# TRANSFER FUNCTIONS

By "CATHODE RAY"

WHEN I first mentioned to the Editor that I was thinking of saying something about transfer functions, his reaction was as if I had been offering to write a do-it-yourself manual on how to get into orbit. This worried me rather, because up to that moment nothing had looked simpler than explaining what a transfer function was. One could start by defining it as the ratio of output to input of any system that has an output and input. One would then add a bit of padding to make this clear to the dimmest intellects, and there one would be, ready to knock off for the week. Even the impressively named and highly expensive Transfer Function Analysers are basically simple. But now, was there a snag?

Well, there are a lot of very difficult books mainly on transfer functions, at about £5 each, and perhaps the Editor thought I was going to paraphrase them into half an hour's pleasant reading. To put it in terse if not very sensitive or sincere contemporary speech, he'll be lucky.

The fact that transfer functions usually do reside in such highly mathematical and complicated contexts may however, as he suggested, well make readers fight shy of them. My object, then, is to show that they are something you probably know quite well already but didn't recognize under such a pompous name.

I had better start by clearing away a possible cause of confusion. Transfer functions are not the same as transfer characteristics, though both are relationships between input and output. A transfer characteristic is a graph of instantaneous output voltage, current or whatever, plotted against the input ditto. Fig. 1 is an example. A quick way of obtaining this kind of graph is to connect a suitable

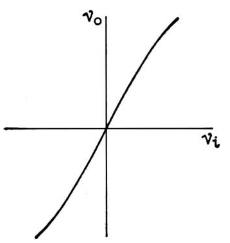



*Fig. 1. This is an example of a transfer characteristic, which is a graph of instantaneous output (voltage, etc.) against instantaneous input (voltage, etc.). It is useful for showing up non-linearity. A transfer function is the "complex" ratio of r.m.s. or peak output to input at any frequency.*

oscillator to the input, in parallel with the X plates of a cathode-ray tube, and connect the Y plates of the same tube to the output. A perfect amplifier (or whatever it is being examined) gives a straight diagonal line. Curvature shows non-linearity, and looping shows phase shift.

Note I said *instantaneous* voltage, etc. There is another type of output/input graph in which r.m.s., peak or average values are plotted. This sort does show up distortion, but is less useful for analysing it.

Now that we know what transfer functions aren't, we can go on to see what they are. I have already given one definition. In the world of amplifiers, the transfer function is often called simply the gain. To be more precise it would have to be called complex gain, and I would hasten to add for the benefit of beginners that "complex" here doesn't mean what it does in ordinary language. I dealt with that at length in the February 1953 issue (and in "Second Thoughts on Radio Theory"). It just means taking account of phase as well as magnitude. But of course there is an output/input ratio in many systems which don't yield a gain in the literal sense. Such things as filters, attenuators and transformers. And the same idea can be—and nowadays commonly is, hence the proliferation of difficult books—applied to mechanical systems, especially those with feedback, such as servomechanisms. It is even being applied to chemical engineering.

## Forms of Expression

There is no difficulty at all, then, in understanding what a transfer function is. But of course if you have a complicated system, its transfer function is likely to be complicated too. And all sorts of very sophisticated mathematical techniques have been devised—some of them comparatively recently—for dealing with such. Another thing that leads to difficulty is non-linearity of the system. And still another is non-sinusoidal waveform of the signal applied to the system. Just now we are going to stick to simple sine-wave signals and linear systems. Even a so-called ultra-linear amplifier is not perfectly linear, but it should be at least a good approximation to the ideal linear system.

First let us review the various forms in which transfer functions can be expressed. You may know them already, but one can hardly go over it too often.

Suppose you put a signal of, say, 0.1V r.m.s. into an amplifier and get 23V out. Then the ratio of $V_o$ to $V_i$ is 23/0.1 = 230. That is what would usually be called the voltage gain. It is also the *magnitude* of the amplifier's transfer function. The other part of the same function is the *phase difference* between $V_i$ and $V_0$. (Mathematicians tend to call these two parts the *modulus* and *argument*.) If $V_o$ lags 32° behind $V_i$, then the phase difference —often denoted by $\phi$—is $-32°$. In more strictly mathematical terms it is $-0.56$ radians. (A whole cycle is 360° or $2\pi$ radians.) One way of writing this particular transfer function is therefore $230\underline{/-0.56}$.

A direct graphical representation would be as two sine waves, one (representing the output) 230 times the amplitude of the other (representing the input) and 32° behind it in phase. But sine waves are difficult to draw, and don't show the important quantities clearly, so a preferred form is a straight line 230 units long, drawn at an angle of 32° clock-

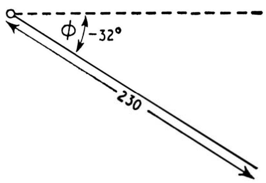



*Fig. 2. Here is the vector (more correctly phasor) form of a transfer function for one particular frequency.*

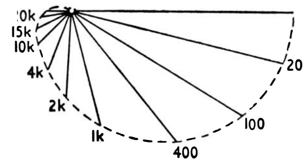



*Fig. 3. If the phasors of a system are plotted for a representative selection of frequencies, their tips trace out a Nyquist type of diagram, as in this example.*

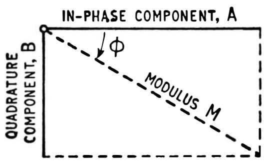



*Fig. 4. Relationship between Cartesian or rectangular transfer-function co-ordinates, A and B, and the corresponding polar co-ordinates, M and φ.*

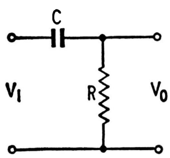



*Fig. 5. Simple example of a system for which a transfer function exists.*

wise with reference to the conventional zero ("3 o'clock")—Fig. 2. This sort of line is commonly called a vector, but since that name is not entirely appropriate the tendency nowadays is to call it a phasor. Sometimes an arrow head is stuck on the end farthest from the point around which the line is slanted to indicate $\phi$, but because the natural interpretation of an arrow head is motion in the direction in which it points, which would be quite wrong here, I prefer a little circle at the other end to mark the centre.

Mathematically, a transfer function is what is called an operator. The thing it operates on is the input signal. The effect of the operation, in the example in Fig. 2, is to multiply the magnitude (voltage, in this case) by 230 and delay its phase by 32°. If this amplifier were followed by another, without any interaction except that the output of the first provided the input of the second, it is pretty obvious (I hope) that the transfer function of the combined amplifiers would be obtained from their separate functions by *multiplying* their magnitudes and *adding* their phase angles. So if the second function was, say, $170/\underline{59°}$, the combined function would be $230 \times 170/\underline{-32° + 59°} = 39{,}100\,/\underline{+27°}$ Alternatively the magnitude can be expressed in decibels, which, being the logarithms of the gains, are combined by adding, like the phase angles. This is one of the reasons for preferring dB.

The most important thing about a transfer function is that it varies with frequency. Often one wishes it didn't, because for many purposes the ideal amplifier is one that treats all frequencies alike. But while it is possible to approach this ideal very closely over quite a useful band of frequencies, there are always limits due to the inevitable stray capacitances and inductances, to say nothing of the more involved effects in transistors. There are instruments—the so-called transfer function analysers—for measuring the complex gain at different frequencies. The question then arises, how to show the results, since both magnitude and phase vary with frequency. The length and angle of a phasor, as in Fig. 2, show these two quantities at a single frequency, so one method is to draw other phasors for other frequencies. Fig. 3 is an example. The figures denote the frequencies, and the dotted line tracing out the end of the phasor as it varies with frequency is particularly valuable in negative feedback problems, where as a Nyquist diagram it shows at a glance whether there is any possibility of the system oscillating.

But when one is interested in how successful or otherwise the system is in working uniformly over a certain range of frequency, it is usually more helpful to plot magnitude and phase separately against frequency. Often the phase graph is omitted. We then have that even better known type of diagram—the amplitude/frequency characteristic.

## Conversion

The Nyquist diagram is a particular kind of polar diagram (Fig. 3), the two parts of the transfer function being specified in polar co-ordinates; angle and radial length. Most graphs are plotted in Cartesian or rectangular co-ordinates, and the two parts of transfer functions can be alternatively specified in them. They are known as the in-phase and quadrature components. The corresponding written form of the transfer function is A + jB, in which j is an instruction to reckon the quantity to which it is attached as a vertical distance (positive upwards) instead of horizontally to the right. The two types of co-ordinates are related as shown in Fig. 4, thus:

$$M = A + jB = \sqrt{A^2 + B^2}$$

$$\tan\phi = \frac{B}{A}$$

$$\text{also } A = M\cos\phi$$

$$B = M\sin\phi$$

A reason why it is necessary to be able to convert from one form to the other is that some methods of measuring transfer functions give the results in one form and some in the other, and the particular method available—or most convenient—may not be the one that fits best into one's calculations.

Obviously (looking at Fig. 4) the A and B values

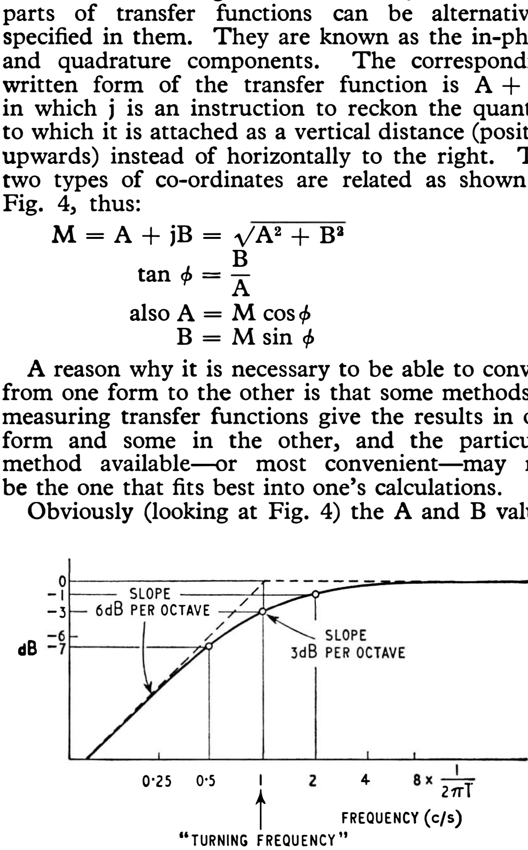



*Fig. 6. Method of quickly drawing a frequency characteristic curve for the Fig. 5 system. This graph has been "normalized" by writing the scale in terms of 1/2πT, so that the "turning frequency" is at 1.*

can be used for plotting phasors and therefore a Nyquist diagram. Alternatively the polar quantities can be plotted separately against frequency as Cartesian co-ordinates, giving gain/frequency and phase/frequency characteristic curves.

In case the frequent reference to amplifiers has given anyone the impression that transfer functions refer mainly to them, let us take as an example of a " system " the simple one shown in Fig. 5. This, of course, often occurs as part of an amplifier. One of its roles is as a device for coupling one valve stage to the next without imposing the d.c. component of the anode voltage on the grid. In this role the variation of transfer function with frequency, though necessary for separating signal frequency from zero frequency, is undesirable within the working frequency range, and the aim is to avoid it as much as practicable. But the same device with different values of C and R is used deliberately in tone-control arrangements as a bass-cut device.

If the input signal is reckoned as the voltage $V_i$ across the input terminals, and the output is the voltage $V_o$ across the output terminals, which are not called upon to supply any appreciable current to a load (i.e., they are practically an open-circuit), then the system can be regarded as a potential divider, in which the transfer function is the ratio of R to the whole impedance of C and R in series:

$$\frac{V_o}{V_i} = \frac{R}{R + 1/j\omega C}$$

Because the transfer function is a function of $j\omega$ (= $j2\pi f$) it is often written as $F(j\omega)$. So

$$F(j\omega) = \frac{R}{R + 1/j\omega C} = \frac{j\omega CR}{1 + j\omega CR}$$

The thing to notice here is that the values of C and R don't matter individually; it is their product CR that counts. This CR is well known as the " time constant " of the system, reckoned in microseconds if R is in ohms and C in microfarads. (CR in the formula must then be multiplied by $10^{-6}$ to bring it to seconds.) The value of the transfer function at any given frequency depends on it alone. So the tendency nowadays is to work in time constants, and accordingly we will substitute T for CR.

In looking at the above transfer function again to see how it varies with frequency, we note first that when the frequency is zero ($\omega = 0$), $F(j\omega)$ is zero. That, of course, is as it should be for blocking d.c. At the other end of the scale, when $\omega$ approaches infinity, $F(j\omega)$ approaches 1; so at very high frequencies the system passes practically the full signal voltage. But the most significant frequency is the one that can be regarded as a sort of change-over point between these two extremes. It is conveniently defined by

$$\omega = \frac{1}{T} \text{ or } f = \frac{1}{2\pi T}$$

because that makes $\omega T = 1$, and $1 + j1 = \sqrt{2}$, and $1/\sqrt{2}$ is 0.707, which in terms of voltages or currents is almost exactly 3 dB down on 1. In terms of power, it is exactly a half. The effective frequency bandwidth of an amplifier is usually reckoned between the upper and lower half-power points.

At frequencies above zero but so much below $1/2\pi T$ that there is little difference between $1 + j\omega T$ and 1, the magnitude of $F(j\omega)$ is very nearly proportional to the frequency. If the frequency is halved the voltage amplitude is halved. The way this is usually put is that the amplitude/frequency characteristic slopes 6 dB per octave. To be a uniform slope, both amplitude and frequency scales must be logarithmic, which is usually achieved by using special graph paper obtainable for plotting dB against logarithmic frequency.

## Simple Rules

Putting all the foregoing information together provides us with simple rules for making a quick sketch of the amplitude/frequency characteristic of any circuit comprising one resistance and one reactance. Fig. 6 shows them applied to the Fig. 5 circuit. Locate the point on the frequency scale where $f = 1/2\pi T$ ($= 1/2\pi CR$); the " turning frequency " or " corner frequency ". If C and R are in microfarads and megohms, T is in seconds and $f$ in c/s, but usually it is more convenient to divide $10^6$ by $2\pi$ times the number of microseconds. Draw a horizontal line at the 0 dB level from this point rightwards, and a line sloping to the left downwards at 6 dB per octave, as shown dotted. Mark the − 3 dB point at the turning frequency and draw a smooth curve through it to approach the two straight lines. Incidentally, the slope at this − 3 dB point should be 3 dB per octave. As a further guide to drawing the curve, it is helpful to note (as can easily be shown by calculation) that at half and double the turning frequency the curve is very nearly 1 dB below the straight lines, as shown .

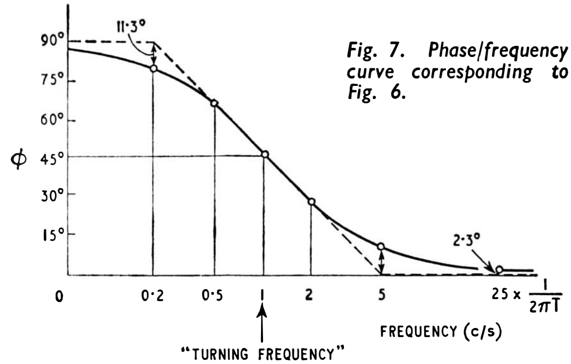



*Fig. 7. Phase/frequency curve corresponding to Fig. 6.*

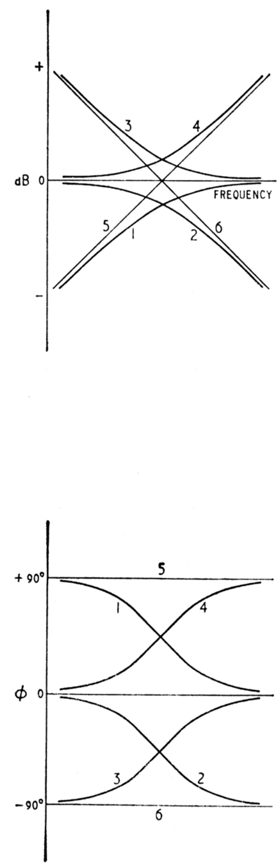
This type of diagram owes its convenient straight-line-approximation feature, with the facility for filling in the curve accurately enough for practical purposes without any actual plotting, to the fact that both scales are logarithmic. As we shall see, it has other advantages.

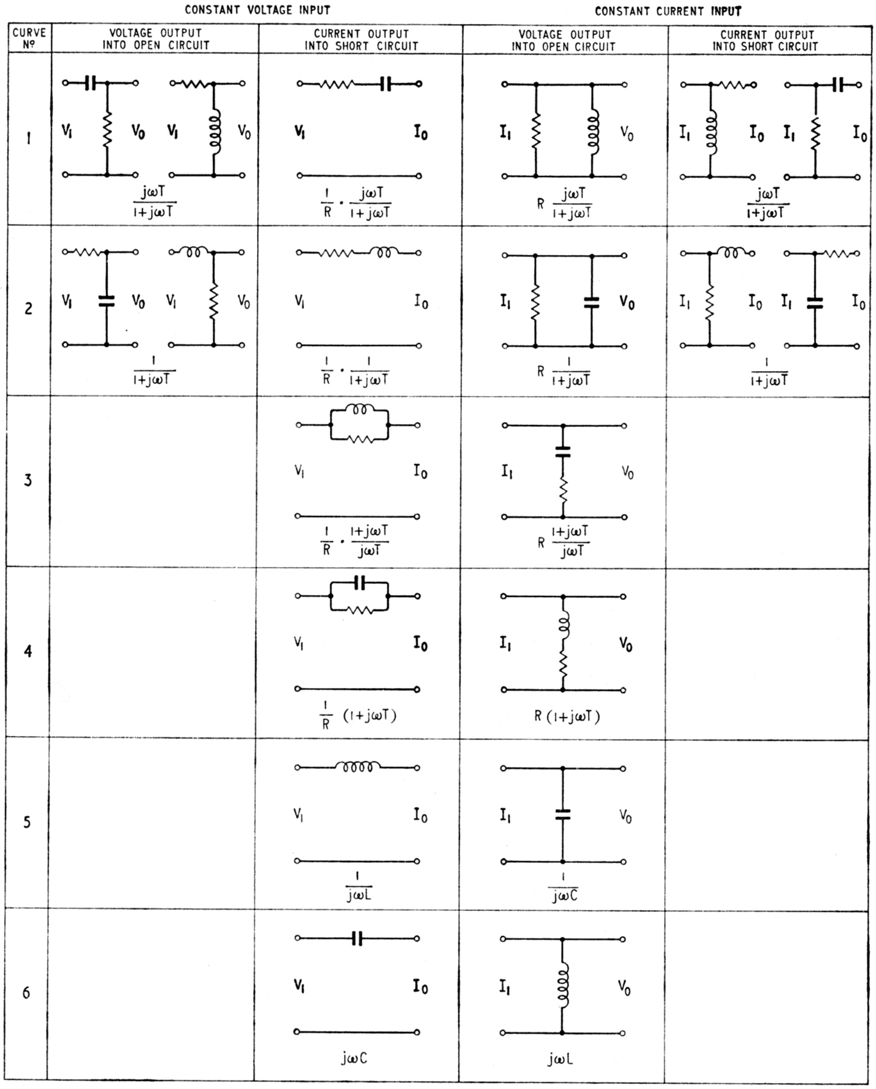
What now about phase angle? As we have seen, the formula is

$$\tan\phi = \frac{B}{A}$$

where B is the quadrature component and A the in-phase component. The catch is that B can be either negative (as in Fig. 4) or positive, and this decides the sign of $\phi$. It can be found mathematically by multiplying the transfer function (in this case $\frac{j\omega T}{1+j\omega T)}$ above and below the line by $1-j\omega T$ to eliminate j from the denominator, and then noting the sign of the j term in the numerator. In this case it is positive. But the easier way is to

CHARACTERISTIC CURVES




| CURVE Nº | CONSTANT VOLTAGE INPUT: VOLTAGE OUTPUT INTO OPEN CIRCUIT | CONSTANT VOLTAGE INPUT: CURRENT OUTPUT INTO SHORT CIRCUIT | CONSTANT CURRENT INPUT: VOLTAGE OUTPUT INTO OPEN CIRCUIT | CONSTANT CURRENT INPUT: CURRENT OUTPUT INTO SHORT CIRCUIT |
|---|---|---|---|---|
| 1 | $\frac{j\omega T}{1+j\omega T}$ | $\frac{1}{R}\cdot\frac{j\omega T}{1+j\omega T}$ | $R\frac{j\omega T}{1+j\omega T}$ | $\frac{j\omega T}{1+j\omega T}$ |
| 2 | $\frac{1}{1+j\omega T}$ | $\frac{1}{R}\cdot\frac{1}{1+j\omega T}$ | $R\frac{1}{1+j\omega T}$ | $\frac{1}{1+j\omega T}$ |
| 3 | | $\frac{1}{R}\cdot\frac{1+j\omega T}{j\omega T}$ | $R\frac{1+j\omega T}{j\omega T}$ | |
| 4 | | $\frac{1}{R}(1+j\omega T)$ | $R(1+j\omega T)$ | |
| 5 | | $\frac{1}{j\omega L}$ | $\frac{1}{j\omega C}$ | |
| 6 | | $j\omega C$ | $j\omega L$ | |

*Fig. 8. Table of all the combinations of one resistance and one reactance, and of one reactance only, and their frequency characteristics (magnitude and phase) and transfer functions.*

remember that the current through a capacitive circuit (such as Fig. 5) *leads* the applied voltage, and as the output voltage—being taken across a resistance—is in phase with the current, the output voltage must lead the input voltage; so its phase angle is, according to convention, positive.

It is convenient to plot the phase angle against the same logarithmic frequency scale as for the "gain," not only for making the same scale do for both, but also because the curve turns out symmetrical that way; one half of it is exactly the same as the other turned upside down and left to right—Fig. 7. Note that the angle scale is linear; this is not only to preserve the said symmetry but also because (as we have seen) the overall phase angle of a combination of transfer functions is the simple algebraical sum of the separate angles.

As Fig. 7 shows, a rough approximation to the curve is obtainable by drawing a straight line from one-fifth to five times the turning frequency (where the angle is always 45°). The true curve almost exactly coincides with this line from half to double the turning frequency. It can be sketched in by plotting points at five times (11.3°) and one-fifth (90°−11.3°) as shown. Farther away still, the departure of the curve from 0° and 90° tails off in the same ratio as the frequency varies.

If C in Fig. 5 is replaced by R, and R by L, and F (jω) is calculated, precisely the same result will be found, if it is remembered that the time constant of an LR circuit is L/R. So Figs. 6 and 7 hold good for this system too.

If in either of these two systems the resistor and reactor are interchanged, the appropriate graph is the mirror image of Fig. 6, falling from 0dB at zero frequency to −3dB at $1/2\pi T$ and thereafter tending towards a downward slope of −6dB per octave. The actual formula is

$$F(j\omega) = \frac{1}{1 + j\omega T}$$

In all these four systems we assumed that the input voltage was constant and the load impedance was infinite. If we pass a constant *current* through R and C or L in parallel, the voltage across these varies in a similar manner; Fig. 6 for R and L, and its image for R and C. By passing constant current through R and C or L in series, we get the same output voltage curves upside down. Because we are converting a current signal into a voltage signal the " dimensions " of the transfer functions we have been using for voltage-to-voltage (a pure number) would be wrong; on working it out we find they must be multiplied by the constant factor R. So 0 dB corresponds to the output voltage obtained by passing the constant input signal current through R.

If you remember my enthusiasm for the principle of duality (" two formulæ for the price of one ") you will expect me to include the duals of the foregoing arrangements, obtained by exchanging current and voltage, inductance and capacitance, resistance and conductance, and series and parallel. The duals of constant current input and open-circuit voltage output are therefore constant voltage input and short-circuit current output, and the transfer functions of the systems (as modified by the dual exchanges) are the same as before, except that of course the factor R for resistance becomes 1/R for conductance. The original series fits the conditions of feeding into a negatively biased valve—practically an open circuit—and the new one approximates the situation when driving a transistor, if the resistances and reactances of the " system " are relatively high.

By now we have accumulated quite a variety of systems covered by only four stock shapes of frequency characteristics—or, rather, one stock shape, which can be made up as a template, turned this way or that. A single curve likewise serves for all the phase graphs. Fig. 8 displays all the possible combinations of one resistance and one reactance, with the addition (for the sake of completeness) of the still simpler systems made up of one reactance only. These give continuous 6 dB-per-octave slopes and constant 90° phase shifts.

The systems shown, although so simple, are of real practical value, as any tone-control designer knows. They are also real practical nuisances, as any feedback-amplifier designer knows.

Next time we shall see how more elaborate transfer functions can be analysed into combinations of these simple ones.

# MANUFACTURERS' PRODUCTS

## NEW ELECTRONIC EQUIPMENT AND ACCESSORIES

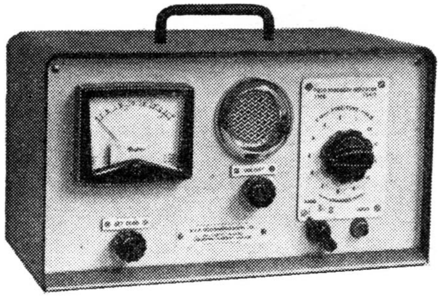
### *Suction Handling Aid*

AN AID to the handling of small pieces of, say, semiconductor material has been produced by the Solo-Seeda Limited. In both of the models air is drawn in through a nozzle which is designed to receive and hold the object to be lifted; thus special types are available for individual purposes. The makers claim that lifting, moving and dropping again at the rate of 100 objects per minute is relatively simple. Releasing the held object with the model fitted with a pump is accomplished by uncovering a small suction-release hole in the nozzle holder. Retail prices are: hand model with squeeze bottle 5s; model with water-operated pump 15s; both including a standard nozzle.

Solo-Seeda Limited, Spencer Road, Berkswell, Coventry, Warwickshire.

### *Field Strength Indicator*

CARRIER level measurement in Bands I, II and III can be undertaken with the "Telecomm" Indicator Type FS4/T. This instrument, which uses transistors throughout, employs a 13-position turret tuner covering all f.m. and television channels, and indicates field strength directly on a moving-coil meter calibrated from 10μV to 30mV. The Indicator can be supplied to cover other channels in the range 30 to 220Mc/s. The makers of the FS4/T are R.E.E. Telecommunications Ltd., 15a Market Square, Crewkerne, Somerset.



*R.E.E. "Telecomm" field strength indicator.*

### *V.H.F. Frequency Meter*

ONE result of the Wayne Kerr-Gertsch agreement is the appearance on the British market of the v.h.f. frequency meter FM-7, which will measure and generate frequencies to an accuracy of 0.0002%. Both amplitude and frequency modulation are available—30% a.m. and

*Wayne Kerr-Gertsch FM-7 frequency meter, with a frequency converter in the lower half.*

# POLES AND ZEROS

By "CATHODE RAY"

## MORE ABOUT TRANSFER FUNCTIONS

NOTWITHSTANDING their higher-mathe-
matical sounding name,
saw last month) are nothing more than output-input ratios in which phase as well as magnitude is taken into account. These both vary with frequency,
function of frequency,
concerned) is usually denoted by $F(j\omega)$,
stands for "function of,"
the frequency,
"complex"; i.e.,
magnitude. So at any particular frequency its specification consists of *two* numbers.

In one type of specification these numbers are simply the magnitude and phase angle. They are then often written in the form $A \angle \phi$. A and $\phi$ are polar co-ordinates. So they can be represented graphically by a straight line,
at an angle $\phi$ to the zero position,
tionally "3 o'clock." Alternatively, a transfer function can be expressed in the familiar cartesian or squared-paper co-ordinates,
phase" or "real" or horizontal one,
the "quadrature" or (as the j indicates) the "imaginary" or vertical one.

Among the many systems that have transfer functions are practically everything with an input and output—amplifiers,
ers,
Last time we considered every possible combination of one resistance and one reactance—called first-order systems—and also of one reactance only,
voltage or current input and output. We found that one simple shape of amplitude/frequency characteristic and one shape of phase/frequency characteristic,
all 16 first-order systems. If last month's issue has mysteriously vanished, Fig. 1 may serve as a reminder. Each of these four positions corresponds to a transfer function containing only the terms 1 and $j\omega T$, where T is the time-constant of the system—CR or L/R.

Often several of these first-order systems occur combined,
the standard playback characteristic for disk records, now laid down in BS 1928:1960 to end a long era of confusion. As we have seen, there
in which a transfer function can be expressed,
one form of this particular sample was given by T. M. A. Lewis on p. 121 of the March 1961 issue, preliminary to a description of a transistor amplifier providing the characteristic so specified:

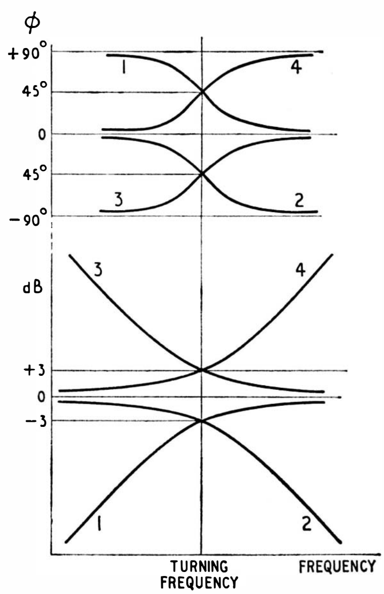
$$F(j\omega) = \frac{A(1 + j\omega T_2)}{(1 + j\omega T_1)(1 + j\omega T_3)}$$

where A is the amplification at zero frequency—as can be seen by putting $\omega = 0$ throughout.

Here we obviously have three of the simple first-order transfer functions multiplied together,
means that signals are subjected to the effects of all three. One of them,
type 4 (Fig. 1), and the other two,
type 2. The standard time-constants for long-playing records are: $T_1 = 3{,}180\ \mu$sec, $T_2 = 318\mu$ sec., $T_3 = 75\ \mu$sec. The first two numbers look a little odd; they were chosen so that when multiplied by $2\pi$ the results are round numbers: 20,000 and 2,000. The turning frequencies—those that make $\omega T = 1$—are $10^6$ divided by these numbers,
and 500 c/s. The 75 $\mu$sec., presumably chosen to be the same as for f.m. de-emphasis, gives 2,120 c/s as the third turning frequency.

To draw the graphs of gain and phase against frequency (Fig. 2) just use the simple rules given last month for each of these three turning frequencies, and combine them by adding. For both these purposes it is necessary to use a logarithmic scale of frequency—normal practice anyway—and also for gain—which again is normally achieved by reckoning it in dB. Phase angles are additive,
linear scale is correct. Fig. 2 is an example of what is sometimes called the Bode diagram.

From 50 c/s to 500 c/s the linear approximation is a 6 dB-per-octave downward slope. Incidentally, if you have made the dB scale so that 6 dB on it occupies the same distance as a 2 : 1 frequency ratio you can draw these slopes with a 45° set-square. From 500 c/s onwards the downward slope still applies but is exactly cancelled out by the upward (type 4) slope that begins there. So the resultant is a horizontal section. This ends at




| CURVE TYPE | TRANSFER FUNCTION |
|---|---|
| 1 | $\frac{j\omega T}{1+j\omega T}$ |
| 2 | $\frac{1}{1+j\omega T}$ |
| 3 | $\frac{1+j\omega T}{j\omega T}$ |
| 4 | $1+j\omega T$ |

*Fig. 1. All systems comprising one resistance and one reactance, with an input and output, voltage and/or current, have transfer functions in this list of four, with characteristic curves as shown.*

2,120 c/s, the second type-2 turning frequency. It is so near 500 c/s that when the 3 dB and 1 dB points are plotted those at and near 1,000 c/s are on opposite sides of the line. The curve must therefore be drawn to pass midway between them, as shown.

All this need take only a couple of minutes, and the result is practically as good as (and more likely to be correct than) a curve drawn from points obtained by the very tedious straightforward computation of the whole transfer function. (Try it and see!)

The obvious but inadvisable method of obtaining the composite phase-angle curve is to draw the individual type 2 and type 4 curves and add their ordinates. It is easier (and uses half the space) to draw the negative of the type 4, which is the same as a type 2, and then to add the *difference* between the curves for $T_2$ and $T_3$ (picked out by dividers) to the one for $T_1$. This difference is greatest around 1,000 c/s and causes the hump in the resultant curve.

The performance represented by Fig. 2, need one say, is obtained by combining one type-4 resistance/reactance pair with two type-2. And that is where the difficulties begin. The pairs can't just be connected in cascade, as in Fig. 3 for example, because each is supposed to work from and into infinite or zero impedance. In practice one can allow impedances that are respectively much larger or much smaller than the impedance of the transfer circuit itself. But working into or from another transfer circuit generally won't do. One solution is to use suitably arranged valves or transistors between each. Some are needed anyway to amplify the gramophone signals. But not usually as many as four.

So this is the cue to introduce the slightly more complicated system shown in Fig. 4. It is easy to work out its transfer function:

$$F(j\omega) = \frac{V_0}{V_1} = \frac{\frac{1}{j\omega C} + R_2}{\frac{1}{j\omega C} + R_1 + R_2} = \frac{1 + j\omega C R_2}{1 + j\omega C(R_1 + R_2)} = \frac{1 + j\omega T_2}{1 + j\omega T_1}$$

where $T_1 = C(R_1 + R_2)$ and $T_2 = CR_2$. The third time constant can be brought in with an independent simple type-2 circuit. This, in fact, is basically what T. M. A. Lewis did in his amplifier already mentioned, though he used a counterpart of Fig. 4 in a negative feedback connection. His type 2 was in the output circuit of the amplifier. Measurements showed the whole thing worked extremely well. Please don't ask me to tabulate every possible combination of one reactance and two resistances; try Heinz. The 16 varieties with one reactance and one resistance were enough of an effort. Suffice to say that Fig. 4 is one of many possible examples of what is known as a step circuit, for a reason that is clear if one looks at the straight-line approximation to its amplitude/frequency characteristic—Fig. 2 below 2,120 c/s. The limited phase shift is one reason for its use in negative feedback amplifiers, as I explained in my treatise on the use of the Nyquist diagram in the January, 1956, issue. In this connection it is worth noting that if the two curves of the Fig. 2 type are arranged on the paper so that the $-180°$ level for one coincides with the 0 dB level for the other, this common level corresponds with the critical point in the Nyquist criterion for instability. The vertical distance below it of the gain curve at the point where the phase curve cuts it (if it does) is the gain margin of stability. And the vertical distance above it of the phase curve at the point where the gain curve cuts it (if it does) is the phase margin of stability. So there is no need to

(*Continued on page 227*)

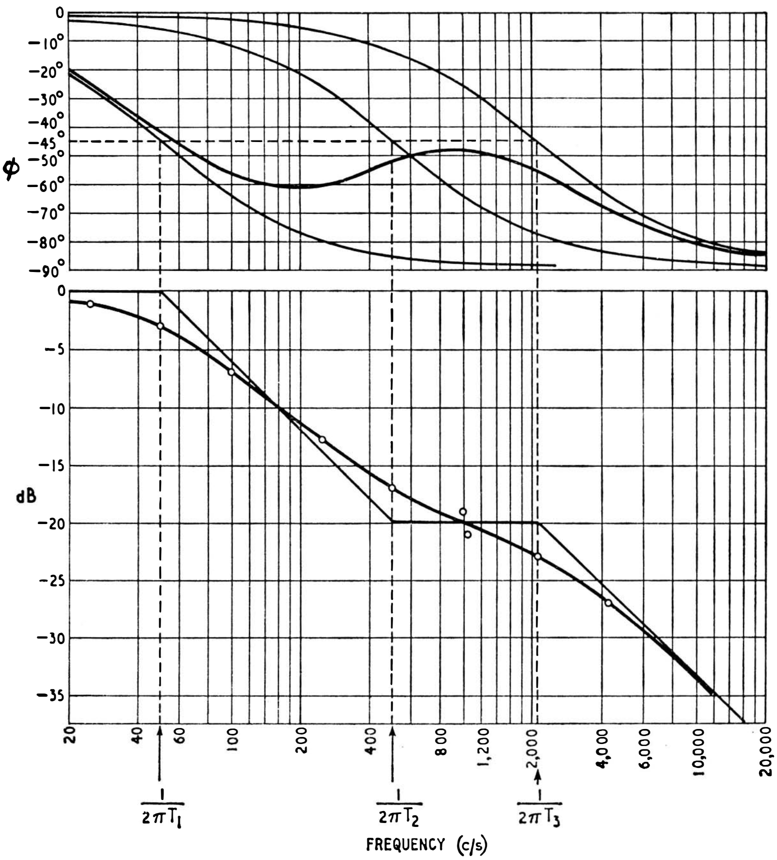



*Fig. 2. Characteristic curves (Bode diagram) for standard fine-groove disk play-back equalizer, drawn from inspection of the transfer function. The straight lines are the "skeleton," giving a first approximation to the curves.*

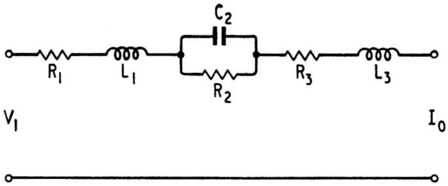



*Fig. 3. This way of lumping together three elementary systems would not give a result equal to the product of their separate functions, because the systems would be incorrectly terminated.*

draw a Nyquist diagram to find these quantities.

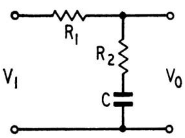
For the fun of it I tried extending the same principle as in Fig. 4 to include all three time constants in one potential divider, shown in Fig. 5. The working-out, in case anyone is interested, appears as an appendix. A practical snag about this circuit is the inductor. Being of the order of henries, it is relatively expensive and liable to pick up hum. A minor comfort is that one doesn't have to keep its resistance low, for that can be anything up to and including $R_1$
to calculate, all the components being in series.

As many as three alternative three-in-one networks using two resistors and two capacitors were shown by W. H. Livy in the January, 1957, issue, p. 29. These are easier to provide in practice, but the formulæ are a little less simple.

The transfer function for any of these, or for Fig. 5, has two $1 + j\omega T$ factors in the denominator, so squared terms are involved and the description "second-order" applies to it and the corresponding network.

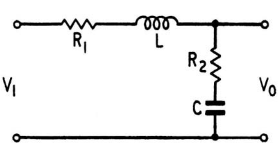
And so one could go on, if necessary into higher orders, with at each step a vast extension in the variety of system and the area of paper and length of time occupied in computation by the straightforward complex algebra we have been using. So several more advanced mathematical techniques have been developed to shorten the work. We begin to see, then, that although transfer functions are basically simple there can be quite a lot to learn about them; hence the expensive books thereon.

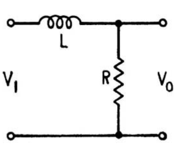
The method of poles and zeros has become standard study in America at a fairly early stage in communications, but I have never seen it mentioned in *Wireless World*. So I shall have to be very introductory.




*Left: Fig. 4. Example of a step circuit, equivalent to two elementary systems.*

*Right: Fig. 5. This second-order system is equivalent to three elementary systems, and can be designed to yield the Fig. 2 curves at one go.*



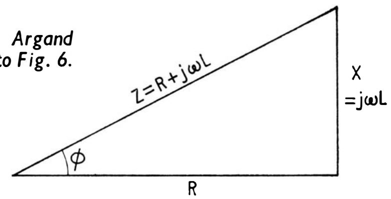

*Left: Fig. 6. Example of an elementary first-order system for pole-zero practice.*

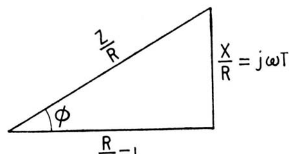
We have become used to seeing transfer functions with numerator or denominator, or both, made up of factors of the form $(1 + j\omega T)$. The variable quantity is $\omega$ $(= 2\pi f)$ and T is a measure of the component values. If $\omega$ is made equal to $j/T$ (or $f = j/2\pi T$), than $j\omega T = j^2 = -1$, and the factor is zero. So if the factor is in the numerator the whole transfer function is equal to zero. If on the other hand it is in the denominator the transfer function goes to infinity.

"So what?" you may say. Imaginary frequency (as denoted by the j) is nonsense, so these situations can never arise and the whole exercise is pointless.

Not quite. The foregoing is a device for constructing a type of diagram that is helpful in the study of networks. It is closely related to what is known as the Argand diagram, which we used last month

*Right: Fig. 7. Argand diagram relating to Fig. 6.*




*Left: Fig. 8. Same as Fig. 7 but reduced in scale by dividing all the sides by R.*

for our approach to transfer functions. The relationship will probably be easier to follow if we take a simple example, Fig. 6. Fig. 7 is its Argand diagram for impedance. The value of R is measured along the positive "real" axis, and the reactance of L is measured along the positive "imaginary" axis. (Capacitive reactance is measured along the negative imaginary axis.) The impedance, Z, of the two in series is represented in magnitude by the length of the sloping line, and in phase by its angle of slope, $\phi$. It is, of course, the vectorial sum of R and $j\omega L$. The transfer function is

$$F(j\omega) = \frac{V_0}{V_1} = \frac{R}{Z}$$

Substituting $R + j\omega L$ for Z gives us $F(j\omega) = \frac{1}{1 + j\omega T}$, the standard form, as in Fig. 1. R and Z must be reckoned vectorially, to give us the complete transfer function—magnitude and angle. The magnitude alone, denoted by $|F(j\omega)|$, is the ratio of the magnitudes of R and Z, represented by the lengths of two sides of the triangle.

X is the variable quantity, directly proportional to frequency, $f$. One can visualize the length of the vertical line increasing from nothing (at zero frequency) and the resulting changes in the reciprocal length and angle of the Z line, which we represented by the type 2 curves in Fig. 1.

The transfer function, being the ratio of two sides of this triangle, is quite unaffected by altering the scale of the diagram. So we are in order in dividing all the sides by R, giving Fig. 8. This brings the length of the R line to 1, so that F(j ω) is simply the reciprocal of the length of the sloping line.

In these Argand diagrams, "real" values of ω are measured upwards, because they are prefixed by a j. Now note what would happen if (without bothering to consider whether it makes physical sense) we put $\omega = j/T$, as suggested earlier. This brings in a second j, rotating the direction another quarter-turn anticlockwise. An alternative way of looking at it is to multiply the two js, making $-1$; either way it means that ω must be measured along the negative real axis; i.e., to the left. And as its magnitude is 1, its length coincides with the horizontal side of the triangle, and the sloping line collapses to nothing, making $Z/R = 0$ and $R/Z$ infinity, as predicted.

This seems a fantastic way of approaching the Argand diagram, which in our simplicity we have been using happily from the start without any of this crazy mathematical philosophy. But if we alter the scale of our triangle once more, as in Fig. 9, some method begins to be discernible in the madness. Here we have divided all round by 2πT as well as by R, with the result that the vertical axis is now definitely a scale of frequency, not depending (as in Figs. 7 and 8) on any particular circuit values. This removes the advantage of Fig. 8, that the reciprocal of the length of the sloping line (to scale) alone gives the magnitude of the transfer function. In Fig. 9 we must revert to taking the ratio of the two lines—$1/2\pi T$ to $Z/R2\pi T$. We can, if we like, regard $1/2\pi T$ as a scale factor that has to be brought in. To make this scale factor stand out clearly we can write the transfer function in this form:

$$\frac{1}{2\pi\,T} \cdot \frac{1}{\dfrac{1}{2\pi\,T} + jf}$$

If we were concerned only with systems as simple as our example, Fig. 6, we would obviously have made an indifferent bargain in exchanging Fig. 8 for Fig. 9. The reward comes with more complicated systems. Take our gramophone equalizer, for example. It has three time constants, so there will be three corresponding points where $f = j/2\pi\,T$. One of them, $T_2$, gives a zero value of the transfer function, and this point is called a zero and marked O. The others, $T_1$ and $T_3$, give infinite values, as in our simple example, and are called poles and marked X (Fig. 10).

We now see the advantage of using a frequency scale, for it is common to all three—and to as many others as a still more complicated system might require.

Fig. 10 has been drawn for one particular value of frequency, $f_1$. As the frequency varies, all three slopes vary accordingly, those starting from poles giving reciprocal measures of the corresponding component transfer functions of the system, and the

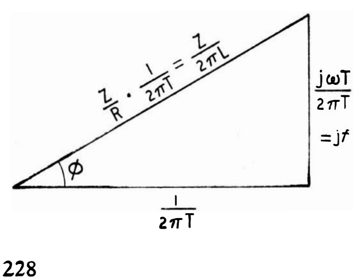



*Fig. 9. Same as Fig. 8, multiplied by 1/2πT.*

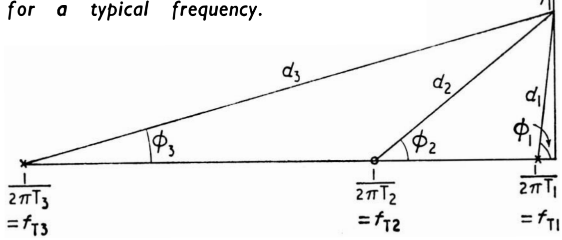



*Fig. 10. Pole-zero diagram for system corresponding to Fig. 2, with d lines drawn for a typical frequency.*

one starting from a zero giving a direct measure of its transfer function. $1/2\pi\ T_1$ is of course 50, $1/2\pi\ T_2$ is 500, and $1/2\pi\ T_3$ is 2,120, those being the turning frequencies of the system.

If we remember that the function for the system as a whole is obtained by *multiplying* the magnitudes of the separate ones, and *adding* their phase angles, we can make general rules for using a pole-zero diagram for this purpose:

(1) The *magnitude* of the overall transfer function is obtained by multiplying together the distances of all zeros from the point representing the selected value of frequency, and the reciprocals of the distances of all poles, each distance having been divided by the appropriate $1/2\pi$ T.

For instance, in Fig. 10, at the frequency for which the lines are drawn,

$$|F(j\,\omega)| = \frac{d_2}{d_1\,d_3} \cdot \frac{2\pi\ T_2}{2\pi\ T_1\ .2\pi\ T_3} = \frac{d_2}{d_1 d_3} \cdot \frac{f_{T1}\ \ f_{T3}}{f_{T2}}$$

The $f_{T1}\ f_{T3}/f_{T2}$ is of course a constant, the same for all signal frequencies, so only needs to be computed once for any circuit.

(2) The *phase angle* of the overall transfer function is obtained by adding together all the separate phase angles, those at zeros being reckoned positive and those at poles negative.

So in Fig. 10

$$\phi = \phi_2 - \phi_1 - \phi_3$$

Fig. 10 is a pole-zero diagram to which *d* lines have been added to apply it to a particular frequency, represented by the typical point marked $f_1$ on the j$f$ scale, from which they radiate. The pole-zero diagram itself includes only the axes and the O and X points. From what has gone before you might easily assume that these points must all lie on the "negative real" axis. Certainly they do for all systems of the types we have been considering. But complex frequency has physical meaning and the poles and zeros can appear almost anywhere. Fig. 11 is an example.

But let's not be led away by complexities of this sort before we have seen how our more familiar Fig. 10 works. If we start from zero frequency, we see that all the angles begin to increase from zero, but the rate they do so is inversely proportional to the distance of the poles or zeros from the origin. So $\phi_1$ increases comparatively fast, $\phi_2$ only one tenth the rate (on a diagram drawn to scale), and $\phi_3$ at less than one fortieth. Similarly $d_1$ begins to lengthen much faster than $d_2$ and $d_3$. In other words, at frequencies up to and somewhat beyond $f_{T1}$ the whole system behaves almost as if it were a simple type-2 circuit.

As can easily be proved, each angle varies fastest at its turning frequency, and likewise the ratios of the sides. So at 500 c/s the type-4 part of the system is the major influence, overhauling the negative phase angle $\phi_1$, which is now growing comparatively slowly, and reducing the rate at which the " gain " is falling off. At still higher frequencies, around 2,000 c/s, the second type 2 comes into its own, speeding up the fall-off and phase lag again.

But we see that the total phase shift can't exceed $-90°$ however high the frequency goes; in fact, it can't quite reach it. By that time all three $d$ lines are practically equal in length, so $d_2$ cancels out $d_1$ and $d_3$ is increasing linearly, indicating an inverse linear fall-off in gain (at 6 dB per octave, as we know).

With a little practice one can quickly visualize the performance of a system from its pole-zero diagram. It is easily seen that a pole and a zero tend to cancel one another out as they come closer together on the diagram. And several poles (or zeros) close together indicate a very rapid rate of fall-off (or growth) in gain and increase in phase shift at frequencies about equal to their distance from the j$f$ axis.

A further stage of progress in the art is to draw a pole-zero diagram to represent the performance you want, and then use it to work back to the network that will give it.

Still heeding the warning against trying to run before learning to walk, however, we make quite sure we know how to deal with a simple type 1 or type 3 circuit. Fig. 12 is a type-1 example, which incidentally we studied last month. If we take its transfer function in the form

$$\frac{1}{1+\frac{1}{j\omega T}} \quad \cdots \quad \cdots \quad \cdots \quad (1)$$

we will be disconcerted to find that the condition for making the denominator zero (and thereby establishing a pole) is the same as for

$$\frac{1}{1+j\omega T} \quad \cdots \quad \cdots \quad (2)$$

which is a type-2 circuit. This is because $1/-1 = -1$. Yet type 2 has an exactly opposite performance, frequencywise. The explanation of this paradox is that (1) has a zero too, the condition being $f = 0$. When that is substituted, the denominator goes to infinity. Alternatively, you can multiply (1) above and below by $j\omega T$ to obtain the form I preferred:

$$\frac{j\omega T}{1+j\omega T} \quad \cdots \quad \cdots \quad (1a)$$

To make the next step inescapably clear, I'll write this

$$\frac{0+j\omega T}{1+j\omega T} \quad \cdots \quad \cdots \quad (1b)$$

So besides the pole at $j/2\pi T$ there must be a zero at $f = 0$. The complete pole-zero diagram is therefore Fig. 13, and is in accordance with known fact by making the gain zero at zero frequency.

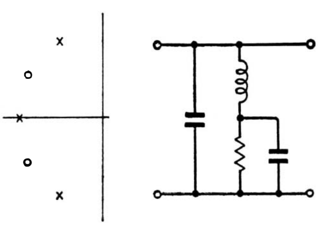
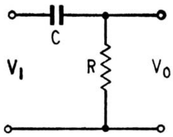
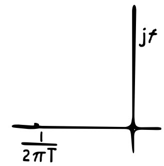



*Fig. 11. Example of a more involved pole-zero diagram (a), and the system to which it refers, (b).*




*Left: Fig. 12. Our original elementary system once more, for re-examination in the light of the pole-zero technique.*




*Right: Fig. 13. Pole-zero diagram for the Fig. 12 system. Because the transfer function is a simple ratio between the distances to zero and pole, scale factors (1/2πT) cancel out, so do not come into the calculation.*

The same line of reasoning will establish that the type 3 function is distinguished from type 4 by a pole at the origin.

If you have a sense of symmetry you will be crying out that it isn't fair. Types 1 and 3 have been handed out a pole and a zero each, while the even numbers have had only a pole and a zero between them.

Well; let's see what we can do for them. The type-4 function, $1 + j\omega T$, doesn't seem to have much room for hiding a pole besides its obvious zero ($f = j/2\pi T$). But it does go to infinity if $f = \pm j\infty$, so that is a sort of pole. No wonder we couldn't see this pole, right at the back of beyond and therefore equally far to the east as to the west. The same goes for the far-flung zero in the type 2 empire. At that distance, the corresponding $\phi$ never starts to grow, and $d$—being infinitely long at the start—has no scope for growth either, so these infinite poles and zeros are of no more concern to us than a tax collector situated in an extra-galactic nebula. Still, symmetry is satisfied, all types now have a pigeon pair each, and no further complaints will be entertained.

## APPENDIX

Derivation of time constants, $T_1$, $T_2$ and $T_3$, to fit the system shown in Fig. 5 to the transfer function

$$\frac{V_0}{V_1} = \frac{1+j\omega T_2}{(1+j\omega T_1)(1+j\omega T_3)} \quad \cdots \quad \cdots \quad (1)$$

According to the circuit this must be equal to

$$\frac{\frac{1}{j\omega C}+R_2}{\frac{1}{j\omega C}+R_2+R_1+j\omega L} \quad \cdots \quad \cdots \quad (2)$$

The resemblance can be made more obvious by multiplying (2) by $j\omega C$:

$$\frac{1+j\omega CR_2}{1+j\omega C(R_1+R_2)+j^2\omega^2 LC}$$

This compares with (1) multiplied out:

$$\frac{1+j\omega T_2}{1+j\omega(T_1+T_3)+j^2\omega^2 T_1 T_3}$$

So $T_2 = CR_2$

$T_1 + T_3 = C(R_1 + R_2)$

$T_1 T_3 = LC$

Solving the last two simultaneous equations gives:

$$T_3 = \frac{C(R_1+R_2)}{2} \pm \sqrt{[C(R_1+R_2)/2]^2 - LC}$$

$$T_1 = \frac{LC}{T_3}$$

So if one were given the time constants one would choose suitable values of C and $R_2$ to make $CR_2 = T_2$ and at the same time make their impedance about right for the circuit. Then $L = \frac{T_1 T_3}{C}$ and $R_1 = \frac{T_1+T_3}{C} - R_2$