UNIVERSAL
LIBRARY

INTERNATIONAL TRACTS IN

COMPUTER SCIENCE AND TECHNOLOGY AND THEIR APPLICATION

VOLUME 5

EXPERIMENTAL CORRELOGRAMS AND FOURIER TRANSFORMS

EXPERIMENTAL CORRELOGRAMS AND FOURIER TRANSFORMS

N. F. BARBER, M.Sc.

DOMINION PHYSICAL LABORATORY
D.S.I.R., WELLINGTON, NEW ZEALAND

PERGAMON PRESS

OXFORD · LONDON · NEW YORK · PARIS

1961

PERGAMON PRESS LTD.
Headington Hill Hall, Oxford
4 & 5 Fitzroy Square, London W.1

PERGAMON PRESS INC.
122 East 55th Street, New York 22, N.Y.
1404 New York Avenue, N.W., Washington 5 D.C.
Statler Center 640, 900 Wilshire Boulevard
Los Angeles 17, California

PERGAMON PRESS S.A.R.L.
24 Rue des Écoles, Paris Ve

PERGAMON PRESS G.m.b.H.
Kaiserstrasse 75, Frankfurt am Main

Library of Congress Card Number 61-10009,

Printed in the German Democratic Republic

CONTENTS

PREFACE

MANY devices have been invented and used from time to time for finding Fourier transforms and correlograms. This book is intended as a brief summary of them. Fourier theory is also developed from a "physical" point of view, to help a student to gain a working knowledge of it.

Though electronic digital computers are being used more and more, there still seems to be room for relatively inexpensive analogue machines. They can be very simple if no great accuracy is needed. Analogue machines have not the flexibility of a digital computer and each needs to be devised for a special purpose. Yet they can be very speedy and serviceable when data are recorded from the start in the form that the machine accepts.

In the matter of Fourier theory, the writer owes much to G. B. Madella who has put to use in electrical practice, concepts that many physicists and engineers have treated as mathematical fictions.

N. F. BARBER

CHAPTER 1

FOURIER SERIES

WHEN the voice of a singer is caught by a microphone and displayed on a cathode ray oscilloscope, the picture appearing on the screen may resemble the curve in Fig. 1.1. It shows the fluctuating air

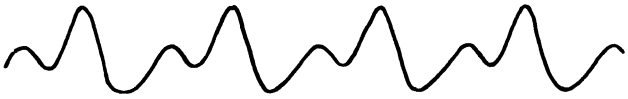

FIG. 1.1. Cyclic variations of air pressure in a musical note.

pressure in the sound. Positive and negative pressures, that is pressures slightly above and slightly below the mean atmospheric pressure, follow in rapid succession, perhaps only 100, perhaps as many as 1000 cycles per second. The pressure goes through a similar cycle of values again and again, though the pattern varies with every

FIG. 1.2. A pure tone.

modulation of the singer's voice. Such repetition is characteristic of a musical sound; a fluctuating air pressure that is not repetitive is a discordant noise.

When a pure tone from an electronic oscillator is examined in the same way, the curve of its fluctuating air pressure resembles that in Fig. 1.2. Again the pattern is repetitive but it has a simpler form. Indeed the ear recognizes this simplicity of the pure tone. By contrast with it, a voice seems to carry "overtones" or additional tones that are higher in the musical scale.

Figure 1.3 shows a number of patterns of pure tones such as might come from different oscillators. If these musical notes were all sounded at the same time, the various pressures would add together giving a sound with a more complicated wave form. The sum of all the simple curves in Fig. 1.3 is in fact the curve shown as a broken line. It closely resembles the pattern shown in Fig. 1.1 for a singer's voice. This idea works in practice; pure tones generated

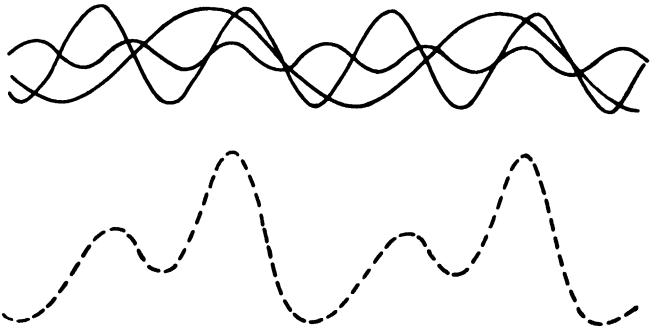

FIG. 1.3. A combination of three pure tones whose frequencies are in simple numerical ratios.

by electronic circuits can be combined to imitate a human voice or the sound of any particular musical instrument. Of course the curves in Fig. 1.3 were specially chosen so that they added up to the curve in Fig. 1, and to imitate any particular sound the pure tones must each have the right pitch and the right strength.

A Geometrical Example of a Pure Tone

When a point such as P in Fig. 1.4 moves at uniform speed around a circular path of radius A, its height above the centre of the circle varies smoothly with time. A curve showing this variation can conveniently be drawn by setting off a number of positions equally spaced along the circular path and then plotting the heights of these points at uniform intervals along a straight line. The curve resembles that of the "pure tone" shown in Fig. 1.2.

The height PQ can be expressed in the formula

$$PQ = A \sin \Phi$$

where the angle Φ is shown in Fig. 1.4. Alternatively if the angle is α at zero time and if point P takes a time T to make one revolution (2π radians) the angle at a time t is $(2\pi t/T + \alpha)$ and the formula for PQ is

$$A \sin (2\pi t/T + \alpha) \tag{1.1}$$

Or one may know that the point makes f revolutions per unit time (f is equal to $1/T$) and the formula is

$$A \sin (2\pi f t + \alpha) \tag{1.2}$$

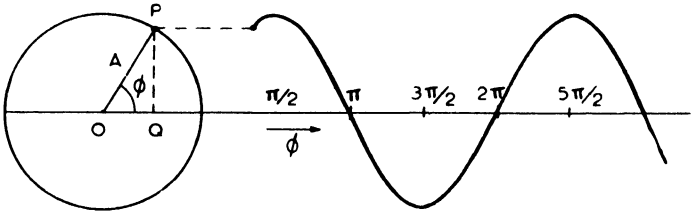

FIG. 1.4. Graphical construction of a sinusoid.

Because these formulae use the "sine" function, the curve in Fig. 1.4 is often called a sinusoidal curve. Three coefficients are enough to describe it; the "amplitude" A, the "phase constant" α, and the "frequency" f (or else the "period" T). In other applications the symbol t need not necessarily represent time. It may for instance be a distance or an angle. In what follows it will be convenient to speak of time as being the variable but the arguments of course are quite general.

The Harmonic Series

In many electrical or mechanical systems it is easy to foretell the effect of applying a sinusoidal voltage or force. It is also often true that, when the applied signal is the sum of a number of sinusoids, the total effect is the sum of the effects they would have if they were

applied separately. One is therefore enabled to predict the effect of a quite complicated signal, providing one can decide what sinusoids compose it.

If one selects sinusoids that individually make one, two or any exact whole number of cycles in a given interval T, it is clear that when added they will give a curve that may look complicated but must be repetitive with period T. If this were done with audible tones the sum would be a musical note and the individual tones could be said to harmonize. So such sinusoids are called "harmonics" and the set of them is a "harmonic series". The characteristic is that their frequencies are integral multiples of some basic frequency. A sinusoid with zero frequency, in other words a constant term, should be included and there is no need as yet to consider sinusoids with negative frequencies. So a complicated signal $g(t)$ that repeats in each interval T can be written as

$$g(t) = A_0 + 2A_1 \sin(2\pi t/T + \alpha_1) + 2A_2 \sin(4\pi t/T + \alpha_2) + \\ + \cdots + 2A_n \sin(2\pi n t/T + \alpha_n) + \cdots \qquad (1.3)$$

The amplitudes A and phase constants α remain to be decided. A factor of 2 has been included in all terms but the first, so the amplitude of the nth sinusoid is actually $2A_n$. This will be found to make later formulae more convenient. It is also a convenience to expand each sinusoid into two terms, writing a_n and b_n for $A_n \sin x_n$ and $A_n \cos x_n$

$$g(t) = a_0 + 2a_1 \cos 2\pi t/T + 2a_2 \cos 4\pi t/T + \cdots \\ + 2b_1 \sin 2\pi t/T + 2b_2 \sin 4\pi t/T + \cdots \qquad (1.4)$$

The unknowns are now the coefficients a and b. The trick for finding them makes use of the "orthogonal" property. This is illustrated in Fig. 1.5, which represents two sinusoids, one making 7 and the other making 8 complete cycles in the interval represented. At the start they are in step, they are out of step near the middle of the interval and get in step again near the end. The *product* of the two is represented in the third curve of this figure. It tends to be positive near the start, negative near the middle and positive again at the

end of the interval. Its average value over the whole interval is zero. The same is true for any pair of different sinusoids as the following equations show, (points indicate "multiply" here).

$$\left.\begin{aligned} \frac{1}{T}\int_0^T \cos 2\pi nt/T \cdot \cos 2\pi mt/T \cdot \mathrm{d}t &= 0, \quad \text{for } n \neq m \\ \frac{1}{T}\int_0^T \cos 2\pi nt/T \cdot \sin 2\pi mt/T \cdot \mathrm{d}t &= 0, \quad \text{for all } n, m \end{aligned}\right\} \quad (1.5)$$

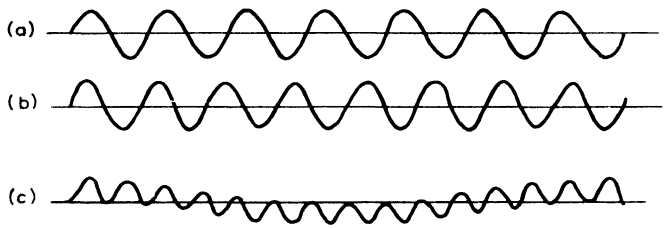

FIG. 1.5. The "orthogonal" property; the product of any two harmonics has a zero average.

But the square of any sinusoid is always positive; its mean is not zero

$$\frac{1}{T}\int_0^T \cos^2 2\pi nt/T \cdot \mathrm{d}t = \tfrac{1}{2} = \frac{1}{T}\int_0^T \sin^2 2\pi nt/T \cdot \mathrm{d}t \quad (1.6)$$

There is an exception if n happens to be zero; the first integral is unity and the second is zero.

To find a_n one may therefore multiply $g(t)$ by $\cos 2\pi nt/T$ and average the product. One can find b_n by a similar process.

$$\left.\begin{aligned} a_n &= \frac{1}{T}\int_0^T g(t) \cdot \cos 2\pi nt/T \cdot \mathrm{d}t \\ b_n &= \frac{1}{T}\int_0^T g(t) \cdot \sin 2\pi nt/T \cdot \mathrm{d}t \end{aligned}\right\} \text{ for } n = 0, 1, 2, \text{etc.} \quad (1.7)$$

These formulae can be verified by substituting the series (1.4) for $g(t)$ in the integrals. It may also be remarked that the limits in the

integrations (1.5), (1.6) and (1.7) need not be 0 and T but could equally well be $-\frac{1}{2}T$ and $\frac{1}{2}T$ or indeed any values that differ by a complete period T.

If the a and b coefficients are calculated in this way the corresponding A and z constants in Eqn. (1.3) can be calculated from the relations

$$\begin{aligned} a_n &= A_n \sin \alpha_n & a_0 &= A_0 \\ b_n &= A_n \cos z_n & b_0 &= 0 \end{aligned} \tag{1.8}$$

Mathematical Examples

The wave forms that an experimenter may wish to resolve into Fourier series do not usually lend themselves to being expressed by a mathematical formula. When they do so, however, it is often possible to calculate the coefficients of the Fourier series. The following example will be of use in later chapters.

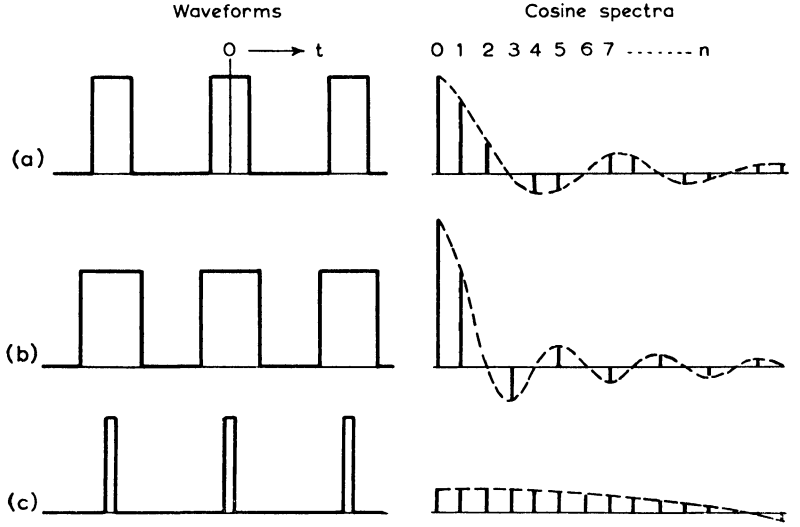

FIG. 1.6. Rectangular waveforms and their spectra. The origin of time has been chosen symmetrically so that only the cosine spectra need be represented.

A square-topped wave form. The wave form is pictured in Fig. 1.6. The function takes a value 1 during a fraction r of the cycle, but for the rest of the cycle it is zero. For analysis it is convenient to measure time from the mid point of the "unity" period, as the figure shows. The function is symmetrical about this point in the sense that its value at any instant t is the same as its value at the preceding time $-t$. Because sine functions are antisymmetrical, that is

$$\sin \theta = -\sin(-\theta)$$

every sine coefficient b_n that might be calculated from formula (1.7) would be zero.

For the cosine coefficients integration only proceeds from $-\frac{1}{2}rT$ to $\frac{1}{2}rT$ since $g(t)$ is zero before and after.

$$a_n = \frac{1}{T}\int_{-\frac{1}{2}rT}^{\frac{1}{2}rT} \cos 2\pi nt/T \cdot \mathrm{d}t \qquad r \cdot \frac{\sin \pi rn}{\pi rn} \tag{1.9}$$

This quantity is graphed as the broken line in Fig. 1.6 (a). It is greatest at n zero and dies away with oscillations. The zeros occur at every whole number value of rn. The Fourier coefficients are the the ordinates drawn where n is any whole number. The figure illustrates the spectrum for the case where r is $\frac{1}{3}$, so ordinates are zero where n is 3, 6, 9 ... etc.

Two special cases are of interest. They are shown in Fig. 1.6 (b) and (c). When r is $\frac{1}{2}$, the even order harmonics are absent. When r is very small, the harmonics all tend to be of equal amplitude; the function itself is a succession of very brief pulses and the spectrum has that form also, so the spectrum of a "comb" is another "comb".

Some Typical Devices for Calculating Fourier Coefficients:

MAX HARTENHEIM (1917)

This is a purely mechanical analyser. Hartenheim was studying the cycle of pressure in an internal combustion engine and his arrangements illustrate how the processes of recording can be ar-

ranged to facilitate the process of analysis. The varying pressure was detected electrically and made to deflect the spot of an oscilloscope. The spot was photographed on a sheet of photographic paper rotating about a centre which was offset along the line of deflection of the spot. The paper was arranged to rotate at the same frequency as the explosive cycles. The record then shows a polar coordinate

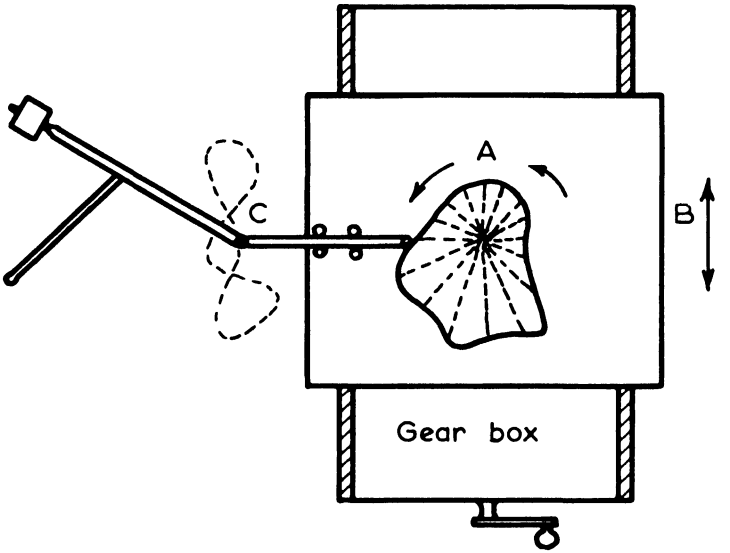

FIG. 1.7. Hartenheim (1917). The function is represented by the cam *A* which rotates once while the carriage *B* moves to and fro in cycles of simple harmonic motion. A planimeter measures the area of the curve traced out by the end point of the rod *C* that rides on the cam.

plot of the cyclic pressure variation. A cardboard cam is copied from the photograph and mounted on the turntable shown in Fig. 1.7.

This turntable is given two motions through gears driven from a single hand wheel. One motion is a rotation and the other is a simple harmonic translation by which it moves bodily to and fro. By a change of gears the lateral motion can be made to execute up to ten complete oscillations during one rotation of the table. The pointer of a planimeter is attached to a rod moved by the cam, and the area recorded by the planimeter is read after one complete rotation of

the table. This area is a measure of the amplitude of the appropriate harmonic. If t is the fraction of a whole turn made by the table, the lateral oscillation which is n times as frequent gives the pointer a displacement in, say, the x direction equal to $A \sin 2\pi nt$. Here A is the amplitude of the lateral motion and its magnitude is always the same. At the same time the cam gives the pointer a deflection in a direction y at right angles to x, proportional to the pressure $g(t)$ that has been recorded say

$$y = B \cdot g(t)$$

Here the constant B depends upon the deflectional sensitivity of the oscilloscope. The total area is proportional to the amplitude a_n in the harmonic series, for

$$\text{Area} = \int y \, \mathrm{d}x = \int_0^{t=1} Bg(t) \cdot A \, 2\pi n \cos 2\pi nt \cdot \mathrm{d}t = A B \, 2\pi n a_n$$

If the process is repeated with the carriage geared to have a displacement $A \cos 2\pi nt$, the planimeter measures an area equal to $A B \, 2\pi n b_n$ proportional, that is, to the harmonic amplitude b_n. The areas must be divided by the constant factor $2\pi A B$ and by the integer n to get the actual amplitudes a_n and b_n.

VASILESCO, V. (1934)

This is an electrical method. A signal in the form of a fluctuating voltage is got from a rotary switch that picks up in succession the outputs of potentiometers adjusted to represent the successive ordinates of the curve that is being analysed. One cycle of this voltage signal is applied to a ballistic galvanometer through a varying shunt resistance and a reversing switch that cause the current in the galvanometer to be the product of the applied voltage and a sinusoidal factor. The cycle is executed in such a short time that the throw of the galvanometer is a measure of the integrated current and is consequently a measure of one Fourier harmonic of the signal. Different harmonics are examined in succession by altering the gear ratios that drive the varying resistance.

G. V. Bekesy (1937)

This is an optical method, and the apparatus is sketched in Fig. 1.8. The wave cycle which is to be analysed is represented as an illuminated profile upon a dark background. A real image of this display is formed by a cylindrical lens which has its axis at right angles to the time axis of the profile. The image is therefore quite unfocussed along lines perpendicular to the time axis and the wave cycle is represented by bands of varying light intensity instead of by a profile. This image is arranged to be quite small, about an inch square, and a series of sinusoidal masks marked on successive frames of 35 mm film are arranged in turn to coincide with the image. At every point along the "time axis" the light that passes the mask is the product of the height of the mask and the intensity of the image formed by the cylindrical lens. The total light that passes is in effect the integrated product of the original wave form and a sinusoid. It is detected by a photocell and indicated on a recording milliammeter.

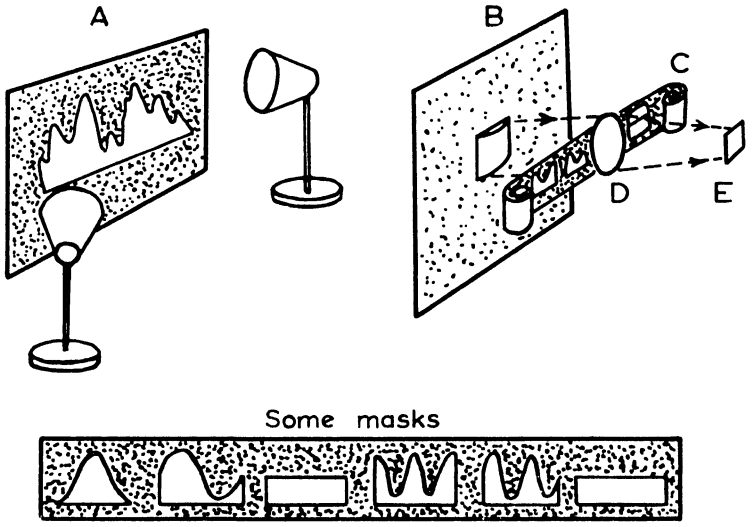

Fig. 1.8. Bekesy (1937). The illuminated profile *A* is viewed by the cylindrical lens *B* and gives an image in which the intensity of illumination varies in imitation of the profile height. This image falls on a sinusoidal mask *C* and a photoelectric cell *E* catches the total transmitted light.

It should be remarked that the sinusoidal masks can only be drawn to represent quantities such as $1 + \cos 2\pi nt/T$ and $1 + \sin 2\pi nt/T$ which are essentially positive. Similarly, the illuminated profile representing a wave form $g(t)$ which may take negative values, is actually drawn to represent $g_0 + g(t)$, where the constant g_0 is large enough to make the total quantity always positive. The product at any position on the t axis is then

$$[g_0 + g(t)]\,[1 + \cos 2\pi nt/T] = g_0 + g(t) + g_0 \cos 2\pi nt/T + \\ + g(t) \cos 2\pi nt/T$$

When all the transmitted light is added on the photocell, the third term in this expansion disappears, since its contributions are alternately positive and negative. The output of the photo electric cell is proportional to

$$\frac{1}{T}\int_0^T [g_0 + g(t)]\,[1 + \cos 2\pi nt/T]\,\mathrm{d}t = g_0 + a_0 + a_n \quad (1.10)$$

To allow for these terms Bekesy follows each sinusoidal mask by a rectangular one representing unity. The integrated product is then

$$\frac{1}{T}\int_0^T [g_0 + g(t)]\,\mathrm{d}t = g_0 + a_0 \quad (1.11)$$

So these unity masks all produce equal deflections of the milliammeter and they provide an artificial zero against which the deflections indicating a_n or b_n are read off from the pen record.

When the silhouette wave form has been prepared and the apparatus has been set up, the process of introducing the various masks and reading off the deflections in the pen record need occupy only a few minutes. Masks can conveniently be made to represent harmonics up to say the twelfth order, that is showing up to 12 complete cycles. After calibrating the apparatus against some wave pattern such as a sinusoid or a square wave whose harmonic content may be calculated, the deflection of the milliammeter provides a measure of the various harmonic amplitudes.

CHAPTER 2

TRANSIENT SIGNALS

THE motion felt in an earthquake builds up and dies away in some twenty or forty or a hundred irregular cycles that show little or no similarity to one another. It will be shown that such a transient

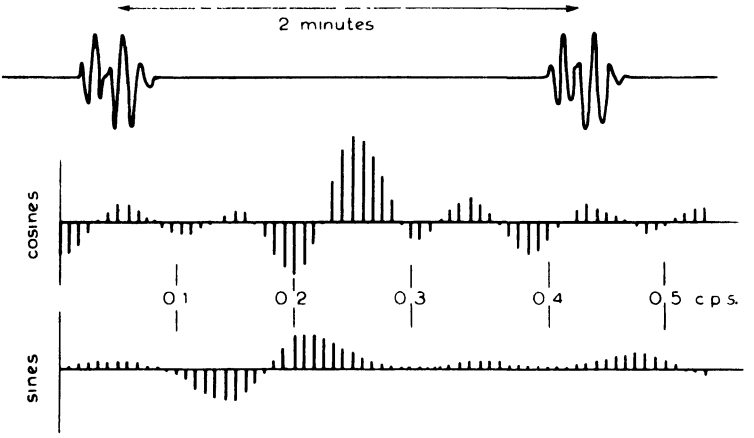

FIG. 2.1. If a brief burst of activity is imagined to repeat itself continually after long equal intervals of time, the harmonics are close together on the frequency scale. They are individually of small amplitude and adjoining harmonics have very similar amplitudes.

activity can also be represented by the sum of sinusoids even though each sinusoid must be thought of as continuing for all time.

Suppose first that the whole earthquake motion were to repeat itself at regular intervals of time; if the duration of the earthquake were 20 sec, suppose that it were to repeat itself every 2 min. This cyclic behaviour is indicated in Fig. 2.1, where activity lasts for

20 sec with a dead period of 100 sec following it, and so on. This cyclic behaviour could of course be built up from a series of harmonic sinusoids. Two things should be noted about this series. First, the harmonics crowd together closely on the scale of frequency; their frequencies must be whole number multiples of the repetition frequency, namely 1/120 cycles/sec, so successive harmonics differ in frequency only by this small amount. Secondly, successive harmonics will tend to have similar amplitudes. To see this, consider two cosine curves that make respectively 50 and 51 cycles in the repetition interval; they are in step at the start of the interval and are only one cycle out of step at the end of 120 sec. During the first 20 sec, when the earthquake is active, they will not differ very much. Since the amplitude of a harmonic is estimated from the mean product between the earthquake motion and the sinusoid, it is seen that the amplitudes of successive harmonics, say the 50th and 51st cosine harmonics, will differ little.

Consequently both the cosine spectrum and the sine spectrum will be a smoothly graded set of amplitudes closely and regularly spaced along the scale of frequency, as Fig. 2.1 suggests. Consider now what happens if the earthquake is supposed to repeat only every 4 min instead of every 2 min. First, the harmonics will crowd together

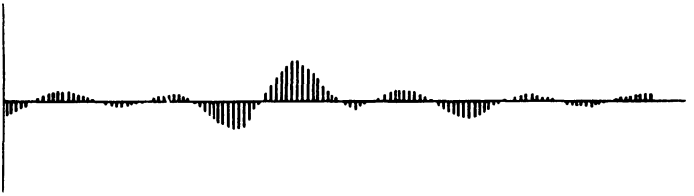

FIG. 2.2. Doubling the length of the repetition interval doubles the number of harmonics occurring per unit interval on the frequency scale but halves the amplitude of those near to any specific frequency.

twice as closely on the scale of frequency, the frequency step being only 1/240 cycles/sec. Secondly, the amplitude at any given frequency will be half what it was before; the product of the earthquake and any given sinusoid is unchanged of course, but its *mean* value is halved when the repetition interval is made twice as long.

Figure 2.2 indicates the new appearance of the cosine spectrum with this longer repetition interval.

When the repetition interval is extended still further, the earthquake activity becomes more and more isolated, more like the transient disturbance it really is. At the same time the harmonics present in any small part of the frequency scale become more and more numerous while the amplitude of each individual harmonic grows less. In the limit it is appropriate to think not of individual harmonics but of the combined effect of all the sinusoids that fall in any small range of frequency. If the repetition interval is T the harmonic frequencies rise in steps of $1/T$ and the number falling in any frequency range from f to $f + \delta f$ will be

$$T\delta f \tag{2.1}$$

They will all have very similar amplitudes. In Chapter 1, it was shown that the nth cosine harmonic has an amplitude $2a_n$

$$2a_n = \frac{2}{T}\int_{-\frac{1}{2}T}^{\frac{1}{2}T} g(t)\cos 2\pi nt/T \cdot \mathrm{d}t \tag{2.2}$$

If the frequency is actually f the order n is $f \div 1/T$ or fT or the nearest integer. Making this substitution the amplitude of the harmonic of frequency f is close to

$$\frac{2}{T}\int_{-\frac{1}{2}T}^{\frac{1}{2}T} g(t)\cos 2\pi ft\,\mathrm{d}t \tag{2.3}$$

Now all the harmonics in the frequency range f to $f + \delta f$ have very similar amplitude and frequency so, using Eqns. (2.1) and (2.3), the sum of them approximates to a sinusoid of amplitude

$$T\delta f\,\frac{2}{T}\int_{-\frac{1}{2}T}^{\frac{1}{2}T} g(t)\cos 2\pi ft\,\mathrm{d}t$$

Of course this amplitude is proportional to the width of the particular frequency range δf. It is reasonable to think in terms of an

"amplitude density", written $2a(f)$, which is such that $2a(f)\delta f$ equals the amplitude of the sinusoid arising from a frequency range f to $f + \delta f$. Comparing this with the previous expression and allowing the repetition interval T to tend to infinity, the amplitude density for cosine components is found to be given by

$$a(f) = \int_{-\infty}^{\infty} g(t) \cos 2\pi f t \, dt \tag{2.5}$$

Similar arguments for the amplitude density of the sine components $b(f)$ gives the result

$$b(f) = \int_{-\infty}^{\infty} g(t) \sin 2\pi f t \, dt \tag{2.6}$$

By definition, of course, a cosine wave of amplitude $2a(f)\delta f$ and a sine wave of amplitude $2b(f)\delta f$ arise from the frequency range f to $f + \delta f$, so the original transient signal $g(t)$ can be written

$$g(t) = 2\int_{0}^{\infty} [a(f) \cos 2\pi f t + b(f) \sin 2\pi f t] \, df \tag{2.7}$$

It is therefore possible to represent a transient signal as the sum of sinusoids by the *Fourier integral representation* summarized in Eqns. (2.5), (2.6) and (2.7).

Symmetry

There is a reciprocal relation between a function and its spectrum that is not obvious from the relation given above, Eqns. (2.5), (2.6) and (2.7). One step towards showing this reciprocity is to think of the function $g(t)$ as being the sum of two functions $g_1(t)$ and $g_2(t)$ of which the first is antisymmetrical about the zero instant of time, and the second is symmetrical. They can be calculated merely by

$$g_1(t) = \tfrac{1}{2}[g(t) - g(-t)] \qquad g_2(t) = \tfrac{1}{2}[g(t) + g(-t)] \tag{2.8}$$

It is evident that $g_1(t)$ must contain all the sine terms in the expansion of $g(t)$, Eqn. (2.7), since sines are antisymmetrical. On the other hand, the symmetrical part $g_2(t)$ contains all the cosine terms.

The second step is imagine the possibility of negative values for the frequency. As the equations stand, the function $g(t)$ extends into both negative and positive time and the integrals in (2.5) and (2.6) have limits $\pm\infty$. On the other hand, the frequency f has been thought of as being always positive. However, if one is able to accept that for purposes of calculation, one can consider negative values of f; it is evident that one may write

$$2\sin 2\pi ft = -\sin 2\pi(-f)t + \sin 2\pi ft$$

The sine terms in Eqn. (2.7) can therefore be thought of as if they ranged in frequency from $-\infty$ to ∞, on the understanding that for negative values of frequency the coefficient is $-b(f)$; that is $b(f)$ is an antisymmetrical function of f. On the other hand

$$2\cos 2\pi ft = \cos 2\pi(-f)t + \cos 2\pi ft$$

so the cosine spectrum extends to negative frequencies on the understanding that the coefficient at negative frequency is $a(f)$, making $a(f)$ a symmetrical function.

Consequently, the antisymmetrical parts of the function and transform are related in the way

$$g_1(t) = \int_{-\infty}^{\infty} b(f)\sin 2\pi ft\,\mathrm{d}f \qquad b(f) = \int_{-\infty}^{\infty} g_1(t)\sin 2\pi ft\,\mathrm{d}t \qquad (2.9)$$

These equations have precisely the same form. They show that one might equally well think of $b(f)$ as being the antisymmetrical function and $g_1(t)$ as being its spectrum.

In a similar way, it follows that the symmetrical parts have a similar relation

$$g_2(t) = \int_{-\infty}^{\infty} a(f)\cos 2\pi ft\,\mathrm{d}f \qquad a(f) = \int_{-\infty}^{\infty} g_2(t)\cos 2\pi ft\,\mathrm{d}t \qquad (2.10)$$

In short, $g_1(t)$ and $b(f)$ are "sine transforms" of each other, while $g_2(t)$ and $a(f)$ are "cosine transforms" of each other.

In Chapter 4 it will be seen that negative frequencies can have a physical meaning when one deals with polyphase electrical signals or with vector quantities that have a direction in a plane as well

as a magnitude. In most experimental applications, the variables one deals with are scalar quantities like lengths, pressures or potentials. The idea of negative frequency is then merely a device to emphasize the mathematical symmetry. In actual calculations the symmetry of the functions means that one need only evaluate the integrals in Eqns. (2.9) and (2.10) between 0 and ∞ and double the result.

Mathematical Examples

When the function can be given by a mathematical expression, the transform can often be deduced. Two examples are discussed in detail below.

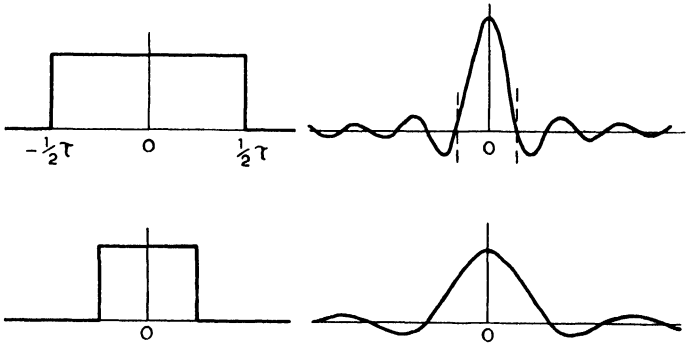

FIG. 2.3. A single rectangular pulse and its cosine transform. As the pulse is made more brief, the transform spreads more widely over the scale of frequency.

A square topped pulse. If the function is unity in the range of t from $-\frac{1}{2}\tau$ to $\frac{1}{2}\tau$ the sine transform is zero and the cosine transform is

$$a(f) = \int_{-\frac{1}{2}\tau}^{\frac{1}{2}\tau} \cos 2\pi f t \, \mathrm{d}t = \tau \frac{\sin \pi f \tau}{\pi f \tau}$$

When the pulse is narrow, the transform is wide, and conversely, as Fig. 2.3 shows. The shape of the transform is like the envelope of the Fourier series that represents a square-topped repetitive wave and which was discussed in Chapter 1. Indeed it will be shown in

Chapter 6 that there is a quite simple connection between the series spectrum representing any repetitive wave form and the integral spectrum representing a transient that shows one whole cycle of the wave form.

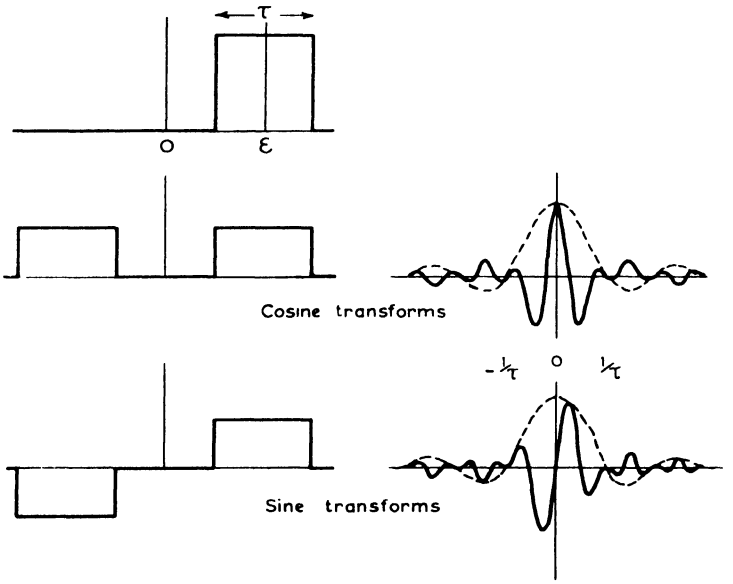

FIG. 2.4. The effect of delaying the pulse by a time interval ε is to modulate the transforms by factors $\cos 2\pi f\varepsilon$ and $\sin 2\pi f\varepsilon$

If the pulse does not happen to occur symmetrically about the time origin but lasts from $-\frac{1}{2}\tau + \varepsilon$ to $\frac{1}{2}\tau + \varepsilon$, it can be separated into symmetrical and antisymmetrical parts as shown in Fig. 2.4. The symmetrical and antisymmetrical parts both have a value $\frac{1}{2}$ in the range $-\frac{1}{2}\tau + \varepsilon$ to $\frac{1}{2}\tau + \varepsilon$. Their transforms are

$$a(f) = 2 \int_{-\frac{1}{2}\tau+\varepsilon}^{\frac{1}{2}\tau+\varepsilon} \tfrac{1}{2} \cos 2\pi f t \, \mathrm{d}t$$

which after some reduction gives

$$a(f) = \tau \cos 2\pi f\varepsilon \cdot \frac{\sin \pi f\tau}{\pi f\tau}$$

and similarly $b(f)$ which varies with frequency f in the manner

$$b(f) = \tau \sin 2\pi f\varepsilon \cdot \frac{\sin \pi f\tau}{\pi f\tau}$$

The second factor has the shape seen in Fig. 2.3 but it is modulated by factors $\cos 2\pi f\varepsilon$ and $\sin 2\pi f\varepsilon$. As the origin is taken further from the mid point of the pulse, these oscillations become more rapid.

The "complex" notation developed in Chapter 4 allows this result to be arrived at more easily.

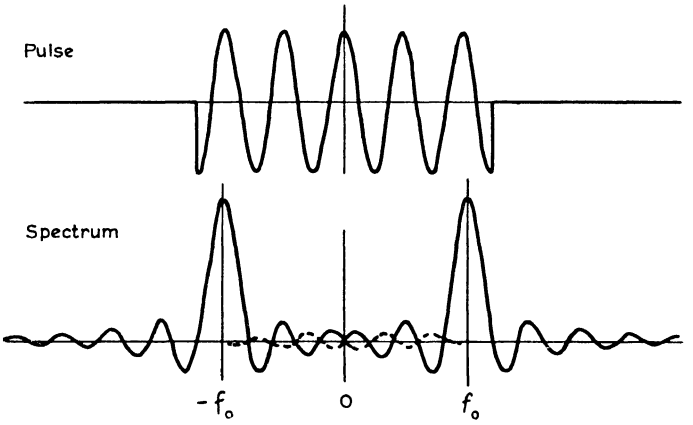

FIG. 2.5. A pulse of waves of frequency f_0 has a transform consisting of two patterns centred on f_0 and $-f_0$.

A wave pulse. Suppose a train of waves represented by $\cos 2\pi f_0 t$ starts abruptly at $-\frac{1}{2}\tau$ and ends abruptly at $\frac{1}{2}\tau$ as in Fig. 2.5. This has symmetry about the time origin and only the cosine transform need be considered. It is

$$\begin{aligned} a(f) &= \int_{-\frac{1}{2}\tau}^{\frac{1}{2}\tau} \cos 2\pi f_0 t \cdot \cos 2\pi f t \, \mathrm{d}t \\ &= \tfrac{1}{2} \int_{-\frac{1}{2}\tau}^{\frac{1}{2}\tau} [\cos 2\pi (f \quad f_0)t + \cos 2\pi (f + f_0)t] \, \mathrm{d}t \\ &= \tfrac{1}{2}\tau \left[\frac{\sin \pi (f - f_0)\tau}{\pi (f - f_0)\tau} + \frac{\sin \pi (f + f_0)\tau}{\pi (f + f_0)\tau} \right] \end{aligned}$$

The transform is therefore the sum of two patterns like that from a square-topped pulse, but centred on the frequencies f_0 and $-f_0$. The effect of cutting off the sinusoid has been to spread out the spectrum around the ordinates at $\pm f_0$ that would represent a pure sinusoid. The central peak of each pattern extends as far as frequencies $\pm 1/\tau$ on either side of its mid point. A short pulse therefore has a wide spread in frequency. A pulse of radio waves lasting for 0·001 sec contains quite a lot of energy in frequencies more than 1000 cycles/sec different from the nominal radio frequency. This is why scientific radio equipment that emits pulses is apt to interfere with radio telephony on a neighbouring frequency-band.

Some Typical Devices for Analysing Transient Signals

Michelson and Stratton

This is a purely mechanical instrument. The transient signal $g(t)$ is measured at a large number of equally spaced instants, $0, \pm\tau, \pm 2\tau, \ldots \pm 80\tau$. The instrument is given the values of g_1 or g_2 (Eqn. 2.8) at τ, $2\tau, \ldots 80\tau$, multiplies them by the appropriate values of $\sin 2\pi ft$ or $\cos 2\pi ft$ and adds the products. The sum, which approximates to $a(f)$ or to $b(f)$ is presented as a curve drawn by a pen on moving paper.

The apparatus is sketched in Fig. 2.6. A crank handle is geared to drive a long shaft on which are fixed side by side some eighty gear wheels graduated in diameter. These each engage with gear wheels that are free to rotate independently on a second shaft. The gear ratios are chosen so that the speeds of rotation of the wheels in this second set are proportional to the integral numbers $1, 2, 3, \ldots, 80$. Arms rising from eccentric mountings on the wheels of this second set move levers that pivot freely about a third shaft. The rocking motion of these levers imitates the sinusoidal factors $\cos 2\pi ft$ or $\sin 2\pi ft$. Cosine factors are represented if the eccentric mountings are adjusted so that at the start of the analysis the levers all have their maximum displacement. If their displacements are all made zero at the start, they imitate sine factors. The motions of these levers are picked up by a set of long light rods whose lower ends

can be set to rest on the levers at any desired distances from the pivot either in front of it or behind it. These distances are in fact made proportional to the measured values of the quantity g_1 or g_2.

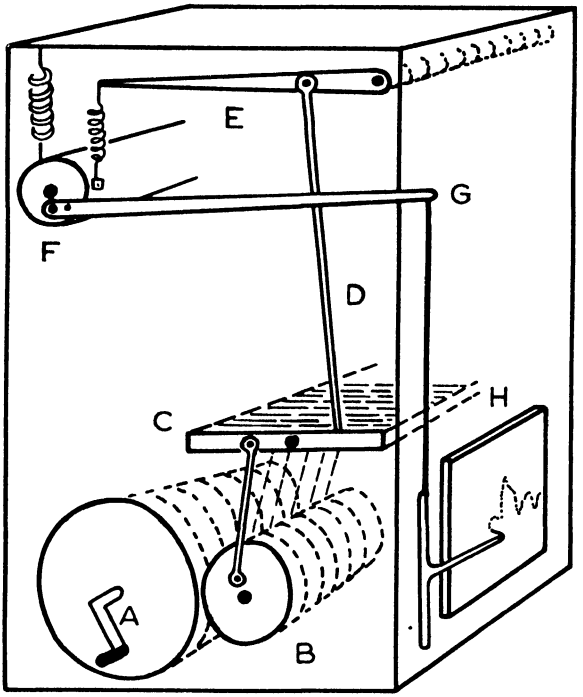

FIG. 2.6. Michelson, Stratton (1898). The handle A drives through different gears the set of eighty wheels B, that rotate at speeds proportional to the integers 1...80. Each wheel rocks its associated lever C and vertical simple harmonic motion is produced in a light rod D whose lower end can be set resting at any point along the lever. The rods extend springs attached to an axle F whose rotation lifts a pen by a lever and cord G. The pen writes on a tablet H that is driven horizontally as the handle A is turned.

The vertical displacements of these rods, now proportional to the product of g_1 or g_2 and a sinusoidal factor, are made to extend short springs all attached to one side of a common axle, whose rotation is opposed by the pull of a single large spring on the other side. If

all the light springs have the same elastic constant it can be shown that the rotation of the common axle is proportional to the sum of the vertical displacements of the long rods. The axle carries an arm that moves a pen vertically up the face of a paper sheet which at the same time is geared to move horizontally through a gearing to the crank handle.

In a sine analysis the pen deflection after the slowest lever has executed a fraction F of a cycle is

$$b(F) = \sum_{0}^{n=80} g_1(n\tau) \sin 2\pi nF$$

In a cosine analysis it is

$$a(F) = \sum_{0}^{n=80} g_2(n\tau) \cos 2\pi nF$$

Comparison with Eqns. (2.9) and (2.10) shows that the frequency f is expressed by F/τ, so turning the handle of the machine corresponds to exploring the scale of frequency. The machine is stopped after the slowest lever has made half a cycle and a linear scale of frequency is drawn on the record between 0 at the start and $\frac{1}{2}\tau$ at the finish.

The true scale of $a(f)$ or $b(f)$ is obtained by knowing that at the start of the cosine analysis the pen deflection is a true $a(f)$ value of

$$a(0) = 2\tau \sum_{1}^{n=80} g_2'(n\tau)$$

In cosine analysis the machine ignores $g_2(0)$ so an amount $\tau\, g_2(0)$ should be added throughout by suitably shifting the zero of the $a(f)$ scale.

Many analysing machines act by synthesizing harmonic sinusoids in this way. The summation error is discussed on p. 89.

Maxwell (1940)

Analysis by synthesis can be represented in terms of electrical voltages. Maxwell built an apparatus in which sinusoids were represented by the fluctuating amplitude of a.c. voltages, there being some advantage in using a.c. rather than d.c. in a laboratory instrument. The voltages were got from small coils turning at the desired

frequency within air-cored solenoids, as in Fig. 2.7, supplied with alternating current whose strength could be adjusted. A stationary pick-up coil was included to avoid negative values appearing in the final sum, for this would have called for a method of distinguishing a reversal in phase of the final a.c. voltage. Maxwell's apparatus added only two sinusoids, but the principle was carried much further by later workers (see Chapter 3).

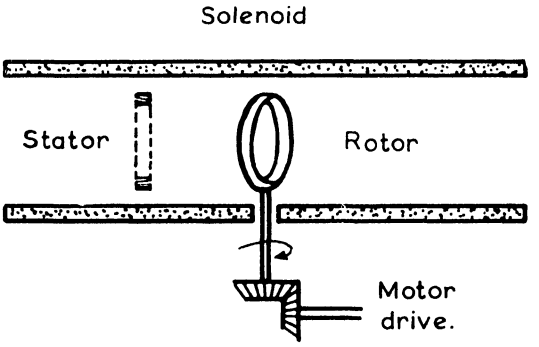

FIG. 2.7. Maxwell (1940). A small coil turning in a solenoid carrying a. c., gives an a. c. voltage whose amplitude varies sinusoidally with a maximum depending on the solenoid current. Such modulated a. c. signals can be added, provided they strictly agree in a. c. phase, to imitate the sum of sinusoids.

BORN, FURTH, PRINGLE (1945); FURTH, PRINGLE (1946)

This is an optical system, very rapid in action, which takes the transient function in the form of a variable density or variable area trace and displays its sine or cosine transform on the face of a cathode ray tube.

A circular photographic plate (Fig. 2.8) shows a pattern of closely spaced parallel fringes across which the transmission coefficient varies sinusoidally. An image of this plate is projected upon a mask pierced with a long fine slit. The transmitted light is caught by a cylindrical lens with its axis at right angles to the slit and is imaged on the trace whose transmission represents the function under examination. By this arrangement, the illumination on the trace varies sinusoidally along its length, the separation of the fringes depending upon the secant of the angle between the slit and the fringes on the original plate. The total transmitted light is evidently a measure of one Fourier amplitude of the function represented on the trace. In analysis, the plate rotates so that the fringes expand

and contract and the transmitted light is caught by a photocell whose electrical output is displayed on a cathode ray tube with its sweep synchronized with the rotation of the first plate. The cathode ray tube therefore displays the spectrum of the trace during each rotation of the plate.

The cosine spectrum is displayed if the axis of rotation of the plate passes through a point of maximum or minimum transmission,

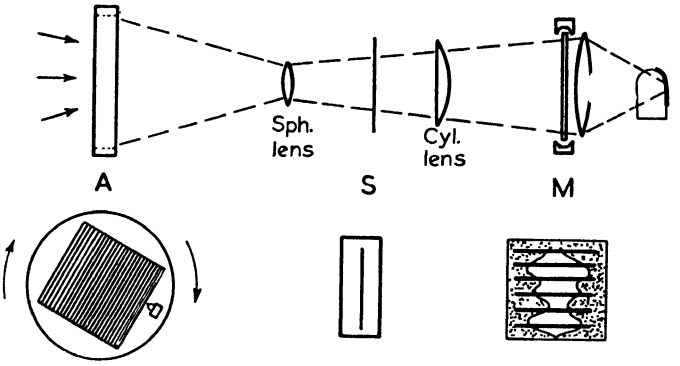

FIG. 2.8. Born, Furth, Pringle (1945). Light passes through a plate marked with twenty parallel fringes that rotates in a journal bearing, A. A lens forms a reduced image of this plate on a vertical slit S, and a cylindrical lens with its axis perpendicular to the slit forms an image on the mask M. The image is a set of horizontal fringes whose spacing varies as the plate A rotates at 10 cycles/sec. The transmitted light is collected on a photocell and its output is displayed on a cathode ray tube which draws the sine or cosine spectrum during every rotation of the plate.

and the sine spectrum is displayed if the axis passes through a point midway between the positions of maximum and minimum transmission. The position of the plate relative to the axis can be adjusted to achieve these conditions. The measured transforms have been shown to agree closely with their calculated values in a number of examples.

The original apparatus was designed to deal with relatively simple functions capable of being described by harmonics up to the 20th order, and the plate carried 20 fringes.

CHAPTER 3

ANALOGUE DEVICES

THE mathematical process of finding Fourier coefficients has been imitated experimentally in such a variety of ways that it is difficult to be entirely systematic in describing them. There are however the broad divisions of numerical calculation, of graphical techniques and of mechanical, electrical and optical analogues: the account that follows is arranged in that order. Discussion of the instruments that involve modulation or filters or diffraction will be deferred till later chapters.

In some cases, the analogue machines evidently put into practice the ideas involved in graphical methods or numerical calculation. Attention will be drawn to such similarities, since they help to make the principle of the machine more obvious.

Computation of Fourier series

It is appropriate first to draw attention to a trick of analysis that is useful in computation generally. It makes use of the symmetrical properties of the harmonics. The wave form that is being analysed may show no particular symmetry but it can be thought of as the sum of two curves, of which one is symmetrical and the other is antisymmetrical about the middle point of the interval. This is illustrated in Fig. 3.1. The cosine components of the harmonic series, being all symmetrical about the middle point of the interval are all contained in the symmetrical curve. Conversely, the sine components of the series are all contained in the antisymmetrical curve. Both curves may now be analysed but their symmetry allows the numerical averages to be found by calculation over only half the length of the interval. This halves the numerical work. The process can be carried further by separating the symmetrical curve into two

parts which are respectively symmetrical and antisymmetrical about the 3/4 point of the interval and which must therefore be made up of cosine components of even order and odd order respectively. The antisymmetrical curve is treated similarly. Four curves are then to be analysed, seeking the appropriate components present

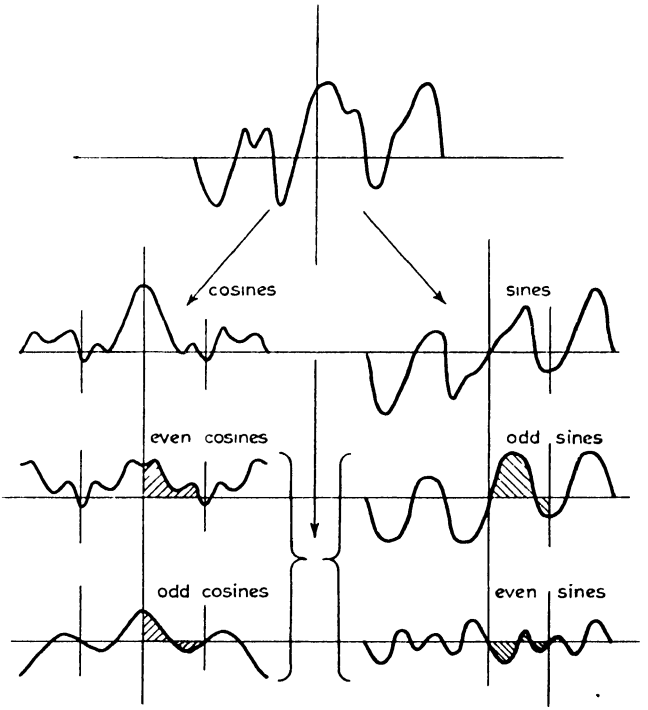

FIG. 3.1. Symmetry can reduce the work of analysis. Only the hatched portions of the four derived curves need enter the calculations.

in each, but their symmetry means that the numerical averages need extend over only a quarter of the whole interval. The numerical work is reduced to one quarter.

In computation one cannot of course take the entire curve but only selected values. Representing a curve by the height of its ordinates at equally spaced intervals along the abscissa is an approximation and a possible source of error.

Discussion of these errors will be reserved for Chapter 6 or may be read in papers by Cochran (1948) and Fellget (1958). The errors arise when high order harmonics are present in the curve. It can be shown that if there are no harmonics with significant amplitude of higher order than, say, the nth, then no significant error is introduced, provided that not fewer than $2n$ equally spaced ordinates are used to represent the curve.

The following will serve to illustrate a calculation with 12 ordinates, effective so long as harmonics up to only the sixth order are present in the curve. The ordinates are taken at intervals of 30° where 360° represents the length of the complete wave cycle.

t	−165°	135°	−105°	75°	−45°	−15°	15°	45°	75°	105°	135°	165°
$g(t)$	2	−1	−3	0	4	2	6	10	2	−8	1	3
Fold							2	4	0	−3	−1	2
(1) Add to give symmetrical part							8	14	2	−11	0	5
Fold							5	0	−11			
(a) Add for even order cosines							13	14	−9			
(b) Subtract for odd cosines							3	14	13			
(2) Subtract to give antisymmetrical part							4	6	2	−5	2	
Fold							1	2	−5			
(c) Subtract for even order sines							3	4	7			
(d) Add for odd order sines							5	8	−3			

Each of the short series (a) (b) (c) and (d) may now be taken in turn and used to find the appropriate harmonics. For instance, series (a) contains only the cosines of even order. To find the amplitude of the fourth order cosine harmonic, the series must be multiplied by cos $4t$ and averaged, t having the values shown. Division by 12, the original number of ordinates, gives −1.0 for the amplitude a_4. The procedures for finding other coefficients are exactly analogous.

t	15°	45°	75°	
Series (a)	13	14	—9	
cos $4t$	0·50	—1·00	0·50	
Products	6·5	—14	—4·5	Total —12·0

The process has been described by Terman (1948) for 6 ordinates and by Manley (1945) for 48 ordinates. Werenskiold (1942) describes a summary method in harmonic analysis using 24 ordinates.

Aids to Calculation

The most tedious part of calculation is probably that of finding products. As an aid to calculation, one may have suitable products already tabulated. Thus Bartels (1931) recommends the use of a table. Beevers and Lipson (1936) produced a set of cardboard strips suitable for the Fourier analysis of a curve represented by 60 ordinates. The strips for cosine calculations are arranged in a box with divisions corresponding to each 6° of t between 0° and 90°. Each division contains strips headed by the integral numbers 1 to 99, with negatives on the reverse side, and below are the products with the values of cos nt where n ranges from 1 to 30. A similar box holds strips with sine products. To analyse a curve represented by 60 ordinates, one may use the symmetry rules to reduce this to 4 curves of 15 (Ross, 1943).

The first use of strips carrying products is attributed to L. Herman in 1890 (see Robertson, 1948). Robertson (1948) devised a more numerous set of cards with three-figure accuracy and describes also a mechanical device for selecting them. Further modifications have been described by Beevers (1952), Alexander (1953) and Timbrell (1958).

The process is of course quite suitable for digital machines, and Beevers (1939) describes a circuit of uniselectors and counters that imitates the action of the cards but is much more rapid. Other

machines have been described by MacEwan and Beevers (1942), and Robertson (1955).

Programmes of Fourier analysis have been devised for machines working on punched cards or for electronic digital computers, such as E.D.S.A.C. Some references are given below.

Punched card machines. Shaffer, Schomaker, Pauling (1946); Cox, Gross, Jeffrey (1947); Nowacki (1948); Hodgson, Clews, Cochran (1949); Greenhalgh, Jeffrey (1950); Kartha (1953);

Digital computers. Bennet, Kendrew (1952); Mayer, Trueblood (1953).

Graphical Methods

It has been pointed out by Manley (1945) that when only a very few harmonics are present in strength, their frequency, amplitude and relative phase can often be picked out by inspection. As an

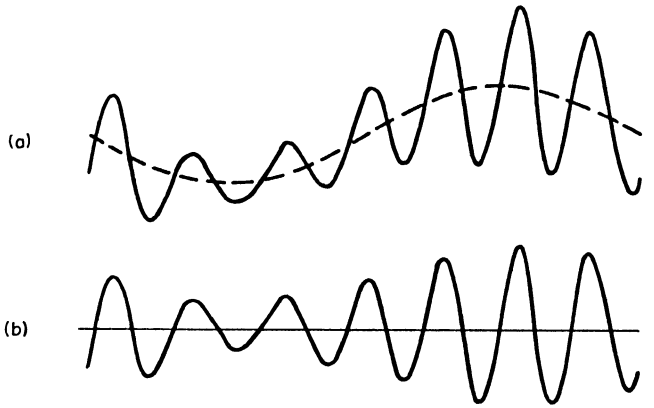

FIG. 3.2. Manley (1945). Analysis by inspection.

example, the curve (a) in Fig. 3.2 repeats its behaviour after the interval shown and an averaging of the more rapid oscillations suggests the broken sinusoid as being the fundamental. Added to this is a higher frequency tone showing beats as indicated in (b). Since the higher frequency is continuously present and shows seven cycles for each one of the fundamental, the most prominent harmonic

is the seventh and its amplitude is the mean of the maximum and minimum that appear in the beating. The amplitude of the next most important harmonic is half the difference between the maximum and minimum amplitudes that appear in the beating. This harmonic must be either the 6th or 8th because it produces only one beat per cycle with the 7th; it can be seen to be the 8th, because the high frequency oscillation is slightly more rapid when its amplitude is large than when it is small. Manley shows techniques of this kind applied to many complicated wave forms.

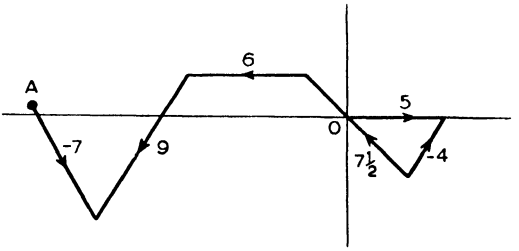

FIG. 3.3. Ashworth, Harrison. Successive ordinates of the curve being analysed are plotted at uniformly increasing angles to find the vector total. This diagram shows six ordinates plotted at successive multiples of 60° to evaluate the first order harmonics.

A more systematic method that detects harmonics of quite small amplitude is due to Ashworth and Harrison (1906). The successive ordinates of the curve are replotted as a series of vectors at angles that successively increase as illustrated in Fig. 3.3. The x and y coordinates of the end point are evidently the summed products of the ordinates and the cosines or sines of the angles and, when the angles are properly chosen, measure the amplitude of the cosine and sine harmonic. A fresh plotting must be made for every different harmonic order. Harvey's (1930) process of analysis by weighing imitates this principle.

Another method, used by Hermann (1909) is to replot the curve against $-\cos 2\pi nt$ or $\sin 2\pi nt$ as abscissa instead of t, as in Fig. 3.4. The area of the figure so formed can be measured by a planimeter and is a measure of a harmonic amplitude. A new plot must be made for each cosine and sine harmonic. The mechanical instruments of

Mader (1909), Chubb (1914) and Hartenheim (1917) use this principle. Straiton and Terhune (1943) avoided the replotting by deforming the graph sheet of the curve on a former made up of tangential semi cylinders and rephotographing it from a sufficient distance. Areas were then measured on the photographs. Johnson (1941) presented the plots on the screen of a cathode ray tube.

Multiplication can be avoided by choosing a different set of ordinates for evaluating each harmonic (Fischer-Hinnen, 1909). To find for example the third cosine harmonic amplitude of a curve graphed

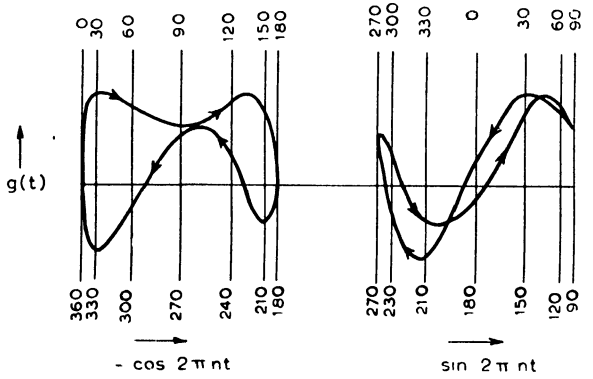

FIG. 3.4. Herman (1909). The curve is replotted against $\sin 2\pi nt$ or $-\cos 2\pi nt$ to find a_n or b_n from the areas so formed. The diagram represents replots for finding the first order harmonics.

between 0° and 360°, ordinates are measured at the six points where $\cos 3t$ takes values 1 and −1 as in Fig. 3.5. These are averaged after alternate reversals of sign. It can be shown that the result is not a_3 but

$$a_3 + a_6 + a_9 + \cdots \text{etc.}$$

The value of a_3 can be deduced, however, when the higher harmonics have been measured in a similar way. Lombardi (1920) made an apparatus working on these principles to detect harmonics in the output of an alternator.

A variant of this method has been used by Daniels (1952) as a means of graphical analysis. To find the nth cosine harmonic, he erects ordinates where the zeros of the harmonic fall (that is at 0°,

$180/n$, $360/n$, etc., for the sines or at $90/n$, $170/n$, etc., for the cosine. A planimeter is then used to total the areas below the curve, alternate areas between ordinates being counted negative. One traces over the "positive" parts of the curve first, returning the planimeter down an ordinate to the zero line in the "negative" intervals. The "negative" regions are then included on the return stroke when the "positive" regions are omitted. New ordinates must of course be erected for every different sine or cosine harmonic. Daniels proposes the use of a succession of transparent sheets on which the proper ordinate lines are marked. The area after division by the length of

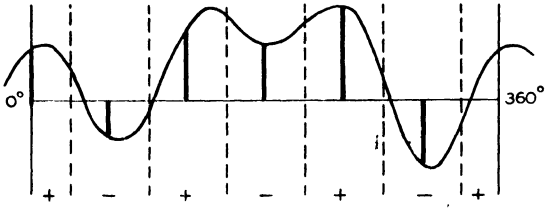

FIG. 3.5. Fischer, Hinnen. Ordinates are measured wherever the harmonic sinusoid takes values -1 or 1 (the heavy lines) and are meaned with appropriate sign. Daniels (1952) erects ordinates where the harmonic sinusoid is zero (the broken lines) and totals the intervening areas with appropriate changes of sign. A correction for higher harmonics is necessary in both cases. The figure represents the construction for finding a_3.

the base line of the curve includes a fraction of the amplitudes of higher harmonics. Thus the process aimed at the nth cosine amplitude, a_n actually gives

$$a_n - \tfrac{1}{3}\, a_{3n} + \tfrac{1}{5}\, a_{5n} - \cdots \text{etc.}$$

but the proper correction can be applied when the higher harmonics themselves have been measured in the same way. Barber (1952) suggested an amended process which would avoid the contribution produced by the harmonic a_{3n} when measuring a_n, and Crease and Tucker (1954) showed how to avoid the contribution from a_{5n} also. Heurta and Casals (1954) find it an aid to have sinusoids with a range of amplitudes graphed on a single chart. Ordinates are taken off from this plot, using the sinusoid whose amplitude is equal to the ordinate of the original curve. The ordinate lengths are added graphically.

Mechanical Methods

Weighing. Harvey (1930) describes a modified chemical balance (Fig. 3.6). A set of horizontal bars are arranged to pivot one above the other about a vertical axis that is attached to the balance directly below the knife edges. Equal weights riding on each of the bars can be set out to distances proportional to the ordinates of the curve that is to be analysed. The bars are then rotated through angles proportional to the corresponding abscissae of the curve, and the out-of-balance is found by adding weights to the balance pan. A

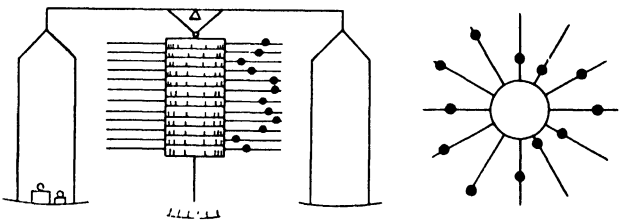

FIG. 3.6. Harvey (1930). Analysis by weighing. The idea was extended by Vand (1952) to an automatic machine synthesizing harmonics to the 140th order.

different setting of angles must be made for each harmonic that is examined. The method recalls the graphical method of Ashworth and Harrison. As described, the device has twelve bars and by use of symmetry would be able to analyse a curve of 48 ordinates up to the 24th order harmonic.

Vand (1952) describes a machine that synthesizes or analyses harmonics up to the 140th order. It is formed in a rigid rectangular framework pivoted on knife edges about the horizontal diameter. A small motor on the top of the frame drives gear trains arranged along the top of the frame and down one side and these turn shafts crossing the frame, fourteen vertically and ten horizontally. Where the shafts cross, gears take the sum or the difference so as to drive 140 vertical spindles rotating at speeds in arithmetic progression. Discs mounted on these spindles each carry a spigot to which a variety of weights can be attached, each proportional to the amplitude of the

corresponding sinusoid. When the motor is started, the changing positions of the weights produce gravitational moments that cause the frame, which is bottom heavy, to deflect very slightly. The deflections are recorded optically and the resultant curve is the synthesis (or analysis) of the original data that the assembly of weights represents.

Curve-tracing instruments. These provide measures of the harmonic amplitudes after tracing the machine over a curve that is to be analysed. That of Mader (1909) in effect produces a moving point whose path is the original curve replotted against $\cos nt$ or $\sin nt$ instead of t. It is therefore a realization of the graphical method used by Hermann (1909). A planimeter is attached to the moving point and records the area of its path, that is the amplitude of the harmonic. Successive harmonics are found by successive tracings after a suitable change of gears. The instrument is commercially available, giving harmonics up to order 14. Its accuracy is discussed by Baer (1937) and Nystrom (1938).

The instrument of Henrici (1894) evaluates five harmonic amplitudes at a single tracing and can work up to the 50th order (Miller, 1916A).

An instrument working on the principle of Harvey (1934) and Yule (1895) is described in the *Z. Instrumentenk.* **59**, 288–92, 1939. It gives harmonics up to the 6th order, or with additional gearing up to order 13. Harvey's 1934 paper is reprinted in Manley (1945), pages 213–218.

Charp (1949) describes an analysing instrument using 4 ball and disc mechanisms.

The Addition Process

Chain and pulleys. The weakness of the Michelson-Stratton machine described in Chapter 2 lies in its method of summing the sinusoidal motion by spring tension. Springs are apt to change their elastic constant with time. A better way of adding displacements was used by Kelvin (1878) in his tide predicting machine. Here a set of pulleys works on a long chain or metal strip that is under con-

stant tension. The principle is illustrated in Fig. 3.7. The machines of Kranz (1927) and Vand (1949) use this method of addition, but friction limits the number of harmonic elements.

The amplitudes of the sinusoids are usually controlled by adjusting the radius of motion of an eccentric pivot, but Caimann and Hoppe (1953) use pairs of eccentrics fixed in amplitude but adjustable in relative phase to give any desired amplitude on addition.

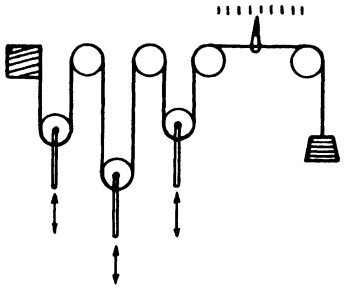

FIG. 3.7. Addition by pulleys, after the manner of Kelvin's tide predicting machine.

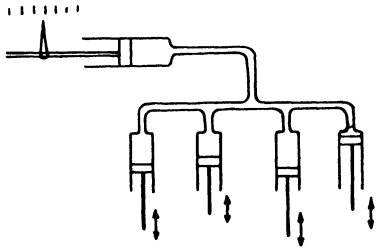

FIG. 3.8. Shilton (1944). Hydraulic addition.

Oil pistons. Shilton (1944) summed displacements by applying them to pistons whose cylinders were all connected together and filled with oil. The piston in an additional cylinder connected to the system, moved a pen by an amount proportional to the sum of the sinusoidal motions.

Sand. MacLaughlan and Champaygne (1946) have described an interesting mechanical device that is suitable for synthesis in two dimensions (Fig. 3.9). It uses sand as the measuring material. The flow from a long hopper is controlled by the rate of rotation of a long horizontal roller blocking the orifice. After being scattered in a mesh so that its rate of fall is uniform over a long rectangular area, it falls on a cut-out mask whose opening varies sinusoidally along its length. The apparatus is moved slowly sideways and the falling sand accumulates in a grid. By fitting masks with different sinusoidal profiles and by adjusting the rate of flow by the speed of the roller, sinusoids

of different wave number, amplitude and orientation can be accumulated in the grid. The depth of accumulated sand is then a measure of the synthesized two-dimensional function.

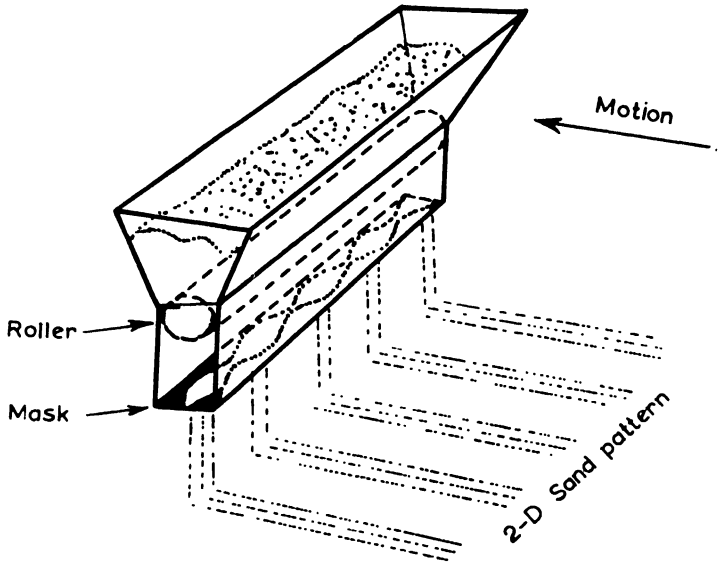

FIG. 3.9. McLaughlan, Champaygne (1946). A two-dimensional synthesis of sand patterns. The diagram is somewhat simplified; provision is made for disposing of the sand that strikes the mask as it falls.

Electrical

Harmonics in power supplies. The wave shape of alternating current in electric power supplies has been of importance to engineers for a long time. Des Coudres in 1898 (see Dina, 1916) advocated the Fourier system of measuring the harmonics by multiplying by a harmonic sinusoid and averaging the product. Dina obtained this multiplication by a wattmeter, the inertia of the instrument providing the necessary smoothing. Labouret (1921) fed the alternating current to the field coils of a converter and drove its rotor at the speed appropriate to the harmonic being sought; the d.c. output from the converter indicated the amplitude of that harmonic.

The signal and the sinusoid need not explicitly be multiplied. It is sufficient if the two are added and the total is squared, for the square of the sum involves the product. It is true that other terms appear and contribute to the d.c. output, but Liable and Binder (1930) overcame this difficulty by making the frequency of the sinusoidal slightly different from that of the harmonic being sought. When the sum of the signal and the sinusoid was passed to a square-law voltmeter, its pointer showed slow beats as the phase of the sinusoid slowly changed in relation to the phase of the harmonic. The amplitude of the beats is then a measure of the product, and of the harmonic amplitude when the amplitude of the sinusoid is known. Nicholson and Perkins (1932) used a dynamometer to show the beats and Greenwood (1932) used an electronic valve with a square-law characteristic.

An alternative way of measuring a harmonic is to use tuned circuits that select it. and P. de la Gorce (1914) gives practical details of the construction of suitable inductances and capacitors. Blondel (1915) reviews his own and Gorce's work in this matter. Roth (1918) got improved selection by using a dynamometer built with two stationary coils tuned a little above and a little below the frequency of the harmonic being sought. But since there is some difficulty in relating the original harmonic amplitude to the response of a tuned circuit, Blondel (1925) and Rogers (1948) found it better to use a null method. At the same time they avoided the difficulty of creating harmonic frequencies in sinusoidal form, by using a calibrating signal whose harmonic amplitudes were known; thus Blondel used a square wave. An adjustable fraction of the calibrating signal was subtracted from the signal under test until a detector tuned to the frequency of the harmonic showed no response.

A quite different idea was used by Lombardi (1920), who arranged for momentary contacts to be made 1, 3, 5 or 7 times per wave cycle and the voltages added in a bank of condensers. At three contacts per cycle the condenser voltage is independent of the fundamental but depends upon the 3rd, 6th and 9th order harmonics. Correction for the higher harmonics can be made after these have been measured in a similar way. Gates (1932) reduced the necessary correction by

making each contact last for a third of the contact cycle. The basic idea is very like the graphical method listed under the name Fischer-Hinnen. Gates' modification shows the likeness between this method and the "chopping" method of Barlow and Keene, which is discussed in Chapter 5 as a process of modulation.

The harmonics are often of small amplitude. They are enhanced in proportion to their order, and so are more easily measured, if the current first passes through a differentiating network (Blondel, 1926).

Electrical multiplication. Ohm's law (Fig. 3.10) affords a convenient means of forming the product of two quantities, one represented by the magnitude of the current and the other by the magnitude of the resistance. The machine used by Vasilesco (1934) employed a resistance that varied sinusoidally (or a stepwise approximation to a sinusoid) and was supplied with a current showing a rapid and transient fluctuation in imitation of the function that was to be analysed. This current was derived from a rotary switch that selected in rapid succession from a bank of rheostats adjusted to represent the ordinates of the function. The voltage that was developed in the fluctuating resistance passed to a ballistic galvanometer which in effect integrated the voltage and gave a throw proportional to the harmonic amplitude. The gears driving the sinusoidal resistance could be changed to assess the harmonics of different order.

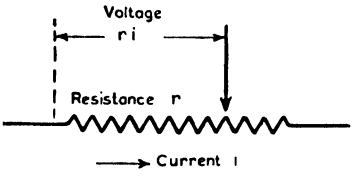

FIG. 3.10.
Ohm's law for multiplication of quantities represented by a current and a resistance.

Vasilesco's method is comparable with Bekesy's optical method in the sense that only one sinusoid is introduced at one time to determine the amplitude at one point of the spectrum. A second operation with a different sinusoid calculates a second point on the spectrum and so on. This attack on the problem is often less useful than the idea employed by Michelson and Stratton (Chapter 2) to get the spectrum by synthesis. This alternative approach calls for many sinusoids to be generated at once, a sinusoid for each different ordinate that is used to represent the original function. On the other

hand, the process corresponding to the integral sign in the Fourier formula presents little difficulty, for it is only a matter of adding similar quantities all generated at the same time. One is not then driven, as is Vasilesco, to perform the whole process of modulating the original function by a sine wave in so short a time that the integrating device does not "forget" the early part of the process. Many of the electrical analogue machines work by synthesis. Most of them were directed at the study of crystal structure by X rays. The multiplication is commonly brought about in potentiometers where the current is made to vary sinusoidally and the contact is set at the start of the analysis.

Ryme and Butler (1944) used a direct current apparatus. Voltages proportional to the values of a sinusoid at 20° intervals were made by a cascade of resistance across a battery. These were connected through mercury cups to a set of potentiometers whose sliding contacts were adjusted to represent the ordinates of the curve being measured. The sum of the output voltages was observed on a galvanometer. The interconnection between the mercury cups carrying the voltages to the potentiometers was made by one of a number of plywood strips fitted with interconnected metal studs, a different strip being put in place for every different coefficient that was being measured. The description the authors give refers to an instrument with 18 potentiometers. By the use of the symmetry rules this should allow a curve to be analysed to harmonics of the 36th order. This would, however, require seventy different plywood jigs to be made and used successively.

Hagg and Laurent (1946) describe a potentiometer device using alternating current, which is preferable to direct current in some ways. A transformer with multiple tappings gives voltages proportional to the successive ordinates of a sine wave. The connections between these points and the potentiometers whose settings represent the curve that is being analysed, are made by uniselector switches. This arrangement no doubt speeds up the process of analysis. It is arranged to provide harmonics up to the 16th order. A very similar system is described by Braun (1958).

The instrument described by Bowen and Burnup (1951) generates

only one sinusoidal voltage at any one time, this being got by applying simple harmonic motion to the sliding contact of a potentiometer. This sinusoidal voltage is applied in succession to 24 potentiometers whose sliding contacts are set to represent the ordinates of the curve to be analysed. The voltages that develop on these potentiometers successively pass to an integrating circuit. The analysis proceeds rather slowly and one has to adjust the speed of the mechanical drive to make it execute some whole number of oscillations during the time taken for the uniselector to run through the rank of potentiometers.

Frank (1957) describes an analyser modelled upon that of Hagg and Laurent but arranged to include harmonics up to order 20. He emphasizes the importance of using components that are commercially available and states that the instrument has given very little trouble.

Maxwell (1940) proposed a synthesizing apparatus in which the different sinusoids were represented by the sinusoidal variation of amplitude in alternating currents. He produced a model using variable air cored transformers whose secondary coils rotated at the speed of the sinusoid to be represented, though his model includes only two such elements. The amplitude of the sinusoid is adjusted through the amplitude of the current fed to the primary. The output voltages are added and displayed. Shimiz, Elsey and McLachlan (1950) adopted this idea in an instrument adapted for synthesis to the 8th order. The variable inductances were "selsyns", the rotors of which were turned at different speeds by gear trains. The gearing was made to facilitate synthesis of a two-dimensional pattern. Provision was made for extending the apparatus to include harmonics up to the 16th order. Beauclair (1949) uses a similar principle in an apparatus built with fifteen rows and thirteen columns of instruments like selsyns to synthesize functions in two and three dimensions.

Mohanti and Booth (1955) synthesize sinusoids that are stored ready for use on a magnetic drum.

Optical Synthesis and Analysis

The light transmitted by a variable density or variable area mask is the product of the incident intensity and the function that the mask represents. Addition of light fluxes can be made either photographically or by a photocell.

The apparatus of Bekesy (1937) has already been described (Chapter 1). Montgomery (1939) used a similar idea but the functions he analysed were on variable area sound track. The image of the sound track was projected on to slides whose transmission varied sinusoidally along the length of the sound track image. The total transmitted light was caught in a photocell. The phase of the particular harmonic was found by moving the slide lengthwise to find the positions of maximum and minimum transmission. A different slide was used in determining each different order of harmonic.

Furth and Pringle (1944) describe a synthesizing apparatus in which sinusoids up to the 9th order are represented as transparent traces of varying width on successive annular zones of a rotating disc. The light passing through a radial slit in front of each trace is controlled in intensity by the adjustment of a neutral optical wedge. The light accumulated from all the traces is passed to a photocell for measurement. A method is described for analysis in which the light from any one sinusoid can be multiplied electronically by the signal that is to be analysed (which is also reproduced photoelectrically). This seems to be an unnecessary addition since the machine can be used for analysis by adjusting the optical wedges to represent the ordinates of the curve. The synthesis or analysis lends itself to display on an oscilloscope screen.

Lohmann (1959) points out that two similar grids with their lines inclined at an adjustable angle, create a system whose transmission varies approximately sinusoidally (Moiré fringes). He suggests that such fringes could be imaged on a cut-out mask representing the function whose spectrum is required. The spacing of the fringes can be adjusted by altering the angle between the two grids, and Fourier coefficients would be determined in succession.

A rotating disc printed with black and white sectors (a strobo-

scope) can be used to pick out frequencies in the fluctuating intensity of light. It is usually convenient to drive the disc at a constant speed and to have sector patterns of a variety of different angular spacings on different annular zones of the disc. T. de Nemes (1934) points out some advantage in having two annular zones for each frequency, the black sectors in one zone corresponding with the white sectors in the other; fringes are then more obvious by contrast. The stroboscopic disc is a useful way of detecting the phase and frequency of fluctuations in lighting but the amplitude of the fluctuations is less easy

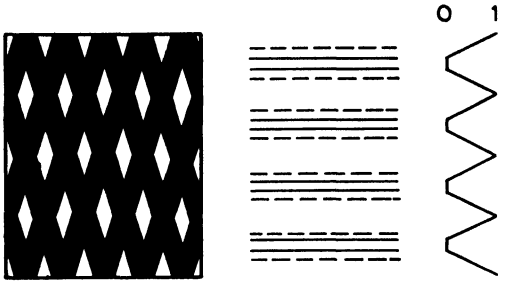

FIG. 3.11. Lohmann (1959). Moiré fringes, formed by inclined grids in front of a slightly extended source of light. The intensity variation is sketched on the right and is not truly sinusoidal.

to assess. Railsback (1937) built an apparatus by which a number of stroboscopic discs were driven at different speeds by a tuning fork and amplifier. It was able to determine the pitch of musical notes more accurately than they could be set by ear and was directed at improved piano-tuning. The sound was of course detected by microphone and arranged to produce synchronous flashes in a neon lamp. Fischer (1950) has described the use of both rotating discs and moving rasters in detecting the frequencies in complex acoustic wave forms.

Robertson (1950) has used the deflection of an optical beam to add harmonic sinusoids. It is successively reflected between a set of fixed mirrors and a set of deflecting mirrors whose simple harmonic oscillations are produced mechanically.

Optical synthesis in two dimensions. A great advantage of optical processes is that they can simulate Fourier processes in two dimensions. Bragg (1929) proposed the idea of adding sinusoids in two dimensions by exposing a photographic plate in succession to sinusoidal fringes having different spacings and orientations. The amplitude appropriate to each fringe pattern was introduced by adjusting the length of exposure. The density of the plate after development

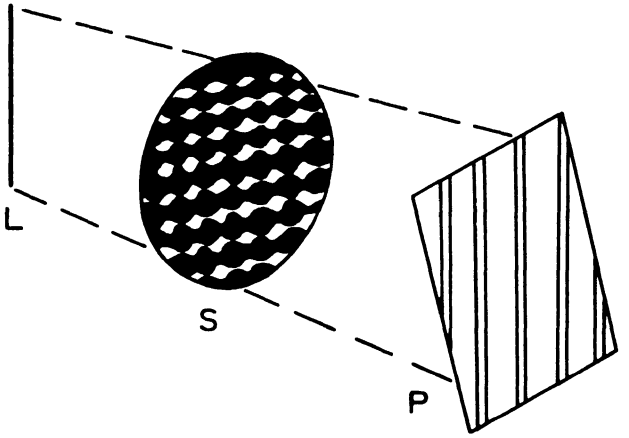

FIG. 3.12. Eller (1951). A long slit L acts as an extended source of light to illuminate a mask of sinusoidal patterns, S. The shadow thrown on the photographic plate P shows a sinusoidal variation of intensity. The fringe spacing is controlled by adjusting the distance SL. The plate P is turned to record fringes with the desired orientation.

is not of course precisely proportional to the illumination it has received so the synthesis is not exact. Huggins (1941, 1944) developed this method and Howell, Christiensen and McLachlan (1951) produced a set of microfilm carrying sinusoidal fringe patterns with all the desired spacings. Eller (1951) showed, however, that by using a line source of light and a single mask of variable area patterns one could produce fringes with all the necessary wavelengths in quite a simple way. The scheme is indicated in Fig. 3.12. His idea was put to use by Bru, Rodrigues and Cubero (1952) and by Bru, Garcia and Roderigues (1955). A review of optical devices is given by Howell (1959).

Optical interference fringes have been used in Fourier synthesis by Hanson and Lipson (1952), Hanson (1952) and Mertz (1956).

Pepinsky (1947) has presented two-dimensional syntheses on an oscilloscope screen by causing the spot to sweep rapidly (1000 cycles/sec) in the X direction and more slowly (1 cycles/sec) in the Y direction after the manner of a television display. A sinusoidal signal that modulates the intensity of the spot then produces a sinusoidal fringe pattern. Different sinusoidal signals can be added before applying them to the control of the intensity, and the corresponding fringe patterns appear added on the screen. The obvious advantage of the method is that the synthesized pattern can be seen at once, and the trial-and-error methods that are necessary in X-ray studies to decide the proper sign of all the Fourier elements are less time-consuming. It is essential that the sinusoidal frequencies should synchronize precisely with the rates of scan. Pepinsky achieves this by using the 1000 cycles/sec signal to produce a rotating field in the stators of a number of selsyns. If the rotor turns at exactly 1, 2, ... etc. cycles/sec, the frequency it gives out is 1000 ± 1 or ± 2 etc. cycles/sec, the sign depending upon the sense of rotation. The rotors are driven mechanically by gearing that makes their rotation some whole-number multiple of the frequency of the Y sweep. This instrument very adequately answers the need for a rapid preliminary synthesis in X-ray crystallography. The author emphasizes that it is not a replacement for accurately calculated contour maps of electron density in the final stages of a successful structural determination.

FURTHER REFERENCES

Graphical and Numerical Processes

(1) Using special characteristics of the curve: Puget (1919), Abason (1932), Espley (1939), Cunningham (1947), Svoboda (1950).

(2) Using ordinates: Kemp (1920), Terebesi (1934), D. Robertson (1935), Lewis (1935), Labrouste (1936), Chaffee (1936), Vercelli (1934, 1937), Williams (1944), Lovera (1948), Mideendorf (1956).

(3) Use of strips: Lipson, Beevers (1936), Ross (1942), Robertson (1958).

Analogue Machines

(1) Mechanical: Charp (1949), Robertson (1955).

(2) Optical: Kent (1943), a stroboscope.
Bernasconi (1950), photo electric.

Review Papers

GROVER F. W. (1913) *Bull. Bur. Standards* **9**, 568–642.
RUSSELL A. (1915) *Proc. Phys. Soc.* **27**, 149–169.
MILLER D. C. (1916) *J. Frank. Inst.* **182**, 285 322.
DELLENBAUGH F. S. Jr. (1921) *T. Amer. Inst. elect. Engrs* **40**, 451–475.
MAXWELL R. (1940) *Rev. Sci Instrum.* **11**, 47–54.
BALDOCK G. R., WALTER W. G. (1946) *Electronic Engng* **18**, 339–44.
ROBERTSON J. M. (1948) *J. Sci. Instrum.* **25**, 28–30, 216–18.
HOWELL B. J. (1959) *J. opt. Soc. Amer.* **49**, 1012–21.

CHAPTER 4

THE UNDERLYING THEORY

THE previous chapters have spoken of methods of analysing the varying pressure in a musical note or the varying displacement of the ground in an earthquake. These are fluctuating quantities but they are either directionless like air pressure or their direction is prescribed, as when one decides to study an earthquake by recording, say, the vertical displacement of the ground. Mathematicians however have long since shown that the theory underlying Fourier analysis is essentially a theory of vectors. For example, the sine wave itself has been defined in Chapter 1 as the varying vertical displacement of a point which is actually moving on a circular path. Circular motion, not sinusoidal motion is really the basic idea.

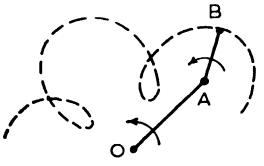

FIG. 4.1. The sum of two rotating vectors.

This chapter will outline the fuller theory, for if we adopt the proper point of view, it becomes possible to give quite simple answers to some curious questions which arise in using analysing machines.

The basic idea is uniform circular motion. The number of revolutions per second is the frequency of the motion. The radius of the circular path is called the amplitude of the motion. Circular motions may have different frequencies and different amplitudes. Two or more circular motions can be added. Thus, in Fig. 4.1, a rod OA, pivoted at O, may have attached to it a second rod AB, pivoted to the first at A. The displacement of B from O is then the sum of the vector distance OA and AB. If the rods both turn about their pivots at different fixed angular speeds, the point B moves on a complicated path which is the sum of two circular motions. Still more com-

plicated motions can be found by combining a greater number of circular motions.

Indeed it will be assumed that any complicated motion of a point such as B can be created by adding together a sufficient number of circular motions having the proper frequencies and amplitudes. The problem is to discover what these frequencies and amplitudes are. To make the question more definite, suppose that the moving point continually repeats a complicated motion, each cycle occupying a time T. Evidently, if it is to be represented as the sum of circular motions, each of these themselves must make some whole number of rotations in each cycle of the complicated motion. Their frequencies will be integral multiples of $1/T$. Here it should be noted that circular motions executed with the same speed but in opposite directions around the circle are quite different motions. If angles measured in one sense are positive, angles measured in the other sense are negative, and it is quite reasonable in this context to speak of positive and negative frequencies, however unusual this may seem. Zero frequency is possible too. If the rod OA in Fig. 4.1 were held stationary, the point A would have zero frequency and remain at a constant vector displacement from the centre O. In short, the circular motions can have any of the frequencies 0, $\pm 1/T$, $\pm 2/T$ etc.

What is the average position of a point which is executing circular motion? By this is meant a time average, or the centre of gravity of all the positions the point assumes at, say, 100 equally spaced instants throughout the interval. If the average is taken during a complete rotation, the mean position is the centre; its displacement from the origin is zero. Consider now the position of a point executing a complicated motion, averaged over one complete cycle. All the inherent circular motions (with one exception) have executed a whole number of rotations and make no contribution to the average position. The one exception is that component which has zero frequency. This component is a constant: the average of the whole complicated motion is merely the component of zero frequency. Thus the value of this component, its amplitude and direction, could be found numerically merely by averaging the positions of the moving point throughout its history.

This idea can be carried further. Suppose we wish to find the amplitude of the component of frequency $+3/T$ (the one which produces a rotation in the anticlockwise sense three times during each cycle of the complicated motion). The complicated motion can be thought of as being produced mechanically with respect to some supporting plane that has hitherto been kept still. Suppose that at zero time this plane itself started to rotate in the negative sense at $3/T$ cycles per unit time. The motion now looks quite different to the observer. Indeed every circular component appears to rotate at a frequency numerically less by the amount $3/T$. In particular, the component at frequency $+3/T$ seems to be at rest. If the observer averages the motion from his new point of view, the value he gets is the amplitude of the component whose frequency was $+3/T$.

One might imagine this process repeated to find in succession the amplitude of every circular component in the complicated motion. This is the mathematical principle of Fourier analysis.

The Complex Algebra

In seeking to represent these quite simple ideas it is perhaps unfortunate that one is driven to use an algebra that involves the "imaginary" quantity i whose square is -1. However, there is no need to try to imagine what number this symbol i may be. One can use the algebra like a machine without knowing in detail how the machine is made.

A vector in a plane can be specified in various ways. One way is to specify its rectangular components x and y in two directions at right angles as pictures in Fig. 4.2. In the complex algebra this is written as

$$x + iy \tag{4.1}$$

Whenever the algebra leads to some expression like this, the part which does not include i represents a displacement parallel to the x axis (the "real" axis) and the part which has the factor i represents a displacement parallel to the y axis (the "imaginary" axis).

Alternatively one may specify the radial length (or the "modulus") of the vector r and the angle θ (the "argument") that it makes

with the real axis. Figure 4.2 shows this with a positive angle θ. Evidently the quantities $r \cos \theta$ and $r \sin \theta$ are the displacements parallel to the real and imaginary axes so the "complex" notation for this vector is

$$r \cos \theta + ir \sin \theta \qquad (4.2)$$

Some writers abbreviate this to r cis θ and it is often convenient to refer to the elements as cis functions rather than as "rotating vectors".

There is another short notation for the vector, namely

$$r e^{i\theta}, \quad \text{or} \quad r \exp(i\theta) \qquad (4.3)$$

where e is 2·718..., the numerical base for natural logarithms, but once again there is no need to wonder how a real number can be raised to the power i. Even the actual numerical value of e rarely needs to be known in using the algebra for the present purpose.

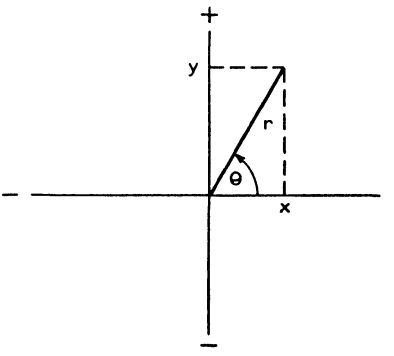

FIG. 4.2. In complex notation the vector is written as $x + iy$, $r \cos \theta + ir \sin \theta$, r cis θ, $r e^{i\theta}$

When a vector has a fixed length A and starts at zero time at an angle α radians to the real axis and rotates about the origin at f cycles per unit time, the angle after time t is $(2\pi ft + \alpha)$ radians and the vector is represented as

$$A e^{i2\pi ft + \alpha} \quad \text{or} \quad A \exp[i(2\pi ft + \alpha)] \qquad (4.4)$$

If now an additional rotation is impressed on this vector by causing its reference frame or "base plane" to rotate at $-f_1$ cycles per unit time, starting at zero time, the vector that the observer sees is got by multiplying the above expression by the factor

$$e^{-i2\pi f_1 t} \qquad (4.5)$$

According to the ordinary rules of algebra this gives

$$A \exp[i(2\pi ft + \alpha)] \exp(-i2\pi f_1 t) = A \exp\{i[2\pi(f - f_1)t + \alpha]\} \qquad (4.6)$$

The result states that the vector now rotates with frequency $f - f_1$, a result which obviously accords with what the observer would see.

It is also worth noting that if one expresses the vector in the rather longer form shown in expression (4.2) the result is just the same, a vector rotating with frequency $f - f_1$ showing that the two kinds of notation, (4.2) and (4.3), are equivalent. The ordinary rules of algebra apply, bearing in mind that i^2 is -1.

$$[A\cos(2\pi ft + x) + iA\sin(2\pi ft + x)][\cos 2\pi f_1 t \quad i\sin 2\pi f_1 t]$$
$$= A\cos[2\pi(f - f_1)t + x] +$$
$$+ iA\sin[2\pi(f \quad f_1)t + x] \tag{4.7}$$

The Fourier Series

The complex algebra is therefore able to summarize all the physical arguments made earlier. A complicated motion that has the complex value $\mathbf{g}(t)$ at any instant t and that in fact repeats its behaviour in a periodic time T can be thought of as the sum of vectors having frequencies that are positive or negative integral multiples of $1/T$. This statement is expressed as

$$\mathbf{g}(t) = \sum_{-\infty}^{n=\infty} A_n \exp[i(2\pi nt/T + x_n)] \tag{4.8}$$

The statement that the amplitude A_n and phase x_n can be found by averaging the motion after impressing a rotation of frequency $-n/T$ is expressed as

$$\frac{1}{T}\int_0^T \mathbf{g}(t)\exp(\quad i\,2\pi nt/T)\,dt = A_n\exp(ix_n)$$

The fact that A_n and α_n appear together as a complex number suggests that it might be convenient always to express them by a single symbol $\mathbf{A}_n$ for the nth complex amplitude

$$\mathbf{A}_n = \frac{1}{T}\int_0^T \mathbf{g}(t)\exp(-i\,2\pi nt/T)\,dt \tag{4.9}$$

The Fourier Integral

In Chapter 2 it was shown how the harmonic series could be modified so as to be capable of representing a transient function by means of a Fourier integral. Just the same arguments can be used with the complex notation; any transient complex signal can be represented by an integral of cis functions with complex amplitude density $\mathbf{G}(f)$. Thus

$$\mathbf{g}(t) = \int_{-\infty}^{\infty} \mathbf{G}(f) \exp(i\,2\pi f t)\,\mathrm{d}f \tag{4.10}$$

and $\mathbf{G}$ is found from

$$\mathbf{G}(f) = \int_{-\infty}^{\infty} \mathbf{g}(t) \exp(-i\,2\pi f t)\,\mathrm{d}t \tag{4.11}$$

Some Deductions

The briefer statement that

$$\mathbf{g}(t) \quad \textit{has the transform} \quad \mathbf{G}(f) \tag{4.12}$$

will be used to imply the relationship shown in Eqns. (4.10) and (4.11). It can be seen that the form of the two equations is interchanged if either f or t (but not both) is written with reversed sign; consequently statement (4.12) implies the statements

$$\left.\begin{aligned} &\mathbf{g}(-t) \textit{ has the transform } \mathbf{G}(-f) \\ &\mathbf{G}(f) \textit{ has the transform } \mathbf{g}(-t) \\ &\mathbf{G}(-f) \textit{ has the transform } \mathbf{g}(t) \end{aligned}\right\} \tag{4.13}$$

It is also useful to consider the conjugate functions got by changing the sign of i. This will be indicated by an asterisk when the complex number is not written explicitly to show its imaginary part. Thus if $\mathbf{g}(t)$ takes some value $r\,\mathrm{e}^{i\theta}$ then $\mathbf{g}^*(t)$ is understood to have the value $r\,\mathrm{e}^{-i\theta}$. Equations remain valid if the sign of i is changed everywhere. If this is done to Eqns. (4.10) and (4.11) their forms are interchanged so statement (4.12) implies the statement

$$\mathbf{G}^*(f) \quad \textit{has the transform} \quad \mathbf{g}^*(t) \tag{4.14}$$

From this, by analogy with Statements 4.13 follow

$$\left.\begin{array}{lll} \mathbf{G}^*(\ \ f) & \textit{has the transform} & \mathbf{g}^*(-t) \\ \mathbf{g}^*(t) & \textit{has the transform} & \mathbf{G}^*(\ -f) \\ \mathbf{g}^*(-t) & \textit{has the transform} & \mathbf{G}^*(f) \end{array}\right\} \qquad (4.15)$$

The Use of Complex Notation

It must be admitted that in physical problems more significance usually attaches to simple harmonic motion than to circular motion. Cis functions on the other hand are much better behaved mathematically than are sinusoids. Fortunately it is quite easy to change from one to the other using the rules

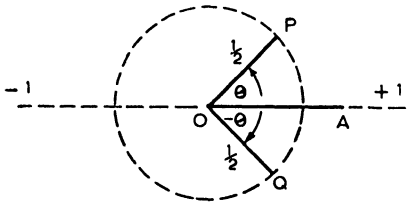

FIG. 4.3. Simple harmonic motion along the "real" axis can be regarded as the sum of two circular motions. The diagram illustrates the rule $\cos\theta = \frac{1}{2}\,e^{-i\theta} + \frac{1}{2}\,e^{i\theta}$.

$$\left.\begin{array}{l} \cos\theta = \frac{1}{2}\,e^{-i\theta} + \frac{1}{2}\,e^{i\theta} \\ \sin\theta = \frac{1}{2}\,i\,e^{-i\theta} - \frac{1}{2}\,i\,e^{i\theta} \end{array}\right\} \qquad (4.16)$$

These can all be given geometrical interpretations. Thus Fig. 4.3 illustrates how a simple harmonic motion along the "real" axis can be regarded as the sum of two opposing circular motions. It has been seen in previous chapters how a scalar variable can be expressed as the sum or integral of sinusoids. Suppose that the sinusoidal element at frequency f is

$$2a\cos 2\pi ft + 2b\sin 2\pi ft$$

This can be written as two functions of frequency f and $-f$,

$$(a + ib)\exp(-i\,2\pi ft) + (a - ib)\exp(i\,2\pi ft)$$

The complex transforms of all "real" variables show this conjugate symmetry. The real part of the complex transform is symmetrical and is equal to $a(f)$, the cosine transform, and represents the symmetrical part of the original function. The imaginary part is antisymmetrical and is $-b(f)$, the negative of the sine transform, and represents the antisymmetrical part of the original function.

Visualizing Complex Spectra

Variables generally, whether they are real or complex, have a transform that is complex. Sometimes it is profitable to visualize separately the "real" and "imaginary" parts of the transform. This is true when the original variable is real, for one can then profit by the rules of symmetry mentioned above. An alternative way is to picture each complex value plotted at its proper angle in a plane at right angles to the frequency axis. But one may merely visualize the transform simply as a plotted curve, so long as one remembers that each ordinate represents not merely a length or an ordinary number, but has an angle associated with it that should not be forgotten when one comes to add or multiply two such quantities.

Mathematical Examples

The transform can often be calculated as a mathematical expression when the function itself is presented in this way. Campbell and Foster (1948) give a table of over 700 mathematical functions and their Fourier transforms. The next example illustrates the square waveform that was discussed in terms of sines and cosines in Chapters 1 and 2.

A square topped pulse. It will be supposed that the function takes the value 1 between $(-\frac{1}{2}\tau + \varepsilon)$ and $(\frac{1}{2}\tau + \varepsilon)$. The mid point of the pulse is therefore at $t = \varepsilon$. The transform is

$$\mathbf{G}(f) = \int_{-\frac{1}{2}\tau+\varepsilon}^{\frac{1}{2}\tau+\varepsilon} \exp(-i\,2\pi f t)\,\mathrm{d}t$$

$$= \tau \frac{\sin \pi f \tau}{\pi f \tau} \cdot (\cos 2\pi f \varepsilon - i \sin 2\pi f \varepsilon)$$

The real and imaginary parts are the sine and cosine transforms $a(f)$ and $-b(f)$ that were derived and pictured in Chapter 2, Fig. 2.4.

A decaying pulse. The formula

$$A\,\mathrm{e}^{-nt}$$

represents a quantity that decays to 1/e of its former value in each time interval $1/n$. If it rises abruptly from zero at the instant $t = 0$, the pulse has the form shown in Fig. 4.4(a). Its transform is

$$\mathbf{G}(f) = \int_0^\infty A\, e^{-nt} e^{-i2\pi ft}\, dt = \int_0^\infty A\, e^{-t(n+i2\pi f)}\, dt = A/(n + i\, 2\pi f)$$

The real and imaginary parts give the cosine and sine transforms

$$a(f) = An/(n^2 + 4\pi^2 f^2) \qquad b(f) = 2\pi fA/(n^2 + 4\pi^2 f^2)$$

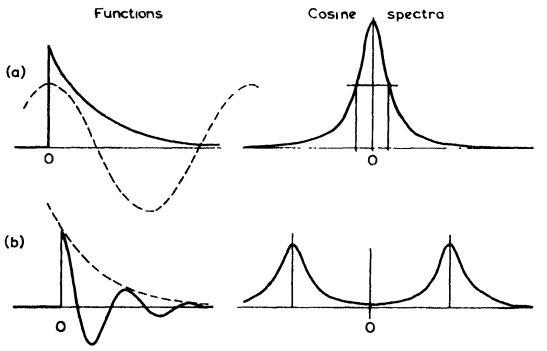

FIG. 4.4. (a) A decaying pulse and its cosine transform. (b) An oscillatory decaying pulse and its cosine transform.

The cosine transform is plotted in Fig. 4.4(a). It peaks at zero frequency and falls to half its maximum at frequencies

$$f = \pm n/2\pi$$

A sinusoid of this frequency increases in phase by one radian in the time that it takes the pulse to decay by a factor 1/e.

An oscillatory decay. The formula

$$A\, e^{-nt} \cos 2\pi pt$$

represents a sinusoid of frequency p decaying in amplitude in the manner shown in Fig. 4.4(b). If it is supposed to rise abruptly from

zero at the instant $t = 0$ the pulse has the transform

$$\begin{aligned}\mathbf{G}(t) &= \int_0^\infty A\, e^{-nt} \cos 2\pi pt \cdot e^{-i2\pi ft}\, dt \\ &= \tfrac{1}{2}A \int_0^\infty [e^{-t(n + i2\pi f - i2\pi p)} + e^{-t(n + i2\pi f + i2\pi p)}]\, dt \\ &= \tfrac{1}{2}A/[n + i\,2\pi(f + p)] + \tfrac{1}{2}A/[n + i\,2\pi(f - p)]\end{aligned}$$

Evidently it is the sum of two patterns like the transform of a pulse, but centring on frequencies $\pm p$. The real part is pictured in Fig. 4.4(b).

The Law of Products

It often happens that one has to think about the effect of multiplying two variables. This is where cis functions are useful. In Bekesy's analyser (Chapter 1) the function that is being analysed is presented as a white trace of varying depth, but a wrong analysis will be made if the trace is not uniformly illuminated. Suppose for instance that the white trace represents the function

$$1 + \cos 2\pi\, 4t/T \tag{4.17}$$

in the range $t = 0$ to $t = T$, and that the illumination fluctuates in the manner

$$1 + \tfrac{1}{3} \sin 2\pi t/T \tag{4.18}$$

These two variations are plotted as curves (a) and (b) in Fig. 4.5. The instrument examines and analyses the reflected light, approximately the product of these two variables as sketched in curve (c). How then will the spectrum of curve (c) differ from that of curve (a)?

When expressed in cis functions the quantities (4.17) and (4.18) become respectively

$$\tfrac{1}{2} \exp(-i\,2\pi\, 4t/T) + 1 + \tfrac{1}{2} \exp(i\,2\pi\, 4t/T) \tag{4.19}$$

$$\tfrac{1}{6}\, i \exp(-i\,2\pi t/T) + 1 - \tfrac{1}{6}\, i \exp(i\,2\pi t/T) \tag{4.20}$$

These are summarized in the "spectra" shown beside the curves in Fig. 4.5. For Expression (4.19) it is three ordinates of value $\frac{1}{2}$, 1 and

$\frac{1}{2}$ at frequencies $4/T$, 0 and $4/T$. For expression (4.20) it is ordinates of value $\frac{1}{2}i$, 1 and $-\frac{1}{2}i$ at frequencies $-1/T$, 0 and $1/T$, illustrating the fact that a spectrum commonly takes complex values even when it represents a variable that is real. It is not very convenient to represent complex values properly in a sketch such as Fig. 4.5, so the ordinates are there drawn as if the values were real.

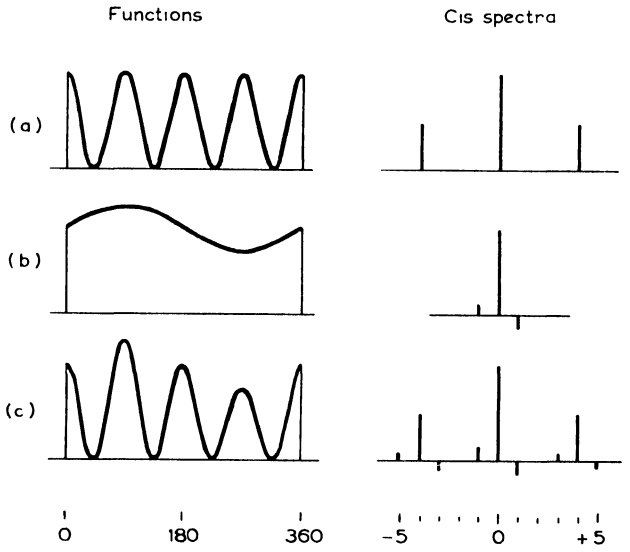

FIG. 4.5. How an uneven illumination spoils the action of Bekesy's analyser; an instance of the convolution theorem.

Cis functions form products in a very simple way: the amplitudes multiply and the frequencies add. So in forming the product of expressions (4.19) and (4 20) it is evident that the first term in (4.19) changes when multiplied by (4.20) into three terms of amplitude $\frac{1}{16}i$, $\frac{1}{2}$ and $\frac{1}{16}i$ at frequencies $-5/T$, $-4/T$ and $-3/T$. Similar remarks apply to the other two terms of (4.19). Curve (c) therefore has the spectrum that is shown beside it. In this spectrum, *every ordinate of the spectrum of curve* (a) *has spread out into a pattern like the spectrum of* (b). This is a simple principle that applies quite generally when thinking of the spectrum of products.

One may of course read sinusoids out of this spectrum drawn for curve (c) by combining the cis functions of equal positive and negative frequency according to Eqn. (4.16). One finds that the sinusoids in (c) are

$$1 + \tfrac{1}{4} \sin 2\pi t/T - \tfrac{1}{8} \sin 2\pi\, 7t/T + \cos 2\pi\, 8t/T + \\ + \tfrac{1}{8} \sin 2\pi\, 9t/T \qquad (4.21)$$

In the present case one might also get expression (4.21) by multiplying expressions (4.17) and (4.18) using trigonometric rules, but this method is difficult to use in more general cases.

The Convolution Theorem

As another instance, suppose that Born's analyser (Chapter 2) is being used to analyse the transient function pictured in curve (a) of Fig. 4.6. Suppose however that the plotted function is physically

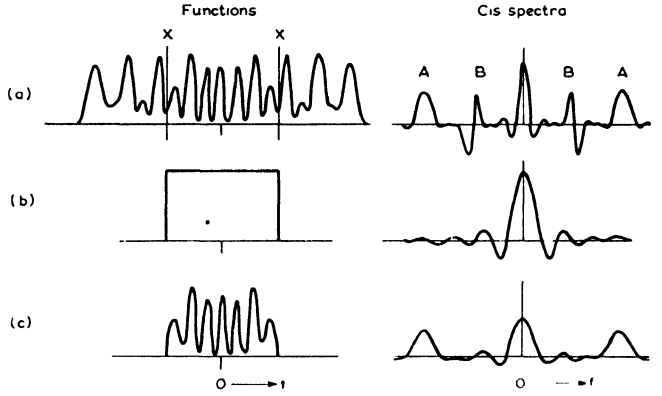

FIG. 4.6. Ignoring the early and late parts of the function has the effect of "blurring" its spectrum. This is an example of the convolution theorem. In the above example, it is seen that the broad peaks AA of the spectrum are little affected, the rapid oscillations BB almost disappear and the narrow peak at zero frequency is reduced. This conclusion is physically reasonable since in cutting off the ends of the function (a) one has retained most of the high frequency oscillations, cut off most of the slow ones, and about halved the area under the curve. The function has been chosen symmetrically in order to avoid the need to represent complex values in the transform.

a little too long to be included entirely in the apparatus. It may be possible to include only the middle portion between the points marked $X\,X$. Analysis will then proceed as if the function were actually zero outside these limits as pictured in curve (c). Indeed curve (c) may be regarded as the product of (a) and a factor that is unity in the range $X\,X$ and is zero before and after. This factor is sketched in curve (b). How then will the spectrum of (c) differ from that of (a)? For purposes of notation, let the values in curves (a), (b) and (c) at any point t be called $\mathbf{g}(t)$, $\mathbf{k}(t)$ and $\mathbf{g}(t)\,\mathbf{k}(t)$ respectively.

Spectra are sketched in Fig. 4.6 on the right hand of the curves. Because the functions are transient the spectra are continuous as explained in Chapter 2. Again we deal with cis spectra that extend to both positive and negative frequencies. The spectrum of curve (a) will commonly take complex values but the sketch does not attempt to represent this. Its value at any frequency f will be denoted by $\mathbf{G}(f)$. The factor (b) has a spectrum that has been examined in Chapter 2 and its cis equivalent is shown in Fig. 4.6. There is no need to specify it here other than to say it has a value $\mathbf{K}(f)$ at frequency f. According to the rule explained in the previous example, the spectrum of the product $\mathbf{g}(t)\,\mathbf{k}(t)$ in curve (c) is got by allowing each element in the spectrum $\mathbf{G}(f)$ to expand into a pattern like $\mathbf{K}(f)$. If the element is taken to lie at some frequency f_0 the pattern spreads about this point so its contribution at a neighbouring frequency f is proportional to $\mathbf{K}(f - f_0)$. It is also porportional to the size of the original element $\mathbf{G}(f_0)\,\delta f_0$. Taking account of the patterns arising from every possible f_0 element, one finds that the total contribution made at frequency f is

$$\int_{-\infty}^{\infty} \mathbf{G}(f_0)\,\mathbf{K}(f - f_0)\,\mathrm{d}f_0 \qquad (4.22)$$

Regarded as a function of the frequency f this is the expression for the spectrum of curve (c). It is called the "convolution" of the spectra of curves (a) and (b). The "convolution" theorem states that if $\mathbf{g}(t)$ has a spectrum or Fourier transform $\mathbf{G}(f)$ and $\mathbf{k}(t)$ has a Fourier transform $\mathbf{K}(f)$ then the product function $\mathbf{g}(t)\,\mathbf{k}(t)$ has a Fourier transform that is the convolution of $\mathbf{G}(f)$ and $\mathbf{K}(f)$ according to expression (4.22).

This principle is quite general and very useful, and indeed is the main reason for cis spectra being emphasized in this chapter. The principle has some obvious implications; one can see for instance that the spectrum of curve (c) is a blurred imitation of the spectrum of curve (a), blurred because each element of (a) has spread into a pattern as happens to points of light in an out-of-focus picture. In this connection, it may be of use to take a slightly different view of expression (4.22), regarding it as a weighted integral of the values of **G** in the vicinity of the frequency f. Such an averaging process would tend to confuse or obscure the details in the spectrum. The weighting factor in this process is the pattern $\mathbf{K}(f - f_0)$, that is the spectrum **K** reversed on the frequency scale and shifted to centre upon the wanted frequency f.

The Formal Statement

The convolution theorem follows easily from the statements (4.10) and (4.11) that define a transform pair. If $\mathbf{g}(t)$ and $\mathbf{k}(t)$ have the transforms $\mathbf{G}(f)$ and $\mathbf{K}(f)$, then using (4.10) and its conjugate form it is seen that

$$\mathbf{g}^*(t) \cdot \mathbf{k}(t) = \int_{-\infty}^{p=\infty} \mathbf{G}^*(f)\, e^{-i2\pi pt}\, dp \int_{-\infty}^{q=\infty} \mathbf{K}(q)\, e^{i2\pi qt}\, dq$$

$$= \int_{-\infty}^{\infty} e^{i2\pi ft}\, df \int_{-\infty}^{\infty} \mathbf{G}^*(p)\, \mathbf{K}(p+f)\, dp$$

where f has been written for $(q - p)$. Comparing the form of this with Eqn. (4.10), it is seen that

$$\mathbf{g}^*(t)\, \mathbf{k}(t) \quad \textit{has the transform} \quad \int_{-\infty}^{\infty} \mathbf{G}^*(p)\, \mathbf{K}(p + f)\, dp \qquad (4.23)$$

For a converse statement one may begin with Eqn. (4.11) and deduce in a similar way that

$$\int_{-\infty}^{\infty} \mathbf{g}^*(\tau)\, \mathbf{k}(\tau + t)\, d\tau \quad \textit{has the transform} \quad \mathbf{G}^*(f)\, \mathbf{K}(f) \qquad (4.24)$$

These are the general statements of the convolution theorem but perhaps more interest attaches to the case where the function **g** is real.

Then $g(t)$ can be written for $g^*(t)$ and because the transform $\mathbf{G}$ will have conjugate symmetry, $\mathbf{G}(-f)$ can be written for $\mathbf{G}^*(f)$. Then Statement (4.23) reduces after a slight change to

$$g(t) \cdot k(t) \quad \textit{has the transform} \quad \int_{-\infty}^{\infty} \mathbf{G}(f_0)\, \mathbf{K}(f - f_0)\, \mathrm{d}f_0 \qquad (4.25)$$

as already deduced in (4.22). Statement (4·24) has two real forms that will be of use. They are

$$\int_{-\infty}^{\infty} g(\tau)\, k(t - \tau)\, \mathrm{d}\tau \quad \textit{has the transform} \quad \mathbf{G}(f)\, \mathbf{K}(f) \qquad (4.26)$$

which will be of use in Chapter 5, and

$$\int_{-\infty}^{\infty} g(\tau)\, g(\tau + t)\, \mathrm{d}\tau \quad \textit{has the transform} \quad \mathbf{G}^*(f)\, \mathbf{G}(f) \qquad (4.27)$$

which will be of use in Chapter 7.

Use of the Convolution Theorem

Two illustrations of the convolution process have already been given but the following kind of argument is frequently useful in physical problems.

Suppose that some quantity is distributed with distance or time in some manner such as curve (a) of Fig. 4.7. This may represent the variation of an electrical voltage with time or the distribution of light intensity in an optical image or the distribution of vertical magnetic intensity at the earth's surface or the elevation of water surface produced by sea waves. In the last three cases it would be more realistic to present the distribution in two dimensions but for purpose of simplicity the argument is here restricted to one dimension only.

Suppose that it represents a one-dimensional optical image. At a little distance from the true focal plane each point of the image may spread out into an intensity pattern like that in the curve in (b). The out-of-focus image is the sum of such patterns and is the "convo-

lution" of curves (a) and (b). Curve (c) shows the likely pattern of intensity in this out-of-focus image. At a still greater distance from the focus, each point of the true image would spread out into a still wider pattern, say curve (d), and the convolution of curves (a) and (d) is the new out-of-focus image. It is represented in curve (e).

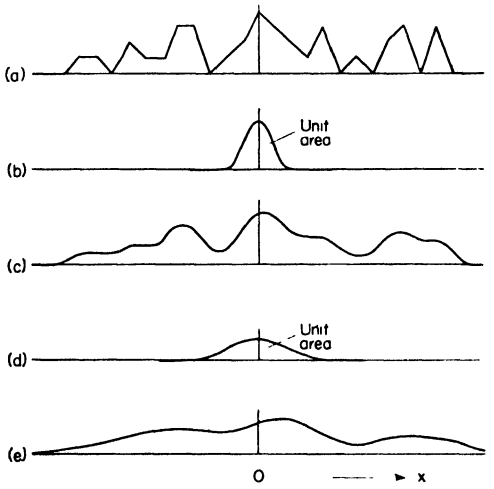

FIG. 4.7

These convolutions may be calculated numerically or found by analogue methods. One is in effect finding a weighted mean or weighted integral of the values of curve (a) near to each point of the abscissa. If one denotes the ordinates of curves (a), (b), etc., by the symbols $a(x)$, $b(x)$, etc., then curve (c) is found from the formula

$$c(x) = \int_{-\infty}^{\infty} a(x_1)\, b(x - x_1)\, \mathrm{d}x_1$$

One might conveniently find $c(x)$ by an optical analogue method like that of Bekesy (Chapter 1) or Montgomery (Chapter 3).

Other questions may arise, however; would it be possible for instance to calculate the very badly focused image (curve (e)) from a knowledge of the rather better focussed image in curve (c)? Might it even be possible to reverse this process and measure the badly

focused image (e) and calculate from it the better image (c) or even the true image (a)?

These questions can be answered by the convolution theorem if one considers the cis spectra of these various curves. We know that the spectrum of the out-of-focus image in curve (c) is merely the product of the spectra of curves (a) and (b). Similarly the spectrum of curve (e) is merely the product of the spectra of curves (a) and (d). The spectra of curves (b) and (d) are sketched in Fig. 4.8. They both fall to low values at high frequency, so high frequencies will be

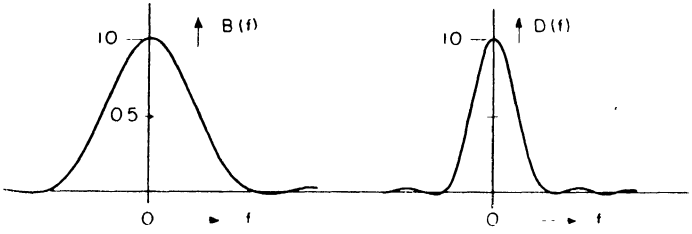

FIG. 4.8

small or absent in the out-of-focus images. But it can also be seen that the spectrum of curve (e) could be got from that of curve (c) by multiplying by the ratio of the two curves shown in Fig. 4.8. One might actually do this, measuring the slightly out-of-focus image shown in curve (c), calculating its spectrum, multiplying the spectrum by the ratio of the curves in Fig. 4.8 and then converting the spectrum back into an intensity distribution which would be curve (e).

There is, however, a shorter way. Figure 4.9 shows on the right the ratio $R(f)$ of the two curves in Fig. 4.8. On the left is a curve $r(x)$ that is calculated as having the ratio curve as its transform.

$$r(x) = \int_{-\infty}^{\infty} R(f) \exp(i\, 2\pi f x)\, \mathrm{d}f$$

To produce curve (e) in Fig. 4.7 one may see from the convolution theorem that it is only necessary to form a convolution between curve (c) in Fig. 4.7 and the curve $r(x)$ in Fig. 4.9.

The reverse process of calculating a better focussed image from one that is less well focused is not so straightforward. Suppose for instance one were attempting to reproduce curve (a) in Fig. 4.7 after

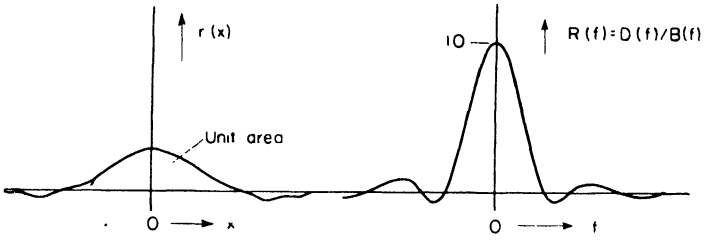

FIG. 4.9

having measured the distribution in curve (c). The finer structure and the more rapid oscillations in curve (a) are much attenuated in curve (c). To reproduce curve (a) one would need to amplify very consi-

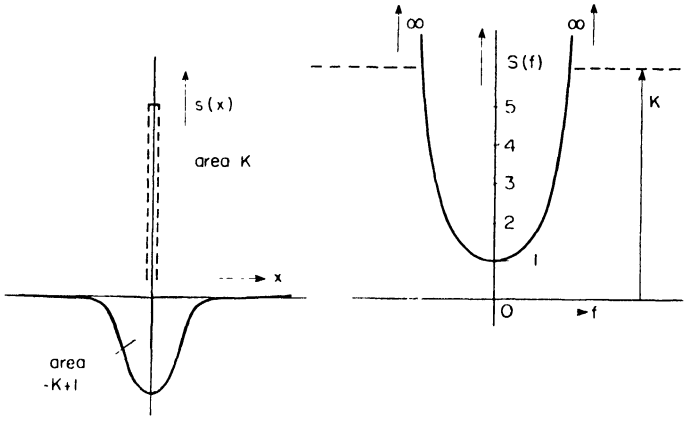

FIG. 4.10

derably such high frequencies as are present in curve (c), and any errors in measurement would be greatly exaggerated. Indeed its spectrum would need to be divided by the spectrum of the spread pattern shown in curve (b), that is multiplied by the reciprocal of the left hand curve in Fig. 4.8.

This reciprocal is sketched on the right of Fig. 4.10. It tends to infinity at certain frequencies indicating that these have become vanishingly small in curve (c). It would therefore not be practicable to restore curve (a) exactly from a knowledge of curve (c). One may, however, make an approximation to curve (a) at least as far as the lower frequencies are concerned. The higher parts of curve $S(f)$ may be approximated by the horizontal broken line. This approximation means that the higher frequencies will not be fully restored. This modified $S(f)$ curve is the spectrum of the function $s(x)$ shown on the left of Fig. 4.10. It is the sum of two parts; the delta function of area K corresponds to a uniform spectrum at the level of the broken line and to this is added a curve which has as its spectrum the negative hump which adds to the uniform spectrum to give the modified $S(f)$ curve. The convolution of $s(x)$ and curve (c) in Fig. 4.7 gives an approximation to the true image shown in curve (a). In effect the convolution takes a weighted integral of curve (c) in vicinity of each position x and subtracts this from K times the actual value $c(x)$.

Photographers will see that this result corresponds to the technique of using an "unsharp mask" to exaggerate the fine details in a photographic print.

CHAPTER 5

FREQUENCY FILTERS AND MODULATION DEVICES

THE fact that a tuned reed or a tuned electrical network will respond to vibrations or signals near its own natural frequency, but remain almost unaffected by other frequencies, is very familiar. The following explanation can be offered. If the reed is struck once, it oscilla-

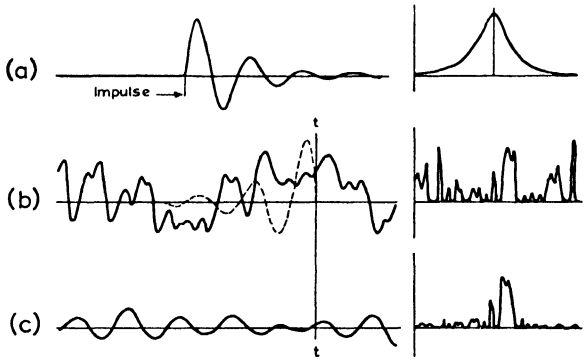

FIG. 5.1. The action of a filter. Curve (a) is the response of the filter to an impulse. Its response to a force with the history of curve (b) is the convolution of the two, curve (c).

tes at its natural frequency after the manner shown in Fig. 5.1 (a). The oscillations die away as the energy is used up in friction. If the reed is subjected to a force that fluctuates irregularly as does curve (b), this force may be thought of as a succession of impulses, each being the product of the momentary value of the force, $g(t)$ and the brief time δt during which it is maintained. Each of these small impulses sets up an oscillatory response like curve (a); the sum of them all is the actual motion of the reed, pictured in curve (c). From

what was said in Chapter 4, it is clear that this motion of the reed is the convolution of the functions shown in curve (a) and curve (b).

$$\int_{-\infty}^{\infty} g(\tau)\, k(t - \tau)\, \mathrm{d}\tau$$

The convolution theorem can therefore be invoked. It states that the spectrum of the motion of the reed is just the product of the spectrum of the force, $\mathbf{G}(f)$, and the spectrum of the response of the reed to a single impulse, $\mathbf{K}(f)$. The spectra are shown on the right hand of the curves. They should be cis spectra, taking complex values, but only the modulus is represented. That of the fluctuating force may be very complicated, while that of the characteristic response of the reed is usually quite simple. Obviously the frequencies that are most prominent in the motion of the reed tend to be those that appear most prominently in its characteristic response to an impulse.

From a slightly different point of view, one can see also that at any instant t the momentary displacement of the reed is the integral of the previous history of the force multiplied by a curve that is the characteristic response of the reed drawn in reverse. This is illustrated by the broken curve drawn on curve (b). This point of view emphasizes that the reed has a "memory". The behaviour at any instant is the effect of forces that were applied shortly before. A sharply tuned reed shows little frictional decay in its response, and such a reed in consequence has a long "memory": a correspondingly long time must elapse before it can respond fully to a change in the character of the fluctuating force. The converse is true of a reed with heavier damping and a broader frequency response.

The foregoing explanation will be useful later in drawing a parallel with the way in which "optical filters" work. For most purposes, however, it is sufficient to think of a reed (or any other resonant system or tuned network) as having a "dynamic characteristic" that is the spectrum of curve (a). The range of frequency within which it is large is called the "pass band" of the system; these are the frequencies to which it most readily responds.

Mechanical Resonators

Iddings (1919) detected harmonics in electrical circuits by using the voltage developed in a search coil to energize a telephone receiver. This was placed near an open-ended pipe which could be tuned to resonate to the harmonic being sought.

For the direct analysis of sound, d'Albé (1926) used a set of open-ended pipes whose tuning could be adjusted. He detected resonance in the pipes by the motion of reeds mounted at the pipe orifices. The reeds carried mirrors to record their motion on photographic paper. Hickman (1934) made an acoustic spectrometer by causing a bank of tuned reeds to be driven electrically from a microphone and amplifier.

Walter's encephalograph (Walter, 1943) used a bank of 25 reeds tuned at frequencies between 1 and 25 cycles/sec. They were electrically driven by the amplified voltages detected on the patient's body. The motion of the reeds was originally detected by their making an intermittent contact on mercury cups, but in a later model of the instrument, the motion was detected photoelectrically.

Electrical Filters

It is reasonably easy to produce a network that is sharply resonant but much more difficult to adjust its resonant frequency over a wide range. Sacia (1924) overcame this difficulty by taking a film record of the sound he wished to analyse and replaying it through a light beam and a photocell, as in Fig. 5.2 the film being driven as an endless belt between two small drums. A filter of fixed frequency was fed from the photocell, and to explore the frequency range the speed of rotation of the drums was adjusted in suitable steps. Dietsch and Fricke (1932) and Milatz, Wapstra and Wieringen (1953) describe similar devices. In both cases they describe the function as being graphed and cut out of black paper that is wrapped into a cylinder and spun by a motor. A light within the cylinder illuminated a

narrow part of the profile and the light that passed was caught by a photocell and the signal passed to a filter.

In analysing records of ocean waves, Barber, Ursell, Darbyshire and Tucker (1946) arranged the record on a large drum mounted freely in ball bearings. The drum was turned at about 4 cycles/sec by hand and then its speed was allowed to decrease slowly by friction. It is apparently much easier to "glide" the speed in this way than to control the speed accurately at set values. A sharply tuned electromechanical filter was actuated by a photocell working on

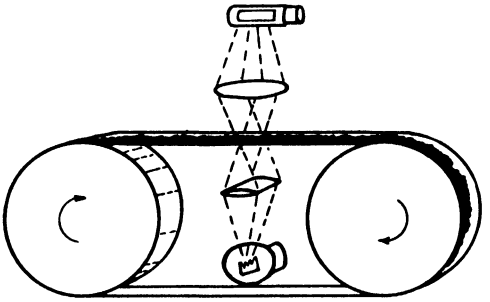

FIG. 5.2. Sacia (1924). The film record is driven at various speeds, and scanned by the light-beam and photocell.

light reflected from the moving trace. The use of reflected light rather than transmitted light is a defect of the system. The arrangement has been further described by Darbyshire and Tucker (1953). In the application of this instrument, interest lay in harmonics between the 50th and 150th. As with the other filter analysers, amplitudes can be estimated but phases are not so readily found.

Whitely and Aldridge (1952) used transmitted light to scan a film record in a small Lucite drum whose speed was allowed to glide from a maximum of 10 cycles/sec under friction. The amplitude spectrum of the trace was given as the history of response of a sharply tuned filter during the decay of speed of the drum.

When using gliding systems it is necessary to enquire how a resonant circuit responds to a tone whose frequency is slowly changing. This has been examined by Hok (1948) and by Barber and Ursell

(1948). The amplitude of the resonance tends to build up smoothly but the amplitude decays in a series of beats if the rate of change of frequency is too great. The response got from a mechanical filter is shown in Fig. 5.3. These "ringing" responses take various forms depending upon the ratio of the rate of change of frequency to the bandwidth of the filter. Curves appropriate to simple resonant filters (Q circuits or reeds) have been given by Barber and Ursell. It appears that if the frequency is changing at a rate of $\mathrm{d}f/\mathrm{d}t$ cycles/sec per sec, the best discrimination that can be achieved corresponds to

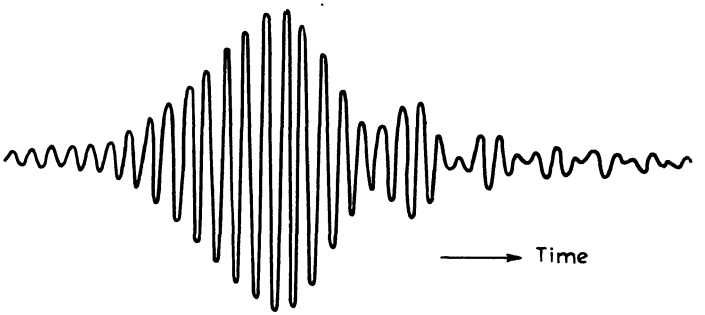

FIG. 5.3. A ringing response to a gliding tone.

a band width of about $(\mathrm{d}f/\mathrm{d}t)^{\frac{1}{2}}$ cycles/sec. To achieve this, one needs to use a filter whose band width under steady conditions is about one third of this value. There is considerable ringing in these conditions. It can be avoided at the cost of some loss of discrimination by increasing the bandwidth of the filter, or at the cost of time by making the frequency change more slowly. Similar rules hold for "panoramic" receivers (Thomasson, 1948) where the tuning is continuously swept over some wide range of radio frequencies in order to show the frequency and intensity of any radio transmissions occurring in that range (Barber, 1949).

Banks of tuned filters are very appropriate when a signal needs to be analysed whilst it is being received. Thus, Freystedt (1935) analysed sound with electric filters supplied in parallel from a microphone and amplifier. Each filter covered a third of an octave and the output potentials were displayed on an oscilloscope. Baldock and

Walter (1946) used twenty-four filters that were in fact high-gain amplifiers, with negative feedback at all but the frequency which they were set to detect. An outstanding example of the use of a bank of filters is an apparatus made by the Raytheon Mfg. Co. for displaying the spectrum of radio atmospherics ("whistlers"). Here 400 magnetostrictive tuned circuits are fed with the radio signal and are sampled in rapid succession by a rotating arm carrying a capacitative pick-up. The activity of these circuits is recorded by a stylus sweeping synchronously across moving recording paper.

Electrical Heterodyne Filters

It is difficult to vary the resonant frequency of a sharply tuned filter over a wide range and yet keep control of its sensitivity and selectivity. On the other hand, it is easy to vary the frequency of an oscillator. This has led to "heterodyne" systems. The signal under examination is added to the oscillator signal in an electronic tube that produces sum and difference frequencies in addition to the original ones. These sum and difference frequencies pass to a sharply tuned filter circuit. If one suitably adjusts the frequency of the oscillator, any frequency component present in the original signal can be arranged to give rise to the frequency to which the filter is sensitive. The amplitude of the response of the filter is proportional to the amplitude of the original signal component.

The method of "beats" used by Liable and Binder (1930) and referred to in Chapter 3 can be looked upon as a heterodyne process, though in that case the filter is one that selects frequencies near to zero. Frequency analysers that use beats have also been reported by Carriere (1939) and Partridge (1938). If the oscillator frequency is changed smoothly and continuously, any fixed frequency in the signal generates a difference tone that can be represented as $\cos [\pi a (t - t_0)^2 + \alpha]$. The frequency of this is $a(t - t_0)$ and the rate of change of frequency is a cycles/sec per sec. The frequency reaches zero at the instant t_0 and at this point the tone has a phase α. The oscillation takes various shapes depending on the phase angle α. Two possible forms are shown in Fig. 5.4(a). In order to be able

to pick out and measure the tones produced from different frequencies in the original signal, one uses a low pass filter to select only the central low-frequency part of these tones. These are the "beats". In order to estimate their amplitude properly, it would seem necessary to retain two or three oscillations unchanged. The effect of a low pass filter on tones such as those in Fig. 5.4(a) is to produce signals which generally look like the curves in Fig. 5.4(b) but these have been drawn merely by guesswork. To use the method of beats most efficiently, the form of the curves in Fig. 5.4(b) should be

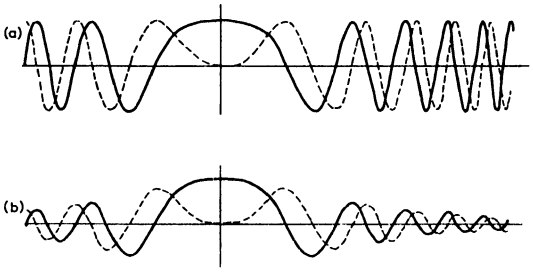

FIG. 5.4. (a) An oscillation gliding through zero frequency.
(b) The possible effect of a low pass filter.

calculated to determine the best width for a low pass filter, but it does not appear that such calculations have been made.

An early form of heterodyne wave analyser was described by Kobayashi (1930). The signal that is being analysed controls the amplitude of the output of a variable frequency oscillator. The output is fed to a mechanically tuned vibrating strip whose tuning is further sharpened by a resonant electrical circuit. A mirror on the vibrating strip records the motion on photographic paper: since the displacement of the paper is locked to the tuning dial of the oscillator, the amplitude spectrum of the signal is obtained automatically. Schuck (1934) used a heterodyne analyser to display the spectrum of incoming sound by the deflections of a spot of light continuously sweeping across a translucent screen. Hall (1935) made an audio frequency recording analyser using a magnetostrictive filter at 20 kc/s and an oscillator tuneable between 20 and 30 kc/s. Tamm and Pritching (1951) report a highly selective heterodyne system in

which two electrically coupled vibrating steel bars act as a sharp band-pass filter capable of distinguishing two notes differing by 25 cycles/sec even when one is 40 db less than the other. Grierson (1957) describes an apparatus designed to present the changing spectrum of radio atmospherics, in particular "whistlers" which are electromagnetic signals with frequencies of some few kilocycles per second: these are audible if earphones are connected through an amplifier to a large loop of wire. From the point of view of upper-atmospheric studies, it is important to observe their rate of change of frequency. To do this, Griersch takes each short sample of the incoming signal, stores it electronically, and replays it many times at a greatly increased speed; while a heterodyne filter scans over the relevant frequency range, and the spectrum is displayed. That sample is then replaced by the next brief portion of signal and the process continues in rapid succession.

In contrast, Smith, McClure and Bostick (1957) deal with signal frequencies in the range 0·01 to 1·0 cycles/sec. The signal is recorded on a loop of magnetic tape and replayed continuously a hundred times faster. Difference frequencies got between this signal and the one from a tuneable oscillator go to a fixed electrical 50 cycles/sec filter. The frequency of the oscillator is made to change in small steps to measure the spectrum at 100 selected points.

The Modulation Process

Heterodyne analysers depend on the production of signals having the sum-frequency or the difference-frequency between a tuneable oscillator and the incoming signal. The process is essentially one of multiplication by a sinusoid

$$\cos 2\pi f_1 t \cdot \cos 2\pi f_2 t = \tfrac{1}{2} \cos 2\pi (f_1 - f_2)t + \tfrac{1}{2} \cos 2\pi (f_1 + f_2)t$$

This may be brought about in a wattmeter or dynamometer, in an electronic multigrid tube, or by causing the incoming signal to vary the amplitude of the signal from the oscillator. Alternatively, one may merely add the two signals and square the sum since this involves the formation of the product. The sum can merely be rect-

ified but this produces higher order products and multiple sum- and difference-tones. Ideally the process is that of multiplication by a sinusoid. When this condition is not fully met, one must be prepared to introduce suitable corrections.

Barlow and Keene (1922) picked out different frequencies in an electrical signal by chopping it in a motor-driven switch formed of brass brushes bearing on conducting segments arranged around the driving shaft. Four different arrangements were available using 1, 4, 16 and 64 conducting segments around the periphery. Since the motor speed could readily be controlled between 3 and 30 rev/sec the arrangement could examine frequencies between 3 and 2000 cycles/sec. Figure 5.5 shows their interrupter switch.

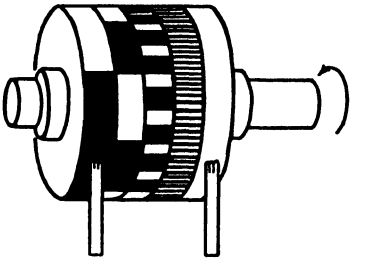

FIG. 5.5. Barlow, Keene (1922). The rotary switch, able to operate from 3 to 2000 cycles/sec.

In this system the switch alternately blocks and transmits the signal, any component having the same frequency as the switch tends to be rectified and its amplitude is estimated by the amount of d.c. that is produced. Indeed this system approximates to the mathematical procedure of Fourier analysis outlined in Chapter 1. The proper procedure however would be to multiply the signal by a sinusoid, and it may be recalled that this was achieved in Vasilesco's electrical machine. Here the signal is being multiplied by a factor that is not a sinusoid, but a "square-topped" factor that is alternately zero and unity. This factor contains many sinusoids and it simultaneously rectifies other frequencies in the signal than the one that is desired. The square topped factor is drawn as curve (a) in Fig. 5.6 and its spectrum in real sinusoids is shown on the right. The "fundamental" has the frequency of the switch. There is however a zero frequency component that would permit any d.c. in the electrical signal to pass direct to the output and be included in the measurement. In the application intended by the inventors the incoming signal might be free from d.c. but in other applications of this method it might be better to

make the switch reverse the signal rather than merely block it. The factor would then be alternately + and −. This is pictured in curve (b).

The spectrum of curve (b) is shown beside it. Zero frequency is absent but harmonics of order 3, 5, 7...etc. are present as before. The factor can be represented as

$$\cos 2\pi ft \quad \tfrac{1}{3}\cos 6\pi ft + \tfrac{1}{5}\cos 10\pi ft \quad \cdots \text{ etc.}$$

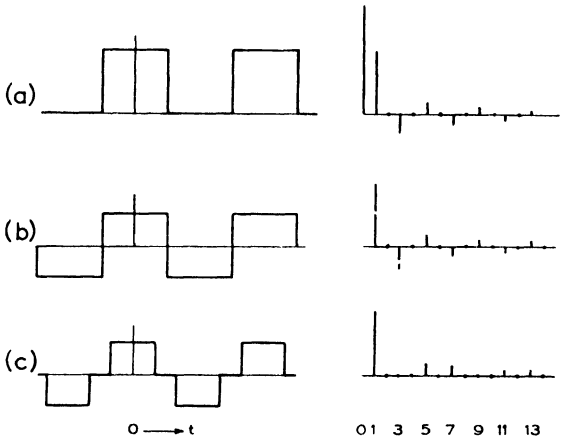

FIG. 5.6. Switching cycles and their transforms.

so the action of the machine approximates not to the simple Fourier process indicated in Eqn. (2.10) but to the process

$$\int g(t)\,[\cos 2\pi ft - \tfrac{1}{3}\cos 6\pi ft + \cdots]\,dt$$
$$= a(f) - \tfrac{1}{3}a(3f) + \tfrac{1}{5}a(5f) - \cdots$$

The harmonics in the factor contribute errors proportional to their amplitudes $-\frac{1}{3}$, $+\frac{1}{5}$, etc., and proportional to the signal components at the frequencies of the harmonics.

A slightly more complicated switching cycle can give a factor a little closer to a sinusoid and so give smaller errors. This is pictured in curve (c). It has contacts, direct and reversed, lasting for $\frac{1}{3}$ of a cycle separated by dead intervals lasting $\frac{1}{6}$ of a cycle. The spectrum is shown on the right; harmonics of order 3, 9 and all multiples of 3

are missing, but the other harmonics remain. Such a switching cycle could be of use in avoiding errors due to components of the signal having three times the frequency that is being examined. Although harmonics of order 5 and 7 are still present in the factor, it might in general be possible to attenuate such frequencies from the signal itself before analysis, by means of a low-pass network. Crease and Tucker (1954) have discussed more complicated switching cycles that are still better approximations to a sinusoid.

Barber (1947) has proposed a system of modulation in two channels. The input signal is multiplied by $\cos 2\pi ft$ in one channel and by $\sin 2\pi ft$ in a parallel one. After a low pass filter in each channel the

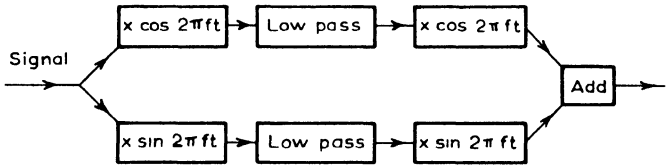

FIG. 5.7. A two-channel modulation filter.

outputs are multiplied respectively by $\cos 2\pi ft$ and $\sin 2\pi ft$ and added. The result is to extract from the original signal a band of frequencies near to frequency f and to present them in their correct frequency and (for the mid frequency at least) in their correct phase. The system is blocked in Fig. 5.7.

The correct explanation of this method was advanced by Madella (1947). The two channels can be regarded as representing the "real" and "imaginary" parts of a complex electrical signal so that the first pair of multiplications is in effect equivalent to multiplying the input by $\exp(-i2\pi ft)$. The whole complex spectrum of the signal is shifted a distance f along the scale of frequency and the part that now falls near zero frequency is transmitted by the low pass filters. The next stage is a multiplication by $\exp(i2\pi ft)$ that returns the selected band to its true frequency.

Madella (1944 B). Two phase or three phase or multiphase electrical supplies are commonly represented by a rotating vector but Madella has gone further than most writers in making explicit analogy

with the complex algebra in describing the behaviour of motors and networks. Figure 5.8 is an interesting instance given also by Barber (1948). If three-phase power is applied to the terminals 1, 2, 3 of this network, the phases being in that order, the same three-phase voltage develops at the terminals 1*, 2*, 3*. If however the connections to 1 and 3 are interchanged so that the phase sequence of the input becomes 3, 2, 1 in that order, a different three-phase voltage (not necessarily reversed in phase) is developed at the output terminals.

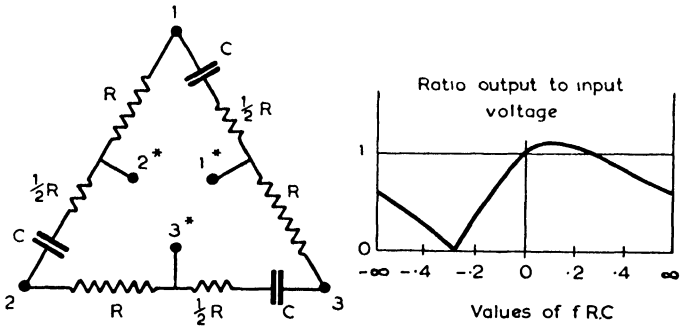

FIG. 5.8. A network that will distinguish phase sequence. Zero transmission occurs where the applied frequency f is $1/|3\pi RC$.

Indeed for a certain applied frequency the output voltage falls to zero. When plotted against frequency the output voltage varies in the manner shown in Fig 5.8. The continuity of this curve justifies the idea of treating a reversal of phase sequence as being a reversal in sign of the frequency of the supply.

Madella (1944A) has used his idea of complex processes in improving the design of a heterodyne analyser. Such instruments, of which there are numerous commercial types, select a narrow frequency band by means of a very selective filter, but because it is impracticable to vary the resonant frequency of such a filter the incoming signal is first changed in an adjustable way, multiplying it ("mixing" it) with the sinusoidal voltage from a tuneable oscillator. The difficulty is that this process creates two frequencies for every one that is present in the input signal. If the selective filter is not to accept signals from both of these sidebands at once and so confuse

signals that were of different frequency in the input signal, it must be given a resonant frequency that is either lower or higher than any present in the input. This is not always convenient and Madella introduced the idea of "mixing" in two channels with frequencies $\cos 2\pi f_0 t$ and $\sin 2\pi f_0 t$. The resulting complex (two channels) signal is the input signal with its complex spectrum shifted a distance f_0

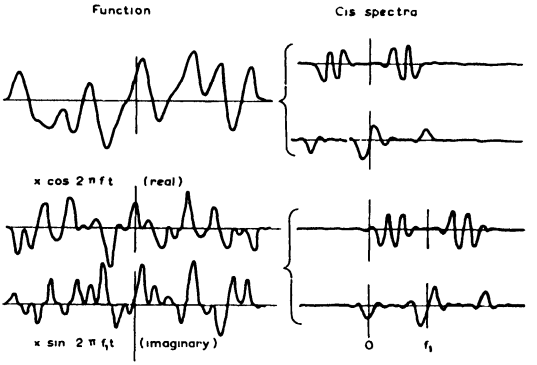

FIG. 5.9. Madella (1944 A). If a signal is modulated by $\cos 2\pi f_1 t$ and $\sin 2\pi f_1 t$ in parallel channels, the two outputs can be regarded as being respectively the real and imaginary parts of a complex signal whose spectrum is identical with that of the original signal except that it has been translated an amount along the frequency scale.

along the frequency scale, as indicated in Fig. 5.9. The selective filter can then be made to pick out some intermediate frequency so long as it is possible to distinguish between complex frequencies $-(f + f_0)$ and $(f + f_0)$. This is done by feeding the complex signal to a network that behaves like that in Fig. 5.8 but is designed for two-phase signals. A single-channel output that is free from any contribution from the $(f \quad f_0)$ frequency in the complex signal is then passed to the selective filter.

Filtering Many Signals

It is sometimes desirable to select the same frequency from a number of different signals, with a view to comparing the amplitudes and perhaps the phases of the frequency components. A heterodyne

or modulation method then appears much preferable to the idea of selective filters all having the same central frequency and the same pass band. The various signals can be separately mixed with the same sinusoidal signal got from an oscillator. The mixed signals then pass through separate low pass filters that determine the passband. It is relatively easy to produce low pass filters whose transmission characteristics are closely alike. The outputs are then comparable both in amplitude and in phase. The fact that the selected frequency may be changed merely by altering the frequency of the oscillator makes the system very flexible.

In dealing with coherent signals, the amplitude transmitted from the mixer depends on the relative phases of the input signal and the signal used for mixing. There may therefore be some advantage in using "complex" mixing in these cases. In dealing with "noise" signals where the amplitude and phase in any narrow frequency band vary randomly with time a one-channel mixing is sufficient to measure the input power. But the comparison of two different "noise" signals must take account of not only the power, or mean square amplitude, and the average phase difference, but also the degree of coherence between the two signals. These qualities are better summarized by a complex correlation coefficient which is discussed in Chapter 8 and can be determined by "complex" mixing.

D.G.Tucker (1958) describes an electrical system that, on being supplied with the values of N ordinates of a curve that is to be analysed, will display its spectrum on the screen of a cathode ray tube. Perhaps its most obvious parallel is the graphical method (Ashworth–Harrison) in which ordinates are added vectorially. Tucker achieves the vector addition by introducing the N values as the amplitudes of N carrier waves that all have the same frequency but are given different a.c. phases. Summation then takes place vectorially. The most novel feature of his apparatus is the way in which the phase shifts are introduced and caused to vary systematically. The a.c. signals, originally all in the same phase and with amplitudes proportional to the N ordinates are fed to N successive points along a delay line as indicated in Fig. 5.10. The output at the end of the line is then the sum of the first signal without delay, the se-

cond with a delay φ, the third signal with phase delay 2φ, and so on to $(N-1)\varphi$. The amplitudes therefore add vectorially and give the amplitude and phase of the transform for a frequency that would show a phase change $-\varphi$ between successive ordinates of the curve. To explore the spectrum, the phase delay φ must be made to vary by an amount π (or 2π if one is to display the negative range of frequency as well as the positive one). It may for instance vary from 0 to π, or from π to 2π, etc. Tucker achieves this variation by vary-

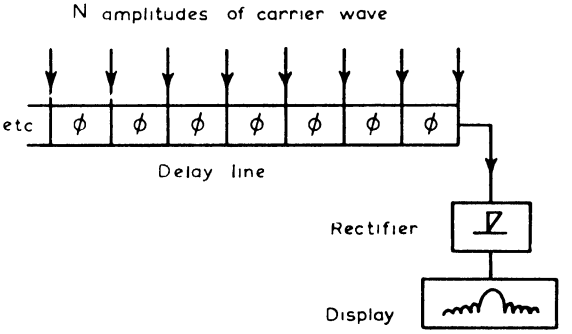

FIG. 5.10. D. G. Tucker (1958). Displaying the transform of N ordinates.

ing the frequency of the carrier wave; a transmission line that delays signals by a definite time will of course lead to a greater delay in phase if the carrier frequency is increased. It is not essential, however, that the phase delay should vary linearly with frequency so long as one knows what delay corresponds to what frequency. A linear relation is convenient, however, if the spectrum is to be displayed on the screen of a cathode ray tube.

The method offers a number of possibilities. One may merely rectify the output carrier signal and so find the amplitude spectrum ignoring the phase. One may measure the mean square (the "power") of the output and so find the power spectrum. One may use a "phase sensitive" rectifying system, measuring the amplitude of the components in phase or in quadrature with the original carrier frequency, to find respectively the cosine or sine transform of the function represented by the N ordinates. If it were desirable, one might

control both the amplitude and the a.c. phase of the N input signals and display the amplitude and phase of the output; this would correspond to making the complex transform of a complex signal.

BICKEL (1960)

A very interesting method of presenting the spectrum of an incoming electrical signal uses the simple-looking arrangement of Fig. 5.11

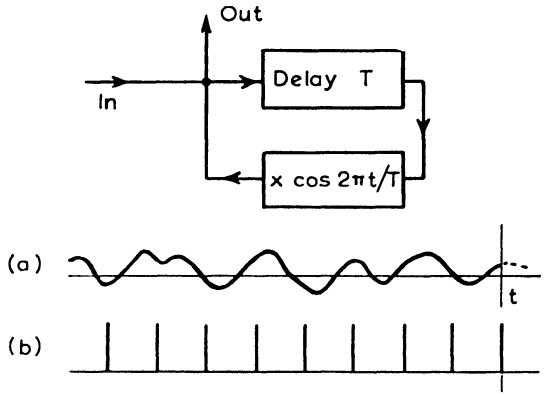

FIG. 5.11. Bickel (1960). The output of this circuit is a cyclic wave whose waveform is the spectrum of the signal that has been accumulated by the input.

(Bickel, 1960). Ignoring first the heterodyne element, the signal is delayed by a fixed time interval T and is returned to the input after correcting for any attenuation. Once an impulse has entered this loop it continues to cycle round it, appearing at the input point after times T, $2T$, $3T$ and so on. It is convenient first to consider the voltage that appears at the output of the loop at an instant that will be treated as the time origin, t zero. The voltage is the sum of the input voltages that occurred at 0, $-T$, $-2T$, $-3T$ and so on. It can be thought of as the integrated product of the input signal pictured as curve (a) in Fig. 5.11 with a comb of ordinates or unit delta functions pictured in curve (b). If this factor were a sinusoid with frequency $1/T$ or f_0 the momentary voltage would be a measure of

Fourier amplitude in the signal

$$a(f_0) = \int_{-\infty}^{\infty} g(t) \cos 2\pi f_0 t \, dt$$

However, the comb has many frequencies and is expressible as

$$1 + 2 \cos 2\pi f_0 t + 2 \cos 4\pi f_0 t + \cdots$$

Thus the momentary voltage is the sum of many Fourier amplitudes

$$a(0) + 2a(f_0) + 2a(2f_0) + 2a(3f_0) + \cdots$$

To avoid confusing these different components, the input signal is first put through a band pass filter, not shown in the sketch, that is arranged to transmit frequencies in one of these ranges only, say from above $(n - 1)f_0$ to include nf_0. The output voltage is a measure of $a(nf_0)$.

The heterodyne element allows this idea to be applied to all the other frequencies in the range selected by the band pass filter. The heterodyne introduces a phase advance of $2\pi t/T$ into all the frequencies that pass it. When a portion of signal is cycling round the loop, it will be observed that it experiences the same phase advance each time it passes the heterodyne, since its times of passage differ by the delay interval T. Consequently, the portions of the signal that are detected simultaneously at the input of the loop at any instant t_0 are in fact the values of the signals fed to the apparatus at times t_0, $t_0 - T$, $t_0 - 2T$... etc. and they have had impressed on them phase advances of 0, $2\pi t_0/T$, $4\pi t_0/T$, etc. This reduces the natural phase increase of the signal by $2\pi t_0/T$ in each interval T. Such a rate of decrease of phase, if it were applied continuously, would amount to decreasing the frequency of every component of the signal by an amount t_0/T^2. Actually, these particular phase advances occur only in the portions of signal selected by the delay loop, but the overall effect is the same as if first the frequency change had been applied to the entire signal and then the delay loop had selected and added its equally spaced values. Thus, the delay line selects frequencies nf_0 in the modified signal, and these are frequencies $nf_0 - t_0/T^2$ in the actual signal.

As time increases, from 0 to T, the apparatus responds to the frequency components starting at nf_0 and falling to $nf_0 - 1/T = (n - 1)f_0$. At this stage the phase shifts have all become multiples of 2π so the process in effect begins again. The output is therefore a repetitive wave whose form is the graph of Fourier amplitudes between nf_0 and $(n - 1)f_0$ in the input signal; and this waveform can be displayed on a cathode ray tube as a picture of the spectrum in this range of frequency.

It should be emphasized that this system is very superior in principle to the process of exploring the spectrum by a filter whose frequency sweeps slowly over the desired frequency range. In the present system, the full spectrum (in any range $(n - 1)f_0$ to nf_0) is got from a single adequate sample and is displayed immediately the sample is completed. With a tuneable filter, every independent point on the spectrum is based on its own adequate sample of signal, so the time taken to explore the spectrum is much greater.

Full details of the circuit have not been published. Since the modulator must only increase the phase and not reduce it as well, it seems likely to be a two-channel modulator of the type discussed in Figs. 5.7 and 5.9.

Optical Filters

Some optical devices for frequency analysis act in a manner closely like that of a mechanical or electrical filter.

Foster (1946) describes the very simple apparatus shown in Fig. 5.12. The function that is to be examined is presented as a white or translucent illuminated profile. In front of it is set a grid of equally spaced rods or bars, all perpendicular to the abscissa of the profile. Beyond this again is a translucent screen. Fringes that appear on this screen show the existence of periodicities in the illuminated profile. In explanation of its action it can be seen that each ordinate in the illuminated profile will cast a shadow of the grid on to the screen. The aggregate of all these elementary patterns is the pattern seen on the screen. This is clearly analogous to the action of a reed or tuned circuit, though the pattern develops in distance instead of in time.

From another point of view, any particular place on the translucent screen "sees" the profile only through the gaps in the grid. If there happens to be a periodicity in the profile, such that all the high parts are simultaneously obscured by the bars of the grid, the place on the screen will be darker than the average. Places on the screen at a little distance to the left or right will however be brighter than the average because, as seen from those positions, the grid obscures the successive low parts of the profile. Thus fringes appear on the screen if there is a periodicity in the profile whose wavelength

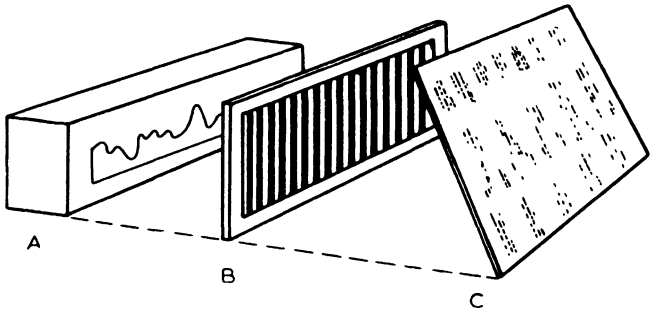

FIG. 5.12. Foster (1946). A simple optical method of detecting periodicities in the illuminated trace by fringes that appear on the translucent screen.

exceeds the spacing of the grid by just the same factor as the distance of the profile from the screen is greater than that of the grid from the screen. The position of the grid between the profile and the screen is made adjustable, so that fringes can be sought. It helps to have the screen tilted as the diagram shows, so that a range of different fringe patterns can be seen at the same time.

A defect of the system is that the grid has a transmission factor that is zero or unity, whereas it should ideally show a sinusoidal variation. The higher harmonics of the grid pattern cause additional fringes having only $\frac{1}{3}$, $\frac{1}{5}$ the spacing of the proper ones. These are obvious when they occur, and can be ignored.

IMAHORI (1914), BARBER (1949). This optical analyser is illustrated in Fig. 5.13. The signal is made to vary the intensity of a source of light, either through a Kerr cell or through a vane moving over the

face of a lens. The light illuminates photographic film or paper that moves steadily past the field. In front of the film or paper is a stationary mask carrying a fan-shaped arrangement of alternate black and transparent zones. The mask can be placed so that the shadow it casts on the film is a hazy one, and ideally the illumination in the

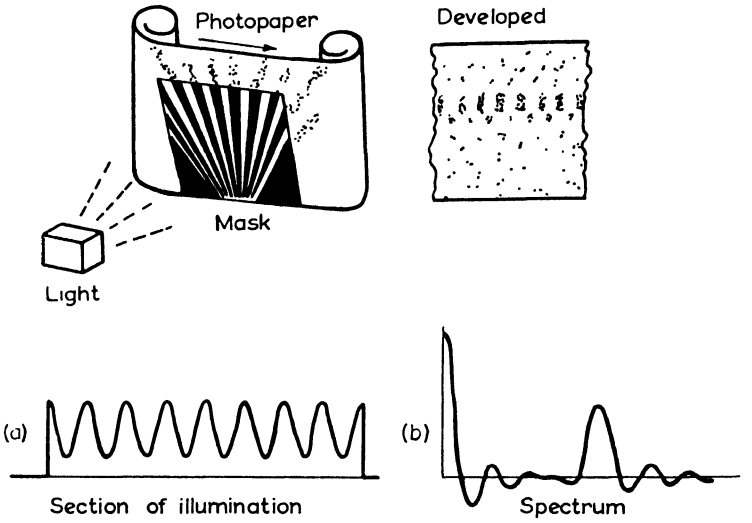

FIG. 5.13. Imahori (1941). Sinusoidal periodicities in the fluctuation of the light source appear as fringes on the moving photo paper after it has been developed. Different frequencies produce fringes at different positions on the photo paper. Imahori used masks of a size suitable to record the fringes on 35 mm film and applied the system to the analysis of speech sounds.

shadow should show a sinusoidal variation. When the film has passed through the apparatus, it shows on development a pattern indicating the frequency spectrum of the signal presented by the fluctuating lamp. The presence of a certain frequency is indicated by fringes appearing on the film. It is clear that a given small patch of film will pass the apertures of the mask in quick succession if it happens to lie on the side of the mask where they are closely spaced, or will pass them in slower succession where the apertures are wider. Some point

on the paper will pass the apertures at the same frequency as that at which the lamp happens to be fluctuating. It may "see" the lamp through the apertures whenever the lamp is bright, in which case it receives more total light than the average, or else it is always in shadow when the light is brightest, in which case its exposure is less than the average. Alternate light and dark fringes therefore appear in this region of the film. If a number of frequencies are present in the modulation of the light, then fringes appear at a number of zones on the film.

Taking a different point of view, one sees that each momentary emission of light from the lamp casts a pattern of illumination on the film. Because the film is moving, the patterns produced by successive emissions are displaced along the paper. The final pattern which the paper shows is the sum of all these elementary patterns. The action of the apparatus therefore resembles the action of a tuned circuit, or rather a bank of tuned circuits, and the fringes correspond to the oscillatory output of the tuned circuits. The frequency passband is the spectrum of the pattern thrown on the paper by the mask and it can be designed to have a variety of characteristics. The intensity of the pattern thrown by the mask along one particular line parallel to the progress of the paper may approximate to a sinusoid, cutting off fairly sharply at the ends of the mask. This is illustrated in curve (a) of Fig. 5.13. The spectrum of the characteristic is sketched in curve (b). It includes a band near zero frequency, since the illumination cannot be negative. As a result, the apparatus also selects zero frequency from the signal and displays it as infinitely long fringes, that is a grey background upon which the other frequencies are seen. The diagram indicates some side bands in the spectrum: these appear as flanking fringes on either side of the main ones, when the apparatus examines a sinusoidal signal. They can be prevented by grading the transmission of the mask towards its beginning and end.

It is desirable of course that the shadow of the mask on the paper should be sufficiently hazy to be sinusoidal. Fringes may appear on the paper at places where the signal frequency matches some harmonic of the pattern. But even if such false fringes occur in analysis,

they can be recognised because they have a spacing only $\frac{1}{2}$, $\frac{1}{3}$, etc., of the usual spacing appropriate to the position where they lie.

A virtue of the system is that the phasing of the fringes reveals the phase of the signal. A dual system can be made comprising two light sources and two masks recording their spectra side by side on the same sheet of paper or film, and one may compare the phases of equal frequencies in the two signals. The system therefore detects frequencies and their phases. Since amplitudes however are displayed as the intensity of fringes, relative amplitudes are difficult to assess with confidence.

CHAPTER 6

THE SPECTRUM OF "NOISE"

The Problem

Previous chapters have shown how one may find the spectrum (or Fourier transform) of a signal that either continually repeats itself or is a transient. Some studies however call for the examination of a signal which is not repetitive and yet continues indefinitely; a continuous "noise". For example, Fig. 6.1 shows a record made by an

FIG. 6.1. A "noise" function.

instrument lying on the sea bed at a depth of 20 m. It shows the fluctuating pressure in the water resulting from the passage of ocean waves. Much information about the storms from which the waves come can be got from such records but it is evident that the records have no real beginning and end and are certainly not repetitive.

Analysing a Sample

One attack is to select a portion of the signal, say the signal occurring between times $-\frac{1}{2}T$ and $\frac{1}{2}T$, and analyse this sample. But although one does not know what happened before the start of the sample or after its end, one has to invent some sort of history to cover these periods. It is impossible to make a Fourier analysis of a function unless its behaviour is known (or assumed) for all instants between $-\infty$ and ∞. There are two possible assumptions one might make, both certainly wrong. One might suppose the sample were one

cycle of a repetitive function and adopt the methods of Chapters 1 and 2 to represent it as a Fourier series. Alternatively, one might suppose that the signal was actually zero at all times before and after the interval that the sample represents, and use the methods of

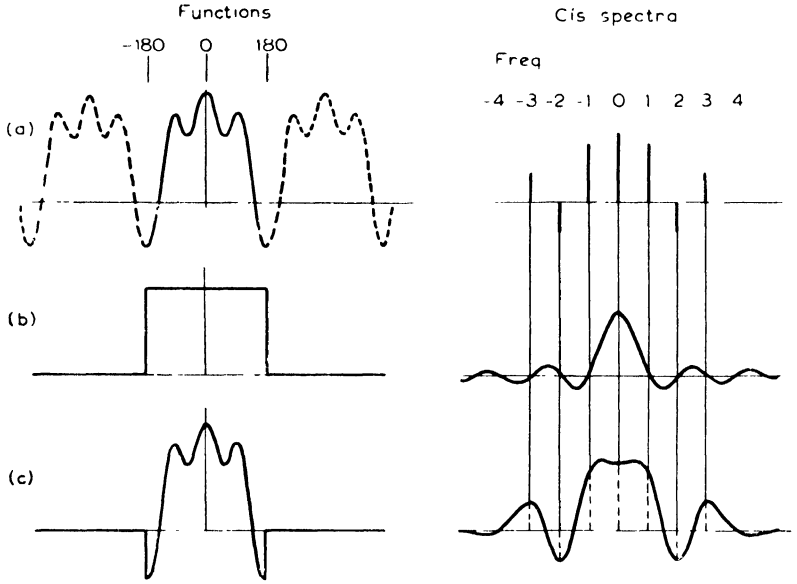

FIG. 6.2. For simplicity, the sample shown as a full line has only three oscillations and is symmetrical to avoid the need of representing complex values in the transform. A suitable noise sample would normally contain several hundred oscillations. The sample may be transformed to a Fourier series as in (a) or to a continuous spectrum as in (c) but the two aspects are closely related.

Chapter 2 to represent it by a continuous spectrum. Although the resulting spectra look difficult and are both certainly wrong, it can be shown that they contain the same real information. The curve (a) in Fig. 6.2 shows the sample repeating itself, and on the right is the corresponding Fourier series spectrum, expressed this time in cis functions, as explained in Chapter 4. Curve (c) shows the sample extended by zero values, and its spectrum also is pictured in complex form on the right. In both these spectra the amplitudes would

normally have complex values, but for simplicity the diagram shows a symmetrical function whose transform is all real. Curve (c) can be got from curve (a) by multiplying it by the square topped factor represented in curve (b). Consequently, by the convolution theorem Eqn. (4.25), the spectrum or transform of (c) is the convolution of the transforms of (a) and (b). That is, each ordinate in the transform of (a) expands into a pattern like the transform of (b), and these add. It so happens that the zeros in the transform of (b) are separated exactly by the harmonic interval $1/T$; in this convolution process, the value got at any harmonic frequency is quite unaffected by the other Fourier harmonics; in other words, the transform of (c) is like the transform of (a) with the tops of the ordinates joined into a smooth curve. Of course the numerical values are different, because in the series spectrum one plots amplitude and in the continuous spectrum one plots amplitude density. The amplitude of any Fourier harmonic in the transform of (a), when divided by the frequency interval $1/T$, gives the corresponding amplitude density in the transform of (c). This is evident from the formulae for the cosine amplitudes in the two cases.

$$a_n = \frac{1}{T} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} g(t) \cos 2\pi n t/T \cdot \mathrm{d}t \tag{1.7}$$

$$a(f) = \int_{-\frac{1}{2}T}^{\frac{1}{2}T} g(t) \cos 2\pi f t \, \mathrm{d}t \tag{2.10}$$

If f is equal to one of the harmonic frequencies n/T then

$$a(f) = a_n T$$

Representing a Curve by a Series of Ordinates

This is perhaps a convenient place to remark that some analysing machines and all numerical methods of computing the spectrum are unable to represent a continuous function as it really is, but must represent it as a set of ordinates, its values at times 0, τ, 2τ etc., where τ is sufficiently small. At all intervening times, the analysis

proceeds as if the function were zero. Figure 6.3 illustrates the difference between a continuous function curve (a) and its very "spiky" representation by ordinates, curve (c). One could get curve (c) by multiplying curve (a) by a "comb" of equally spaced ordinates (b) of value unity, the intermediate values being zero. The convolution theorem will therefore show how the transforms are related. The transform of (c) is the convolution of the transforms of (a) and (b).

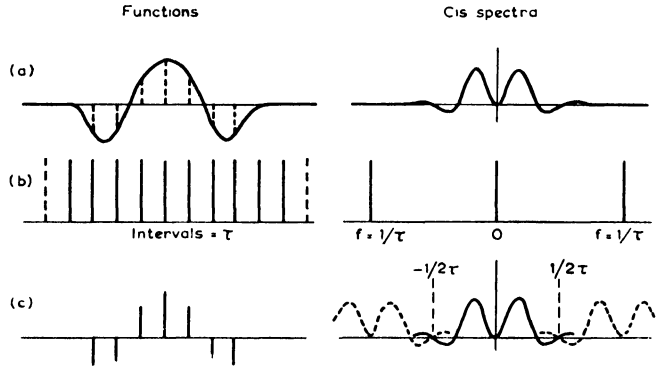

FIG. 6.3. When a continuous function is represented by equally spaced ordinates τ, replicas of the true spectrum arise centred on points $\pm 1/\tau \pm 2/\tau \ldots$ on the frequency scale. The original spectrum is not confused so long as it contains no components of frequency higher than $\frac{1}{2}\tau$.

Now the transform of (b) is another "comb" whose ordinates lie at frequencies 0, $\pm 1/\tau$, $\pm 2/\tau$, etc. The convolution is the superposition of a number of patterns like the transform of curve (a), their origins lying at the frequencies 0, $+ 1/\tau$, $+ 2/\tau$ etc. Analysers that work on ordinates give a repetitive spectrum like this. The Michelson-Stratton analyser is an example. It uses 80 ordinates and 80 gear trains with speeds in the proportions 1:2:3: etc.:80, but there comes a time when the slowest gear has made a complete cycle so that it and all the others have returned to their initial positions. The entire spectrum is then drawn again.

One must of course ignore this repetition of the spectrum. In computations, there is no point in exploring the spectrum to frequencies

beyond $\pm 1/\tau$, since it is a mere repetition of the values in this range. But it should be noticed that even the portion of the spectrum lying between $\pm 1/\tau$ is not necessarily correct; if the original continuous function really has frequencies outside this range, the image spectra that centre upon the frequencies $-1/\tau$ and $1/\tau$ will extend into this range. In fact, the amplitude calculated for any frequency f is the sum of the true amplitudes for frequencies f, $f \pm 1/\tau$, $f \pm 2/\tau$, etc. This effect is called "aliasing". To avoid it, the internal τ between the successive ordinates must be made so small that $\frac{1}{2}\tau$ is greater than any frequency present in the original trace. Another way of saying this is that τ must be smaller than half the period of the shortest wave component in the signal.

Long and Short Samples

It is time to return to the original problem. Having analysed a sample, what significance is one to attach to the resulting spectrum? Is it better for instance to analyse a long sample or a short one? Figure 4.6 illustrates this point. Curve (a) represents a long sample, and on the right is its transform or spectrum. Now if a short sample is selected from the long one, this is equivalent to multiplying the record by a factor which is unity within some short range of time and is zero before and after. This factor is pictured in curve (b), and the product of (a) and (b) is the short sample shown as (c). Now the transform of the square-topped function (b) is a fairly narrow slightly oscillatory curve. The convolution theorem says that the transform of the short sample is the convolution of the transforms of (a) and (b). It is in other words a weighted average of one, using the other as a weighting function. It is a blurred copy of the transform of the long sample; it must therefore contain less detail. A short sample will therefore show only the broad outlines of the spectrum, while a long sample will show more detail.

But it is still not certain that the spectral details got from a long sample will be significant. One may take a number of long samples; the signal never repeats itself, so the samples will all differ in detail. Their spectra will consequently differ in detail too, although one

would hope that they would have a resemblance in their "main features", whatever that phrase might mean. It should be borne in mind that these spectra themselves can be regarded as blurred copies of still more detailed spectra got from still longer samples. This all suggests that the technique of Fourier analysis is not appropriate to "noise" signals. It would be better to decide beforehand what sort of information is significant and what is not, and to adopt a technique of analysis that selects the one and ignores the other.

The Statistical Approach

It is possible to make up a series by selecting successive numbers at random from an agreed range, say -99 to $+99$. Less erratic series can be got by suitable averaging. Such random series are reasonable representation of "noise", on the understanding that one thinks of the successive values as representing equally spaced ordinates on a continuous curve.

It is possible to discuss in theory the transform of a series like this, or the transforms of samples from the series. If the selected sample has a duration T, it can be represented by a Fourier series of cis functions.

$$g(t) = \sum_{-\infty}^{n=\infty} \mathbf{A}_n \, \mathrm{c}^{i 2\pi n t/T} \tag{6.1}$$

Theoretical discussion shows that

1. No correlation is expected between the amplitudes $\mathbf{A}_n$, $\mathbf{A}_{n+1}$ of successive terms. These amplitudes are complex numbers whose phases or arguments appear to arise quite at random.

2. No correlation is expected between the different values of any one harmonic amplitude in different samples. Nevertheless a connection exists. It is as if the values $\mathbf{A}_n$ from different samples were random choices from a gaussian family of complex numbers. If the variance or mean square modulus of this family is large, then the modulus of $\mathbf{A}_n$ got from any one sample is likely to be large, though there is no certainty that it will be so. The argument of $\mathbf{A}_n$ is quite random.

The Power Spectrum

In the analysis of noise, therefore, the significant quantity is a statistic; one should aim to measure the variance of the family from which the amplitudes A_n are derived. If the sample is a long one, so that the different Fourier components are very close together in frequency, there is usually good reason to expect that a large number of adjacent harmonic amplitudes A_n, A_{n+1} etc. come from families with the same variance. One may estimate the variance by calculating the mean square modulus of a number of neighbouring harmonic amplitudes. In real sinusoids, the mean square amplitude is closely connected with the energy or power of the sinusoidal motion. The variance is therefore often referred to as the "power", and the way in which the variance depends upon the frequency is usually called the *power spectrum*.

INSTRUMENTS FOR FINDING THE POWER SPECTRUM

Filter Analysers

Filter analysers are particularly suited to recording the power spectrum of an incoming signal. The filter can be adjusted to have a suitable passband, and the fact that it does not conveniently measure the relative phases of the different frequencies is no drawback to the recording of a power spectrum. Consequently the filter analysers described in Chapter 5 are all suitable machines for finding a power spectrum, provided that one can measure or record the power in the signal that passes the filter. For example, Darbyshire (1955) in studying sea waves used the "wheel" analyser described in Chapter 5 to give a Fourier amplitude spectrum of the wave record, and then calculated the power in certain frequency bands by squaring the Fourier amplitudes. A very similar result could be achieved by using a broad filter in the machine and arranging for an instrument to form a running average of the power transmitted by the filter.

In studying the changing power spectrum of speech, workers at the Bell Laboratories produced an instrument in which a portion of

speech 2.4 sec long was recorded on a magnetic drum. This was continually replayed and fed to a fairly broad heterodyne filter whose power output was recorded as a line of varying blackness on the paper surface of another drum turning synchronously with the first. As the tuning of the heterodyne filter was slowly changed, the recording stylus moved slowly down the recording drum, so that the final result was a plot of frequency against time, with the power displayed as a picture of varying blackness. This display has been called "visible speech" (Potter 1945, Steinberg and French 1946,

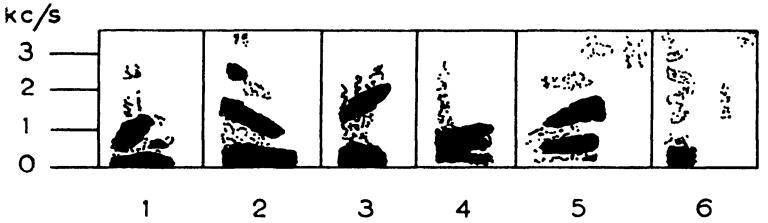

FIG. 6.4. "Visible speech" The power spectra of the numerals spoken in English (after Koenig, Dunn, Lacy 1946).

Riesz and Schott 1946, Gruenz 1951). Figure 6.4 represents one of many examples given by Koenig Dunn and Lacy (1946). In these studies the precise value of the power is not of first importance, so the rather crude effect given by the varying blackness of the trace is quite sufficient to show the character of the changing spectrum associated with different syllables. Attempts have been made to reconvert such analyses back to speech sounds (Schott, 1948; Cooper, 1950). The patterns of "visible speech" have proved valuable in helping congenitally deaf persons to develop clear speech. For this purpose, an immediate display of the spectrum pattern is very desirable, and a bank of filters displaying their power outputs simultaneously on an oscilloscope screen is more suitable than the scanning analyser.

Edgardh (1951) and Kaule and Johne (1956) describe electrical heterodyne instruments which analyse the sound as it is caught in a microphone and display the spectrum on the screen of a cathode ray tube.

BROWN, LYTTLETON (1947); BROWN, COLEMAN, LYTTLETON (1949). This apparatus is strikingly like that of Brown, Furth and Pringle (Chapter 3) though it was being developed independently at about the same time. Here the aim was to get high resolution in displaying the frequencies present in speech, and since the records took the

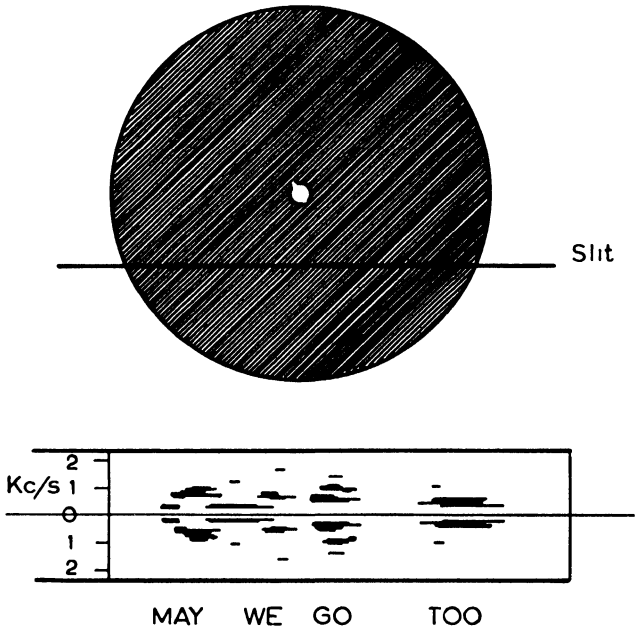

FIG. 6.5. Brown, Coleman, Lyttleton (1949). A speedy method of recording the power spectrum of noises recorded on "sound track".

form of sound track on 35 mm film, the fringe spacing had to be very fine. The fringe plate shown in Fig. 6.5, was 6 inches in diameter, carrying 55 lines to the inch. It was desired, too, to show the changing character of the voice spectrum. Consequently, the sound track was drawn slowly through the analyser and the spectrum was displayed as the varying intensity of a cathode ray spot; this was swept in synchronism with the rotation of the plate and the display was itself photographed on moving film. A sound track

carrying a single pure tone gives a recorded spectrum in which a central line represents the mean unmodulated transmission of the track and it is flanked by two lines representing the single frequency. Speech was expected to show a number of coexisting frequencies corresponding to the various harmonics of the pitch of the vocal cord, certain of the harmonics being augmented by the resonances of vocal cavities. These characteristics are evident in the spectrum of the spoken word "there"; an example given by the authors in the 1947 paper.

The Conflict between Speed Resolving Power

In all these instruments that make an analysis of the signal as it is presented in real time, there is an unavoidable conflict between getting the spectrum quickly and getting it in sufficient detail and accuracy. Suppose the filter has a bandwidth A cycles/sec; this limits the amount of detail that can be resolved in the spectrum. Suppose that the instrument is arranged to average the power during a time interval T. How accurately will this estimate the true mean that would be got from a much longer averaging time? During the prescribed interval T, the noise signal can be thought of as made up of sinusoids that are harmonics of the total length and have frequencies that are multiples of $1/T$; that is their frequencies lie at intervals $1/T$ on the frequency scale. Consequently the number of harmonics which are transmitted by the filter is TA.

The theory of random noise shows that in a large number of samples, all this duration T, the amplitude (and phase) of the harmonic of any specified frequency will vary randomly from one sample to the next, but all can be regarded as samples from a family whose variance is the power density at this frequency. If it can also be assumed that the power density does not vary significantly over the pass band A, all the harmonics that pass can be regarded as having amplitudes that are independent samples from the same family. There are TA of them. How will their mean square modulus differ from the variance of the whole family; differ, that is, from the true value of the power spectrum?

The question is answered by Kendall (1948). When the number of harmonics N, is large, the error tends to have a gaussian distribution and the probable error is a fraction $1/N^{\frac{1}{2}}$. An accuracy of 5 per cent would call for 400 harmonics; that is, the sampling time should not be less than

$$T = 400/\Delta \tag{6.2}$$

The necessary sampling time differs greatly in different studies. For radio work at several megacycles per second, a band width of 10,000 cycles/sec is probably quite narrow enough; the sampling time need be only 0·04 sec. For work in acoustics, the band width should perhaps not exceed 100 cycles/sec, and the necessary sampling time is 4 sec. For work on ocean waves whose frequency is of the order of 0·1 cycles/sec, a band width as narrow as 0·02 cycles/sec may be necessary to distinguish different kinds of wave and swell; here the sampling time rises to 20,000 sec or 6 hr.

This uncertainty in measuring power implies an uncertainty in deciding at what frequency a noise has its greatest power. Indeed it was pointed out in Chapter 2 that a purely sinusoidal signal has a wide spectrum if it continues for only a short time. This "uncertainty principle" is important in the theory of communications (Gabor, 1946).

Diffraction

Optical diffraction offers a very neat way of finding the power spectrum of a signal, when this can be represented as the variation in the optical transmission of a transparency. The scale should be quite small, that is each typical fluctuation in the signal should occupy a length of a millimetre or less.

When a collimated beam of monochromatic light (wavelength λ) falls on the transparency, the ray emerging from a point x at the diffraction angle θ has a shorter path, and a phase diminished by $2\pi\theta x/\lambda$. The total amplitude of this diffracted beam is therefore proportional to

$$\mathbf{A}(\theta, \lambda) = \int_{-\infty}^{\infty} g(x)\, \mathrm{e}^{-i2\pi\theta x/\lambda}\, \mathrm{d}x$$

Here $g(x)$ is the factor by which the transparency changes the amplitude of light passing through it at position x: it will be a complex number if changes in optical path occur in the transparency. The cyclic factor represents change in phase from the extra path length in the diffracted beam. Evidently, the amplitude **A** represents in magnitude and phase the exact Fourier transform of the signal $g(x)$. Measurements are made at various angles θ, not of the amplitude of the light, but of its squared modulus, the intensity, and this evidently shows the power in the spectrum.

In practice it is convenient to start with a line source of monochromatic light and to use a large long focus lens to form an image. The transparency is then placed over the face of the lens, with its x axis perpendicular to the slit. Schouten (1938) has used a "variable area" trace whose varying width represents the signal (including some additive constant, since the width cannot be negative). A "variable density" trace can also be used, but since transmission is reckoned on terms of intensity, the transmission should be made proportional to the square of the width of the corresponding variable area trace. In both cases, the additive constant, being a component of zero frequency, produces an undiffracted beam. It is desirable to keep this small, that is, modulate the transparency as much as possible.

The boundaries of the lens limit the sample trace. A sharp boundary should be avoided because the additive constant then appears as a square-topped component whose spectrum of diffracted beams spreads into that of the signal and confuses it.

Diffraction readily lends itself to the analysis of two-dimensional patterns. Figure 6.6 for example represents a photograph of the sea taken vertically from an aircraft, when the sun was low. The pattern of surface waves could readily be seen. When this transparency was used to cover the face of the lens, the light source in this case being a pinhole, the image of the pinhole became enlarged. The wave pattern can of course be thought of as the sum of many long crested sinusoidal wave trains having a variety of wavelengths and a variety of directions. Its Fourier transform and its power spectrum also, must represent these different waves, and are therefore two

dimensional. The diffracted image in Fig. 6.6 is the power spectrum. At a point having polar coordinates $(r, 0)$ relative to the centre of the image, the intensity of the image denotes the power in the transparency of those sinusoids whose normal lies in the direction 0 and whose frequency (cycles per unit distance) is proportional to r (actually $r/\lambda D$ where D is the distance of the image from the transparency).

The difficulty of the method lies in ensuring that the transparency faithfully represents the quantity that is to be analysed. The trans-

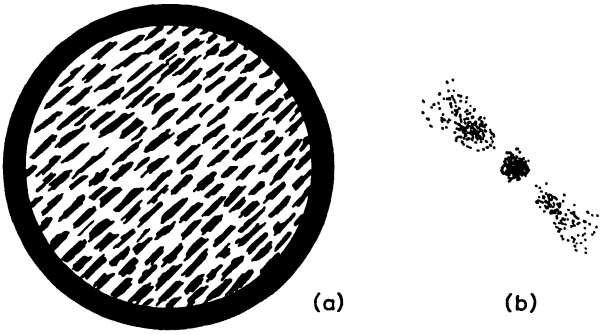

FIG. 6.6. Barber (1949). The transparent photograph of the sea gives a diffraction pattern that is the power spectrum of the function represented by the *amplitude* transmission coefficient of the transparency.

mission coefficient of a sea photograph for instance is unlikely to depend in any predictable way upon water elevation or water slope.

To obtain reasonable angles of diffraction, it is necessary for the diffracting pattern to be miniaturized. This can be done by making a much reduced copy of the plotted function, using plates rather than film since celluloid is a poor optical material. Germansky (1930) got over the difficulty by masking a surface silvered mirror with non-reflecting pigment and using it very obliquely to reflect parallel monochromatic light. The foreshortening effect makes the function behave as if it were drawn on a much smaller scale.

Debye and Sears (1932) and Lucas and Biquard (1932) appear to have hit independently upon the idea of using diffraction to detect the wavelength of supersonic waves generated in clear liquids by a

vibrating crystal. A collimated beam of monochromatic light passes through the liquid in a direction parallel to the wave fronts of the sound. Diffraction occurs because the spatial variations in the pressure, and consequently in the density, give spatial variations in phase to the emergent beam. The angle of diffraction is the ratio of the optical to the supersonic wavelengths. Diffracted beams also occur at multiple angles if the induced phase variations much exceed a radian (Raman and Nath, 1936).

The principle is useful for measuring the wavelength and hence the velocity of high frequency sound. Bergmann (1934) demonstrated the presence of harmonics in a high frequency electrical voltage that was used to drive the crystal. Becker (1939) used diffraction to study the power spectrum of audible sound between 100 and 6000 cycles/sec by causing it to generate capillary ripples on the surface of mercury which then diffracted light that was reflected from the surface.

Meyer (1935) measured the power spectrum of sound in air by reflecting it from parallel rods arranged as a concave grating. This is a converse use of diffraction, corresponding to optical spectroscopy using finely ruled gratings. One may use a known regular grating to find the power spectrum of the incident radiation or one may use monochromatic radiation to examine the power spectrum of the diffracting object.

L. Bragg (1944) described a new use of optical diffraction in helping to decide crystal structure from X-ray diffraction diagrams. A regular pattern of spots on a photographic plate was produced in what was thought to be a reasonable guess of the crystal structure. This was produced by a "fly's eye", a pinhole camera with a regular array of multiple holes, forming images of an illuminated model of the supposed molecule. The photographic plate, in association with a lens and point source of light, gave a diffracted image that agreed with the X-ray diffraction pattern if the guess at structure was a good one. Some further descriptions of the system are given by Bragg and Stokes (1945), Taylor, Hinde and Lipson (1951), Bru, Cubero and Montis (1953) and Hanson, Lipson and Taylor (1953).

The diffraction of X-rays from crystals is a Fourier process, of course, each diffracted beam corresponding to one particular periodicity of the crystal. The small wavelength makes it possible to study finer structure. For this reason electron beams and beams of neutrons have also been used to study molecular structure (see, for instance, Brockway, 1936; Zinn, 1947; Keating and Antal, 1955; Richter and Steeb, 1959).

A recent paper formulating the principles of processing and filtering information by optical systems, in terms of diffraction and Fourier transforms, is given by Cutrona, Leith, Palermo and Porcello (1960).

CHAPTER 7

CORRELOGRAMS

Power Spectra from Correlograms

In analysing noise, it was shown in Chapter 6 that one should aim to measure not the amplitude of any Fourier component but its square modulus or "power". That is, the signal $g(t)$ has a transform $\mathbf{G}(f)$, but one aims to measure $\mathbf{G}^*(f) \cdot \mathbf{G}(f)$. Equation (4.27) on page 60 states that this quantity is the transform of the convolution

$$\varrho(\tau) = \int_{-\infty}^{\infty} g(t)\, g(t + \tau)\, \mathrm{d}t \tag{4.27}$$

This principle is very commonly put to practical use in studying noise. The reason is that the convolution $\varrho(\tau)$ is usually a very much simpler function than the noise, and it is usually quite a simple matter to calculate its Fourier transform, either numerically or with the help of some analogue machine. Indeed the convolution itself is so characteristic that in many studies one scarcely needs to transform it to a "power spectrum".

It is worthwhile discussing the idea in a little more detail. If a noise $g(t)$ is sampled from $-\frac{1}{2}T$ to $\frac{1}{2}T$, one can think of it as one cycle of a repetitive signal and describe this by a Fourier series

$$g(t) = \sum_{-\infty}^{n=\infty}{}' \mathbf{A}_n\, \mathrm{e}^{i2\pi nt/T} \tag{7.1}$$

The convolution, like the transform $\mathbf{A}_n$ (Eqn. 4.9), is then best expressed by a mean product

$$\varrho(\tau) = \frac{1}{T}\int_0^T g(t)\; g(t+\tau)\, \mathrm{d}t \tag{7.2}$$

Then ϱ will itself be repetitive and be represented by the series

$$\varrho(\tau) = \sum_{-\infty}^{n=\infty} \mathbf{A}_n^* \mathbf{A}_n \, e^{i 2\pi n \tau / T} \tag{7.3}$$

Thus the convolution is made up of the same frequency components as the original noise signal, but their amplitudes now are all positive

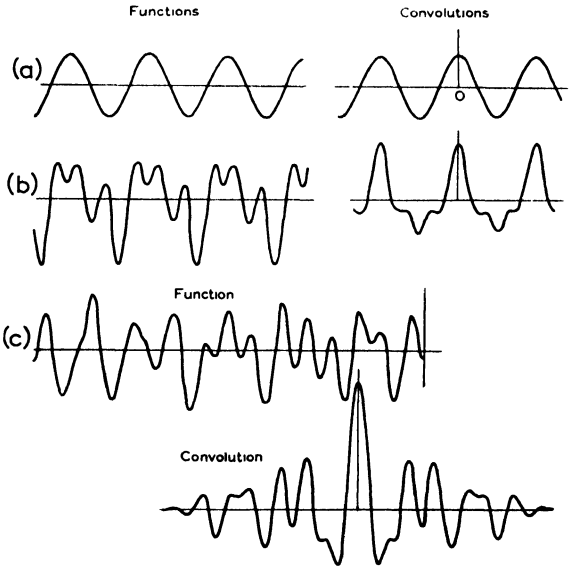

FIG. 7.1. Functions and their convolutions. The convolution of a sinusoid (a) is a sinusoid. Curve (b) contains three sinusoids and its convolution contains the same three sinusoids but their amplitudes are now the squares of what they were in the curve. Because the phase constants are all zero in the convolution, the sinusoids add to give the convolution a central large positive peak. The effect is more marked when eight sinusoids are present as in curve (c).

real quantities $\mathbf{A}_n^* \mathbf{A}_n$. It follows therefore that near its origin all the components of $\varrho(\tau)$ take real positive values and tend to add together making $\varrho(\tau)$ large and positive. At points further from its origin, the cis functions take a variety of different phases owing to their different frequencies, and there is much destructive interference making $\varrho(\tau)$ small as may be seen in Fig. 7.1. Far from the origin,

$\varrho(\tau)$ is usually small, except that near points $\pm T$, $\pm 2T$ etc. the behaviour in the vicinity of the origin is repeated.

It will be recalled that in analysing a noise sample the aim was not to measure the power of each individual harmonic, such as $\mathbf{A}_n^* \mathbf{A}_n$, but to estimate the mean power of a number of neighbouring harmonics. This is very easily brought about: suppose one were to take notice of the relatively large values of $\varrho(\tau)$ occurring in the

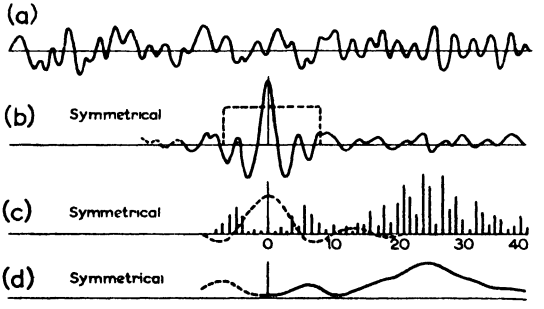

FIG. 7.2. Curve (a) represents a "noise" sample and curve (b) its convolution. If (b) is calculated on the assumption that the noise sample repeats itself, then the convolution (b) also has the same repetition interval. In this case, its transform is the Fourier series spectrum suggested in (c). A transform made only of the central portion of the convolution is a smoothed version of the series spectrum, that is the continuous spectrum (d). This approximates to the power spectrum of the noise. Blackman and Tukey (1959) suggest other ways of weighting the central part of the convolution in order to smooth the spectrum by other weighting curves.

vicinity of the origin, but to ignore and treat as zero both the small values remote from the origin and all the repetitions at $\pm T$ etc. Figure 7.2 indicates that this would be equivalent to multiplying $\varrho(\tau)$ by a factor that was unity for small values of τ but was zero for all large ones. This modified $\varrho(\tau)$ would have a transform that was a weighted mean of the true transform of $\varrho(\tau)$. At any frequency f, it would be an average of all the amplitudes $\mathbf{A}_n^* \mathbf{A}_n$ in the vicinity. This is exactly what is required and it can properly be called the "power spectrum". Moreover if only the relatively large values of ϱ occuring near its origin are to be taken account of, only these

values of ϱ need be calculated and only these values need be used in calculating the transform. Thus one gets just the right information with much less labour than would at first sight seem necessary.

Two sorts of error can arise. The values of the power spectrum may be "unstable", that is they may vary somewhat erratically from sample to sample, or from frequency to frequency in any single sample. This comes about if too few individual amplitudes $\mathbf{A}_n^* \mathbf{A}_n$ are included in the averaged power estimates. The instability can be overcome by transforming an even smaller region of the convolution near its origin: if the multiplying factor is made narrower, then its transform is made wider and will include a greater number of individual amplitudes. This leads however to the second sort of error, by which an excessive smoothing obscures real detail in the power spectrum. Both errors can be reduced by using a more extensive sample of the noise signal; if T is larger, the harmonics crowd together more closely on the frequency range. As a general rule no significant detail in the spectrum need be lost, providing that all the significant range of the convolution is included when calculating its transform. The calculated power spectrum will be stable to a fraction $1/N$ providing that the overall duration of the original sample of noise is N^2 times the largest interval it is necessary to include in the convolution. Thus five per cent accuracy would demand a sample 400 times as long as the length of the convolution.

What has been hitherto called the convolution is often referred to as the "correlogram" though it is perhaps better to use this word to describe the ideal, stable form of convolution that would be got from a very long sample of the noise signal. Formula (7.2) for the convolution $\varrho(\tau)$ shows that it can be thought of as the mean product of the functions $g(t)$ and $g(t + \tau)$, in fact their unnormalized correlation coefficient as discussed in Chapter 8. The word "correlogram" is therefore a natural one for the plot of ϱ as a function of the interval τ.

It will be appreciated that the signal $g(t)$ may sometimes be a distribution in space, not time, in which case t represents a distance, τ is a space interval, and the average extends over a long distance. In the case of waves on the surface of the sea the elevation of the

water surface depends both on time t and the space coordinates x and y. It can be visualized as a distribution in these three dimensions. The correlogram is then the assembly of correlation values determined after the manner of Eqn. (7.1) for combinations of different intervals in all three coordinates x, y and t. The power spectrum, in principle at least, is a three-dimensional Fourier transform of this three-dimensional correlogram.

Space and Time Correlograms

The eddies in a turbulent stream of air or water are a good example of random noise. The momentary velocity at any point, or rather its components in three dimensions at right angles, can be measured by suitable instruments. The velocity in the downstream direction will be a fluctuating value u added to a steady value U. Velocities v and w at right angles to this will fluctuate about zero and have no steady component. If two instruments that both measure u (or v or w) are set close together, the velocities they record will be very similar; the correlation will be high. If they are set further apart, their outputs will show a smaller correlation because some eddies will have too small a diameter to affect both instruments at once. In principle, it is possible to estimate the correlation of the signals for a variety of separations and to display the results as a plot, as in Fig. 7.3. Plots that show merely displacements downstream or across the stream are often used. Since the direction of separation is important, as well as the actual distance, the complete plot would be in two dimensions. Such a plot can be called a "space correlogram". It serves to indicate the general dimensions of the eddies in the stream. The actual measurement of correlation can conveniently be done by the "method of signs" in Chapter 8, for the fluctuating velocities are found to have a normal distribution. Otherwise an electronic mixing circuit can be made to form the mean product of the two signals.

Because the eddies each persist for a brief time and are carried with the stream, an eddy which has affected one instrument is likely to affect another instrument further downstream at a later time.

This suggests that one might make measurements at points separated by a distance Y downstream and compare the signals after delaying that of the upstream instrument by some fixed time inter-

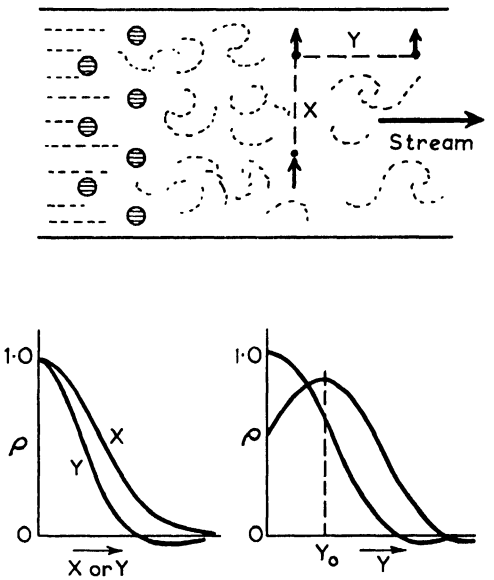

FIG. 7.3. The diagram illustrates measurements being taken of velocity at right angles to the mean flow of a turbulent stream. Measurements at points separated in the same direction as that of the velocity component being measured, give the "longitudinal" correlogram marked X. Points separated in a direction at right angles to the component being measured give the "transverse" correlogram marked Y. Measurements may be compared after one has been delayed by a fixed time interval τ_0. Highest correlation then occurs if the points are separated by a distance $U\tau_0$ downstream, U being the stream velocity.

val τ_0. In this case the plot of correlation against time interval has the form shown in Fig. 7.3. Highest correlation occurs at a distance Y_0 and the ratio Y_0/τ_0 is a good measure of the stream velocity. The correlation values are less than before, showing that the pattern of eddies changed somewhat while it moved downstream during the interval τ_0.

Cross Correlograms

The last curve discussed in Fig. 7.3 is an example of a cross correlogram. The velocities measured at points separated downstream are different functions of time and are correlated according to a formula

$$\varrho(\tau) = \frac{1}{T}\int_{-\frac{1}{2}T}^{\frac{1}{2}T} g(t)\, k(t+\tau)\, \mathrm{d}t \tag{7.4}$$

However, the velocity that exists can be thought of as a continuous function of two variables, the downstream position y and the time t. In this sense it is a two-dimensional distribution $g(y, t)$ and will have a two-dimensional correlogram

$$\varrho(Y, \tau) = \frac{1}{T}\int_{-\frac{1}{2}T}^{\frac{1}{2}T} g(y, t)\, g(y + Y, t + \tau)\, \mathrm{d}y\, \mathrm{d}t \tag{7.5}$$

Evidently the cross correlogram in Eqn. (7.4) is merely a section of the two-dimensional autocorrelogram in Eqn. (7.5), got by specifying a particular value of the interval Y. It is profitable to take this view of "cross correlograms", for the two- or three-dimensional picture often tells the physical story more clearly. It is useful for example in deducing ionospheric winds from the pattern of fading seen in radio signals (Yerg, 1955; Barber, 1956) or in deducing the directional spectrum of waves from measurements made at several points on the sea (Pierson, 1952; Barber, 1954).

Finding Correlograms

Calculation processes. According to Eqn. (7.1), it is required to calculate the mean product of pairs of values t and $t + \tau$, repeating this for a sufficient number of different intervals τ. It is sometimes easier to add the functions and form the mean square of the sum or difference, for this involves the mean product. By this method

$$2\varrho(\tau) = \overline{[g(t) + k(t+\tau)]^2} + \overline{g^2(t)} - \overline{k^2(t)} \tag{7.6}$$

The second and third terms do not of course vary with τ.

A further simplification that avoids much squaring is to add or subtract the functions and then to form an average taking no notice of negative signs. One can deduce the correlation from the relation

$$\pm 4\varrho/\pi = (\text{Mean}\,|\,g \pm k\,|)^2 - (\text{Mean}\,|\,g\,|)^2 - (\text{Mean}\,|\,k\,|)^2 \quad (7.7)$$

which is true so long as the two variables show a jointly normal distribution and each has a zero average. Readers will see some parallel between this idea and that of estimating the r.m.s. voltage of an alternating power supply merely by rectifying the voltage and smoothing it.

Another simplification is the "method of signs" described later in Chapter 8. This method allows one to ignore the magnitudes of g and k taking notice only of their signs. One decides, for each different delay, the fraction of occasions on which the variables show the same sign, both being positive or both being negative. This method, like the previous one, requires the two variables to have a jointly normal distribution and each to have a zero average. In order to get the same accuracy as with methods that take account of magnitudes, more data are necessary. This is illustrated in Fig. 8.6 of Chapter 8.

Glock (1942) advocates a "trend method" for the correlation of continuous time series, that has the advantages of simplicity and speed.

McNicol (1949) deals with the autocorrelation of a varying quantity that is essentially positive and has a Rayleigh law of distribution. Following Furth and MacDonald (1947), he finds the standard deviation σ_τ of the differences between successive values R of the variable at regular intervals τ. He shows that the correlation $\varrho(\tau)$ for this time interval τ is given by

$$\varrho(\tau) = 1 - \frac{\pi^2}{4(4-\pi)} \frac{\sigma_\tau^2}{(\overline{R})^2}$$

This is analogous to the method of squares applied to a variable having a normal gaussian distribution, Eqn. (7.6).

Time Lags

Correlograms can be formed directly between incoming signals that vary with time, but the introduction of a time delay is difficult. Of course if the data can be recorded on punched tape or magnetic tape or as traces of variable area or variable density on photographic film, it is easy to introduce a relative delay in later analysis. Short delays have been achieved in electrical signals by the use of suitable transmission lines, or in acoustic signals by transmission through an adjustable column of mercury. Stone and Dandl (1957) describe a lag introduced by a synchronously switched bank of condensers.

The following instruments for forming correlograms have been classified mainly with respect to the means used for introducing the delay.

Devices with Mechanical Delay

Seiwell (1950) used a machine built with two disc and ball elements. The functions $g(t)$ and $k(t)$ were presented as curves on paper. Two operators, one for each curve, moved a stylus following the curve as the paper records passed through the machine. One disc is driven synchronously with the paper records, and its ball carriage moves with stylus following $g(t)$. The output rate of rotation is therefore proportional to $g(t)$ and it drives the second disc. The second carriage moves with the second stylus, so the output from this stage has a rate of rotation proportional to the product $g(t)\,k(t)$. The total rotation after the records have been through the machine is shown on a dial and is a measure of the correlation. Delays τ can of course be introduced by pre-setting the positions of the paper records, and the curves need to be retraced for each different value of τ.

Stephens (1950) used an air tube, as in Fig. 7.4, to delay the signal and form an autocorrelation. The signal was recorded and reproduced from a loop of magnetic tape, and produced as sound from a cone "speaker". The sound passed down an air pipe to a non-reflecting termination, and pick-ups at the side of the tube could be adjusted in relative position to introduce the desired delays. The mean product of the picked-up signals was found electrically.

Shire and Runcorn (1951) built a machine out of standard components for forming the autocorrelogram of a stationary time series.

Brook and Smith (1952) obtained auto- or cross-correlation by recording signals on two channels on the same magnetic tape and by changing the relative positions of the two pick-up heads to effect the delay. The mean product was found electrically and recorded automatically. The machine described by Hastings and Meade (1952) uses magnetic tape in a similar way and finds the mean product by an electrodynamic wattmeter.

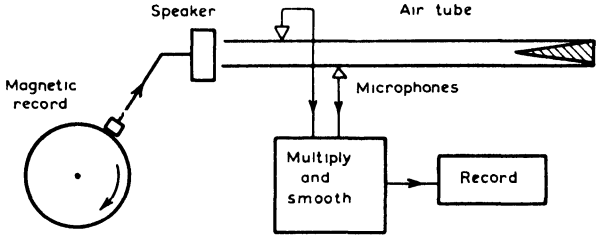

FIG. 7.4. Stevens (1950). An acoustic delay.

Revesz (1954) also obtained a delay by magnetic tape. On replay, the integrated product was got by a d.c. wattmeter. Goff (1955) delayed acoustic signals on a magnetic drum as part of an electronic correlator.

Tucker (1950, 1952) used records presented as white silhouettes on paper, as in Fig. 7.5, and scanned these by an optical beam rotating before the records. A photocell detected the reflected light and gave an output proportional to the sum of the functions on the records. An early model of this apparatus introduced delays by slowly shifting one record past the other. A later modification used an optical system with two beams arranged at an angle that slowly changed as the beams continually scanned the records.

Electrical Delays

Lee, Cheatham and Weisner (1950) form auto- and cross-correlograms by a pulse technique. An oscillator generates brief pulses at an interval greater than any delay that is required and uses them to

sample one incoming signal and to create a series of long pulses whose height is proportional to the values of the signal at the instants that are sampled. The original brief pulses also pass in a second channel through a delay line and are used to sample the second signal. In this case, the circuit generates pulses whose height is constant but whose duration is proportional to the sampled values of the signal. This second set of pulses is used to "gate" the first set

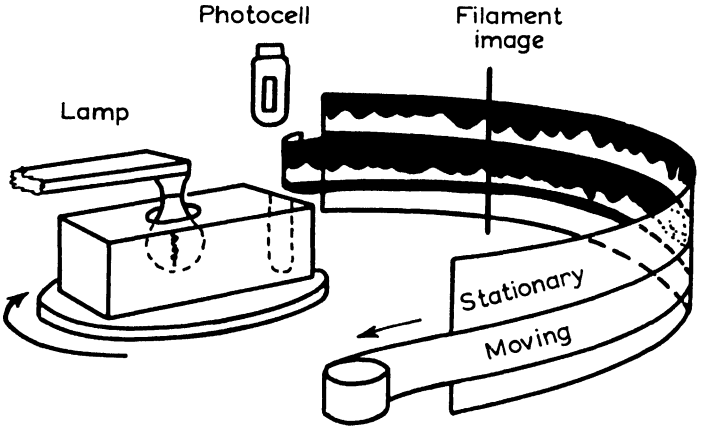

FIG. 7.5. Tucker (1950, 1952). An optical system on a turntable scans a pair of traces whose relative position is slowly changed.

and so produce pulses whose height is proportional to values of one signal and whose duration is proportional to values of the second signal. These are integrated to give a measure of the correlation. The time delay is increased in small steps and the correlogram is recorded automatically.

Kraft (1950) refers to a digital electronic correlator which samples the signals by pulses whose height expresses the value of the signal on a binary code. The machine automatically computes 110 points on the correlation curve. Bell and Rideout (1954) used a delay line variable in discrete steps in a high speed correlator for audio signals, getting 41 points on the autocorrelation curve in only 5 sec.

An interesting method has been used by Lampard (1954). Suppose

that two electrical signals $g(t)$ and $k(t)$ are developing in real time. The aim is to find the correlogram, the mean product of values that occur at a range of different relative delays. It will be recalled from the discussion of filters in Chapter 5 that if signal $g(t)$ is fed into a filter, the output at every instant t is the sum of numerous earlier values of the signal g, in fact a weighted average of the earlier values, the weighting factor being the impulse response of the filter, but reversed on the time axis. If one forms the mean product of the signal $k(t)$ as it arrives and the signal $g(t)$ as it appears after the filter, in effect the signal $k(t)$ is being multiplied by numerous earlier values of $g(t)$. The resulting mean is a weighted average of the correlogram, or at least that half of it corresponding to delays that make g earlier than k. One could obtain different weighted averages of the correlogram by using different filters.

An unknown curve, in this case the correlogram, can be discovered if a large number of different weighted averages of it are known, and if the weighting functions have been chosen on a right principle. Indeed the Fourier series is an example of this. Formula (1.7) in Chapter 1 can be read as showing that each coefficient a_n or b_n is a weighted average of the function $g(t)$, the weighting functions being "harmonic" sinusoids. If the values of a sufficient number of these coefficients are known, the function $g(t)$ may be reconstructed by plotting the value of the series 1.4 against time.

Other families of curves besides sinusoids have this property. Lampard has used curves derived from the Laguerre polynomials and they are shown in Fig. 7.6. The analysis then develops in two successive steps. One signal is fed simultaneously to many filters so designed that their responses to a single pulse have the forms of the curves in Fig. 7.6. The outputs of these filters are all multiplied by the second signal and the products are all averaged over a sufficient length of time. As soon as the averages have been obtained, they are supplied as short impulses to a second set of filters designed like the first set. The outputs of these filters are added, and the sum is displayed on the face of cathode ray tube. The trace is the correlogram of the two signals, or at least the correlogram can readily be got from it. The whole process can be automatic so that the correlo-

gram is displayed at regular short intervals of time and any changes in its form can be seen at once.

This idea has been followed up by Choudbury and Chakrabarta (1958).

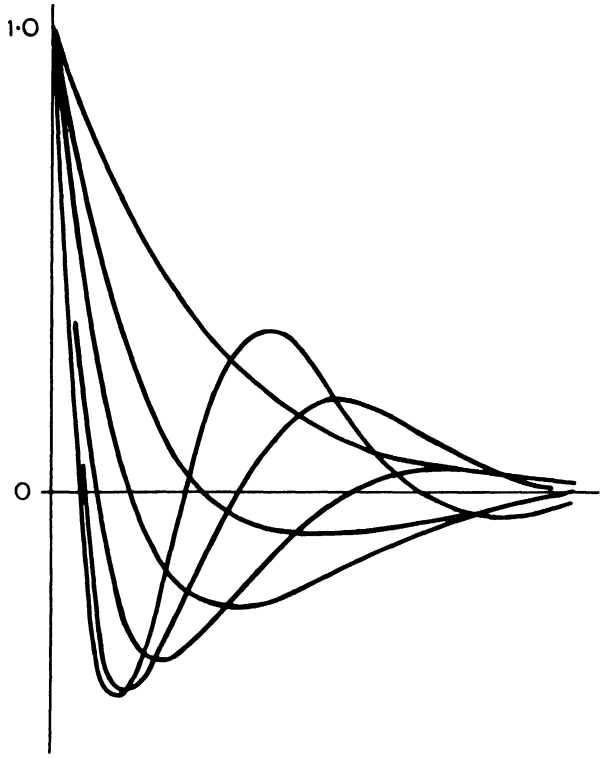

FIG. 7.6. The family of functions used by Lampard (1954).

A very neat optical method of producing space correlograms in one or two dimensions is reported by Kovaznay and Arman. It requires that the function that is to be analysed be represented as a transparency whose intensity transmission is the variable. Two transparencies can conveniently be used. One is placed in front of a uniform extended source of light. The second is placed parallel to it, a distance L away as in Fig. 7.7 and is followed by a lens large enough

to cover the transparency. The image lying in the focal plane of the lens, say a distance F from it, is then examined. Any one point of the image at coordinates αF, βF is illuminated by rays which take paral-

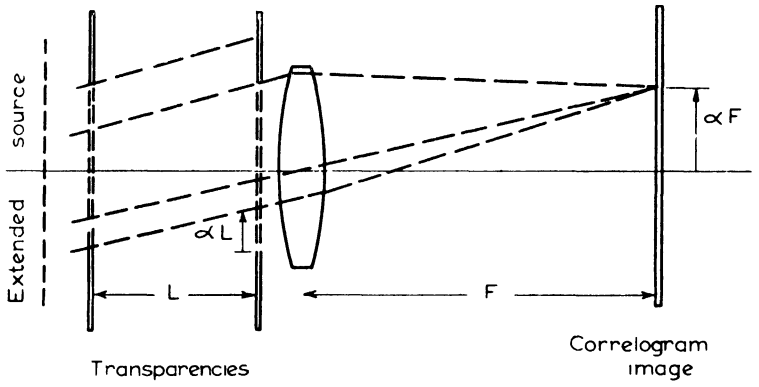

FIG. 7.7. Kovasznay, Arman (1958). An optical method of producing two-dimensional correlograms from two-dimensional functions represented by the intensity transmission of a transparency.

lel paths between the plates and consequently pass through pairs of points whose coordinates have some fixed difference X, Y given by

$$X = \alpha L \quad Y = \beta L$$

Consequently the total illumination at this point is proportional to

$$\varrho(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)\, g(x + X, y + Y)\, \mathrm{d}x\, \mathrm{d}y$$

and is one value of the space correlogram of the transparency. The whole image at the focal plane represents the two-dimensional space correlogram, and its intensity distribution can be measured directly, or photographed directly if this gives a sufficiently accurate copy.

It should be pointed out that the transparency will usually need to represent the sum of the significant pattern and an additive constant. The final image is near to the sum of their correlograms, for though the cross correlogram between the constant and the signal is

present also, it is probably small enough to be ignored. This is suggested by the full and broken curves in Fig. 7.8. The broad distribution due to the additive constant (the broken line) may be found by rotating one of the transparencies through an angle of 90°, so long as its shape is square or circular, for in this relative position the signals on the transparencies are unlikely to show any correlation. The correlogram of the signal will then be the difference

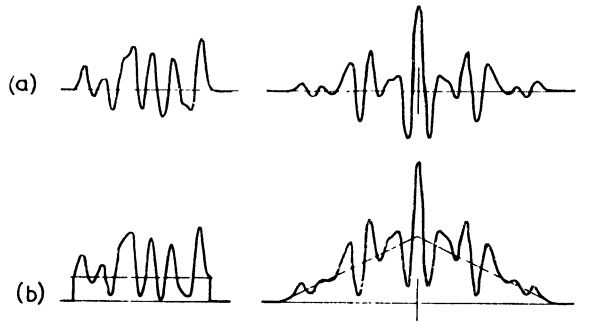

FIG. 7.8. When the analogue demands an additive constant, the true function and its convolution shown in (a), appear riding on the additive constant and its convolution shown by broken lines in (b).

between the images when the transparencies are in this position and when they are given the same orientation.

The method obviously lends itself to the use of different transparencies. The image in that case will be the cross correlogram of the two signals, as defined by

$$\varrho(X,Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)\, k(x + X, y + Y)\, \mathrm{d}x\, \mathrm{d}y$$

Signals that depend upon only one variable, distance or time, can be represented by transparencies whose transmission varies with x but not with y.

CHAPTER 8

THE CORRELATION COEFFICIENT – SCATTER DIAGRAMS

Scalar Variables

The correlation coefficient may be thought of as indicating the degree of resemblance between the histories of two fluctuating quantities, say $g(t)$ and $k(t)$ which both vary with time. It is clear that if the two quantities behave in a somewhat similar way it should be possible, by subtracting a certain proportion of one from the other, to produce a quantity whose fluctuations are in general smaller than the original. That is, one may combine the two in the manner

$$h(t) = g(t) - Ck(t) \tag{8.1}$$

and by a suitable choice of C make $h(t)$ appreciably smaller in general than $g(t)$. One has to decide how to assess the size of a fluctuating quantity. It is common to adopt the mean square or "variance" as a measure of size. The aim then is to choose C so as to make $\overline{h^2}$ as small as possible and then to compare this with $\overline{g^2}$. Now

$$\overline{h^2} = \overline{(g - Ck)^2} = \overline{g^2} + C^2\overline{k^2} - 2C\overline{g \cdot k} \tag{8.2}$$

By differentiating for C it appears that the smallest value of $\overline{h^2}$ is achieved by making

$$C = \overline{g \cdot k}\big/\overline{k^2} \tag{8.3}$$

Then $\overline{h^2}$ is

$$\overline{h^2} = \overline{g^2}\left[1 - \overline{(g \cdot k)}^2\big/\overline{g^2 \cdot k^2}\right] \tag{8.4}$$

Evidently the reduction in size, the ratio of $\overline{h^2}$ to $\overline{g^2}$, depends merely on the value of the quantity ϱ,

$$\varrho = \overline{g \cdot k}\big/\left(\overline{g^2} \cdot \overline{k^2}\right)^{\frac{1}{2}} \tag{8.5}$$

which is called the correlation coefficient. If it is unity, then $\overline{h^2}$ is zero, which means that the two variables are alike in the sense that they differ only by a constant factor of proportionality C. The sign of ϱ serves to show the sign of the proportionality factor C. If ϱ is negative, it indicates that g and k have a tendency to be opposite in sign. The coefficient can range from -1 to $+1$.

It is possible of course that one may be interested not in the actual values of g and k but in their departures from their mean values. In this case, one would calculate the means $\bar{g}$ and $\bar{k}$ and subtract these constants from g and k before evaluating the correlation coefficient.

It is worth noting, that the correlation coefficient can also be written

$$\varrho = \frac{\overline{(g+k)^2} - \overline{g^2} - \overline{k^2}}{2(\overline{g^2} \cdot \overline{k^2})^{\frac{1}{2}}} \tag{8.6}$$

This expresses the coefficient entirely in terms of the mean squares of the quantities g, k and $(g + k)$. In calculation or in the design of analogue machines, it is usually easier to evaluate a mean square than a mean product.

Scatter Diagrams

In many physical studies, particularly studies of wave motions of natural origin, such as microseisms, ocean waves or extra-terrestrial radio noise, it is to be expected that the variable will show a "normal" distribution. The proportion of time during which the variable shows values in any small range g to $g + \delta g$ is expected to be

$$\frac{1}{\sqrt{2\pi}\,\sigma_g} e^{-g^2/2\sigma_g^2} \delta g \tag{8.7}$$

It is the Gaussian error curve sketched in Fig. 8.1. When a variable has this distribution, its mean is zero and its mean square is σ_g^2. (These are the first and second moments of the curve.)

Two such random variables can also be expected to show a "jointly normal" distribution. That is, the chance that one variable lies in a small range g to $(g + \delta g)$ while at the same time the second

variable lies in a small range k to $(k + \delta k)$ is expected to follow a law

$$\frac{\delta g \delta k}{2\pi\sigma_g\sigma_k(1-\varrho^2)^{\frac{1}{2}}}\exp\left|-\frac{1}{2(1-\varrho^2)}\{g^2/\sigma_g^2 - 2\varrho gk/\sigma_g\sigma_k + k^2/\sigma_k^2\}\right| \quad (8.8)$$

Indeed, it is in these cases that the correlation coefficient has its simplest and most useful properties. The two variables both have zero averages. The joint distribution can conveniently be displayed by a two dimensional plot using g and k as rectangular coordinates

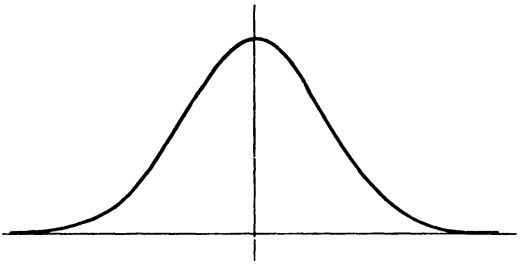

FIG. 8.1. The "error curve" of Gauss.

as in Fig. 8.2. When a sufficiently large number of representative pairs of (g, k) values have been plotted, one may proceed to draw contours of equal density of points. If the distribution is normal, these contours should all be similar ellipses following the law

$$g^2/\sigma_g^2 - 2\varrho gk/\sigma_g\sigma_k + k^2/\sigma_k^2 = \text{const.} \quad (8.9)$$

If there is high correlation, the ellipses are long and narrow. If it is low, the contours are nearly circular. It is possible to deduce the correlation coefficient from the shape of a contour using the formula

$$\varrho^2 = 1 - (k_2/k_1)^2 \quad (8.10)$$

where k_2 is the diameter of the contour measured along the k axis and k_1 is the overall width between tangents taken perpendicular to the k axis. This is illustrated in Fig. 8.2. The correlation in this instance is positive because the major axis of the ellipse lies in quadrants where g and k have the same sign. It would be negative if the axis lay in the other pair of quadrants. The interpretation is

most easily made if the variables have approximately equal variance so that the major axis has an inclination of about 45°. Figure 8.3 illustrates some contour shapes for different correlations.

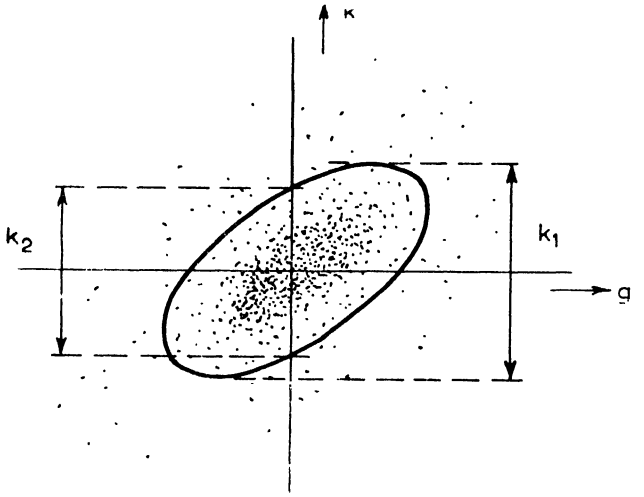

FIG. 8.2. A gaussian distribution in two dimensions. Contours of equal density are concentric similar ellipses. The correlation coefficient can be estimated as $\varrho^2 = 1 - (k_2/k_1)^2$.

Sugar (1954) gives examples of scatter diagrams, both for variables following the normal distribution and for variables each having a "Rayleigh" distribution, like the *amplitude* of radio noise picked up by a tuned aerial.

Display

If the signals are converted to electrical voltages, the scatter diagram can be displayed on the face of cathode ray tube. Its form can be seen by inspection if the noise has a high frequency. Otherwise it is best to record the display by photographing it with an exposure time that is sufficient to collect an adequate sample. The contours of equal density are rather difficult to estimate by inspection, and there is much advantage in using a photographic emulsion

of high contrast in order to make the form of the diagram more obvious.

Goodman (1956) has used the scatter diagram as a null indicator. This avoids the difficulty of measuring a density contour in a diffuse

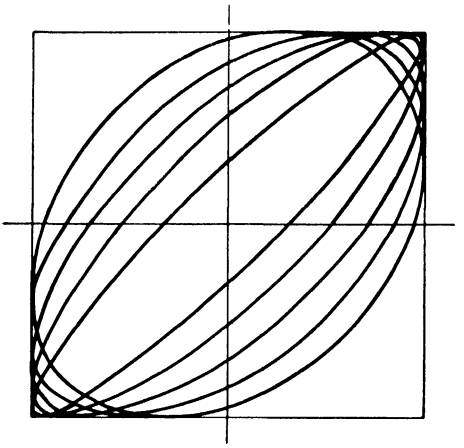

FIG. 8.3. The elliptic shapes corresponding to different magnitudes of correlation, 0·1, 0·3, 0·5, 0·7 and 0·9 in the special case where the variables have equal variance.

picture. The procedure is as follows. A display on a cathode ray tube is formed from the two noise signals, and the size of one or both signals is adjusted till the major axis of the display is approximately at 45° to the two axes. A measured fraction C of one signal is then added to the other, so that a display is formed between $(k - Cg)$ and g. The fraction C is adjusted till by inspection the major axis of the display lies along one or other of the axes, as illustrated in Fig. 8.4. This shows that the correlation has fallen to zero. The fraction C is then an indication of the correlation coefficient of the original signals. Thus if

$$0 = \overline{(k - Cg) \cdot g} = \overline{g \cdot k} - C\overline{g^2}$$

then

$$\varrho = \overline{g \cdot k} \Big/ \left(\overline{g^2} \cdot \overline{k^2}\right)^{\frac{1}{2}} = C \left(\overline{g^2} \Big/ \overline{k^2}\right)^{\frac{1}{2}} \tag{8.11}$$

Criterion of Signs

If the "noise" signals have a jointly normal distribution, and this is often the case, it is possible to assess the correlation coefficient merely by comparing the number of occasions n on which the signals

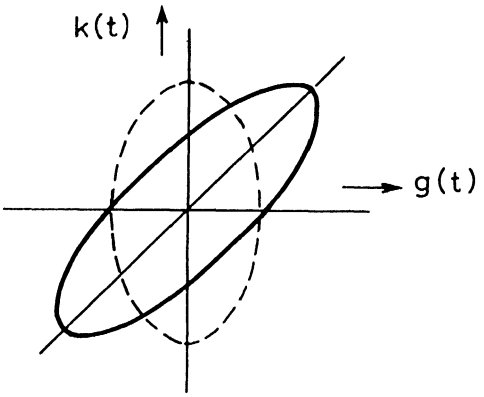

FIG. 8.4. Goodman (1956). A null method. The scatter diagram is displayed on the screen of a cathode ray tube. A fraction of one signal is added to or subtracted from the other till the major axis of the diagram becomes parallel with one or other coordinate axis, indicating zero correlation (broken curve).

have the same sign with the total number of occasions N that are sampled. The coefficient is given by (Cramér, 1946), as

$$\varrho = -\cos \pi n/N \tag{8.12}$$

and the relationship is plotted in Fig. 8.5. It is necessary of course that the mean of each signal shall be near to zero. This procedure has the great attraction that only the signs of the signals are needed, not their actual values, and very simple switches and counters are sufficient to record the total number of instants that are sampled and the number of these instants at which the sign of the signals is the same.

The accuracy of the estimate got from a given noise sample is of course less than if the estimate were made by taking account of the actual values of the signals according to Eqn. (8.5). The standard

deviation of ϱ when calculated from Eqn. (8.5) using N independent sets of g, k values is (Fisher, 1946)

$$(1 - \varrho^2)/N^{\frac{1}{2}} \tag{8.13}$$

while that of ϱ calculated from Eqn. (8.12) is

$$\pi \sin(\pi n/N) \left| \frac{n}{N}\left(1 - \frac{n}{N}\right) \right|^{\frac{1}{2}} N^{\frac{1}{2}} \tag{8.14}$$

The ratio of these is plotted in Fig. 8.6 and at a correlation of $\frac{1}{2}$ it appears that method (8.12) requires a sample about three times as long as method (8.5) to achieve a similar accuracy.

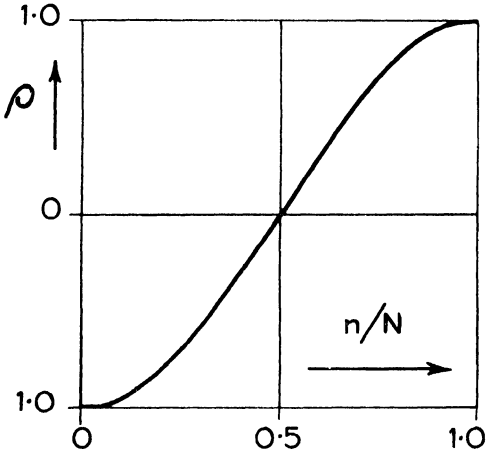

FIG. 8.5. The relation between the correlation coefficient ϱ and the ratio n/N (see text) in using the "method of signs". The variables must have a jointly normal distribution.

Complex Variables

It occasionally happens that one has to discuss variables that are represented by complex numbers. For instance, the velocity and direction of a wind are conveniently indicated by the modulus and angle of a complex number. One may wish to compare the histories of two winds, perhaps winds at different levels above the earth's

surface. An argument very like that used in comparing the histories of two voltages (scalar variables) can be developed to show that the resemblances in the histories of two winds is indicated by a coefficient ϱ

$$\varrho = \mathbf{g}^*(t)\,\mathbf{k}(t)/\sigma_g\sigma_k$$

The numbers here are complex. For instance, $\mathbf{k}(t)$ can be written

$$\mathbf{k} = k_1 + ik_2 = V\mathrm{e}^{i\theta}$$

where k_1 and k_2 are the components of velocity in directions at right angles, or alternatively where V and θ are the total velocity and the

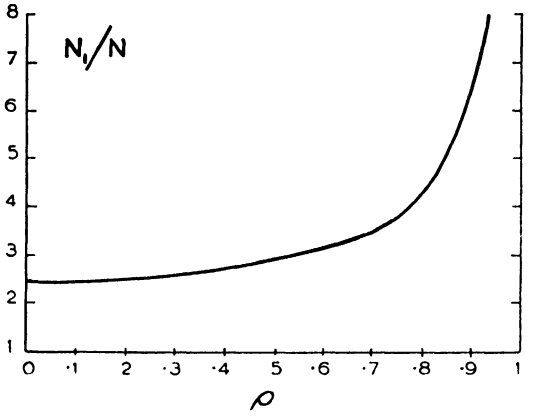

FIG. 8.6. To get equally accurate estimates of the correlation, the orthodox method of mean products may require N pairs of observations while the method of signs requires a larger number N_1. The diagram plots the ratio N_1/N for different values of correlation coefficient.

direction. An asterisk indicates the complex conjugate got by changing the sign of i. Thus

$$\mathbf{g}^* = g_1 - ig_2 = U\mathrm{e}^{-i\varphi}$$

Again, the mean square modulus σ_g^2 is got by averaging the product of $\mathbf{g}$ and its conjugate $\mathbf{g}^*$ and in terms of the components g_1 and g_2 or U and φ it is

$$\sigma_g^2 = \overline{\mathbf{g}^* \cdot \mathbf{g}} = \overline{(g_1 - ig_2)\,(g_1 + ig_2)} = \overline{g_1^2} + \overline{g_2^2} = \overline{U^2}$$

Consequently σ_g^2 and σ_k^2 do not involve the "imaginary" i but are ordinary scalar numbers equal to the mean square velocity, regardless of direction.

Consequently, for purposes of numerical calculation the coefficient can be written

$$\varrho = \left|g_1 k_1 + g_2 \bar{k}_2 + i\,(\overline{g_1 k_2} - g_2 \bar{k}_1)\right| \Big/ (\overline{U^2} \cdot \overline{V^2})^{\frac{1}{2}} \qquad (8.16)$$

Evidently ϱ itself is a complex number. It best lends itself to physical interpretation if one expresses its value in the alternative way using modulus R and argument α

$$\varrho = Re^{i\alpha} \qquad (8.17)$$

Then the angle α is the average angle which the direction of the wind **k** bears to the direction of the wind **g**. The size of the modulus R indicates the similarity in the fluctuations of the two winds and may range from 0 to 1. In comparing winds at 50 ft and at 2000 ft above the earth's surface, it is usual to find that R is large and positive, indicating a very similar history, but the angular difference α is usually appreciable, perhaps 15° over open sea, and is brought about by the frictional drag on the air at the sea surface.

A geometrical explanation of the complex correlation coefficient is given by Linn and Poschl (1956).

BARBER (1957)

It is possible to form a scatter diagram that displays the complex correlation coefficient of two complex variables merely by displaying the sum of the two on an Argand diagram, the angle of one being reversed in sign before adding. That is, one forms the sum

$$g^* + k = g_1 + k_1 + i\,(-g_2 + k_2) \qquad (8.18)$$

so that $g_1 + k_1$ and $-g_2 + k_2$ become the coordinates of the moving point on the "real" and "imaginary" axes at right angles. The distribution tends to be elliptical as in Fig. 8.7, like the jointly normal distribution discussed before. The major axis of the distribution lies at an angle $\frac{1}{2}\alpha$ to the "real" axis and if the major and minor

diameters of any elliptic contour are measured as A and B the modulus R of the correlation coefficient is

$$R = (A^2 - B^2)/(A^2 + B^2). \tag{8.19}$$

The process requires that the two variables themselves obey certain conditions. They must be adjusted to be of equal "size", that

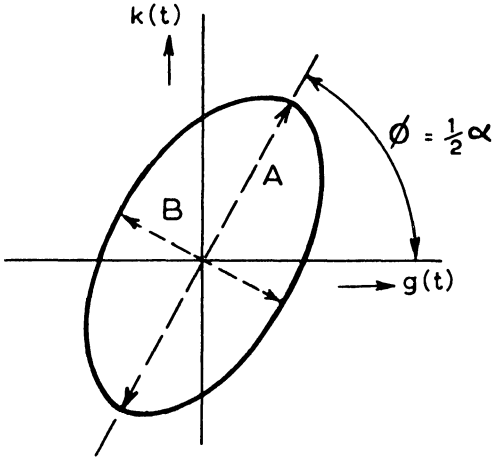

FIG. 8.7. Barber (1958). A scatter diagram that gives the modulus R and argument α of the complex coefficient of two complex variables as

$$R = (A^2 - B^2)/(A^2 + B^2) \quad \alpha = 2\Phi$$

The scatter diagram is an Argand diagram displaying the sum of one variable $k(t)$ and the conjugate of the other, $g^*(t)$. The variables must have equal variance and fulfil other conditions (see text).

is to have the same mean square modulus before the addition. Each variable alone should have a statistically normal distribution. Each variable g and k, and also their sum $(g + k)$ should separately show a distribution having circular symmetry, so that all directions are equally probable. The need for this can be overcome by a modification of the method.

SUGGESTIONS FOR FURTHER READING

BLACKMAN and TUKEY (1959) *The Measurement of Power Spectra*, Dover.

BOOTH, A. D. (1955) *Numerical Methods*, Butterworth, London.

CHURCHILL, R. V. (1941) *Fourier Series and Boundary Value Problems*, McGraw-Hill, New York.

FISHER, R. A. (1941) *Statistical Methods for Research Workers*, Oliver & Boyd, Edinburgh.

GOLDMAN, S. (1948) *Frequency Analysis, Modulation and Noise*, McGraw-Hill, New York.

JENNISON, R. C. (1960) *Fourier Transforms and Convolutions for the Experimentalist*, Pergamon Press, Oxford.

KHARKEVICH, A. A. (1960) *Spectra and Analysis*, Consultants' Bureau Enterprises, New York.

REFERENCES

ABASON, E. (1932) *Congrès International d'Electricitie, Paris*, Sect. 4, Comm. No. 402, 9 pages.

D'ALBE, E. E. F. (1926) *Optical Convention Proc. II*, p. 894–898.

ALEXANDER, L. (1953) *Acta Cryst.* **6**, 727–731.

ALLEN, R. A. (1955) *Amer. J. Phys.* **23**, 297–8.

ARMAN (see Kovasznay and Arman, 1958).

ASHWORTH, HARRISON (1906) *Engineering* **81**, 201–202, 204.

AZAROFF, L. V. (1954) *Rev. Sci. Instrum.* **25**, 471–7.

BAER, H (1937) *Z. Inst.* **57**, 225–235.

BALDOCK, G. R., WALTERS, W. G. (1946) *Electronic Engng.* **18**, 339–44; (1947) *Wireless Engr.* **24**, 132–134.

BARBER, N. F. (1947) *Wireless Engr.* **24**, 132–4.

BARBER, N. F. (1948) *Nature* **161**, 685.

BARBER, N. F. (1949a) *Electronic Engng.* **21**, 175–9.

BARBER, N. F. (1949b) *J. Sci. Instrum.* **26**, 185–7.

BARBER, N. F. (1949c) *Nature* **164**, 485.

BARBER, N. F. (1953) *Rev. Sci. Instrum.* **24**, 329.

BARBER, N. F. (1954) *Nature* **174**, 1048–50.

BARBER, N. F. (1956) *J. Atmos. Terr. Phys.* **8**, 318–330.

BARBER, N. F. (1957) *N. Z. J. Sci. Tech. B.* **38**, 366–74.

BARBER, N. F., URSELL, F. (1948) *Phil. Mag.* **39**, 345–61.

BARBER, N. F., URSELL, F., DARBYSHIRE, J., TUCKER, M. J. (1946) *Nature* **158**, 329–332.

BARLOW, G., KEENE, H. B. (1922) *Phil. Trans. Roy. Soc. Lond.* **222**, 131–166.

BARTELS, J. (1930) *Gerlands Beitr. z. Geophys.* **28**, 1–10.

BEAUCLAIRE, W. DE (1949) "Investigations on Fourier synthesis of the charge distribution in crystals. I. Methods and apparatus for multidimensional Fourier Synthesis." Akademie Verlag, Berlin, pp. 71 and VIII (in German).

BECKER, H. E. R. (1938) *Ann. d. Physik* **36**, 585–608.

BEEVERS, C. A. (1939) *Proc. Phys. Soc.* **51**, 660–667.

BEEVERS, C. A. (1952) *Acta Cryst.* **5**, 670–673.

BEEVERS, C. A., LIPSON, H. (1936) *Proc. Phys. Soc.* **48**, 772–780 (also McEwan and Beevers, 1942).

BÉKÉSY, G. V. (1937) *E. N. T.* **14**, 157–161.

Bell, H. Jnr., Rideout, V. C. (1954) *I.R.E. Trans; Prof. Group on Elect. Conf.* **14**, 715–23.
Benedict, T. R., Min, H. S., Rideout, V. C., Swift, W. B. (1958) *Proc. Nat. Elect. Conf.* **14**, 715–23.
Bennet, J. M., Kendrew, J. C. (1952) *Acta Cryst.* **5**, 109–16.
Berger, L. (1955) *J. Phys. Radium* **16**, 433–438 (in French).
Bergmann, L. (1934) *Hochfrequ. u. Elektroakustik* **43**, 83–85.
Bernasconi, G. (1950) *Geofis. Pura Appl.* **16**, 181–7.
Bickel, H. J. (1959) *I.R.E. Wescon Convention Record* **3**, Pt 8, 59–67.
Binder (see Liable and Binder, 1930).
Blackman, Tukey (1959) *The Measurement of Power Spectra*, Dover.
Blondel, A. (1915) *Rev. Elec.* **23**, 171–177.
Blondel, A. (1925) *Rev. Gen. d'Elec.* **17**, 363–368.
Blondel, A. (1926) *Rev. Gen. d'Elec.* **20**, 833–837.
Born, M., Furth, R., Pringle, R. W. (1945) *Nature* **156**, 756–7.
Bowen, J. H., Burnup, T. E. (1951) *Electronic Engng.* **23**, 67–69.
Bragg, L. (1944) *Nature* **154**, 69–72.
Bragg, L., Stokes, A. R. (1945) *Nature* **156**, 332–3.
Bragg, W. H. (1929) *Z. Krist.* A **70**, 475–492.
Braun, P. B. (1958) *Philips Res. Rep.* **12**, 491–548.
Brockway, L. O. (1936) *Rev. Mod. Phys.* **8**, 231.
Brook, F. E. Jnr., Smith, H. W. (1952) *Rev. Sci. Instrum.* **23**, 121–126.
Brown, D., Coleman, C. F., Lyttleton, J. W. (1949) *Proc. Phys. Soc. Lond.* B **62**, 149–162.
Brown, D., Lyttleton, J. W. (1947) *Nature* **160**, 709.
Brown, S. L. (1939) *J. Franklin Inst.* **228**, 675–694.
Brown, S. L., Wheeler, L. L. (1943) *J. App. Phys.* **14**, 30–36.
Bru, L., Cubero, M., Montis, V. H. (1953) *An. Real Soc. Espan. Fis. Quim.* **49** A 77–86.
Bru, L., Garcia, J. M., Rodrigues, M. (1955) *An. Real. Soc. Espan. Fis. Quim.* **51**, 163–72.
Bru, L., Rodrigues, M., Cubero, M. (1952) *J. Chem. Phys.* **20**, 1069–70.
Butler (see Ryme and Butler, 1944).

Caimann, V., Hoppe, W. (1953) *Z. angew. Phys.* **5**, 121–30.
Campbell, A. C., Foster, R. M. (1948) *Fourier Integrals for Practical Applications*, van Nostrand, New York.
Carriere, Z. (1939) *J. Phys. Radium* **10**, 14–22.
Chaffee, E. L. (1936) *Rev. Sci. Instrum.* **7**, 384–389.
Charp, S. (1949) *Trans. Amer. Inst. Elec. Engrs.* **68**, Pt I, 644–51.
Choudbury, A. K., Chakrabarti, N. B. (1958) *Indian J. Phys.* **32**, 205–211.
Chubb, L. W. (1914) *Elect. J.* **11**, 91–96.
Cochran, W. (1948) *Acta Cryst.* **1**, 54–56.
Coudres, T. des (1898) *Berlin Phys. Gesell. Verh.* **17**, 129–132.
Cooper, F. S. (1950) *J. Acoust. Soc. Amer.* **22**, 761–2.
Cox, E. G., Gross, L., Jeffrey, G. A. (1947) *Nature* **159**, 433–4.
Cramer, H. (1951) *Mathematical Methods of Statistics*, Princeton University Press.
Crease, J., Tucker, M. J. (1954) *Brit. J. App. Phys.* **5**, 143–5.

CUNNINGHAM, W.J. (1947) *J. Appl. Phys.* **18**, 656–62.
CUTRONA, L.J., LEITH, E.N., PALERMO, C.J. and PORCELLO, L.J. (1960) *I.R.E. Trans. Inf. Theor.* IT-6, No. 3, 386–400.

DANIELS, F.B. (1952) *Rev. Sci. Instrum.* **23**, 369–70.
DARBYSHIRE, J. (1955) *Proc. Roy. Soc.* A **230**, 560–569.
DARBYSHIRE, J., TUCKER, M.J. (1953) *J. Sci. Instrum.* **30**, 212.
DEBYE, P., SEARS, F.W. (1932) *Proc. Nat. Acad. Sci. U.S.A.* **18**, 409–14.
DELLENBAUGH, F.S. Jnr. (1921) *Amer. I.E.E. Journal* **40**, 135–144.
DIETSCHE, Fricke (1932) *Elekt. Nach. Techn.* **9**, 341–345.
DINA, A. (1916) *Elettrotecnica*, **3**, 3–7.

EDGARDH, B.H. (1951) *I.V.A. (Stockholm)* **22**, 134–53.
EILER, G. von (1951) *C.R. Acad. Sci., Paris* **232**, 1122.
EUSEY (see Shimizu 1951).
ESPLEY, D.C. (1939) *Phil. Mag.* **28**, 338–52.

FELLGET, P. (1958) *J. Sci. Instrum.* **35**, 257–8.
FISCHER, F.A. (1950) *Fernmeldetechn. Z.* **3**, 174–80.
FISCHER, F.A. (1951) *Frequenz* **5**, 6–13.
FISCHER-HINNEN, J. (1909) *Elek. u. Mach.* **27**, 335–342.
FISHER, R.A. (1946) *Statistical Methods for Research Workers*, Oliver & Boyd, Edinburgh.
FOSTER, G.A.R. (1946) *J. Stat. Soc. (Supp.)* **8**, 42.
FRANK, V. (1957) *J. Sci. Instrum.* **34**, 210–11.
FREYSTEDT, E. (1935) *Z. techn. Physik*, **16**, 533–39.
FRICKE (see Dietsche, 1932).
FURTH, R., MACDONALD, D.K.C. (1947) *Proc. Phys. Soc.* **59**, 388.
FURTH, R., PRINGLE, R.W. (1944) *Phil. Mag.* **35**, 643–56.
FURTH, R., PRINGLE, R.W. (1946) *Phil. Mag.* **37**, 1–13.

GABOR, D. (1946) *J. Instn. Elec. Engrs.* III, **93**, 429–57.
GATES, B.G. (1932) *J. Sci. Instrum.* **9**, 380–386.
GERMANSKY, B. (1930) *Ann. d. Physik* **7**, 453–69.
GLOCK, W.S. (1942) *Amer. J. Sci.* **240**, 437–42.
GOFF, K.W. (1955) *J. Acoust. Soc. Amer.* **27**, 223–36.
GOLDBERGER, M.L., SEITZ, F. (1947) *Phys. Rev.* **71**, 294–310.
GOODMAN, T.P. (1956) *J. App. Phys.* **27**, 773–5.
GORGE, P. de la (1914) *Soc. Int. Elec. Bull.* **4**, 545–56.
GREENHALGH, D.M.S., JEFFREY, G.A. (1950) *Acta Cryst.* **3**, 311–12.
GREENWOOD, W. (1932) *Wireless Engr.* **9**, 310–313.
GRIERSON, J.K. (1957) *Proc. Inst. Radio Engrs.* **45**, 806–11.
GRUENZ, O. (1951) *Bell Lab. Record* **29**, 256–61.

HAGG, G., LAURENT, T. (1946) *J. Sci. Instrum.* **23**, 155–8.
HALL, H.H. (1935) *J. Acoust. Soc. Amer.* **8**, 257–62.
HANSON, A.W. (1952) *Nature* **170**, 580.
HANSON, A.W., LIPSON, H. (1952) *Acta Cryst.* **5**, 362–6.

HANSON, A. W., LIPSON, H., TAYLOR, C. A. (1953) *Proc. Roy. Soc.* A **218**, 371–84.
HANSON, A. W., MENNARY, A. (1951) *J. Sci. Instrum.* **33**, 24–7.
HANSON, A. W., TAYLOR, C. A., LIPSON, H. (1951) *Nature* **168**, 160.
HARRISON, J. (1906), *Engineering* **81**, 201–202, 204.
HARTENHEIM, M. (1917) *Elektrotech. Z.* **38**, 49–52, 65–67.
HARVEY, J. (1930) *Proc. Phys. Soc.* **42**, 245–49.
HARVEY, J. (1934) *Engineering* **138**, 667–69
HERMANN, H. (1909) *Elec. Rev. and West. Electn.* **55**, 97–102.
HERMANN, L. (1890) *Arch. ges. Physiol.* **47**, p. 44.
HASTINGS, A. E., MEADE, J. E. (1952) *Rev. Sci. Instrum.* **23**, 347–49.
HENRICI, G. (1894) *Phil. Mag.* **38**, 110–25.
HEURTA, F., CASALS, J. M. (1954) *An. Real. Soc. Espan. Fis. Quim.* **48 A**, 238–43.
HICKMAN, C. N. (1934) *J. Acoust. Soc. Amer.* **6**, 108 11.
HINDE (see Taylor, 1951).
HODGSON, M. L., CLEWS, C. J. B., COCHRAN, W. (1949) *Acta Cryst.* **2**, 113–6.
HOK, G. (1948) *J. App. Phys.* **19**, 242–50.
HOWELL, B. J. (1959) *J. Opt. Soc. Amer.* **49**, 1012 21.
HOWELL, CHRISTIENSEN, MCLACHLAN (1951) *Nature* **168**, 282.
HUBER (see Howell, 1960).
HUGGINS, M. L. (1941) *J. Amer. Chem. Soc.* **63**, 66; (1944) *J. Chem. Phys.* **12**, 520; (1945) *Phys. Rev.* **67**, 197.

IDDINGS, F. T. (1919) *Elec. World* **73**, 525–26.
IMAHORI, K. (1941) *J. Fac. Sci. Hokkaido Imp. Univ.* **3** (Ser. 2), p. 103 (in English).

JEFFREY (see Cox, 1947, 1949).
JOHNSON, V. O. (1941) *Trans. Amer. I. E. E.* **60**, 1032–1036.

KARTHA, G. (1953) *J. Indian Inst. Sci.* A. **35**, 332–8.
KAULE, W., JOHNE, A. (1956) *Nachrichtentechnik* **6**, 35–9.
KEATING, D. T., ANTAL, F. J. (1955) *J. App. Phys.* **26**, 1041–4.
KEENE (see Barlow, 1922).
KELVIN, W. T. (1878) *Proc. Roy. Soc.* **27**, 371.
KEMP, P. (1920) *J. Inst. Elec. Engrs.* **57**, 85–91.
KENDALL, M. G. (1948) *The Advanced Theory of Statistics*, Charles Griffin, London.
KENT, E. L. (1943) *Electronics* **16**, 120–121.
KOBAYASHI, M. (1930) *Elect. Commun.* **8**, 315–319.
KOENIG, W., DUNN, H. K., LACY, L. Y. (1946) *J. Acoust. Soc. Amer.*, 19–49.
KOVASZNAY, L. S. G., ARMAN, A. (1958) *Rev. Sci. Instrum.* **28**, 793–7.
KRAFT, L. G. (1950) *J. Acoust. Soc. Amer.* **22**, 762–4.
KRANZ, F. W. (1927) *J. Franklin Inst.* **204**, 245–262.

LABOURET, J. (1921) *Rev. Gen. d'Elec.* **9**, 360–63.
LABROUSTE, H. (1936) *Terr. Mag.* **41**, 15–28.
LAMPARD, D. G. (1954) *Proc. Instn. Elec. Engrs.* **102 C**, 35–41.
LAURENT (see Hägg, 1947).
LEE, Y. W., CHEATHAM, T. T., WIESNER, J. B. (1950) *Proc. Inst. Radio Engrs.* **38**, 1165–71.

LEWIS, F.M. (1935) *J. App. Mech.* **2**, A 137–140.
LIABLE, T., BINDER, E. (1930) *Assoc. Suisse Elec. Bull.* **21**, 365–71.
LINVILLE, W.K., SCOTT, R.E., GUILLEMAN, E.A. (1955) *I.R.E. Trans. Circuit Theory*, Vol. CT 2, No. 3, 243–50.
LINN, H.J., POSCHL, K. (1956) *Arch. Elect. Übertragung* **10**, 105–6.
LIPSON, H., BEEVERS, C.A. (1936) *Phys. Soc. Proc.* **48**, 772–780.
LIPSON (see Hanson, 1951, 1952, 1953).
LIPSON (see Menarry, 1957).
LIPSON (see Beevers, 1936).
LOHMANN, A. (1959) *Optica Acta* **6**, 37–41.
LOMBARDI, L. (1920) *Ellettrotecnica* **7**, 578–80.
LOVERA, G. (1948) *Nuovo Cimento* **5**, 150–3.
LUCAS, R., BIQUARD, P. (1932) *C.R. Acad. Sci., Paris* **194**, 2132–4.

McDONAL, F.J. (1953) *Rev. Sci. Instrum.* **24**, 272–6.
MacEWAN, D., BEEVERS, C.A. (1942) *J. Sci. Instrum.* **19**, 150–56.
MacEWAN, D., BEEVERS, C.A. (1943) *Nature, Lond.* **152**, pp. 302 and 303.
McLACHLAN (see Shimizu, 1950).
McLACHLAN, D., Jnr., CHAMPAYGNE, E.P. (1946) *J. App. Phys.* **17**, 1006–14.
McNICOL, R.W.E. (1949) *Proc. Instn. Elec. Engrs.* **96**, III, 517–524.
MADELLA, G.B. (1944) *Alta Frequ.* **13**, 31–8.
MADELLA, G.B. (1944A) *Alta Frequ.* **13**, 132–49.
MADELLA, G.B (1947) *Wireless Engr.* **24**, 310.
MADER, O. (1909) *Elektrotech. Z.* **36**, 847.
MANLEY, R.G. (1945) *Waveform Analysis*, Chapman Hall, London.
MARCHINGTON (see Kitz, 1953).
MAXWELL, R. (1940) *Rev. Sci. Instrum.* **11**, 47–54.
MAYER, S.W., TRUEBLOOD, K.N. (1953) *Acta Cryst.* **6**, 427.
MENNARY, A., LIPSON, H. (1957) *Acta Cryst.* **10**, 27–9.
MERTZ, L. (1956) *J. Opt. Soc. Amer.* **46**, 548–51.
MEYER, E. (1935) *J. Acoust. Soc. Amer.* **7**, 88–93.
MICHELSON, A.A., STRATTON, S.W. (1898) *Amer. J. Sci.* **5**, 1–13.
MILATZ, J.M.W., WAPSTRA, A.H., WIERINGEN, J.S. van (1953) *Physica* **19**, 175–80.
MILLER, D.C. (1916A) *J. Franklin Inst.* **181**, 51–81.
MILLER, D.C. (1916B) *J. Franklin Inst.* **182**, 285.
MOHANTI, H.B., BOOTH, A.D. (1955) *J. Sci. Instrum.* **32**, 442–44.
MONTGOMERY, H.C. (1939) *Bell Lab. Rec.* **18**, 28–30.

NATH, N.S.N. (1938) *Proc. Indian Acad. Sci.* **9A**, 499–503.
NEMES, T. DE (1934) *Phil. Mag.* **18**, 303–307.
NICHOLSON, M.G., PERKINS, W.M. (1932) *Proc. Inst. Radio Engrs.* **20**, 734–39.
NOWACKI, W. (1948) *Chimia*, **2**, 274–6.
NYSTROM, E.J. (1938) *Soc. Scien. Fennica Comm. Phys. Math.* **9**, 14–24.

PAGE, C.H. (1956) *I.R.E. Trans. Circ. Theory*, CT-2, No. 3, 231–7.
PARTRIDGE, G.F. (1938) *Phil. Mag.* **25**, 505–539.
PAULING (see Shaffer, 1946).

PEPINSKY, R. (1947) *J. App. Phys.* **18**, 601–4.
PIERSON, W. J. Jnr. (1952) *A Unified Mathematical Theory for the Analysis of Ocean Surface Waves*, New York Coll. of Engng.
POSCHL (see Linn, 1956).
POTTER, R. K. (1945) *Science* **102**, 463–70.
PRINGE (see Furth, 1944; 1946).
PUGET, L. (1919) *Rev. Gén. d'Elec.* **5**, 439–442.

RAILSBACK, O. L. (1937) *J. Acoust. Soc. Amer.* **9**, 37–42.
RAMAN, C. V., NATH, N. S. N. (1936) *Proc. Indian Acad. Sci.* **2 A**, 406–20.
REVESZ, G. (1954) *J. Sci. Instrum.* **31**, 406–10.
RIESZ, R. R., SCHOTT, L. (1946) *J. Acoust. Soc. Amer.* **18**, 50–61.
ROBERTSON, D. (1935) *World Power* **23**, 65–71.
ROBERTSON, J. M. (1932) *Phil. Mag.* **13**, 413.
ROBERTSON, J. M. (1948) *J. Sci. Instrum.* **25**, 28–30, 216–18.
ROBERTSON, J. M. (1950) *J. Sci. Instrum.* **27**, 276–78.
ROBERTSON, J. M. (1954) *Acta Cryst.* **7**, 817–22.
ROBERTSON, J. M. (1955) *Acta Cryst.* **8**, 286–88.
ROGERS, W. E. (1948) *Rev. Sci. Instrum.* **19**, 332–35.
ROSS, M. A. S. (1943) *Nature* **152**, 302–3.
ROTH, A. (1918) *Arch. Elektrotech.* **6**, 359–76, 388–406.
RYMER, T. B., BUTLER, C. C. (1944) *Phil. Mag.* **35**, 606–16.

SACIA, C. F. (1924) *J. Opt. Soc. Amer.* **9**, 487–94.
SCHOMAKER (see Shaffer, 1946).
SCHOTT, L. O. (1948) *Bell Lab. Tech. Rec.* **26**, 333–39.
SCHOTT (see Revesz, 1946).
SCHOUTEN, J. F. (1938) *Philips Tech. Rev.* **3**, 299.
SCHOUTEN, J. F. (1940) *Physica* **7**, 101–21.
SCHUCK, O. H. (1934) *Proc. I. R. E.* **22**, 1295–1311.
SEARS (see Debye, 1932).
SEIWELL, H. R. (1950) *Rev. Sci. Instrum.* **21**, 481–84.
SHAFFER, P. A., SCHOMAKER, V., PAULING, L. (1946) *J. Chem. Phys.* **14**, 648–64.
SHILTON, A. (1944) *Proc. Phys. Soc. Lond.* **56**, 130–32.
SHIMIZU, H., ELSEY, P. J., MCLACHLAN, D. (1950) *Rev. Sci. Instrum.* **21**, 779–83.
SHIRE, E. S., RUNCORN, S. K. (1951) *Servomechanisms*, D. S. I. R., London, p. 98–121.
SMITH, H. W., MCCLURE, R. M., BOSTICK, F. X. (1957) *I. R. E. Trans. Instrumentation*, Vol. 1–6. 228–31.
STEINBERG, J. C., FRENCH, N. R. (1946) *J. Acoust. Soc. Amer.* **18**, 4–18.
STEVENS, K. N. (1950) *J. Acoust. Soc. Amer.* **22**, 769–71.
STONE, R. S., DANDL, R. A. (1957) *I. R. E. Trans. Electronic Comput.* Vol. EC-6, No. 3, 187–89.
STRAITON, A. W., TERHUNE, G. K. (1943) *J. App. Phys.* **14**, 535–6.
SUGAR, G. R. (1954) *J. App. Phys.* **25**, 354–57.
SVOBODA, V. (1950) *Elektrotechn. Obz.* **38**, 12–17 (in Czech).

TAMM, K., PRITCHING, I. (1951) *Akust. Beih.* No. 1, AB43-AB48.
TAYLOR (see Hanson 1951; 1953).

TAYLOR, HINDE, LIPSON (1951) *Acta Cryst.* **4**, 261–66.
TAYLOR, C.A., THOMPSON, B.J. (1957) *J. Sci. Instrum.* **34**, 439–47.
TEREBESI, P. (1934) *Arch. Elektrotech.* **28**, 195–250.
TERHUNE (see Straiton, 1944).
TERMAN, F.E. (1948) *Radio Engineers Handbook*, McGraw-Hill, New York.
THOMASSON, D.W. (1948) *J. Brit. Inst. Radio Engrs.* **8**, 171–86.
TIMBRELL (1958) *J. Sci. Instrum.* **35**, 313–18.
TOLSTOV, V.G. (1946) Bull. Acad. Sci. URSS Dep. Sci. Tech. (No. 3) pp. 339–400 (in Russian).
TUCKER, D.G. (1958) *J. Brit. Inst. Radio Engrs.* **18**, 233–35.
TUCKER, M.J. (1950) *Proc. Roy. Soc.* A **202**, 565–73.
TUCKER, M.J. (1952) *J. Sci. Instrum.* **29**, 326–29.
TUCKER, M.J. (see Barber, 1946; Darbyshire, 1953).

URSELL (see Barber, 1946, 1948).

VAND, V. (1949) *Nature, Lond.* **163**, 169–70.
VAND, V. (1950) *J. Sci. Instrum.* **27**, 257–61.
VAND, V. (1952) *J. Sci. Instrum.* **29**, 118–21.
VASILESCO, V. (1934) *Rév. Gen. d'Elec.* **35**, 773–78.
VERCELLI, F. (1934) *Ricerca Scientifica* **1**, 7.
VERCELLI, F. (1937) *Ricerca Scientifica* **8**, 609–21.

WALTER, W.G. (1943) *Electronic Engng.* **16**, 9–13, 236–238.
WALTER, W.G. (see Baldock, 1946).
WALTHER, A., DRYER, H.J., ESTENFELD, H. (1939) *Z. Instrumentenke.* **59**, 162.
WAPSTRA (see Milatz, 1953).
WATSON, W.H. (1953) *Trans. Roy. Soc. Canada* III, **47**, 37–46.
WEDMORE (see Robertson, 1935).
WEINER (see Milatz, 1953).
WERENSKIOLD, W. (1942) *Meteorol. Ann.* **1**, 137–147.
WHITELY, T.B., ALDRIDGE, L.R. (1952) *J. Sci. Instrum.* **29**, 358–62.
WHITTAKER, E.J.W. (1948) *Acta Cryst.* **1**, 165–67.
WILLIAMS, H.P. (1944) *Wireless Engr.* **21**, 108–11.
WOODWARD, P.M. (1951) *Phil. Mag.* **42**, 883–91.

YERG, D. (1955) *J. Geophys. Res.* **60**, 173–85.
YULE (1895) *Phil. Mag.* April.

ZECH, T. (1942) *Arch. Elektrotech.* **36**, 322–28.
ZINN, W.H. (1947) *Phys. Rev.* **71**, 752–57.

INDEX