant to-day. By maintaining this sharp distinction between fact and theory we aim at removing from the subject much of the confusion which, as the literature shows, is a continual source of trouble.

This section begins with an outline of the general principles of measurement. Current, resistance, conductance and voltage are then established independently of a knowledge of any other magnitudes. Ohm's Law and the conception of e.m.f. follow. An examination of the relations between these magnitudes and the geometrical and mechanical magnitudes then leads to theoretical conceptions like potential, and to a consideration of the status of Ampère's two laws. That for the force between two circuits is well established and defines the unit of current. That for the torque on a compass needle near a circuit is less well established; but it leads to the recognition of H in a non-magnetic medium as a defined magnitude, whose significance is found to be independent of the existence of magnets.

§ 1. INTRODUCTORY

A REMARKABLE feature of the literature of the last decade is the number of papers that have been published on the fundamental magnitudes of classical physics. Even more remarkable is the divergence of opinion expressed on such well worn themes as the precise significance of the magnitudes of everyday electromagnetism. The experimentalist must occasionally ask himself whether ideas that have been successfully applied in an ever-widening field for so many years can really rest on such uncertain foundations.

There seems to be fairly general agreement that the working concepts are defined by the processes adopted for their measurement, but there is no sort of agreement as to what is the essential character of these processes. Some magnitudes appear to be associated with many processes, and there is no agreement as to their relative status.

Many difficulties have arisen from the fact that the basis of the subject has almost invariably been sought in processes that are not really practicable—that is to say, in theoretical experiments rather than in experimental facts. Experiments with point charges and magnet poles are well known examples. Every experimentalist soon learns that, whatever part such experiments may have played in the history of the subject, the best approximation he could make to them would to-day be considered quite worthless as evidence in support of his working principles. He realizes also that disputes about such experiments are almost meaningless in relation to science as practised in the laboratory. The very fact that all the doubt and discussion about these theoretical ideas have had no discernible effect on the course of experimental enquiry is sufficient to show that there is no very close connection between the two. The world-wide agreement among physicists in practical matters shows that we must be working on some common basis. It is curious that it should be so difficult to express it in terms on which we are all agreed. The present authors are of the opinion that the difficulty can be overcome by considering those processes only that are actually employed in practice; and in this paper we propose to show that all the working laws of electromagnetism can be soundly based on experimental operations that will be generally recognized as forming part of present-day technique. Every physicist will agree that the basis of the subject is to be found in experimental facts. If, therefore, we can derive our working principles from experiments that are frequently performed in the normal course of physical enquiry, or that are at least representative of existing technique, we shall have established our basis clear of theoretical ambiguities.

Our enquiry shows that the basic experimental laws are not those which are given most prominence in the customary expositions of the subject. It is therefore the more desirable that their importance should be generally recognized, and particularly so in the work of establishing units and standards. The work of the various international conferences has been of great practical value in securing international uniformity, and it is of the utmost importance that it shall not become confused with irrelevant theoretical considerations. The international committees are primarily concerned with securing agreement on the basic facts: their decisions will only have their proper effect on scientific practice so long as the relation between these facts and our working principles is clearly understood.

§ 2. GENERAL PRINCIPLES

By saying that the laws are to be soundly based on the experiments we mean that the development must follow an order that may be described as logical, in the broad sense in which that term is normally used in scientific investigations: we must not assume the truth of any proposition until it has been established. Thus, we must not assume that we can establish any algebraic law relating different properties until we have shown that these properties are measurable by processes that do not depend on the truth of that law.

It follows that we cannot begin with any algebraic law. The subject necessarily starts with qualitative observations, which first become quantitative when we have devised an operation whereby we can make consistent judgments of equality in respect of some observed property, and another operation which possesses the characteristics of addition. We can then, by successive addition, build up a standard scale for that property, and make measurements by establishing judgments of equality between the things measured and determinate parts of the standard scale. The measurement of length is a very familiar example of processes of this type, but it is not generally realized that the basic measurements of electrical properties are of the same kind.

Having established in this way the independent measurement of at least two properties, we are in a position to establish algebraic laws relating them—that is to say, relations between our observations of these two quantities that can be represented by algebraic equations. Such independent observations can never, except by the rarest fluke, show a relation of equality: they may, however, prove proportionality. It follows that the simplest algebraic law that we can establish will take the form $y = kx$, where x and y represent the two magnitudes that are independently measurable and k is a constant of proportionality which can be determined for any given system of the class to which the law applies from the measured values of x and y. We may now find that laws of the same form apply to other classes of similar systems, but that the constant k varies from one class to another, and is characteristic of some recognizable feature of the class. Thus k becomes established as a measure of this new property, which may be described as measurable by application of the law $y = kx$ to direct measurements of x and y. Such processes of derived measurement enable us to establish as magnitudes properties that are not additive, such as density and resistivity, and that may be described as *qualities*, to distinguish them from the additive properties or *quantities*. More complicated algebraic laws may take the form

$$y = kf(x_1, x_2, \ldots x_n),$$

where $f(x_1, x_2, \ldots x_n)$ denotes some definite function of any magnitudes $x_1 \ldots x_n$ that have already been established by either independent or derived measurement. We may regard $f(x_1, x_2 \ldots x_n)$ as a new magnitude z, which we shall call a *defined magnitude*. From an experimental point of view the law reduces to one of proportionality as before.

The value of the constant k in the above laws will obviously depend on the choice of units for x and y, as well as on the property which has been recognized as a derived magnitude. We may symbolize this fact by writing the equation in the form

$$y = S.\ a.\ f(x_1, x_2 \ldots x_n),$$

where a denotes the derived magnitude and S is the *scale factor*, which varies with the choice of units only. Sometimes the constant k turns out to be the same for every system to which the law is applicable. There is then no derived magnitude, for we always use the term " magnitude " to mean something that varies from one measurable system to another. In that case there is no need for our immediate purpose to split k into two factors; it can be regarded simply as a scale factor S; and if we decide to adjust our units of the xs and ys so that this S becomes unity, we can simplify the equation to the form $y = f(x)$. Even when k varies with the system, and a derived magnitude has to be recognized, we can always reduce our equation to this simple form for one particular class of systems by adjusting the units of the xs and ys so that $S . a$ becomes unity for this class. In other words, we adjust the units of the xs and ys until S becomes unity, and define the unit of a as the property possessed by that particular class. Examples will appear later.

We shall follow the customary practice of suppressing scale factors as far as we can consistently do so. If we have established m magnitudes, and n experimental relations between them, then by suitably choosing the units of the m magnitudes we can suppress m scale factors. If $n>m$, the remaining $(n-m)$ scale factors must appear as numerical factors in the equations. Whether derived magnitudes appear or not will depend only on whether such magnitudes have been established by the experiments.

However, it is sometimes desirable to introduce another symbol as well as the scale factor S, even when no derived magnitude has been experimentally established. For example, in a few instances the constancy of k has been established over a range of experiments so wide that it has acquired a special theoretical significance as a ***universal constant***. It is well to guard against the suppression of such a constant by separating it from the scale factor. Again, we may have theoretical reasons for thinking that k may be variable, even though we have not succeeded in obtaining conclusive experimental evidence of significant variation. In such a case we may introduce a factor besides the scale factor so as to leave open the question whether k is actually variable. Obviously we are at liberty to introduce as many factors as we find to be significant in either theory or experiment, so long as we bear in mind the limitations to the significance of each.

There remains for consideration the order in which the various magnitudes should be introduced, and in this connection we note that our algebraic laws must always remain indefinite in their application, to an extent that is represented by the experimental errors of the observations on which they are based. Moreover, it is a well known fact that when the measurement of any property depends on the measurements of others, the experimental error of the dependent measurement is always greater than that of any of the ones on which it depends. A process of measurement that has proved in practice to be significant with high precision cannot therefore be soundly based on a law that has only been established with inferior accuracy. For example, it is absurd to suppose that dielectric constant, a magnitude that has proved to be significant in a very large number of physico-chemical investigations with a precision better than 0·1%, is based on a law for the forces between small charged bodies which has never been established with anything like this accuracy; and it is even more absurd to suppose that voltage, a magnitude that is measured as a matter of routine by electric power companies with an accuracy of 0·01%, and in standardizing laboratories with a precision of 1 part per million, is to be defined in terms of the work done in moving an electric charge from one point to another, an operation about which our knowledge is incomparably more vague. We must therefore recognize that the familiar definitions, common to most expositions of electromagnetism, are the premises from which the mathematical reasoning proceeds, rather than definitions of the magnitudes with which we are concerned in experimental science; and in attempting to build up the subject from its experimental foundations we must begin with the independent measurement of the magnitudes that have been established with the least uncertainty, and then pass on to derived measurements based on them, always giving priority, in any one branch of the subject, to the experiments that are known to be capable of the highest accuracy.

This means that our first stage must include independent measurements of resistance; for there can be no doubt that measurements of resistance can be made with a higher accuracy, and over a greater range, than is attainable for any other property that is recognized as being distinctly electrical in character; and, moreover, the most precise measurements of resistance are in fact independent measurements. Qualitative observations of current are, however, a necessary preliminary; and as the independent measurement of current and voltage is on the same footing as that of resistance, it will be convenient to regard these three magnitudes together as comprising the first stage, and to begin with current and the idea of the electric circuit.

We shall take for granted the existence of all the instruments commonly employed in physical laboratories. From our standpoint they are merely common objects of the modern world, just as easily identified, and far more common, than the traditional amber, catskin, and lodestone, which, indeed, have proved, on closer investigation, to be no less complicated in their structure; and, what is more to our present purpose, far less amenable to precise observation.

§ 3. CURRENT: A MEASURABLE PROPERTY

Construct a circuit of a most general kind by connecting in series a battery, a variable resistor and, say, a filament lamp, a neon lamp, a water voltameter, and indicating instruments of all the common types—moving coil, moving magnet, thermal, electrostatic (with shunt resistor), and electro-dynamometer or current balance.

Varying the resistor, we observe that the indications of all these devices increase and decrease together; there is something characteristic of the whole collection, and not merely of each element of it. By connecting the instruments with different materials we discover the existence of conductors and insulators, and learn that the characteristic " something " is associated with a definite closed path or circuit that must be conducting throughout. Further observation reveals the following facts:—

To any definite indication of one instrument, there are corresponding definite indications of the others, and these corresponding indications are independent of the order in which the various instruments are connected in the circuit, provided that each instrument is always connected in the circuit in the same direction with reference to the rest of the circuit. The reversal of some instruments gives rise to a change in their indications: others are unaffected by reversal. We shall describe instruments of the first class as *direction-sensitive,* and those of the second as *reversible.*

These observations suggest an analogy with the flow of water, the battery being analogous to a pump, the resistor to a throttle, the closed conducting path to a pipe, and the indicators to rate-of-flow meters of various kinds. The analogy further suggests that if we construct more complex circuits, in which several branches lead into a main conductor, the characteristic feature " current " of the main circuit may be the sum of the corresponding features of the branches:

in other words it suggests that there is a property " current ", which is characteristic of electric circuits, and obeys a law of addition in branched circuits. We now establish this law and thereby establish current as a measurable property or magnitude.

We construct two circuits X and Y (figure 1) with a common limb, in which we connect our indicating instruments AA. For the purpose of simplifying the description it will be convenient to imagine that all the instruments possess pointers moving over scales of the ordinary type, which are, however, initially blank, but it will be readily understood that what we say about numbering the scales is equally applicable to scales of any type, such as a set of " weights " for the current balance, which we can mark with the appropriate numbers in due course, and so on. In the branches B and U we have also indicating instruments. U is the instrument we choose to " define our unit ": *in other words, we assign* the numeral 1 to some definite condition of this instrument. For this country

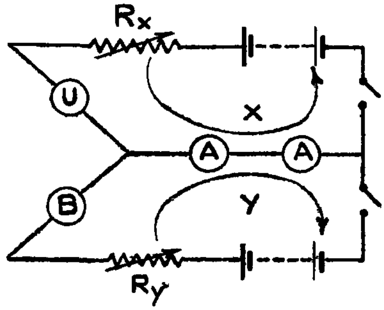

Figure 1.

the legal standard is a current balance, and the definite condition is one of equilibrium with a certain standard weight on one pan of the balance. B, like AA, has a blank scale. We first break the circuit Y, adjust R_X until U is in the standard condition, and then mark all the scales of AA with the numeral 1 at the points corresponding to their indications. We then open X and close Y, adjust R_Y until AA again indicate 1, and mark the corresponding indication of B also 1. Next close both X and Y, adjust R_X and R_Y until both U and B indicate 1, and mark the corresponding indication of AA with the numeral 2. Proceeding in the same way, we use the circuit Y to mark the corresponding indication of B with the numeral 2; then with B at 2 and U in the standard condition to mark AA at 3, and so on. Having marked all the instruments AA in this way, we can insert some in the place of U and others at B, and verify that whatever the indication of the three groups of instruments they always satisfy the law of addition. Further applications of the same principle enable us to subdivide the scales to any desired extent within the limits set by the sensitivity of the instruments, and then to verify the law with high precision for a very wide range of current. We have then established the indication of such an instrument as a unique and precise numerical measure of the current in any branch of a circuit in which it is included.

It is important to notice that the measurement of current in this way is completely independent of the law of operation of the indicating devices. We may notice that many direction-sensitive instruments have scales that lead to the recognition of some " linear law "; and that the scales of many reversible instruments suggest a " square law "; but our ideas of the magnitude " current " are not founded on any such law. The highest precision will be obtained when the indicating devices consist of the combination, well known in all standardizing laboratories, of a shunt resistor and a potentiometer, the indication being the setting of the potentiometer " slider " which corresponds with a state of balance of the potentiometer, but we must suppose that the scale traversed by the slider is initially blank, and that the combination is calibrated as an ammeter by the process we have described.

§4. CURRENT: A DIRECTED MAGNITUDE

The fact that the pointers of some instruments move in the opposite direction when the connections to their terminals are interchanged does not itself introduce any question of negative current; in effect we obtain a different instrument by the reversal, and must re-calibrate it, but that is all. Each direction-sensitive instrument must therefore have its terminals marked, say, red and black, and must always be connected in circuit in the same way with reference to the battery, so that its indications always correspond with those obtained during the calibration. If, now, bearing in mind this condition, we take the circuit of figure 1, and reverse both the battery in the circuit Y and the instrument B, we find, on closing both circuits X and Y, and adjusting R_X and R_Y until both B and U indicate the same current, that A then indicates zero current, whatever may be the current indicated by B and U. If our law of addition is to be generally applicable to branched circuits, we must therefore recognize that the sum of two equal currents may be zero; in other words, that there are such things as negative currents.

We adopt a simple convention to enable us to allocate signs to currents. We use direction-sensitive instruments, and mark their terminals so that during calibration the red terminal of one instrument is always connected to the black one of the next in the same circuit. Then at any junction the currents indicated by instruments with their red terminals connected to the common point are given the + sign, and those of instruments with the black terminals connected to the common point the − sign.

We now establish as an experimental fact Kirchhoff's First Law. The sum of the currents in all the branches meeting at any junction is always zero. This is one of our basic laws—grounded directly on observed facts and independent of any theory of electricity.

§5. RESISTANCE

The idea of resistance as a magnitude characteristic of the various parts of an electric circuit first arises when we observe that the insertion of a resistor in, say, the simple circuit first considered diminishes the indications of all the instruments, and that the insertion of a second in series with the first decreases it still further. We are led to define resistors as equal in resistance when the substitution of one for another causes no change in the indicator, and to establish the operation of connecting resistors in series as addition. It is scarcely necessary to enter into details. Every experimentalist knows that the most accurate measurements of resistance are always made by " simple substitution ", which means by making the above-mentioned judgments of equality; and in standardizing institutions, the standard scale of resistance is always established by constructing resistance "build-ups"—that is to say, groups of resistors connected in series in such a way that it can be shown by experiment that they satisfy the law of addition with high precision. Resistance, then, is a property defined by these operations, and measured independently of any other property; for it may be noted that only qualitative observations of the current indicators are necessary. The various bridges and balancing devices employed for the work are to be regarded as detectors rather than measuring instruments—that is to say, they enable us to make precise judgments of equality of resistance much as a comparator enables us to make precise judgment of equality of lengths. Over a wide range of conditions, resistance proves to be independent of both the circuit employed for its measurement and the property *current.*

§ 6. CONDUCTANCE

It will be convenient to consider also at this point *conductance*, another magnitude independently measurable, but closely associated with resistance; for the test of equality for the two quantities is exactly the same. They differ only in the law of addition, conductance being added by parallel connection instead of series. We find by experiment that for any resistor, conductance and resistance are inversely proportional to one another. The constant of proportionality is merely a scale factor which for convenience is suppressed by defining the unit of resistance and the unit of conductance by means of the same standard resistor. Conductance and resistance are then reciprocals.

§ 7. VOLTAGE

At first sight it would appear that we can establish a property " e.m.f." characteristic of batteries by a process almost identical with that established for resistance and equally fundamental, involving only addition by series connection and judgments of equality based on observations of no change of current on substitution in any circuit. Since the reversal of a battery produces changes of " current ", we must distinguish the two terminals as, say, red and black, and when we find that the series connection of a battery in one direction increases " current ", and in the other direction decreases it, we are prepared to recognize e.m.f. as a directed magnitude, the reversal of the battery changing the sign. Detailed experiments show, however, that the method cannot be established. The judgments of equality are found to depend on the circuit employed as well as on the batteries under comparison; and with any one circuit, the combination (3–2) for example, is never found to be precisely equivalent to the combination (2–1). Subsequently we explain the failure of the method by recognizing another property of batteries, " internal resistance ", but at this stage we can only conclude that the fundamental property *voltage* must be sought elsewhere.

We find that we have two distinct classes of electrical indicating instruments. In addition to those that we have used in order to establish the magnitude *current* (the ammeters), we have others which, unlike the ammeters, cannot form part of any circuit, since on inserting them into any circuit they invariably give rise to zero current, but which nevertheless give definite indications when connected across any component of a circuit. We shall now show that just as the instruments of the first class enabled us to establish the magnitude current in virtue of a law of addition, those of the second class (the voltmeters) enable us to establish voltage by a similar law. As in the first case, the process is quite independent of the law of operation of the instrument itself, but we may note that our voltmeters include electrostatic instruments, thermionic instruments of the electrometer type, and potentiometers. Some of these contain batteries and are found to be direction-sensitive: others are reversible and obey a different law, but all serve to establish the same magnitude. The potentiometer is by far the most accurate of the voltmeters. The fact that the same mechanism also formed part of one of the ammeters need cause no confusion: we must again suppose that the instrument has initially a blank scale, which by appropriate marking and numbering will be established as a scale of voltage.

We observe that if any voltmeter is connected in succession to points XY, YZ, XZ (figure 2), Y being between X and Z and no battery being between X and Z,

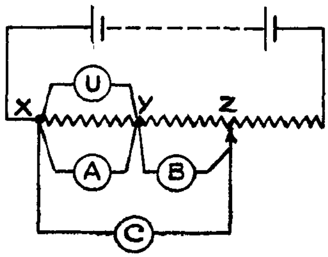

Figure 2.

then it gives a greater reading in the position XZ than in either of the positions XY or YZ. This observation suggests that a law of addition can be established for the property indicated, voltage being additive for circuit elements connected in series. The experimental procedure is closely analogous to the 3-instrument method used for current, and will be sufficiently obvious from figure 2. One voltmeter U is arbitrarily chosen as standard, and its indication when connected to the points XY, say, is marked 1. The indication of a second instrument A, also connected to XY, is also marked 1, and it is then transferred to B, and the point Z adjusted until it again reads 1. At this setting the indication of the third instrument C is marked 2. Proceeding in this way we establish our standard scale of voltage by successive additions of 1, and verify the laws of addition by proving that the indications of the three instruments always satisfy the relation $C=A+B$.

In order that the law of addition may apply perfectly generally to all circuit-elements in series connection we must recognize voltages of both signs. Thus in the circuits of figures 3 and 4 a voltmeter connected to the terminals B and D may indicate zero, even when it gives finite indications in connection to BC and CD.

The sum of the voltages of BC and CD is by the law of addition zero, and, therefore, these two voltages must be equal but of opposite sign. Positive and negative voltages can be recognized by the use of direction-sensitive voltmeters with marked terminals, say one red and one black, as for ammeters. With such an instrument we find that connection to the terminals BC and CD in the case just mentioned gives the same indication provided that the same terminal of the voltmeter is connected to B of the pair BC and to D of the pair CD. Thus, reversal of the

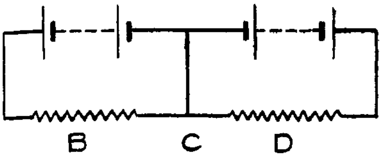

Figure 3.

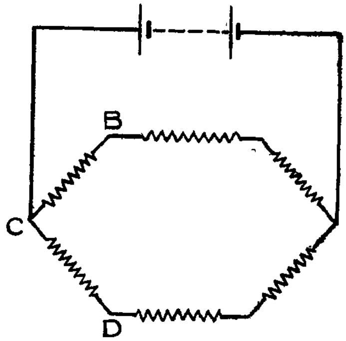

Figure 4.

order of the voltmeter terminals when passing along a circuit indicates a change of sign of voltage. We may note here that the terminals of any component of a circuit are essentially the devices by means of which a voltmeter can be connected to two definite points on that circuit-element.

We are now in a position to establish the following experimental law concerning voltage. The algebraic sum of the voltages measured between all the successive pairs of any set of terminals on any closed circuit is always zero. This law is analogous to Kirchhoff's First Law, and is of the same basic importance.

§ 8. OHM'S LAW

Having now completed our first stage and established the independent measurement of current, voltage and resistance by processes that will give, in favourable circumstances, an accuracy of the order of 1 part per million, we can now proceed to the development of algebraic laws based on these measurements. The most obvious one is Ohm's Law: for a large class of circuit-elements the terminal voltage V is proportional to the product of the resistance R and the current I characteristic of that element:

$$V = kIR. \qquad \ldots\ldots(1)$$

The proportionality constant k proves to be independent of the size and shape of the circuit element and of its composition; metals, non-metals, electrolytic solutions, etc., all have the same value if the law holds. The constant is therefore no more than a scale factor, and is suppressed by adjusting the units of V, I and R so that $k = 1$. Incidentally the fixing of the sign of k determines the relation between the red and black terminals of ammeters and those of voltmeters, which have hitherto been unrelated.

§ 9. ELECTROMOTIVE FORCE

We now find that there is an important class of circuit elements, including batteries, dynamos, thermocouples, etc., which do not obey Ohm's Law, but are characterized by a finite terminal voltage even when they are " on open circuit " and, therefore, have zero current. When these devices are included in a circuit we find that their terminal voltage varies with the current, and we establish the experimental law

$$V_0 - V = kI,$$

where V is the voltage corresponding with the current I, and V_0 is the voltage on open circuit. The constant k proves to vary with the circuit element and to be characteristic of it. We must therefore recognize a new magnitude r as well as the scale factor S and write

$$V_0 - V = SrI \quad \text{or} \quad V = V_0 - SrI. \qquad \ldots\ldots(2)$$

If, then, the circuit external to the generator consists of n circuit elements in series, all obeying Ohm's Law and of resistance $R_1, R_2 \ldots R_n$, we have, by applying Ohm's Law and the law for the summation of the voltages of a closed circuit,

$$IR_1 + IR_2 + \ldots IR_n - (V_0 - SrI) = 0.$$

Note that a negative sign has been attached to the expression for the voltage of the generator in this equation. This is because experiment shows that the order of the voltmeter terminals must be reversed when passing from the resistor elements $R_1, R_2, \ldots R_n$ to the generator. When we examine this relation we perceive that a great simplification can be achieved by choosing the unit of r so that the scale factor S becomes unity, in which case the equation becomes

$$I\left[\sum_1^n R + r\right] = V_0. \qquad \ldots\ldots(3)$$

It is natural therefore to interpret r as a resistance inherent in the generator, and with this interpretation we obtain Kirchhoff's Second Law in the familiar form

$$\Sigma IR = E, \qquad \ldots\ldots(4)$$

the summation now including every element of the circuit, and E, the e.m.f. of the generator, being defined as its *open-circuit voltage*. The introduction of this new defined magnitude is not really necessary, but it has proved to be useful in practice. We can establish that the magnitude so defined is additive in series connection and capable of either sign, so that the general expression for Kirchhoff's Second Law becomes

$$\Sigma IR = \Sigma E.$$

The unit of e.m.f. is, of course, the same as that of voltage, and the well known Weston standard cell can be regarded as a standard of either e.m.f. or voltage, though voltage is unquestionably the more fundamental property. Another point to notice is that we have extended the notion of resistance; for the property r is not measurable by the process by which the magnitude *resistance* is primarily defined. Experimental justification for identifying the two properties can be obtained later by showing that correlation can be established between them by means of measurements with alternating current.

§ 10. ROUTINE METHODS

Having now established the basic laws of direct-current networks we can use them in order to establish all the bridge networks and potentiometer circuits and the great variety of operations by means of which measurements of current, resistance, voltage and e.m.f. are made in everyday practice. We can now use Ohm's Law for example for the purpose of measuring voltage in terms of resistance and current when that procedure proves to be more convenient than the one described. It should be noted here that when we say that two different processes serve to measure *the same magnitude*, we mean that experiment has shown that the two processes give the same result apart from a scale factor.

Routine methods differ very considerably from those we have described, mainly because the operations needed in order to establish a law are not necessarily those most convenient for subsequent applications of that law. When once our basis has been firmly established, our procedure is mainly dictated by consideration of economy of labour and the consistency of our measurements from day to day and year to year. The ease of construction of the required apparatus and its permanence become of dominating importance. In practice we find that a standard series of resistors is the most easily constructed, most permanent, and most complete as regards range and fineness of subdivision of all electrical standards. The general tendency is, therefore, to base most operations on measurements of resistance. Thus the operations we have described may appear at first sight as unfamiliar and not representative of actual practice. The practical application of Kirchhoff's First Law may take the form of a search for " leakage " or a " stray current " in order to explain some discrepancy in the observations. The experiments we have described are admittedly not frequently made deliberately with the idea of establishing a law. We think, however, that the experimentalist will recognize them as the logical equivalent of the mass of indirect experimental evidence for the truth of these laws which he has acquired in the course of his work.

We have now reached a stage at which it becomes necessary to establish relations between the basic electrical magnitudes and those of other branches of physics. Probably every experimentalist will be prepared to take for granted length, mass and time as three independently measurable magnitudes. We find, moreover, that over long periods of time our measurements of these three quantities show greater accuracy and consistency than those of any other physical quantities, and we may well therefore regard them as fundamental. We shall here also take for granted other non-electrical magnitudes that are usually held to be established in terms of length, mass and time by well known relations. The means by which these relations can be established is a separate subject.

§ 11. FARADAY'S LAWS OF ELECTROLYSIS

Among the simplest of the laws that we can now establish are Faraday's Law for the various voltameters and electrolytic cells. We find that if I denotes the current in the cell, the mass, W, of any element liberated in time t can be represented by the relation

$$W = \beta . I . t. \qquad \ldots\ldots(5)$$

The constant of proportionality β is found to be characteristic of the element, and is therefore established as a derived magnitude, measurable for the various elements in terms of mass, current and time. We may fix the unit of β by making the scale-factor unity correspond with existing units of W, I and t, as when we measure β in " grams per ampère-second ". Alternatively we may fix the unit of current by assigning a definite value to β for silver (which is the element for which the most accurate experiments have been made), when the scale factor is unity and the units of W and t are those already chosen. The existing " International ampère " was defined in this way at a time when the silver voltameter was considered to be the best standard for current measurements, being easily available at all times and in most places, and giving very consistent results within a prescribed range of conditions.

§ 12. POWER DISSIPATION

When we study the filament lamp and the neon lamp of our original circuit we find no simple law relating either the luminous flux or the heat developed to the current or voltage alone, but calorimetric measurements show that heat is produced by the lamps at a rate proportional to the product of their current and terminal voltage. Further experiments with other circuit components show that this law is quite general, and that the proportionality constant is no more than a scale factor. The units are therefore chosen so that we may express the relation in the simple form

$$P = VI. \qquad \ldots\ldots(6)$$

By the principle of the conservation of energy the dissipation of energy in a circuit must be accompanied by a uni-directional change in some part of the system, which thereby loses that energy. In our circuit the change is found to be chemical and to occur in the battery. We may say that the battery loses chemical energy equivalent to the thermal energy gained by the rest of the circuit, and that the difference between gain and loss of energy corresponds with the difference in the sign of V and, therefore, of VI for the two parts of the circuit.

§ 13. THE LAWS OF RESISTORS AND CONDUCTORS

Any object that may form part of an electric circuit, so that it may be characterized by a finite current, may be called a *conductor*, but we do not find it possible to determine a definite conductance and resistance for every conductor. Those for which definite values may be found obey Ohm's Law, and may be called *resistors*; but others, like the neon lamp, do not obey Ohm's Law even approximately: they are often called non-linear conductors, since they are characterized by a non-linear relation between current and voltage. It may be remarked that the law of power dissipation holds for all conductors, both linear and non-linear.

As soon as the measurement of resistance and conductance has been established, it becomes possible to establish laws relating these properties to temperature.

The experiments are complicated by the dissipation of power in the resistor, which restricts the range of current that can be employed in the detector circuit, but within limits they are practicable. When Ohm's Law is applied to any such resistor, due account must be taken of the variation of its resistance with temperature, and if this is done the law is found to hold good. On the other hand, the relation between voltage and current is non-linear, except for very small currents, on account of the variation of the temperature, and therefore of the resistance, with current. The filament lamp is therefore an example of a non-linear conductor or resistor which obeys Ohm's Law, while the neon lamp is a non-linear conductor which does not obey Ohm's Law. The most precise experiments show that all resistors are non-linear if we cover the widest possible range of current and voltage, and any measured value of resistance or conductance is therefore of significance only within a limited range of current and voltage.

We can also establish relations between the resistance and conductance of conductors of given material and their size and shape, but we must first notice that we have so far assumed that our circuits and, therefore, their components are " linear " in a sense quite different from that just mentioned. We have regarded the circuit as defined by a succession of points or terminals, such that voltmeters connected across these points, and ammeters connected between them, give definite indications, characteristic of the circuit or its elements. The circuit is therefore linear in the sense that it can be represented by a line having the same value of current at every point on it, and that the location of any voltmeter terminal on the circuit is as definite as that of a fixed point on this line. We immediately think of thin wires as providing the simplest method of achieving the conditions described, and although we have mentioned neon lamps and voltameters, which are certainly not thin wires, we must suppose that they have been joined by thin wires, and that each terminal is some definite point on such a wire. We think of the wire as providing a linear path for the current at the crucial points. When we come to deal with conductors of large cross-section, we think of current as entering terminal areas rather than terminal points, and we find that, by fixing one voltmeter terminal at some definite point on such a conductor and using the other as a probe, we can trace on the surface of a conductor lines of constant voltage called, for reasons to be given later, "equipotentials". Thus a terminal is now regarded as a device by means of which a voltmeter can be connected to an equipotential rather than a point. The experiments of high precision, by means of which the relations between the electrical properties of conductors and their geometrical properties have been established, have been almost entirely confined to straight conductors of constant cross-sectional area. Experiment shows that when these constitute parts of circuits, the equipotentials lie in parallel planes, and that the following law holds good :

$$\frac{I}{V}=\sigma\frac{A}{l}, \qquad \ldots\ldots(7)$$

where I denotes the current in the conductor, V the voltage corresponding with any two equipotentials separated by a distance l, A the cross-sectional area of the conductor, and σ is a constant for any one material, but varies with the material. For completeness we should include a scale factor, but since I/V, by Ohm's Law, measures a conductance, the unit of which is fixed by other considerations, and the units of A and l are also fixed by other considerations, it is usual to make the scale factor unity for this particular set of units, thereby fixing the unit of the derived magnitude σ, conductivity.

By inverting the equation (7) we can show that the same experiments also establish a second derived magnitude, resistivity ρ, which becomes the reciprocal of σ, if the same convention is followed in fixing the unit. We have

$$-\frac{V}{I}=R=\rho\frac{l}{A}. \qquad \ldots\ldots(8)$$

It is obviously desirable that we should be able to deal with conductors of other shapes, and we therefore try to establish laws of greater generality. At this stage we find it necessary to introduce purely theoretical conceptions. We regard the conductor as the field of a vector $\mathscr{E}$ which is equal in magnitude but opposite in direction to the gradient of a scalar function v, the potential. It follows that the line integral of this vector between any two points is independent of the path followed, and simply equal to the difference of v for the two terminal points, and, moreover, for any closed path the line integral is zero. The function clearly has properties akin to those of voltage, and we can regard it as a generalization of voltage if it satisfies the condition that the difference of potential for any two points between which a voltage can be measured is equal to that voltage. As a generalization of current we introduce another vector $\mathscr{I}$, the "current density", which satisfies the two conditions: (*a*) that its flux across any terminal area for which a value of current can be measured is equal to that current, and (*b*) the "law of continuity",

$$\mathrm{Div}\,\mathscr{I}=0, \qquad \ldots\ldots(9)$$

which implies that the flux of $\mathscr{I}$ across any closed surface is always zero, and which we may regard as a generalization of the law that the current is the same along any linear circuit. We now write as a generalization of the law established for straight conductors of uniform cross-section,

$$\mathscr{I}=\sigma\frac{dv}{dn}=\frac{1}{\rho}\frac{dv}{dn}, \qquad \ldots\ldots(10)$$

where n is the normal to an equipotential surface, σ is the conductivity, and ρ the resistivity of the medium at the relevant point. This relation implies that the lines of flow of $\mathscr{I}$ are everywhere normal to the equipotential surfaces and therefore identical with the lines of flow of $\mathscr{E}$. These theoretical relations are postulates, for the truth of which we can never obtain direct experimental evidence; for $\mathscr{I}$ and v, and even $\mathscr{I}\,ds$, the current crossing a small area ds, are not measurable: it is always impossible to insert an ammeter into the conducting medium without altering the conditions appreciably. The postulated relations have, however, been found to lead to all the relations that have been established experimentally for conductors of various shapes. As we have seen already, equipotential lines can be traced experimentally on the surfaces of solid conductors, and in liquid conductors, equipotential surfaces having given voltages from one electrode may be traced out in a similar way. In electrolytes, moreover, it can be shown that the thickness of deposited metal has some relation to current density. The postulated quantities are therefore not entirely without experimental significance, though the experiments in which they are significant are only of low accuracy. We shall later find applications for these conceptions in other branches of the subject.

§ 14. AMPÈRE'S LAW

We have seen that the basic laws in any branch of the subject are those which have been established by, or which govern the operation of, the instruments of highest precision available for work of that kind. It follows that the fundamental law of electrodynamics is the law of operation of the current balance or electrodynamometer, which is, broadly speaking, Ampère's Law, although we shall find it desirable to express it in terms different from those actually employed by Ampère.

Very simple experiments are sufficient to show that the mechanical force between two circuits is proportional to the product of the currents they carry; but the relation between the constant of proportionality and the geometrical properties of the circuits is so complicated that its discovery by Ampère was described by Clerk Maxwell as one of the greatest achievements in physical science. Ampère, by making inspired guesses based on the examination of circuits of simple form, found the form of the constant almost as soon as he discovered the proportionality. Subsequent work has served only to confirm the law with higher and higher precision, and to define the conditions in which it holds good. Ampère expressed his law in terms of the forces between circuit elements, and thereby introduced an ambiguity that has been the foundation of many paradoxes. But this ambiguity can be removed by translating his law into terms of the shapes and geometrical relation of complete circuits. It then becomes

$$F = kI_1I_2(-\partial N/\partial q), \qquad \ldots\ldots(11)$$

where F is the force tending to increase a coordinate q; I_1, I_2 are the currents in the two circuits; and N is defined as

$$N = \oint_1 \oint_2 \frac{dl_1 \,.\, dl_2 \,.\, \cos\theta}{r}, \qquad \ldots\ldots(11.1)$$

where dl_1, dl_2 are infinitesimal elements of the circuits 1, 2; r the distance, and θ the angle, between them. Thus N and $(-\partial N/\partial q)$ can be determined by geometrical measurements on the circuits.

Equation (11) is limited by its form to strictly linear circuits; but it can be extended in an obvious manner to any pair of circuits that can be regarded as each made up of linear filaments, rigidly connected, the proportion of the circuit current passing through each filament being known. But there is another limitation, usually expressed by the statement that the circuits must be composed of, and the space between them occupied by, a medium that is non-magnetic. We are not yet in a position to define " non-magnetic ", and when we are we shall find that none of the materials usually present in the experiments to which (11) is applied are, strictly speaking, non-magnetic. However, a sufficient condition for (11) to be true can be stated definitely; it is that every medium in the neighbourhood of the circuits should be either air or equivalent to air, equivalent meaning that, if any part of the medium that is not air is replaced by air, or *vice versa*, no appreciable change in the force between the circuits is produced. Equivalence thus depends on what changes are appreciable—that is to say, on the sensitivity of the measurement of force.

In this form and with these limitations the law (11) has been confirmed by many investigations with an accuracy that has steadily increased down the years. The most recent experiments at the National Physical Laboratory and the National Bureau of Standards agree and confirm the law to about 2 parts in 100,000.

In most treatises k is written μ, and treated as a derived magnitude whose value may vary. The basis of this practice is the belief that, if the air and its equivalents used in current balances were replaced by a medium not equivalent to air, and therefore magnetic, then, so long as the medium were uniform, (11) would still be true if k were allowed to take a different value. No experiments comparable in accuracy with those in air have been made in any other medium; and it is impracticable to use in such experiments the most important magnetic materials, which are solid and would therefore hamper greatly the measurement of force. Accordingly the direct evidence that k is a derived magnitude, and that it ought to be split into two factors $S\mu$, is of a quite different nature from that for (11), subject to the condition stated above. On the other hand, there are good theoretical grounds, which will be considered later, for thinking that k may be a derived magnitude. In these circumstances it is convenient to take advantage of the liberty noted in § 2 and to write

$$F = S\mu I_1 I_2(-\partial N/\partial q), \qquad \ldots\ldots(12)$$

leaving open for the time the question whether μ is a derived magnitude. It should be noted that (12) still requires that the medium about the circuits should be uniform. If it is not uniform in the sense that an interchange of the medium in one region with that in another might produce a change of the force, a law of the form (11) or (12) can give no account of the matter; for they contain no term that depends on the position of any material that does not form part of the circuits.

The law of the current balance has now been established with such accuracy that it can be used to define the unit of current. The law is (11) for the special case in which $I_1 = I_2 = I$, and we may write it

$$F = S\mu_0 I^2(-\partial N/\partial q), \qquad \ldots\ldots(13)$$

the factor μ_0 serving to remind us that we are limited to an air-equivalent medium. The ampère has been defined by assigning to $S\mu_0$ in this law the value 10^{-2} when F is measured in dynes, or 10^{-7} when F is measured in m.k.s. units. The constant μ_0 in (13) is commonly termed the permeability of free space; it is perhaps worth noting that the relations implied by this term lie quite outside the basic facts that serve to determine the unit of current.

§ 15. MAGNETIC FIELD STRENGTH

Ampère, in his investigation of Oersted's well known discovery, also enunciated the law of the " moving magnet " indicator. Instruments of the type so named have been found to be incapable of the same accuracy as the current balance, and the law has therefore never been established with high accuracy. It has, however, been found to be consistent with the available evidence, and we shall therefore accept it as a hypothesis. It can be written in the form *

$$\boldsymbol{T} = \boldsymbol{m}SI \oint \frac{\boldsymbol{dl} \times \boldsymbol{r}_1}{r^2} = S\boldsymbol{m}I\boldsymbol{G}, \qquad \ldots\ldots(14)$$

where $\boldsymbol{T}$ represents the torque on a small compass needle, about its axis of rotation, arising from the presence in its neighbourhood of a circuit carrying a current I; S is the scale factor; $\boldsymbol{dl}$ represents in length and direction any element of the circuit; $\boldsymbol{r}_1$ is a unit vector in the direction from this element to the magnet; $\times$ represents a vector product; r is the distance between the element and the magnet; the integration is taken completely round the circuit; and $\boldsymbol{m}$ is a derived magnitude characteristic of the needle, its direction always perpendicular to the axis of rotation.

The form of the law requires that the dimensions of the magnet should be small compared with any r, in order that $\boldsymbol{G}$, the integral, should be definite. A limitation is also imposed on the nature of the medium surrounding the magnet and circuit; a sufficient, but perhaps not a necessary, condition is that the medium should be air or its equivalent in the sense defined above.

The experimental evidence for (14) is scanty and usually indirect. Thus, though the law is implied in the well known experiments with sine- and tangent-galvanometers, those experiments do not enable $\boldsymbol{m}$ to be determined in terms of current, torque and the geometrical magnitude $\boldsymbol{G}$; for the torque is measured in terms of that which the needle experiences in the absence of a current circuit, and it is assumed that this torque is also proportional to $\boldsymbol{m}$. In particular, there is very little evidence that $\boldsymbol{m}$ is independent of I and $\boldsymbol{G}$, as the law requires; indeed, there is indirect evidence that, for large values of I, $\boldsymbol{m}$ is not independent of I. However, such doubts are of little consequence, because $\boldsymbol{m}$ is not an important derived magnitude.

The real importance of (14) arises from its implication that two circuits for which $I\boldsymbol{G}$ is the same have the same effect on any compass needle. This law proves to be a special case of one more general; the torque on a compass needle is only one of several effects that can be observed in the neighbourhood of current circuits, e.g., the deflection of a cathode ray and the production of an e.m.f. in a rotating coil. All such effects (so long as the media are non-magnetic) are determined by $I\boldsymbol{G}$ in the sense that two circuits for which $I\boldsymbol{G}$ is the same produce the same effect, and that two circuits whose $I\boldsymbol{G}$s are equal and opposite compensate each other and produce no resultant effect.

The experimental evidence for this statement is much stronger than that for (14). Some of it arises from an intimate mathematical connection between $\boldsymbol{G}$ and N (it will be stated explicitly in a later section), which was used at the National Bureau of Standards in some of the experiments on the current balance. Instead of determining N wholly by geometrical measurements, they determined

* This form is taken from G. P. Harnwell, *The Principles of Electricity and Magnetism* (McGraw-Hill, New York and London, 1938). This book seems to be representative of the working ideas of the experimentalist of today.

an important part of it as the ratio of the $\boldsymbol{G}$s for the fixed and moving coils when these were placed suitably relative to each other; this they did by determining the ratio of the currents in the two coils at which the resultant effect on a compass needle vanished. The fact that they could thereby obtain consistent values of N for different pairs of coils, and that this method of determining N led to results agreeing with those of the National Physical Laboratory, who relied entirely on geometrical measurements, is very complete evidence for the significance attributed to $I\boldsymbol{G}$ as well as for (11).

Another part of the evidence for the significance of $I\boldsymbol{G}$ rests on the "Schuster magnetometer", which is now recognized as the most accurate instrument for the measurement of the earth's magnetic field. (We are not yet in a position to define this quantity; all that we need to know about it here is that it is something that can be compensated by a current circuit.) This instrument consists essentially of a current circuit constructed so that the value of $\boldsymbol{G}$ for a small region fixed relative to it can be determined with high accuracy, some form of field detector being mounted within this region—in practice a small compass needle, a rotating coil, or a vibrating circuit mounted at the centre of a Helmholtz coil-system. The field within the small region is measured by observing what current must flow in the circuit in order that the detector shall indicate zero. Experiments with this instrument establish with high accuracy that the compensating current is inversely proportional to $\boldsymbol{G}$ at the relevant point.

These facts lead us to define a magnitude $\boldsymbol{H}$, characteristic of points in the neighbourhood of a linear circuit, by

$$\boldsymbol{H}=S.I.\boldsymbol{G}.=S.I.\oint\frac{\boldsymbol{dl}\times\boldsymbol{r}_1}{r^2}. \qquad ......(15)$$

We insert the scale factor merely to allow for changes of unit. In accordance with general practice we shall call $\boldsymbol{H}$ the "magnetic field strength due to the current circuit". Notice that it is not a magnitude established by experimental facts, but one laid down by definition; its importance depends on the fact that the quantity so defined proves to be significant in a wide range of experiments of the kind mentioned above. The definition can be extended to a non-linear circuit, so long as it can be regarded as a collection of linear circuits, by making $\boldsymbol{H}$ (which is a vector) mean the vector sum of components each associated with one of the linear circuits. In this extended meaning $\boldsymbol{H}$ retains its significance.

Purely mathematical reasoning is sufficient to show that the vector $\boldsymbol{H}$, defined by (15), is the negative gradient of $I\omega$, where ω is the solid angle subtended by the circuit at the point to which $\boldsymbol{H}$ refers; and $I\omega$ is therefore recognized as the magnetic scalar potential. It follows that the line integral of $\boldsymbol{H}$ around any closed path that does not link a circuit is zero, and that the line integral around a closed path that links n times in the same sense a circuit carrying a current I is given by

$$\oint \boldsymbol{H}.\boldsymbol{dl}=S.4\pi nI. \qquad ......(16)$$

We may remark that the two units of H that are in practical use arise naturally from equations (15) and (16). The oersted, the unit of the c.g.s.μ system, was chosen so as to simplify (15) by making $S = 1$. The "ampère-turn per metre", the unit of the m.k.s. system, was chosen so as to simplify (16) by making $S . 4\pi = 1$. There is no difference in principle between the two choices, and usage alone can show which is the more judicious.

Finally a word about magnetic media. If the medium surrounding the circuit is magnetically heterogeneous, (14) ceases to be true; circuits with the same IG are not in generally equivalent, so that H loses its significance if it is defined by (15), and there is no reason to adopt that definition. Whether in these circumstances there is any other quantity which plays in any sense the same part as H plays when the medium is non-magnetic is a question that must wait until a later section.

If the medium is uniform, but not equivalent to air, the scanty evidence available suggests that circuits with the same IG are still equivalent. Since (15) merely expresses this equivalence, there is no need to introduce a factor μ into it; whether it is desirable to do so, and what modification, if any, is required in (14), are again questions that must be postponed.

At this stage we may regard all the basic principles of circuits under direct current conditions as having been established. We have considered so far only currents, voltages and circuits that are invariable with time, and mainly materials that are non-magnetic. In subsequent parts we propose to deal with varying currents and voltages, which lead to the conceptions of inductance and capacitance, and so to the principles of magnetism, electrostatics and electromagnetic waves.

In concluding the first portion of our enquiry we would emphasize that we are not trying to establish a new system of electrical measurements, but merely to discover the logical basis of the system which is in fact universally practised. It follows that all the relations that we develop will be familiar ones, and the differences between our treatment and that of the standard treatises may appear trivial, especially in this first portion. The detail given is, however, necessary for logical completeness and to establish the method. The differences will become more significant in electrostatics and magnetism, and it is from these branches of the subject that we hope to remove many of the difficulties that frequently disturb the experimentalist. To the mathematician there are many possible ways of developing the relations of electromagnetism, and they are all exactly equivalent in logic and differ only in elegance. The experimentalist has less latitude; his practice is dictated by stubborn facts that defy representation in equations. The one course he is compelled by circumstances to follow is more important to him than the more elegant courses he might have followed in a better world. This is the excuse we offer to those who find our treatment clumsy.