THEORY OF COMMUNICATION*

By D. GABOR, Dr. Ing., Associate Member.†

(*The paper was first received 25th November, 1944, and in revised form 24th September, 1945.*)

PREFACE

The purpose of these three studies is an inquiry into the essence of the "information" conveyed by channels of communication, and the application of the results of this inquiry to the practical problem of optimum utilization of frequency bands.

In Part 1, a new method of analysing signals is presented in which time and frequency play symmetrical parts, and which contains "time analysis" and "frequency analysis" as special cases. It is shown that the information conveyed by a frequency band in a given time-interval can be analysed in various ways into the same number of elementary "quanta of information," each quantum conveying one numerical datum.

In Part 2, this method is applied to the analysis of hearing sensations. It is shown on the basis of existing experimental material that in the band between 60 and 1 000 c/s the human ear can discriminate very nearly every second datum of information, and that this efficiency of nearly 50% is independent of the duration of the signals in a remarkably wide interval. This fact, which cannot be explained by any mechanism in the inner ear, suggests a new phenomenon in nerve conduction. At frequencies above 1 000 c/s the efficiency of discrimination falls off sharply, proving that sound reproductions which are far from faithful may be perceived by the ear as perfect, and that "condensed" methods of transmission and reproduction with improved waveband economy are possible in principle.

In Part 3, suggestions are discussed for compressed transmission and reproduction of speech or music, and the first experimental results obtained with one of these methods are described.

Part 1. THE ANALYSIS OF INFORMATION

SUMMARY

Hitherto communication theory was based on two alternative methods of signal analysis. One is the description of the signal as a function of time; the other is Fourier analysis. Both are idealizations, as the first method operates with sharply defined instants of time, the second with infinite wave-trains of rigorously defined frequencies. But our everyday experiences—especially our auditory sensations—insist on a description in terms of *both* time and frequency. In the present paper this point of view is developed in quantitative language. Signals are represented in two dimensions, with time and frequency as co-ordinates. Such two-dimensional representations can be called "information diagrams," as areas in them are proportional to the number of independent data which they can convey. This is a consequence of the fact that the frequency of a signal which is not of infinite duration can be defined only with a certain inaccuracy, which is inversely proportional to the duration, and vice versa. This "uncertainty relation" suggests a new method of description, intermediate between the two extremes of time analysis and spectral analysis. There are certain "elementary signals" which occupy the smallest possible area in the information diagram. They are harmonic oscillations modulated by a "probability pulse." Each elementary signal can be considered as conveying exactly one datum, or one "quantum of information." Any signal can be expanded in terms of these by a process which includes time analysis and Fourier analysis as extreme cases.

These new methods of analysis, which involve some of the mathematical apparatus of quantum theory, are illustrated by application to some problems of transmission theory, such as direct generation of single sidebands, signals transmitted in minimum time through limited frequency channels, frequency modulation and time-division multiplex telephony.

* Radio Section paper.

† British Thomson-Houston Co., Ltd., Research Laboratory.

(1) INTRODUCTION

The purpose of this study is to present a method, with some new features, for the analysis of information and its transmission by speech, telegraphy, telephony, radio or television. While this first part deals mainly with the fundamentals, it will be followed by applications to practical problems, in particular to the problem of the best utilization of frequency channels.

The principle that the transmission of a certain amount of information per unit time requires a certain minimum waveband width dawned gradually upon communication engineers during the third decade of this century. Similarly, as the principle of conservation of energy emerged from the slowly hardening conviction of the impossibility of a *perpetuum mobile*, this fundamental principle of communication engineering arose from the refutation of ingenious attempts to break the as yet unformulated law. When in 1922 John Carson[1.1] disproved the claim that frequency modulation could economize some of the bandwidth required by amplitude-modulation methods, he added that all such schemes "are believed to involve a fundamental fallacy." This conviction was soon cast into a more solid shape when, in 1924, Nyquist[1.2] and Küpfmüller[1.3] independently discovered an important special form of the principle, by proving that the number of telegraph signals which can be transmitted over any line is directly proportional to its waveband width. In 1928 Hartley[1.4] generalized this and other results, partly by inductive reasoning, and concluded that "the total amount of information which may be transmitted . . . is proportional to the product of frequency range which is transmitted and the time which is available for the transmission."

Even before it was announced in its general form, an applica-

tion was made of the new principle, which remains to this day probably its most important practical achievement. In 1927, Gray, Horton and Mathes[1,5] gave the first full theoretical discussion of the influence of waveband restriction on the quality of television pictures, and were able to fix the minimum waveband requirements in advance, long before the first high-definition system was realized. In fact, in this as in later discussions of the problem, the special Nyquist–Küpfmüller result appears to have been used, rather than Hartley's general but somewhat vague formulation.

The general principle was immediately accepted and recognized as a fundamental law of communication theory, as may be seen from its discussion by Lüschen[1,6] in 1932 before this Institution. Yet it appears that hitherto the mathematical basis of the principle has not been clearly recognized. Nor have certain practical conclusions been drawn, which are suggested by a more rigorous formulation.

(2) TRANSMISSION OF DATA

Let us imagine that the message to be transmitted is given in the form of a time function $s(t)$, where s stands for "signal." Unless specially stated, s will be assumed to be of the nature of a voltage, current, field strength, air pressure, or any other "linear" quantity, so that power and energy are proportional to its square. We assume that the function $s(t)$ is given in some time interval $t_2 - t_1 = \tau$, as illustrated in Fig. 1.1. Evidently

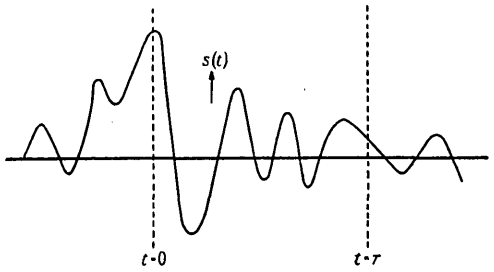

Fig. 1.1.—Signal as a function of time.

this message contains an infinity of data. We can divide τ into, say, N sub-intervals, and define, for instance, the average ordinate in each sub-interval as a "datum." If there is no limit to the sub-division, there is no limit to the number of data which could be transmitted in an *absolutely faithful* reproduction.

As this is impossible, let us see whether it is possible to transmit faithfully at least a finite number N of data. Evidently there is an infinite number of possibilities for specifying the curve $s(t)$ in the interval τ *approximately* by N data. Without knowing the specific purpose of the transmission it is impossible to decide which is the most economical system of selection and specification. Yet, certain methods will recommend themselves by reason of their analytical simplicity. One of these, division into equal sub-intervals, has been already mentioned. Another method is to replace the curve $s(t)$ in the interval τ by a polynomial of order N, to fit it as closely as possible to $s(t)$ by the method of least squares, and to take the coefficients of the polynomial as data. It is known that this method is equivalent to specifying the polynomial in such a way that its first N "moments" M_i shall be equal to those of $s(t)$:—

$$M_0 = \int_0^\tau sdt \quad M_1 = \int_0^\tau tsdt \quad M_2 = \int_0^\tau t^2 sdt \ldots M_{N-1} = \int_0^\tau t^{N-1} sdt$$

Instead of the coefficients of the polynomial, we can also consider these moments as the specified data.

A method closely related to this is the following. Expand $s(t)$, instead of in powers of time, in terms of a set of N functions $\phi_k(t)$, orthogonal in the interval $0 < t < \tau$, and consider as data the N coefficients of expansion. It is known that this is equivalent to fitting the expansion to $s(t)$ by the method of least squares.* How close the fit will be, and how well it will suit the practical purpose, depends on the set of functions selected.

One class of orthogonal functions, the simple harmonic functions sine and cosine, have always played a preferred part in communication theory. It is shown in Appendix 9.1 that there are good reasons for this preference other than their elementary character. Let us now develop the curve $s(t)$ in the interval τ into a Fourier series. This gives an infinite sequence of spectral lines, as shown in Fig. 1.2, starting with zero frequency, all equally spaced by a frequency $1/\tau$. Two data are associated with each frequency, the coefficients of the sine and cosine terms in the expansion. In a frequency range $(f_2 - f_1)$ there are therefore $(f_2 - f_1)\tau$ lines, representing $2(f_2 - f_1)\tau$ data, that is exactly *two data per unit time and unit frequency range.*

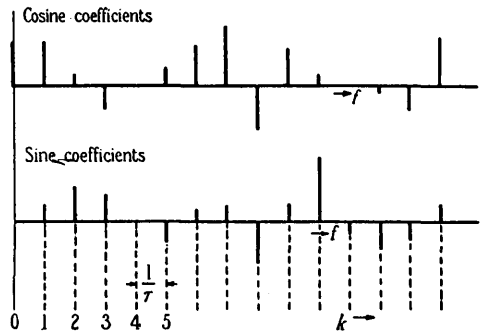

Fig. 1.2.—Fourier spectrum of signal in an interval τ.

This, in fact, proves the fundamental principle of communication. *In whatever ways we select N data to specify the signal in the interval τ, we cannot transmit more than a number $2(f_2 - f_1)\tau$ of these data, or of their independent combinations by means of the $2(f_2 - f_1)\tau$ independent Fourier coefficients.*

In spite of the extreme simplicity of this proof, it leaves a feeling of dissatisfaction. Though the proof shows clearly that the principle in question is based on a simple mathematical identity, it does not reveal this identity in a tangible form. Besides it leaves some questions unanswered: What are the effects of a physical filter? How far are we allowed to sub-divide the waveband or the time interval? What modifications would arise by departing from the rigid prescription of absolute independence of the data and allowing a limited amount of mutual interference? It therefore appears worth while to approach the problem afresh in another way, which will take considerably more space, but which, in addition to physical insight, gives an answer to the questions which have been left open.

(2.1) Time and Frequency

The greatest part of the theory of communication has been built up on the basis of Fourier's reciprocal integral relations†

$$s(t) = \int_{-\infty}^{\infty} S(f) e^{2\pi i f t} df \qquad S(f) = \int_{-\infty}^{\infty} s(t) e^{-2\pi i f t} dt \quad . \quad . \quad (1.1)$$

* Cf. e.g. CHURCHILL, RUEL V.: "Fourier Series and Boundary Value Problems" (McGraw Hill, 1941), p. 40. This book contains an introduction to the theory of orthogonal functions.

† The notations used will follow in the main those of Campbell and Foster.[1,7]

where $s(t)$ and $S(f)$ are a pair of Fourier transforms. We will refer to $S(f)$ also as the "spectrum" of $s(t)$.

Though mathematically this theorem is beyond reproach, even experts could not at times conceal an uneasy feeling when it came to the physical interpretation of results obtained by the Fourier method. After having for the first time obtained the spectrum of a frequency-modulated sine wave, Carson wrote:[1.1] "The foregong solutions, though unquestionably mathematically correct, are somewhat difficult to reconcile with our physical intuitions, and our physical concepts of such 'variable-frequency' mechanisms as, for example, the siren."

The reason is that the Fourier-integral method considers phenomena in an infinite interval, *sub specie aeternitatis*, and this is very far from our everyday point of view. Fourier's theorem makes of description in time and description by the spectrum, two mutually exclusive methods. If the term "frequency" is used in the strict mathematical sense which applies only to infinite wave-trains, a "changing frequency" becomes a contradiction in terms, as it is a statement involving *both* time and frequency.*

The terminology of physics has never completely adapted itself to this rigorous mathematical definition of "frequency." In optics, in radio engineering and in acoustics the word has retained much of its everyday meaning, which is in better agreement with what Carson called "our physical intuitions." For. instance, speech and music have for us a definite "time pattern," as well as a frequency pattern. It is possible to leave the time pattern unchanged, and double what we generally call "frequencies" by playing a musical piece on the piano an octave higher, or conversely it can be played in the same key, but in different time. Evidently both views have their limitations, and they are complementary rather than mutually exclusive. But it appears that hitherto the fixing of the limit was largely left to common sense. It is one of the main objects of this paper to show that there are also adequate mathematical methods available for this purpose.

Let us now tentatively adopt the view that both time and frequency are legitimate references for describing a signal, and illustrate this, as in Fig. 1.3, by taking them as orthogonal co-

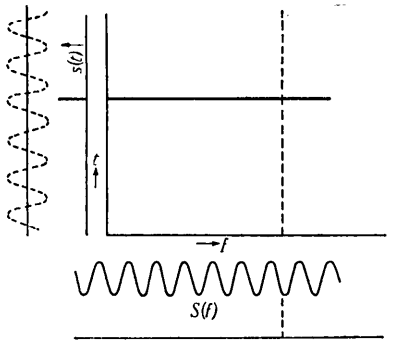

Fig. 1.3.—Unit impulse (delta function) and infinite sine wave in time/frequency diagram.

ordinates. In this diagram a harmonic oscillation is represented by a vertical line. Its frequency is exactly defined, while its epoch is entirely undefined. A sudden surge or "delta function"† (also called "unit impulse function"), on the other hand, has a sharply defined epoch, but its energy is uniformly distributed over the whole frequency spectrum. This signal is therefore represented by a horizontal line. But how are we to represent other signals, for instance a sine wave of finite duration?

In order to give this question a precise meaning we must consider the physical effects which can be produced by the signal. The physical meaning of the $s(t)$ curve, shown at the left of Fig. 1.4, is that this is the response of an ideal oscillograph

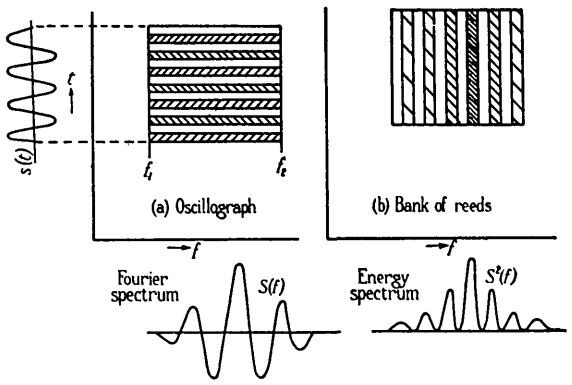

Fig. 1.4.—Time/frequency diagram of the response of physical instruments to a finite sine wave.

which has a uniform response over the whole infinite frequency range. The interpretation of the Fourier spectrum, shown at the bottom of the same figure, is somewhat less simple. It could be obtained by an infinite number of heterodyne receivers, each of which is tuned to a sharp frequency, and connected with an indicating instrument of infinite time-constant. To simplify matters we take instead a bank of reeds, or other resonators, each tuned to a narrow waveband, with equally spaced resonant frequencies. It is known that such an instrument gives only an analysis of the energy spectrum, as it cannot distinguish phases, but this will be sufficient for the purpose of discussion. Let us compare this instrument with a real oscillograph, which responds only to a certain range of frequencies $(f_2 - f_1)$. For simplicity it has been assumed in Fig. 1.4 that the bank of reeds extends over the same range, and that the time-constant of the reeds is about equal to the duration of the signal.

We know that any instrument, or combination of instruments, cannot obtain more than at most $2(f_2 - f_1)\tau$ independent data from the area $(f_2 - f_1)\tau$ in the diagram. But instead of rigorously independent data, which can be obtained in general only by calculation from the instrument readings, it will be more convenient for the moment to consider "practically" independent data, which can be obtained by direct readings. For any resonator, oscillograph or reed, a damping time can be defined, after which oscillations have decayed by, say, 10 db. Similarly one can define a tuning width as, say, the number of cycles off resonance at which the response falls off by 10 db. It is well known that in all types of resonators there is a relation between these two of the form:

Decay time × Tuning width = Number of the order one.

This means that for every type of resonator a *characteristic rectangle* of about unit area can be defined in the time/frequency diagram, which corresponds to one "practically" independent reading of the instrument. In order to obtain their number, we must divide up the (time × frequency) area into such rectangles. This is illustrated in Figs. 1.4(*a*) and 1.4(*b*). In the case of the oscillograph the rectangles are broad horizontally and narrow vertically; for the tuned reeds the reverse. The amplitude of the readings is indicated by shading of different density. Negative amplitudes are indicated by shading of

* Carson proposed the concept of a "generalized frequency" in 1922, and in 1937 elaborated it further with T. C. Fry under the name of "instantaneous frequency" (Ref. No. 1.8). This is a useful notion for slowly-varying frequencies, but not sufficient to cover all cases in which physical feeling and the Fourier integral theorem are at variance.

† Campbell and Foster call this an $\mathfrak{S}_0$ function, but the name "delta function" as used by Dirac has now wider currency.

opposite inclination. We will return later to the question of a suitable convention for measuring these amplitudes.*

Without going into details, it is now evident that physical instruments analyse the time-frequency diagram into rectangles which have shapes dependent on the nature of the instrument and areas of the order unity, but not less than one-half. The number of these rectangles in any region is the number of independent data which the instrument can obtain from the signal, i.e. proportional to the amount of information. This justifies calling the diagram from now on the "*diagram of information.*"

We may now ask what it is that prevents any instrument from analysing the information area with an accuracy of less than a half unit. The ultimate reason for this is evident. We have made of a function of one variable—time *or* frequency—a function of two variables—time *and* frequency. This might be considered a somewhat artificial process, but it must be remembered that it corresponds very closely to our subjective interpretation of aural sensations. Indeed, Fig. 1.4(*b*) could be considered as a rough plan of analysis by the ear; rather rough, as the ear is too complicated an instrument to be replaced by a bank of tuned reeds, yet much closer than either the oscillogram or the Fourier spectrum. But as a result of this doubling of variables we have the strange feature that, although we can carry out the analysis with any degree of accuracy in the time direction or in the frequency direction, we cannot carry it out simultaneously in both beyond a certain limit. This strange character is probably the reason why the familiar subjective pattern of our aural sensations and their mathematical interpretation have hitherto differed so widely. In fact the mathematical apparatus adequate for treating this diagram in a quantitative way has become available only fairly recently to physicists, thanks to the development of quantum theory.

The linkage between the uncertainties in the definitions of "time" and "frequency" has never passed entirely unnoticed by physicists. It is the key to the problem of the "coherence length" of wave-trains, which was thoroughly discussed by Sommerfeld in 1914.† But these problems came into the focus of physical interest only with the discovery of wave mechanics, and especially by the formulation of Heisenberg's principle of indeterminacy in 1927. This discovery led to a great simplification in the mathematical apparatus of quantum theory, which was recast in a form of which use will be made in the present paper.

The essence of this method—due to a considerable part to W. Pauli‡—is a re-definition of all observable physical quantities in such a form that the physical uncertainty relations which obtain between them appear as direct consequences of a mathematical identity

$$\Delta t \Delta f \simeq 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.2)$$

Δt and Δf are here the uncertainties inherent in the definitions of the epoch t and the frequency f of an oscillation. The identity (1.2) states that t and f cannot be simultaneously defined in an exact way, but only with a latitude of the order one in the product of uncertainties.

Though this interpretation of Heisenberg's principle is now widely known, especially thanks to popular expositions of quantum theory,* it appears that the identity (1.2) itself has received less attention than it deserves. Following a suggestion by the theoretical physicist A. Landé, in 1931 G. W. Stewart brought the relation to the notice of acousticians, in a short note†—to which we shall return in Part 2—but apparently without much response. In communication theory the intimate connection of the identity (1.2) with the fundamental principle of transmission appears to have passed unnoticed.

Perhaps it is not unnecessary to point out that it is not intended to explain the transmission of information by means of quantum theory. This could hardly be called an explanation. The foregoing references are merely an acknowledgment to the theory which has supplied us with an important part of the mathematical methods.

(3) THE COMPLEX SIGNAL

In order to apply the simple and elegant formalism of quantum mechanics, it will be convenient first to express the signal amplitude $s(t)$ in a somewhat different form.

It has long been recognized that operations with the complex exponential $e^{j\omega t}$—often called cis ωt—have distinct advantages over operations with sine or cosine functions. There are two ways of introducing the complex exponential. One is to write

$$\cos \omega t = \tfrac{1}{2}(e^{j\omega t} + e^{-j\omega t}) \quad \sin \omega t = \frac{1}{2j}(e^{j\omega t} - e^{-j\omega t}) \quad . \quad (1.3)$$

This means that the harmonic functions are replaced by the resultant of two complex vectors, rotating in opposite directions. The other way is to put

$$\cos \omega t = \mathscr{R}(e^{j\omega t}) \quad \sin \omega t = -\mathscr{R}(je^{j\omega t}) \quad . \quad . \quad (1.4)$$

In this method the harmonic functions are replaced by the real part of a single rotating vector. Both methods have great advantages against operation with real harmonic functions. Their relative merits depend on the problem to which they are applied. In modulation problems, for instance, the advantage is with the first method. On the other hand, the formalism of quantum mechanics favours the second method, which we are now going to follow. This means that we replace a real signal of the form

$$s(t) = a \cos \omega t + b \sin \omega t \quad . \quad . \quad . \quad . \quad (1.5)$$

by a complex time function

$$\psi(t) = s(t) + j\sigma(t) = (a - jb)e^{j\omega t} \quad . \quad . \quad (1.6)$$

which is formed by adding to the real signal $s(t)$ an imaginary signal $j\sigma(t)$. The function $\sigma(t)$ is formed from $s(t)$ by replacing $\cos \omega t$ by $\sin \omega t$ and $\sin \omega t$ by $-\cos \omega t$. The function $\sigma(t)$ has a simple significance. It represents the signal *in quadrature* to $s(t)$ which, added to it, transforms the oscillating into a rotating vector. If, for instance, $s(t)$ is applied to two opposite poles of a four-pole armature, $\sigma(t)$ has to be applied to the other pair in order to produce a rotating field.

If $s(t)$ is not a simple harmonic function, the process by which $\psi(t)$ has been obtained can be readily generalized. We have only to express $s(t)$ in the form of a real Fourier integral, replace every cosine in it by $e^{j\omega t}$, and every sine by $-je^{j\omega t}$. This process becomes very simple if, instead of sine and cosine Fourier integrals, the complex (cisoidal) Fourier integrals are

* *Note added 7th February*, 1946. An instrument called the "Sound Spectrograph" has been developed by the Bell Telephone Laboratories for the recording of sound patterns in two-dimensional form. The first publications have just appeared; POTTER, R. K.: "Visible Patterns of Sound," *Science*, 9th November, 1945, and "Visible Speech," *Bell Laboratories Record*, January 1946.

† SOMMERFELD, A.: *Annalen der Physik*, 1914, **44**, p. 177.
Another field of classical physics in which an uncertainty relation is of great importance is Brownian motion. Cf. FÜRTH, R.: "On Some Relations between Classical Statistics and Quantum Mechanics," *Zeitschrift für Physik*, 1933, 81, p. 143, and BOULIGAND, G.: "Relations d'Incertitude en Géometrie et en Physique" (Hermann et Cie, Paris, 1934).

‡ PAULI, W.: "Handbuch der Physik," vol. 24/1, 2nd ed. (Berlin, 1933). A very lucid exposition of quantum mechanics on these lines is given by TOLMAN, R. C.: "The Principles of Statistical Mechanics" (Oxford, 1938), pp. 189–276. In Dirac's system Pauli's postulates appear as results, derived from another set of postulates. Cf. DIRAC, P. A. M.: "Quantum Mechanics," 2nd ed. (Oxford, 1938), p. 103.

* SCHRÖDINGER, E.: "Science and the Human Temperament" (Allen and Unwin, 1935), pp. 126–129. LINDEMANN, F. A.: "The Physical Significance of Quantum Theory" (Oxford, 1932), pp. 126–127. DARWIN, C. G.: "The New Conceptions of Matter" (G. Bell and Sons, 1931), pp. 78–102.

† STEWART, G. W.: Ref. No. 1.9.
A. Landé has made use of acoustical examples to illustrate the uncertainty relation in his "Vorlesungen über Wellenmechanik" Akademische Verlagsges (Leipzig, 1930), pp. 17–20.

used according to equation (1.1). In this case the passage from $s(t)$ to $\psi(t)$ is equivalent to the instruction: *Suppress the amplitudes belonging to negative frequencies, and multiply the amplitudes of positive frequencies by two.* This can be readily understood by comparing equations (1.3) and (1.4).

Though the Fourier transform of $\psi(t)$ is thus immediately obtained from the Fourier transform of $s(t)$, to obtain $\psi(t)$ itself requires an integration. It can be easily verified that the signal $\sigma(t)$ associated with $s(t)$ is given by the integral

$$\sigma(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} s(\tau)\frac{d\tau}{\tau - t} \quad . \quad . \quad . \quad . \quad (1.7)$$

This is an improper integral, and is to be understood as an abbreviation of the following limit

$$\int_{-\infty}^{\infty} = \lim_{\varepsilon=0}\left[\int_{-\infty}^{t-\varepsilon} + \int_{t+\varepsilon}^{\infty}\right],$$

which is called "Cauchy's principal value" of an improper integral.* To verify equation (1.7) it is sufficient to show that it converts $\cos \omega t$ into $\sin \omega t$ and $\sin \omega t$ into $-\cos \omega t$. Conversely $s(t)$ can be expressed by $\sigma(t)$ as follows:—

$$s(t) = -\frac{1}{\pi}\int_{-\infty}^{\infty} \sigma(\tau)\frac{d\tau}{\tau - t} \quad . \quad . \quad . \quad . \quad (1.8)$$

Associated functions $s(t)$ and $\sigma(t)$ which satisfy the reciprocal relations (1.7) and (1.8) are known as a pair of "*Hilbert transforms.*"†

Pairs of signals in quadrature with one another can be generated by taking an analytical function $f(z)$ of the complex variable $z = x + jy$, which can be expressed in the form $f(z) = u(x,y) + jv(x,y)$. Provided that there are no poles at one side of the x-axis (and if certain other singularities are excluded), $u(x,0)$ and $v(x,0)$ will be in quadrature. The function e^{jz} is an example which gives $u(x,0) = \cos x$ and $v(x,0) = \sin x$. It follows that, as the real axis is in no way distinguished in the theory of analytical functions of a complex variable, we can draw any straight line in the complex plane which leaves all the poles at one side, and the values of the two conjugate functions along this line will give a pair of functions in quadrature.

An example of two functions in quadrature is shown in Fig. 1.5. In spite of their very different forms they contain the

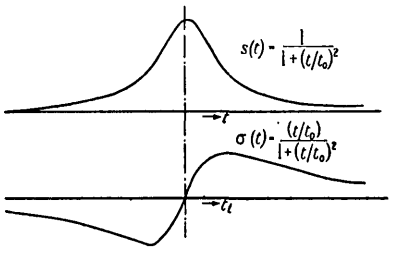

Fig. 1.5.—Example of signals in quadrature.

same spectral components. If these functions were to represent amplitudes of sound waves, the ear could not distinguish one from the other.‡

A mechanical device for generating the associated signal $\sigma(t)$ to a given signal $s(t)$ is described in Appendix 9.2, which contains also a discussion of the problem of single-sideband generation.

* WHITTAKER, E. T., and WATSON, G. N.: "Modern Analysis," 4th ed. (Cambridge), p. 75.

† Cf. TITCHMARSH, E. C.: "Introduction to the Theory of Fourier Integrals" (Oxford, 1937).

‡ Provided that Ohm's law of hearing holds with sufficient accuracy. Such associated signals could be used for testing the limits of validity of Ohm's law.

(4) EXACT FORMULATION OF THE UNCERTAINTY RELATION

By means of the complex signal $\psi(t)$ it is now easy to frame the uncertainty relation in a quantitative manner, using the formalism of quantum mechanics. In order to emphasize the analogy, the same symbol ψ has been chosen for the complex signal as is used in that theory for the "wave" or "probability" amplitudes.

$\psi(t)$ is the time description of the signal. We can associate with this its frequency description by means of its Fourier transform $\phi(f)$, which will also be called the "spectrum" of $\psi(t)$. The two descriptions are connected by the reciprocal Fourier relations

$$\psi(t) = \int_{-\infty}^{\infty} \phi(f)e^{2\pi jft}df \quad . \quad . \quad . \quad . \quad (1.9)$$

$$\phi(f) = \int_{-\infty}^{\infty} \psi(t)e^{-2\pi jft}dt \quad . \quad . \quad . \quad . \quad (1.10)$$

In order to emphasize the symmetry, the first integral has been also written with limits $-\infty$ and ∞, although we have specified $\psi(t)$ in such a way that $\phi(f) = 0$ for negative frequencies; hence we could have taken zero as the lower limit. As in the following all integrals will be taken in the limits $-\infty$ to ∞, the limits will not be indicated in the formulae.

In Section 1 several methods have been discussed for specifying a signal by an infinite set of denumerable (countable) data. One of these was specification by moments, M_0, M_1 . . . This method, with some modifications, will be the best suited for quantitative discussion. The first modification is that it will be more convenient to introduce instead of $s(t)$ the following "weight function":—

$$\psi^*(t)\psi(t) = [s(t)]^2 + [\sigma(t)]^2 \quad . \quad . \quad . \quad (1.11)$$

The asterisk denotes the conjugate complex value. The new weight function is therefore the square of the absolute value of ψ. This can be considered as the "power" of the signal, and will be referred to by this name in what follows. A second convenient modification is that, instead of with the moments themselves, we shall operate with their values divided by M_0, i.e. with the following quotients:—

$$\bar{t} = \frac{\int\psi^* t\psi dt}{\int\psi^*\psi dt} \quad \bar{t}^2 = \frac{\int\psi^* t^2\psi dt}{\int\psi^*\psi dt} \quad . \quad . \quad \bar{t}^n = \frac{\int\psi^* t^n\psi dt}{\int\psi^*\psi dt} \quad . \quad . \quad (1.12)$$

These are the mean values of the "epoch" t of the signal of orders 1, 2 . . . n . . . The factor t^n has been placed between the two amplitude factors to emphasize the symmetry of the formulas with later ones. By a theorem of Stieltjes, if all mean values are known, the weight function $\psi^*\psi = |\psi|^2$ is also determined, apart from a constant factor. The signal ψ itself is determined only as regards absolute value; its phase remains arbitrary. This makes the method particularly suitable, for instance, for acoustical problems. In others, where the phase is observable, it will not be difficult to supplement the specification, as will be shown later.

Similarly we define mean frequencies f^n of the signal as follows:—

$$\bar{f} = \frac{\int\phi^* f\phi df}{\int\phi^*\phi df} \quad \bar{f}^2 = \frac{\int\phi^* f^2\phi df}{\int\phi^*\phi df} \quad . \quad . \quad \bar{f}^n = \frac{\int\phi^* f^n\phi df}{\int\phi^*\phi df} \quad . \quad . \quad (1.13)$$

It now becomes evident why we had to introduce a complex signal in the previous Section. If we had operated with the real signal $s(t)$ instead, the weight function would have been even, and the mean frequency $\bar{f}$ always zero. This is one of the

points on which physical feeling and the usual Fourier methods are not in perfect agreement. But we could eliminate the negative frequencies, only at the price of introducing a complex signal.

As by equations (1.9) and (1.10), ψ and ϕ mutually determine one another, it must be possible to express the mean frequencies by ψ, and, conversely, the mean epochs by ϕ. This can be done indeed very simply by means of the following elegant reciprocal relations:—

$$\int \psi^* \psi dt = \int \phi^* \phi df \quad . \quad . \quad . \quad . \quad . \quad (1.14)$$

$$\int \phi^* f^n \phi df = \left(\frac{1}{2\pi j}\right)^n \int \psi^* \frac{d^n}{dt^n} \psi dt \quad . \quad . \quad . \quad (1.15)$$

$$\int \psi^* t^n \psi dt = \left(\frac{-1}{2\pi j}\right)^n \int \phi^* \frac{d^n}{df^n} \phi df \quad . \quad . \quad . \quad (1.16)$$

The first of these, (1.14), is well known as the "Fourier energy theorem" (Rayleigh, 1889). The other relations can be derived from the identity†

$$\int \psi_1(t) \psi_2(t) dt = \int \phi_1(f) \phi_2(-f) df \quad . \quad . \quad . \quad (1.17)$$

by partial integration, assuming that ψ, ϕ and all their derivatives vanish at infinity.

These very useful reciprocal relations can be summed up in the following simple instructions. When it is desired to express one of the mean values (1.12) by integrals over frequency, replace ψ by ϕ, and the *quantity* t by the *operator* $-\frac{1}{2\pi j}\frac{d}{df}$. This can be called "translation from time language into frequency language." Conversely, when doing the inverse translation, replace ϕ by ψ and the frequency f by the operator $\frac{1}{2\pi j}\frac{d}{dt}$. This corresponds to the somewhat mysterious rule of quantum mechanics: Replace in classical equations the momentum p_x by the operator $\frac{h}{2\pi j}\frac{\partial}{\partial x}$, where x is the co-ordinate conjugate to the momentum p_x. Actually it is no more mysterious than Heaviside's instruction: "Replace the operator d/dt by p," which has long been familiar to electrical engineers.

Applying the rule

$$\bar{f} = \frac{1}{2\pi j} \frac{\int \psi^* \frac{d}{dt} \psi dt}{\int \psi^* \psi dt} \quad . \quad . \quad . \quad . \quad . \quad (1.18)$$

to a simple cisoidal function $\psi = \text{cis}\, 2\pi f_0 t$, we obtain the value f_0 for the mean frequency $\bar{f}$, and similarly $\bar{f^n} = f_0^n$. The mean epochs $\bar{t^n}$, on the other hand, are zero for odd powers, and infinite for even powers $n > 1$. The cisoidal function is to be considered as a limiting case, as the theory is correctly applicable only to signals of finite duration, and with frequency spectra which do not extend to infinity, a condition which is fulfilled by all real, physical signals.

These definitions and rules enable us to formulate the uncertainty relation quantitatively. Let us consider a finite signal, such as is shown, for example, in Fig. 1.6. Let us first fix the mean epoch and the mean frequency of the signal, by means of equations (1.12) and (1.13) or (1.18). These, however, do not count as data, as in a continuous transmission there will be some signal strength at any instant, and at any frequency. We consider $\bar{t}$ and $\bar{f}$ as references, not as data. The first two data will be therefore determined by the mean-square values of epoch and frequency, i.e.

$$\bar{t^2} = \frac{\int \psi^* t^2 \psi dt}{\int \psi^* \psi dt} \quad . \quad . \quad . \quad . \quad . \quad (1.19)$$

† Cf. CAMPBELL and FOSTER: Reference 1.7, p. 39.

$$\bar{f^2} = \frac{\int \phi^* f^2 \phi df}{\int \phi^* \phi df} = -\frac{1}{(2\pi)^2} \frac{\int \psi^* \frac{d^2}{dt^2} \psi dt}{\int \psi^* \psi dt} = \frac{1}{(2\pi)^2} \frac{\int \frac{d\psi^*}{dt} \frac{d\psi}{dt} dt}{\int \psi^* \psi dt}$$

. . . (1.20)

The second of these has been first translated into "time language," as explained, and transformed by partial integration to put its essentially positive character into evidence.

It may be noted that $\bar{t^2}$ and $\bar{f^2}$, and in general all mean values of even order, remain unaltered if the real signal $s(t)$ or its associate, $\sigma(t)$, is substituted in the place of $\psi(t) = s(t) + j\sigma(t)$. Hence in the following we could again use the real instead of the complex signal, but ψ will be retained in order to simplify some of the analytical expressions and to emphasize the similarity with the formulas of quantum mechanics.

We now define what will be called "the effective duration" Δt and the "effective frequency width" Δf of a signal by the following equations

$$\Delta t = [2\pi \overline{(t - \bar{t})^2}]^{\frac{1}{2}} \quad . \quad . \quad . \quad . \quad (1.21)$$

$$\Delta f = [2\pi \overline{(f - \bar{f})^2}]^{\frac{1}{2}} \quad . \quad . \quad . \quad . \quad (1.22)$$

In words, the effective duration is defined as $\sqrt{(2\pi)}$ times the r.m.s. deviation of the signal from the mean epoch $\bar{t}$, and the effective frequency width similarly as $\sqrt{(2\pi)}$ times the r.m.s. deviation from $\bar{f}$. The choice of the numerical factor $\sqrt{(2\pi)}$ will be justified later.

Using the identities

$$\overline{(t - \bar{t})^2} = \bar{t^2} - (\bar{t})^2 \qquad \overline{(f - \bar{f})^2} = \bar{f^2} - (\bar{f})^2$$

Δt and Δf can be expressed by means of (1.19) and (1.20). The expressions are greatly simplified if the origin of the time scale is shifted to $\bar{t}$, and the origin of the frequency scale to $\bar{f}$. Both transformations are effected by introducing a new time scale

$$\tau = t - \bar{t} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.23)$$

and a new signal amplitude

$$\Psi(\tau) = \psi(t) e^{-2\pi j \bar{f} \tau} \quad . \quad . \quad . \quad . \quad (1.24)$$

Expressing t and ψ by the new quantities τ and Ψ, it is found that, apart from a numerical factor 2π, $(\Delta t)^2$ and $(\Delta f)^2$ assume the same form as equations (1.19) and (1.20) for $\bar{t^2}$ and $\bar{f^2}$. Multiplying the two equations we obtain

$$(\Delta t \Delta f)^2 = \frac{1}{4}\left[4 \frac{\int \Psi^* \tau^2 \Psi d\tau \int \frac{d\Psi^*}{d\tau} \frac{d\Psi}{d\tau} d\tau}{[\int \Psi^* \Psi d\tau]^2}\right] \quad . \quad (1.25)$$

But, by a mathematical identity, a form of the "Schwarz inequality" due to Weyl and Pauli,† the expression in brackets is always larger than unity for any function Ψ for which the integrals exist. We obtain, therefore, the uncertainty relation in the rigorous form

$$\Delta t \Delta f \geqslant \tfrac{1}{2} \quad . \quad . \quad . \quad . \quad . \quad (1.26)$$

This is the mathematical identity which is at the root of the fundamental principle of communication. We see that the r.m.s. duration of a signal, and its r.m.s. frequency-width define a minimum area in the information diagram. How large we assume this minimum area depends on the convention for the numerical factor. By choosing it as $\sqrt{(2\pi)} = 2{\cdot}506$ we have made the number of elementary areas in any large rectangular

† WEYL, H.: "The Theory of Groups and Quantum Mechanics" (Methuen, London, 1931), pp. 77 and 393. Cf. also TOLMAN, R. C.: *loc. cit.*, p. 235, and Appendix 9.3 of this paper.

region of the information diagram equal to the number of independent data which that region can transmit, according to the result obtained in Section 1.

Relation (1.26) is symmetrical in time and frequency, and it suggests that a new representation of signals might be found in which t and f played interchangeable parts. Moreover, it suggests that it might be possible to give a more concrete interpretation to the information diagram by dividing it up into "cells" of size one half, and associating each cell with an "elementary signal" which transmitted exactly one datum of information. This programme will be carried out in the next Section.

(5) THE ELEMENTARY SIGNAL

The mathematical developments up to this point have run rather closely on the lines of quantum mechanics. In fact our results could have been formally obtained by replacing a coordinate x by t, the momentum p by f, and Planck's constant h by unity. But now the ways part, as questions arise in the theory of information which are rather different from those which quantum theory sets out to answer.

The first problem arises directly from the inequality (1.26). What is the shape of the signal for which the product $\Delta t \Delta f$ actually assumes the smallest possible value, i.e. for which the inequality turns into an equality?

The derivation of this signal form is contained in Appendix 9.3; only the result will be given here, which is very simple. *The signal which occupies the minimum area $\Delta t \Delta f = \frac{1}{2}$ is the modulation product of a harmonic oscillation of any frequency with a pulse of the form of a probability function.* In complex form

$$\psi(t) = e^{-\alpha^2(t-t_0)^2} \operatorname{cis}(2\pi f_0 t + \phi) \quad . \quad . \quad (1.27)$$

α, t_0, f_0 and ϕ are constants, which can be interpreted as the "sharpness" of the pulse, the epoch of its peak, and the frequency and phase constant of the modulating oscillation. The constant α is connected with Δt and Δf by the relations

$$\Delta t = \sqrt{\left(\frac{\pi}{2}\right)}\frac{1}{\alpha} \qquad \Delta f = \frac{1}{\sqrt{(2\pi)}}\alpha$$

As might be expected from the symmetrical form of the condition from which it has been derived, the spectrum is of the same analytical form

$$\phi(f) = e^{-\left(\frac{\pi}{\alpha}\right)^2(f-f_0)^2} \operatorname{cis}[-2\pi t_0(f-f_0) + \phi] \quad . \quad (1.28)$$

The envelopes of both the signal and its spectrum, or their absolute values, have the shape of probability curves, as illustrated in Fig. 1.6. Their sharpnesses are reciprocal.

Because of its self-reciprocal character, the probability signal has always played an important part in the theory of Fourier transforms. In three recent papers, Roberts and Simmonds have called attention to some of its analytical advantages.[1.11, 1.12, 1.13] But its minimum property does not appear to have been recognized. It is this property which makes the modulated probability pulse the natural basis on which to build up an analysis of signals in which both time and frequency are recognized as references.

It may be proposed, therefore, to call a pulse according to equation (1.27) an *elementary signal.* In the information diagram it may be represented by a rectangle with sides Δt and Δf, and area one-half, centring on the point (t_0, f_0). It will be shown below that any signal can be expanded into elementary signals in such a way that their representative rectangles cover the whole time-frequency area, as indicated in Fig. 1.7. Their amplitudes can be indicated by a number written into the rectangle, or by shading. Each of these areas, with its associated datum, represents, as it were, one elementary

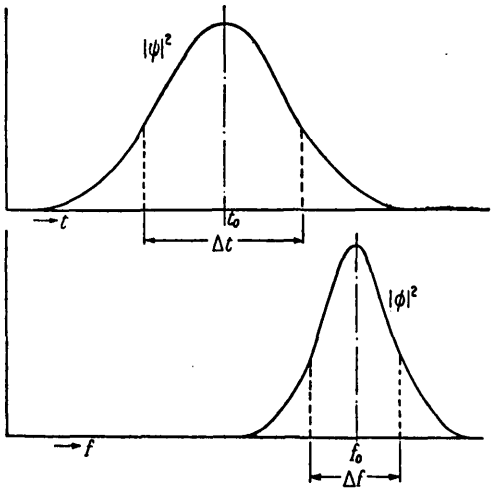

Fig. 1.6.—Envelope of the elementary signal.

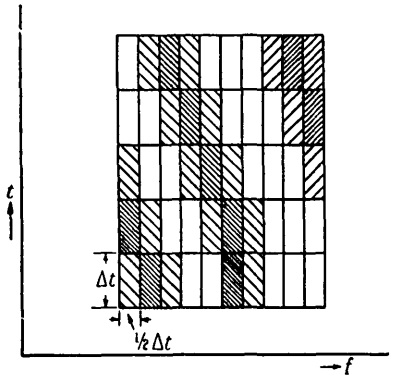

Fig. 1.7.—Representation of signal by logons.

quantum of information, and it is proposed to call it a *logon.* Expansion into elementary signals is a process of which Fourier analysis and time description are special cases. The first is obtained at $\alpha = 0$, in which case the elementary signal becomes a sine wave of infinite length; the second at $\alpha \to \infty$, when it passes into a "delta function."

It will be convenient to explain the expansion into elementary signals in two steps. The first step leads to elementary areas of size unity, with two associated data, but it is simpler and more symmetrical than the second step, which takes us to the limit of sub-division.

This first step corresponds to division of the information area by a network of lines with distances Δt and $1/\Delta t$ respectively, as illustrated in Fig. 1.8.* The elementary areas have suffixes n

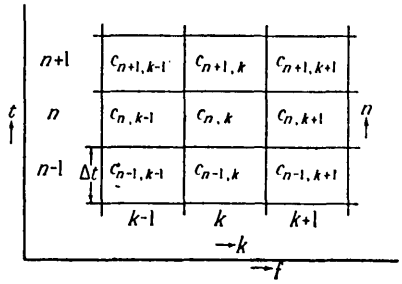

Fig. 1.8.—Representation of signal by a matrix of complex amplitudes.

* For perfect symmetry the spacings in the network ought to have been taken as $(\sqrt{2})\Delta t$ and $1/(\sqrt{2})\Delta t = (\sqrt{2})\Delta f$ respectively.

in the time direction, and k in the frequency direction. The centre lines (horizontally) may be at $t_n = n\,\Delta t$, assuming for convenience that we measure time from the "zero"-th of these lines. The expansion is given by the following formula

$$\psi(t) = \sum_{-\infty}^{\infty} n \sum_{-\infty}^{\infty} k\; c_{nk} \exp - \pi \frac{(t - n\Delta t)^2}{2(\Delta t)^2} \operatorname{cis} (2\pi kt/\Delta t) \qquad (1.29)$$

The matrix of the complex coefficients c_{nk} represents the signal in a symmetrical way, as it is easy to see that if the expansion exists we arrive—apart from a constant factor—at the same coefficients if we expand $\phi(f)$ instead of $\psi(t)$.

As the elementary signals in (1.29) are not orthogonal, the coefficients c_{nk} are best obtained by successive approximations. In the first approximation we consider each horizontal strip with suffix n by itself, and expand the function $\psi(t)$ as if the other strips did not exist, in the interval $(t_n - \frac{1}{2}\Delta t)$ to $(t_n + \frac{1}{2}\Delta t)$, by putting

$$\psi(t) \exp \pi \frac{(t - n\Delta t)^2}{2(\Delta t)^2} = \sum_{0}^{\infty} k\; c_{nk} \operatorname{cis} (2\pi kt/\Delta t)$$

In this formula the exponential function, which is independent of k, has been brought over to the left. We have now a known function on the left, and a Fourier series on the right, which by known methods gives immediately the first approximation for the coefficients c_{nk}. This represents $\psi(t)$ correctly in the intervals for which the series are valid, but not outside them. If the first approximations are added up with summation indices n, there will be a certain error due to their overlap. A second approximation can be obtained by subtracting this error from $\psi(t)$ in eqn. (1.29) and repeating the procedure. It can be expected to converge rapidly, as the exponential factor decays so fast that only neighbouring strips n influence each other perceptibly.

This expansion gives ultimately one complex number c_{nk} for every two elementary areas of size one-half. The real and imaginary parts can be interpreted as giving the amplitudes of the following two real elementary signals

$$\begin{matrix} s_c(t) \\ s_s(t) \end{matrix} = \exp - \alpha^2 (t - t_0)^2 \begin{matrix} \cos \\ \sin \end{matrix} 2\pi f_0 (t - t_0) \quad . \qquad (1.30)$$

where $\alpha^2 = \frac{1}{2}\pi/(\Delta t)^2$. These can be called the "cosine-type" and "sine-type" elementary signals. They are illustrated in Fig. 1.9. We can use them to obtain a real expansion, allocating

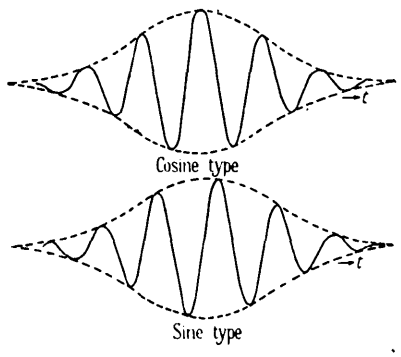

Fig. 1.9.—Real parts of elementary signal.

one datum to every cell of one-half area. But it may be noted that this will have to be necessarily a more special and less symmetrical expansion than the previous one, as the transform of a cosine-type elementary signal, for example, will not in general be of the same type. As always in communication theory, a description by complex numbers is formally simpler than by real data.

We now divide up the information plane as shown in Fig. 1.10

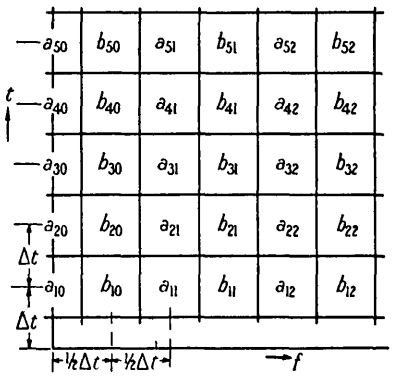

Fig. 1.10.—Expansion of arbitrary signal in cosine-type and sine-type elementary signals.

into cells of size one-half, measuring Δt in the time, and $\frac{1}{2}\Delta t$ in the frequency, direction. Starting from the line of zero frequency, we allocate to these areas in every strip alternately a cosine-type and a sine-type elementary signal. Evidently we must start with a cosine signal at $f = 0$, as the sine-type signal would be zero. This leads us to the following expansion of the real signal $s(t)$:—

$$s(t) = \sum_{-\infty}^{\infty} n \exp - \pi \frac{(t - n\Delta t)^2}{2(\Delta t)^2} \sum_{0}^{\infty} k \left[a_{nk} \cos 2\pi k(t - n\Delta t)/\Delta t \right.$$
$$\left. + b_{nk} \sin 2\pi (k + \tfrac{1}{2})(t - n\Delta t)/\Delta t \right] . \qquad (1.31)$$

In order to find the coefficients a_{nk} and b_{nk} we can carry out the same process of approximation as explained in connection with expansion (1.30), but with a difference. At the first step we arrive at an equation of a form

$$f_n(x) = \sum_{0}^{\infty} k\; a_{nk} \cos kx + b_{nk} \sin (k + \tfrac{1}{2})x$$

with the abbreviations $x = 2\pi(t - n\Delta t)/\Delta t$, and $f(x) = s(t) \exp \frac{1}{2}\pi(t - n\Delta t)^2/(\Delta t)^2$. But the trigonometric series on the right is not a Fourier series. It is of a somewhat unusual type, in which the sine terms have frequencies mid-way between the cosine terms. It will be necessary to show briefly that this series can be used also for the representation of arbitrary functions. First we separate the even and odd parts on both sides of the equation, by putting

$$\tfrac{1}{2}[f_n(x) + f_n(-x)] = \sum_{0}^{\infty} k\; a_{nk} \cos kx$$

$$\tfrac{1}{2}[f_n(x) - f_n(-x)] = \sum_{0}^{\infty} k\; b_{nk} \sin (k + \tfrac{1}{2})x$$

The first is a Fourier series, but not the second. We have seen, however, in Section 3, how all the frequencies contained in a function can be raised by a constant amount by means of a process which involves calculating the function in quadrature with it. Applying this operation to both sides of the last equation we can add $\frac{1}{2}$ to $k + \frac{1}{2}$, and obtain the ordinary Fourier sine series, which enables the coefficients to be calculated.

The expansion into logons is, in general, a rather inconvenient

process, as the elementary signals are not orthogonal. If only approximate results are required, it may be permitted to neglect the effect of their interference. This becomes plausible if we consider that an elementary signal has 76·8% of its energy inside the band Δt or Δf, and only 11·6% on either side. Approximately correct physical analysis could be carried out by means of a bank of resonators with resonance curves of probability shape. It can be shown that if the energy collected by a resonator tuned to f is taken as 100%, the resonators on the right and left of it, tuned to $f+\Delta f$ and $f-\Delta f$, would collect only 0·65% each. Roberts and Simmonds[1.11, 1.12, 1.13] have given consideration to the problem of realizing circuits with responses of probability shape.

Though the overlapping of the elementary signals may be of small practical consequence, it raises a question of considerable theoretical interest. The principle of causality requires that any quantity at an epoch t can depend only on data belonging to epochs earlier than t. But we have seen that we could not carry out the expansion into elementary signals exactly without taking into consideration also the "overlap of the future." In fact, strict causality exists only in the "time language"; as soon as we use frequency as an additional reference the sort of uncertainty occurs which in modern physics has often been called the "breakdown of causality." But rigorous time-analysis is possible only with ideal oscillographs, not with any real physical instrument; hence strict causality never applies in practice. A limitation of this concept ought not to cause difficulties to electrical engineers who are used to the Fourier integral, i.e. to an entirely non-causal method of description.

(6) SIGNALS TRANSMITTED IN MINIMUM TIME

The elementary signals which have been discussed in the last Section assure the best utilization of the information area in the sense that they possess the smallest product of effective duration by effective frequency width. It follows that, if we prescribe the effective width Δf of a frequency channel, the signal transmitted through it in minimum time will have an envelope

$$\Psi(t) = \exp -(2\pi)(\Delta f)^2(t-\bar{t})^2 \quad . \quad . \quad (1.32)$$

and, apart from a cisoidal factor, a Fourier transform

$$\Phi(f) = \exp -\frac{\pi}{2}\left(\frac{f-\bar{f}}{\Delta f}\right)^2 \quad . \quad . \quad . \quad (1.33)$$

But the problem which most frequently arises in practice is somewhat different. Not the effective spectral width is prescribed, but the total width; i.e. a frequency band (f_2-f_1) is given, outside which the spectral amplitude must be zero. What is the signal shape which can be transmitted through this channel in the shortest effective time, and what is its effective duration?

Mathematically the problem can be reduced to finding the spectrum $\phi(f)$ of a signal which makes

$$\Delta t = \frac{1}{(2\pi)^2}\int_{f_1}^{f_2}\frac{d\phi^*}{df}\frac{d\phi}{df}df \Big/ \int_{f_1}^{f_2}\phi^*\phi df \quad . \quad . \quad (1.34)$$

a minimum, with the condition that $\phi(f)$ is zero outside the range f_1-f_2. But this is equivalent to the condition that $\phi(f)$ vanishes at the limits f_1 and f_2. Otherwise, if $\psi(f)$ had a finite value at the limits but vanished outside, the discontinuity at the limits would make the numerator of equation (1.34) divergent. (This is the converse of the well-known fact that a signal with an abrupt break contains frequencies up to infinity, which decay only hyperbolically, not fast enough to make f^2 finite.)

The problem is one of the calculus of variations, and is solved in Appendix 9.4, where it is shown that the signals transmitted in minimum time must be among the solutions of a differential equation

$$\frac{d^2\phi}{df^2} + \Lambda\phi = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.35)$$

where Λ is an undetermined constant. But the possible values of Λ are defined by the auxiliary condition that $\phi(f)$ must vanish at the limits of the waveband.* Hence all admissible solutions are of the form

$$\phi(f) = \sin k\pi\frac{f-f_1}{f_2-f_1} \quad . \quad . \quad . \quad . \quad . \quad (1.36)$$

where k is an integer. We can call this the kth characteristic function of transmission through an ideal band-pass filter. Its effective duration is

$$\Delta t = \sqrt{\left(\frac{\pi}{2}\right)}\frac{k}{f_2-f_1} \quad . \quad . \quad . \quad . \quad . \quad (1.37)$$

and its effective frequency width

$$\Delta f = (f_2-f_1)\sqrt{\left(\frac{\pi}{6}-\frac{1}{\pi k^2}\right)} \quad . \quad . \quad . \quad (1.38)$$

The shortest duration Δt belongs to $k=1$, i.e. to the fundamental characteristic function, which is illustrated in Fig. 1.11.

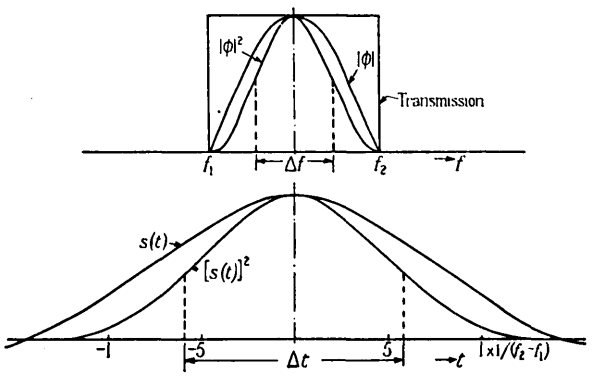

Fig. 1.11.—Spectrum of signal which can be transmitted in minimum time through an ideal band-pass filter, and the signal itself.

The product $\Delta t\Delta f$ is also smallest for $k=1$; its value is 0·571. Though this is not much more than the absolute minimum, 0·5, the transmission channel is poorly utilized, as the effective frequency width is only 0·456 of (f_2-f_1). Practice has found a way to overcome this difficulty by means of asymmetric, vestigial or single-sideband transmission. In these methods the spectrum is cut off at or near the centre more or less abruptly. This produces a "splash," a spreading out of the signal in time, but this effect is compensated in the reception, when the other sideband is reconstituted and added to the received signal.

The advantages of a signal of sine shape, as shown in Fig. 1.11, have already been noticed, as it were, empirically by Wheeler and Loughren† in their thorough study of television images. As in television the signals transmitted represent light intensities, i.e. energies, our definitions must be applied here with a modification. Either the square root of the light intensity must be substituted for ψ, or the square root of the Fourier transform

* Problems of this kind are known in mathematics and theoretical physics as Sturm–Liouville "proper value" problems. Cf. COURANT, R., and HILBERT, D.: "Methoden der mathematischen Physik," Vol. 1 (Springer, Berlin, 1931), or "Interscience" (New York, 1943), p. 249, or any textbook on wave mechanics.

† Ref. No. 14. In comparing the above results with theirs it may be borne in mind that their "nominal cut-off frequency" is one-half of a sideband, and one-quarter of the total channel width.

of the signal for ϕ. The practical difference between these two possible definitions becomes very small in minimum problems. If we adopt the second, we obtain the same "cosine-squared" law for the optimum spectral distribution of energy which Wheeler and Loughren have considered as the "most attractive compromise."

Fig. 1.11 shows also the signal $s(t)$ which is transmitted in minimum time by a band-pass filter. It can be seen that it differs in shape very little indeed from its spectrum. It may be noted that the total time interval in which the signal is appreciably different from zero is $2/(f_2 - f_1)$.

It can be seen from Fig. 1.11, that the optimum signal utilizes the edges of the waveband—in single-sideband television, the upper edge—rather poorly. But this is made even worse in television by the convention of making the electromagnetic amplitudes proportional to the light intensities, so that the electromagnetic energy spectrum in the optimum case has the shape of a $\cos^4$ curve. This means that the higher frequencies will be easily drowned by atmospherics. Conditions can be improved by "compression-expansion" methods, in which, for example, the square root of the light intensity is transmitted, and squared in the receiver.

(7) DISCUSSION OF COMMUNICATION PROBLEMS BY MEANS OF THE INFORMATION DIAGRAM

As the foregoing explanations might appear somewhat abstract, it appears appropriate to return to the information diagram and to demonstrate its usefulness by means of a few examples.

Let us take frequency modulation as a first example. Fig. 1.12

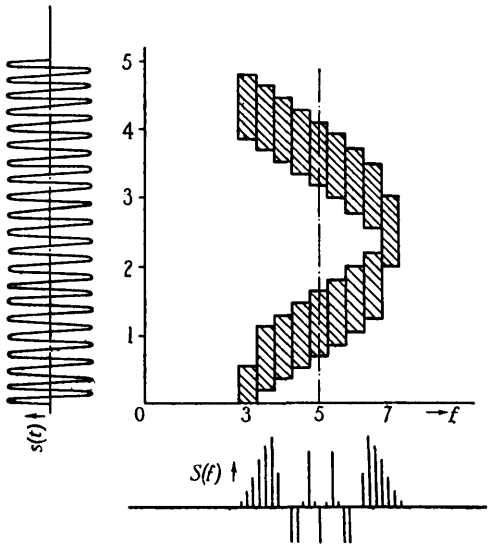

Fig. 1.12.—Three representations of frequency modulation.

contains three different illustrations of the same slowly modulated carrier: the time representation, the spectrum and its picture in the information diagram. It can be seen that the third illustration corresponds very closely to our familiar idea of a variable frequency. The only departure from the naïve expectation that its pictorial representation would be an undulating curve is that the curve has to be thick and blurred. But it appears preferable not to show the blurring, not only because it is difficult to draw, but also because it might give rise to the idea that the picture could be replaced by a definite density distribution. Instead we have represented it by logons of area one-half. The shape of the rectangles, i.e. the ratio $\Delta t/\Delta f$, is entirely arbitrary and depends on the conventions of the analysis. If Δt is taken equal to the damping time of, say, a bank of reeds, the picture gives an approximate description of the response of the instrument. It gives also a rough picture of our aural impression of a siren. How this rough picture can be perfected will be shown in Part 2.

A second example is time-division multiplex telephony, a problem which almost forces on us the simultaneous consideration of time and frequency. Bennett[1.15] has discussed it very thoroughly by an irreproachable method, but, as is often the case with results obtained by Fourier analysis, the physical origin of the results remains somewhat obscure. An attempt will now be made to give them a simple interpretation.

In time-division multiplex telephony, synchronized switches at both ends of a line connect the line in cyclic alternation to a number N of channels. Let f_s be the switching frequency, i.e. the number of contacts made per second. What is the optimum switching frequency if N conversations, each occupying a frequency band w are to be transmitted without loss of information and without crosstalk—i.e. mutual interference between channels—and what is the total frequency-band requirement W?

The information diagram is shown in Fig. 1.13. The fre-

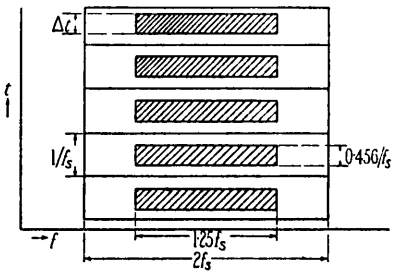

Fig. 1.13.—Information diagram of time-division multiplex-telephony system.

quency band W is sub-divided in the time direction into rectangles of a duration $1/f_s$, i.e. f_s rectangles per sec. If these are to transmit independent data they cannot transmit less than one datum at a time. But one datum, or logon, at a time is also the optimum, as otherwise the receivers would have to discriminate between two or more data in the short time of contact, and distribute them somehow over the long waiting time between two contacts. Hence, if no information is to be lost, the number of contacts per second must be equal to the data of N conversations each of width w, i.e. $f_s = 2Nw$. This is also Bennett's result.

We now consider the condition of crosstalk. This is the exact counterpart of the problem of minimum transmission time in a fixed-frequency channel, considered in the last Section, except that time and frequency are interchanged. Thus we can say at once that the optimum signal form will be the sine shape of Fig. 1.11, and the frequency requirement will be very nearly $2f_s$. The characteristic rectangle $\Delta t \Delta f$ of this signal is shown in every switching period, with the dimensions as obtained in the last Section. The total frequency band requirement becomes $W = 2f_s = 4Nw$. This can be at once halved by single-sideband transmission, i.e. transmitting only one-half of W. But even this does not represent the limit of economy, as the signal is symmetrical not only in frequency, but also in time. In the case of the example treated in the previous Section this was of no use, as the epoch of the signal was unknown. But in time-division multiplex the epoch of each signal is accurately known; hence it must be possible to halve the waveband once more and reduce W to the minimum requirement $W = Nw$. An ingenious,

though rather complicated, method of achieving this, by means of special filters associated with the receiving channels, has been described by Bennett.[1.15]

(8) REFERENCES

(1.1) CARSON, J. R.: "Notes on the Theory of Modulation," *Proceedings of the Institute of Radio Engineers*, 1922, **10**, p. 57.

(1.2) NYQUIST, H.: "Certain Factors affecting Telegraph Speed," *Bell System Technical Journal*, 1924, **3**, p. 324.

(1.3) KÜPFMÜLLER, K.: "Transient Phenomena in Wave Filters," *Elektrische Nachrichten-Technik*, 1924, **1**, p. 141.

(1.4) HARTLEY, R. V. L.: "Transmission of Information," *Bell System Technical Journal*, 1928, **7**, p. 535.

(1.5) GRAY, F., HORTON, J. W., and MATHES, C. R.: "The Production and Utilization of Television Signals," *ibid.*, 1927, **6**, p. 560.

(1.6) LÜSCHEN, F.: "Modern Communication Systems," *Journal I.E.E.*, 1932, **71**, p. 776.

(1.7) CAMPBELL, G. A., and FOSTER, R. M.: "Fourier Integrals for Practical Applications," *Bell Telephone System Monograph* B 584, 1931.

(1.8) CARSON, J. R., and FRY, T. C.: "Variable-frequency Electric Circuit Theory," *Bell System Technical Journal*, 1937, **16**, p. 513.

(1.9) STEWART, G. W.: "Problems Suggested by an Uncertainty Principle in Acoustics," *Journal of the Acoustical Society of America*, 1931, **2**, p. 325.

(1.10) GOLDMARK, P. C., and HENDRICKS, P. S.: "Synthetic Reverberation," *Proceedings of the Institute of Radio Engineers*, 1939, **27**, p. 747.

(1.11) ROBERTS, F. F., and SIMMONDS, J. C.: "Some Properties of a Special Type of Electrical Pulse," *Philosophical Magazine*, (VII), 1943, **34**, p. 822.

(1.12) ROBERTS, F. F., and SIMMONDS, J. C.: "Further Properties of Recurrent Exponential and Probability Waveforms," *ibid.*, (VII), 1944, **35**, p. 459.

(1.13) ROBERTS, F. F., and SIMMONDS, J. C.: "The Physical Realizability of Electrical Networks having Prescribed Characteristics," *ibid.*, (VII), 1944, **35**, p. 778.

(1.14) WHEELER, H. A., and LOUGHREN, A. V.: "The Fine Structure of Television Images," *Proceedings of the Institute of Radio Engineers*, 1938, **26**, p. 540.

(1.15) BENNETT, W. R.: "Time-division Multiplex Systems," *Bell System Technical Journal*, 1941, **20**, p. 199.

(9) APPENDICES

(9.1) Analysis in Terms of Other than Simple Periodic Functions

The discussion in Section 1 suggests a question: Why are we doing our analysis in terms of sine waves, and why do we limit our communication channels by fixed frequencies? Why not choose other orthogonal functions? In fact we could have taken, for example, the orthogonalized Bessel functions

$$\sqrt{(t)}J_n(r_k t/\tau)$$

as the basis of expansion. J_n is a Bessel function of fixed but arbitrary order n; r_k is the kth root of $J_n(x) = 0$; k is the expansion index. These functions are orthogonal in the interval $0 < t < \tau$. The factors r_k/τ have the dimension of a frequency. We could now think of limiting the transmission channel by two "Bessel frequencies," say μ_1 and μ_2. Here the first difference arises. The number of spectral lines between these limits will be the number of the roots of $J_n(x) = 0$ between the limits $\mu_1\tau$ and $\mu_2\tau$. But this number is not proportional to τ. Hence a Bessel channel, or a channel based on any function other than simple harmonic functions, would not transmit the same amount of information in equal time intervals.

In principle it would be possible to construct circuits which transmitted without distortion any member of a selected set of orthogonal functions. But only harmonic functions satisfy linear differential equations in which time does not figure explicitly; hence these are the only ones which can be transmitted by circuits built up of constant elements. Every other system requires variable circuit components, and as there will be a distinguished epoch of time it will also require some sort of synchronization between transmitter and receiver. In competition with fixed-waveband systems any such method will have the disadvantage that wider wavebands will be required to avoid interference with other transmissions. Though this disadvantage —as in the case of frequency modulation—might be outweighed by other advantages, investigation of such systems is outside the scope of the present study, which is mainly devoted to the problem of waveband economy.

(9.2) Mechanical Generation of Associated Signals, and the Problem of Direct Production of Single Sidebands

In order to gain a more vivid picture of signals in quadrature than the mathematical explanations of Section 3 can convey, it may be useful to discuss a method of generating them mechanically. It is obvious from equations (1.7) and (1.8) that, in order to generate the signal $\sigma(t)$ associated with a given signal $s(t)$, it is necessary *to know not only the past but also the future.* Though formally the whole future is involved, the "relevant future" in transmission problems is usually only a fraction of a second. This means that we can produce $\sigma(t)$ with sufficient accuracy if we convert, say, 0·1 sec of the future into the past; in other words, if we *delay* the transmission of $s(t)$ by about this interval. Fig. 1.14 shows a device which might accomplish this.

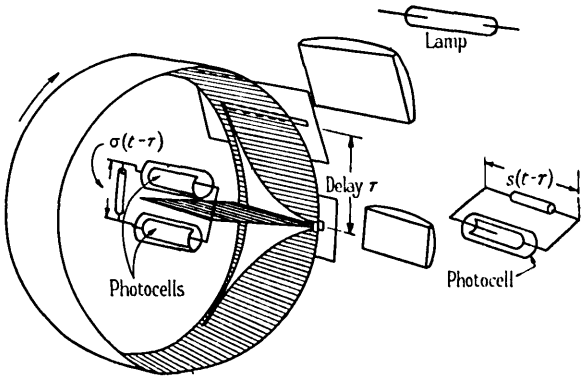

Fig. 1.14.—Device for mechanical generation of a signal in quadrature with a given signal.

The light of a lamp, the intensity of which is modulated by the signal $s(t)$, is thrown through a slit on a transparent rotating drum, coated with phosphorescent powder. The drum therefore carries a record of the signal with it, which decays slowly. After turning through a certain angle the record passes a slit, and here the light is picked up by a photocell, which transmits $s(t)$ with a delay corresponding to the angle.* On the inside of the drum two hyperbolically-shaped apertures are arranged at both sides of the slit opposite to the first photocell. The light from the two hyperbolic windows is collected by two photocells, which are connected in opposition. By comparing this arrangement

* A somewhat similar device (for another purpose) has been described by Goldmark and Hendricks (Ref. No. 1.10).

with equation (1.7) it is easy to see that the difference of the two photocell currents will be proportional to the function in quadrature with $s(t)$.

The complex signal has been discussed at some length as it helps one to understand certain problems of communication engineering. One of these is the problem of single-sideband transmission. It is well known that it is not possible to produce a single sideband directly. The method employed is to produce both sidebands and to suppress one. Equation (1.7) explains the reason. *Direct single-sideband production involves knowledge of the future.* The conventional modulation methods always add *and* subtract frequencies simultaneously. With mechanisms like the one shown in Fig. 1.14 it becomes possible to add *or* subtract them. This means forming the following expression

$$\mathscr{R}[\psi(t) \exp j\omega_c t] = s(t) \cos \omega_c t - \sigma(t) \sin \omega_c t$$

where ω_c is the angular carrier frequency. By substituting a harmonic oscillation for $s(t)$ is is easy to verify that ω_c has been added to every frequency present in the signal. Direct production of single sidebands involves, therefore, the following operations: Modulate the signal with the carrier wave, and subtract from the product the modulation product of the signal in quadrature with the carrier wave in quadrature. It is not, of course, suggested that this might become a practical method; the intention was merely to throw some light on the root of a well-known impossibility.

(9.3) The Schwarz Inequality and Elementary Signals

The inequality

$$(\int \Psi^* \Psi d\tau)^2 \leqslant 4(\int \Psi^* \tau^2 \Psi d\tau)\left(\int \frac{d\Psi^*}{d\tau} \frac{d\Psi}{d\tau} d\tau\right) \qquad (1.39)$$

is valid for any real or complex function Ψ which is continuous and differentiable and vanishes at the integration limits. The following is a modification of a proof given by H. Weyl.†

If a_1, b_1 are two sets of n real or complex numbers, a theorem due to H. A. Schwarz states that

$$|a_1 b_1 + \ldots + a_n b_n|^2 \leqslant (a_1 a_1^* + \ldots + a_n a_n^*)(b_1 b_1^* + \ldots + b_n b_n^*) \quad . \quad (1.40)$$

If a's and b's are all real numbers, this can be interpreted as expressing the fact that the cosine of the angle of two vectors with components $a_1 \ldots a_n$ and $b_1 \ldots b_n$ in an n-dimensional Euclidian space is smaller than unity. This can be easily understood, as in a Euclidian space of any number of dimensions a two-dimensional plane can be made to pass through any two vectors issuing from the origin; hence the angle between them has the same significance as in plane geometry. Equation (1.40) is a generalization of this for "Hermitian" space, in which the components or co-ordinates of the vectors are themselves complex numbers.

By a passage to the limit the sums in (1.40) may be replaced by integrals, so that

$$\Sigma a_1 b_1 \to \int f(\tau) g(\tau) d\tau$$

and similarly for the other two sums. The real variable τ now takes the place of the summation index. The Schwarz inequality now becomes

$$|\int fg\, d\tau|^2 \leqslant (\int ff^* d\tau)(\int gg^* d\tau) \quad . \quad . \quad . \quad (1.41)$$

This remains valid if we replace f and g by their conjugates

$$|\int f^* g^* d\tau|^2 \leqslant (\int ff^* d\tau)(\int gg^* d\tau) \quad . \quad . \quad . \quad (1.42)$$

† WEYL, H.: "The Theory of Groups and Quantum Mechanics" (Methuen, 1931), p. 393.

Adding (1.41) and (1.42) we obtain

$$2(\int ff^* d\tau)(\int gg^* d\tau) \geqslant |\int fg d\tau|^2 + |\int f^* g^* d\tau|^2 \geqslant \tfrac{1}{2}[\int (fg + f^* g^*) d\tau]^2 \quad . \quad (1.43)$$

The second part of this inequality states the fact that the sum of the absolute squares of two conjugate complex numbers is never less than half the square of their sums.

We now put

$$f = \tau\Psi \qquad g = \frac{d\Psi^*}{d\tau} \qquad . \quad . \quad . \quad (1.44)$$

Substitution in (1.43) gives

$$4(\int ff^* d\tau)(\int gg^* d\tau) \geqslant \left[\int \left(\Psi \frac{d\Psi^*}{d\tau} + \Psi^* \frac{d\Psi}{d\tau}\right) \tau d\tau\right]^2 \qquad (1.45)$$

The right-hand side can be transformed by partial integration into

$$\int \left(\Psi \frac{d\Psi^*}{d\tau} + \psi^* \frac{d\Psi}{d\tau}\right) \tau d\tau = \int \tau \frac{d}{d\tau}(\Psi^* \Psi) d\tau = -\int \Psi^* \Psi d\tau \qquad (1.46)$$

where it has been assumed that Ψ vanishes at the integration limits. Substituting this in (1.45) we obtain the inequality (1.39).

In order to obtain the elementary signals we must investigate when this inequality changes into an equality. From the geometrical interpretation of Schwarz's inequality (1.40), it can be concluded at once that the equality sign will obtain if, and only if, the two vectors a, b have the same direction, i.e.

$$b_1 = Ca_1$$

In Hermitian space the direction is not changed by multiplication by a complex number, hence C need not be real.

This condition can be applied also to the inequality (1.39), but with a difference. (1.39) will become an equation only if *both* the conditions (1.41) and (1.42) become equalities; i.e. if the following two equations are fulfilled

$$f = Cg \text{ and } f^* = C'g^* \qquad . \quad . \quad . \quad (1.47)$$

where C and C' are real or complex constants. But these two equations are compatible if, and only if,

$$C' = C^* \qquad . \quad . \quad . \quad . \quad . \quad . \quad (1.48)$$

in which case the two equations (1.47) become identical. On substituting f and g from (1.44) they give the two equivalent equations

$$\frac{d\psi^*}{d\tau} = C\tau\Psi \text{ and } \frac{d\Psi}{d\tau} = C^*\tau\Psi^* \quad . \quad . \quad (1.49)$$

From either of these we can eliminate Ψ or its conjugate Ψ^* and are led to the second-order differential equation

$$\frac{d}{d\tau}\left(\frac{1}{\tau}\frac{d\Psi}{d\tau}\right) = CC^*\tau\Psi \quad . \quad . \quad . \quad . \quad (1.50)$$

Multiplying both sides by $(d\Psi/d\tau)/\tau$, this becomes integrable and gives

$$\left(\frac{1}{\tau}\frac{d\Psi}{d\tau}\right)^2 = CC^*\Psi^2 + \text{const.} \quad . \quad . \quad . \quad (1.51)$$

But the constant is zero, as at infinity both Ψ and $d\Psi/d\tau$ must vanish. We thus obtain the first-order equation

$$\frac{d\Psi}{d\tau} = \pm (CC^*)^{\frac{1}{2}}\tau\Psi \quad . \quad . \quad . \quad . \quad (1.52)$$

with the solution (apart from a constant factor)

$$\Psi = \exp \pm \tfrac{1}{2}|C|\tau^2 \quad . \quad . \quad . \quad . \quad . \quad (1.53)$$

Of the two signs we can retain only the negative one, as otherwise the signal would not vanish at infinity. Putting $\frac{1}{2}|C| = \alpha^2$ we obtain the envelope of the elementary signal. The signal ψ itself results from this by multiplying by cis $2\pi\bar{f}(t - \bar{t})$ and is discussed in Section 5.

It will be useful to sketch briefly the difference between the analysis based on elementary signals and the method of wave mechanics. In the foregoing we have answered the question: What functions Ψ make the product $\Delta f \Delta t$ assume its smallest possible value, i.e. one-half? The question posed by wave mechanics is more general: What functions Ψ makes $\Delta f \Delta t$ a minimum, while fulfilling the condition of vanishing at infinity? This is a problem of the calculus of variations, which leads, instead of to eqn. (1.50), to a more general equation, called the "wave equation of the harmonic oscillator":

$$\frac{d^2\Psi}{d\tau^2} + (\lambda - \alpha^2\tau^2)\Psi = 0$$

where λ and α are real constants. This equation, which contains (1.50) as a special case, has solutions which are finite everywhere and vanish at infinity only if

$$\lambda = \alpha(2n + 1)$$

where n is a positive integer. These "proper" or "characteristic" solutions of the wave equation are (apart from a constant factor)

$$\Psi_n = e^{-\frac{1}{2}\omega^2\tau^2}\frac{d^n}{d\tau^n}e^{-\alpha^2\tau^2}$$

They are known as orthogonal Hermite functions* and form the basis of wave mechanical analysis of the problem of the linear oscillator. They share with the probability function—which can be considered as the Hermite function of zero order—the property that their Fourier transforms are of identical type. The product $\Delta f \Delta t$ for the nth Hermite function is

$$\Delta t \Delta f = \tfrac{1}{2}(2n + 1)$$

That is to say that the Hermite functions occupy in the information diagram areas of size $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}$ . . . Because of their orthogonality Hermite functions readily lend themselves to the expansion of arbitrary signals; hence their importance in wave mechanics. But they are less suitable for the analysis of continuously emitted signals, as they presuppose a distinguished epoch of time $t = 0$, and they do not permit the sub-division of the information area into non-overlapping elementary cells.†

* Also known as parabolical cylinder functions and Weber–Hermite functions Cf. WHITTAKER and WATSON: "Modern Analysis," pp. 231, 347. They are discussed in all textbooks on wave mechanics. Cf. also the study by BABER, T. D. H., and MIRSKY, L.: "Note of Certain Integrals involving Hermite's Polynomials," *Philosophical Magazine* (VII), 1944, 35, p. 532.

† The derivations in this Appendix can be considerably shortened if use is made of the symbolic operator method of quantum mechanics. Cf. MAX BORN: "Atomic Physics" (Blackie, 1935), Appendix XXI, pp. 309–313.

(9.4) Signals Transmitted in Minimum Time through a Given Frequency Channel

It will be convenient to use "frequency language," i.e. to express the signal by its Fourier transform $\phi(f)$. The problem is to make the effective duration Δt of a signal a minimum, with the condition that $\phi(f) = 0$ outside an interval $f_1 - f_2$. Thus

$$\Delta t = \frac{1}{(2\pi)^2 M_0}\int_{f_1}^{f_2}\frac{d\phi^*}{df}\frac{d\phi}{df}df \quad . \quad . \quad . \quad (1.54)$$

must be a minimum, where

$$M_0 = \int_{f_1}^{f_2}\phi^*\phi df$$

This is equivalent to making the numerator in (1.54) a minimum with the auxiliary condition $M_0 = \text{constant}$, and this in turn can be formulated by Lagrange's method in the form

$$\delta\int\left(\frac{d\phi^*}{df}\frac{d\phi}{df} + \Lambda\phi^*\phi\right)df = 0 \quad . \quad . \quad . \quad (1.55)$$

where Λ is an undetermined multiplier. The variation of the first term is

$$\begin{aligned}\delta\int\frac{d\phi^*}{df}\frac{d\phi}{df}df &= \int\left(\frac{d\phi^*}{df}\delta\frac{d\phi}{df} + \frac{d\phi}{df}\delta\frac{d\phi^*}{df}\right)df = \int\left(\frac{d\phi^*}{df}\frac{d\delta\phi}{df} + \frac{d\phi}{df}\frac{d\delta\phi^*}{df}\right)df \\ &= \left[\frac{d\phi^*}{df}\delta\phi + \frac{d\phi}{df}\delta\phi^*\right]_{f_1}^{f_2} - \int\left(\frac{d^2\phi^*}{df^2}\delta\phi + \frac{d^2\phi}{df^2}\delta\phi^*\right)df \quad (1.56)\end{aligned}$$

But at the limits ϕ must vanish, as it is zero outside the interval and must be continuous at the limit, as otherwise the integral (1.54) would not converge. Hence we have here $\delta\phi = \delta\phi^* = 0$, and the first term vanishes. The variation of the second term in (1.55) is

$$\Lambda\int(\phi^*\delta\phi + \phi\delta\phi^*)df \quad . \quad . \quad . \quad . \quad (1.57)$$

The condition (1.55) thus gives

$$\int\left[\left(\frac{d^2\phi^*}{df^2} + \Lambda\phi^*\right)\delta\phi + \left(\frac{d^2\phi}{df^2} + \Lambda\phi\right)\delta\phi^*\right]df = 0 \quad (1.58)$$

and this can be identically fulfilled for arbitrary variations $\delta\phi$ if, and only if,

$$\frac{d^2\phi}{df^2} + \Lambda\phi = 0 \quad . \quad . \quad . \quad . \quad (1.59)$$

This is the differential equation which has to be satisfied by the signal transmitted in minimum time. Its solution is discussed in Section 6.