



Square Wave and Pulse Testing of Linear Systems

AN arbitrary input wave, when transmitted through even an ideally linear system, can be altered in a number of ways. Such a wave can be altered, for example, in size, shape, and time of occurrence. There is, however, one class of waves or functions — the sinusoids — that a linear system can alter in only two ways. The response of any linear system to a sine wave is another sine wave differing from the original at most in amplitude and phase. This cardinal fact gives sine waves their unique position in communication theory and further gives physical significance to Fourier analysis and to the whole concept of frequency spectra. Since any input can be represented as the sum of a num-

ber of sinusoids, the output from a linear system will consist of these same sinusoids modified only in amplitude and phase. Each will be transmitted by the system as if it alone were present (superposition). The simple sum of the output sinusoids will be the output wave.

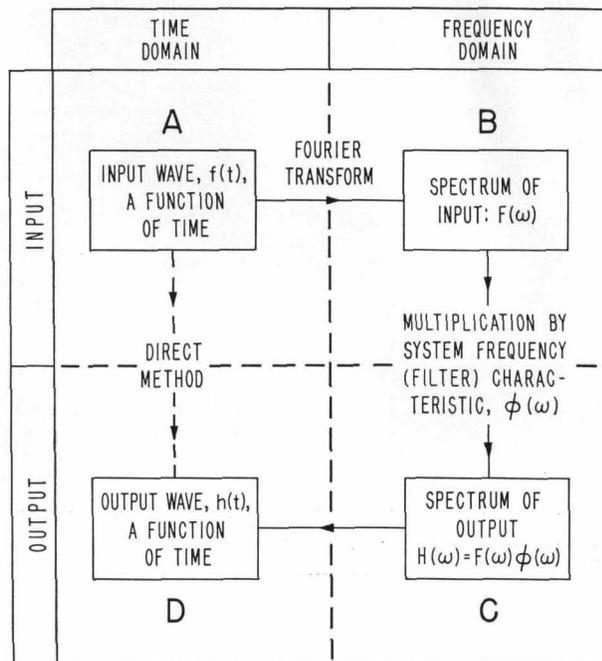
The way that the system modifies sine waves of *all* frequencies (the system amplitude and phase characteristic) thus constitutes a complete description of the system in that it enables the output to be computed for any input. The procedure for such a computation can be represented graphically as shown at left.

If not already known, the spectrum of the input wave is found by evaluating the Fourier transform (A to B in the diagram). The system multiplies the input spectrum by the transmission (amplitude and phase) characteristic $\phi(\omega)$ to give the spectrum of the output (B to C). The inverse Fourier transform of the output spectrum is the output time function (C to D).

Although the above method of computing is an indirect method, it is often actually easier than the direct method. The direct method involves evaluation of the convolution integral (often called the superposition or duHamel's integral):

$$h(t) = \int_{-\infty}^{\infty} f(\tau) \phi(t-\tau) d\tau. \quad (1)$$

$\phi(\tau)$ is the impulse response of the system. $\phi(t-\tau)$ is therefore this same function reversed left to right and displaced by an amount t . What (1) says is that, to find the output, the product of the input and this reversed, displaced impulse response must be integrated. The result will be a function of the displacement, t . In other words, the input wave must



Diagrammatic representation of steps necessary to calculate how a time function is modified by a linear system's frequency characteristic. Direct calculation involves fewer steps but is often more difficult.

be scanned with the reversed impulse response.

Equation (1) is easy to derive by considering the successive ordinates of $f(t)$ to be a succession of impulses and applying the superposition theorem. Equation (1) is also easy to evaluate on occasion, but more often it is difficult. By contrast, rather complete tables of transforms exist, so that getting from A to B and from C to D often involves merely using a table. The intermediate step from B to C is accomplished merely by multiplying the input spectrum $F(\omega)$ by the system's transmission characteristic $\phi(\omega)$. The situation is analogous to the use of logarithms when raising a number to a power. To compute 7^π , for example, involves looking up the logarithm of 7 (A to B), multiplying by π (B to C), and looking up the antilog for the answer (C to D).

Even more impressive is the case where $f(t)$ and $h(t)$ are known and the system frequency characteristic and impulse response are desired. The direct solution involves solving (1) as an integral equation. But by the indirect transform method the answer is simply

$$\phi(\omega) = \frac{H(\omega)}{F(\omega)}.$$

$\phi(t)$ is then the inverse transform of $\phi(\omega)$.

The above has shown how a knowledge of the steady state (sine wave) characteristic of a system enables the output waveform to be computed for any known input waveform. Even if the exact input waveform is not known, however, a knowledge of the steady state performance of the system enables a picture to be gained of how the system will transmit the input wave. All that may be known, for example, is that the input may contain *all* frequencies over a certain band (e.g., speech or music) and that the system must be able to transmit this whole class of inputs without distortion. This requires that the output be a replica of the input except for possible changes of size and delay;

that is: $h(t) = K f(t - t_0)$

where K is a constant and t_0 is a permissible delay. In the frequency domain this requires that the output spectrum be the input spectrum modified only by a constant amplitude factor K , and a linear phase shift, $\phi = -\omega t_0$; that is

$$H(\omega) = K F(\omega) e^{-i\omega t_0} \quad \omega_1 < \omega < \omega_2$$

The system frequency characteristic must therefore be flat, with linear phase over the band of frequencies to be transmitted:

$$\phi(\omega) = K e^{-i\omega t_0} \quad \omega_1 < \omega < \omega_2$$

RESPONSE OF LINEAR SYSTEMS TO IMPULSES

While steady state measurements are very useful for the reasons discussed above, they are also quite time consuming to make. For many purposes the transient response of the system to certain particular types of input waves may provide all the information necessary. In fact, if the input wave is properly chosen, such a transient measurement provides exactly the same information as a steady state measurement but provides it in a different form.

Consider, for example, the case where the input, $f(t)$, is an impulse of negligible duration and, say, unit area. The spectrum of such an impulse contains *all frequencies*. The frequencies all have the same amplitude and are *all in phase* in the sense that they all add at $t = 0$. In other words the spectrum of such an impulse is a constant $F(\omega) = 1$.

Now from the superposition theorem, it makes no difference whether all frequencies are introduced one after another, as in steady state testing, or simultaneously by applying an impulse. In either case any frequency will be modified in the same way by the linear system. Impulse testing thus might be said to be equivalent to an instantaneous steady state test. They both give the same information, but the results must be interpreted in either case. When the input is an impulse, the input spectrum is $F(\omega) = 1$ and the output spectrum is $H(\omega) = \phi(\omega)$. In

other words, the response of a linear network to an impulse is a pulse whose spectrum is the amplitude and phase characteristic of the network.

Obviously, the interpretation of impulse tests involves a familiarity with the spectra associated with a wide variety of time functions and vice versa. A Table of Transforms published* in these pages some time ago illustrated a number of time functions with their associated spectra.

Impulse testing has advantages, but it also has two severe disadvantages. First, the impulse must be short compared with the duration of the finest detail of the output transient which is to be reproduced accurately. In other words the spectrum of the impulse must be flat over the entire frequency range of the device under test. To get appreciable response, then, often requires that the impulse be so large in amplitude that the device under test is driven out of its range of linear operation.

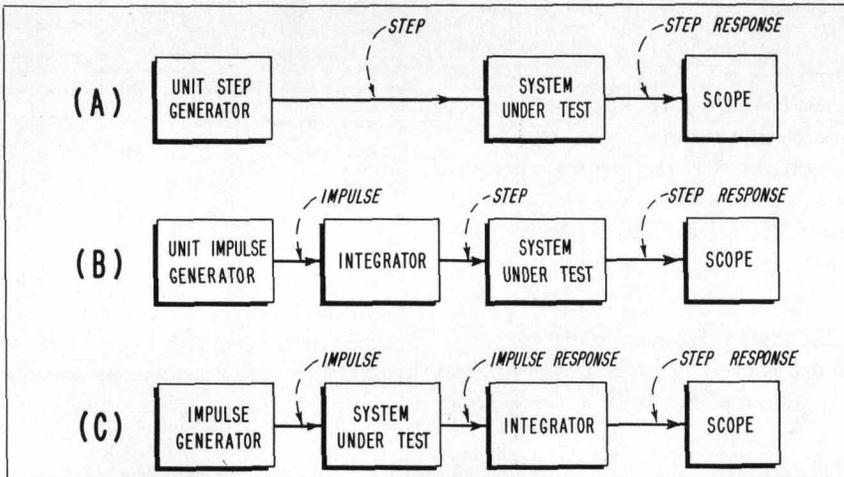
The second disadvantage is that, when testing wide band devices, the low frequency effects are hard to observe. This is because only an insignificant amount of the wide input spectrum is deleted by the low frequency cutoff of the device under test.

RESPONSE OF LINEAR SYSTEMS TO STEP FUNCTIONS

The disadvantages of impulse testing are avoided by using step functions. With a step the rise time can be as short as desired without a resulting increase in amplitude. Since the spectrum of a unit step is $F(\omega) = 1/j\omega$, more energy is concentrated at the low end of the spectrum. Low frequency effects are thus placed on a more nearly equal footing with high frequency effects.

A unit step is the integral of a unit impulse. The step function response of a system might thus be obtained in any of the three ways indicated in the diagram. Part (a) of the diagram shows a straightforward test with a step function generator. In (b) the step generator is replaced by

*B. M. Oliver. "Table of Important Transforms," Hewlett-Packard Journal, Vol. 5, No. 3-4, Nov.-Dec., 1953.



Three possible test arrangements for obtaining a system's step-function response.

a combination of an impulse generator and an integrator. In (c) the integrator has been interchanged with the system under test. Since both are linear systems the output is unaffected.* (c) illustrates the important fact that *the step function response of a linear system is the integral of the impulse response. Further, the spectrum of the step function response is $1/j\omega$ times the steady state amplitude and phase characteristic of the system.*

Conversely, the impulse response is the derivative of the step response.

SQUARE WAVE TESTING

The accompanying table (I) shows the step function responses for some typical common networks. If the condition stated below is met, these responses will also be the response to the (positive) step of a square wave, because a square wave can be considered to be a succession of alternate positive and negative steps. The table includes a few explanatory remarks with each response.

In order for the response to each step of the square wave to be identical with a system's step function response, the square wave frequency must be low enough so that the individual transients do not overlap. In other words the square wave frequency must be less than $1/2t$, where t is the time required for the step response to reach a constant value within the desired accuracy. Cases

*Assuming either a match at all junctions or frequency insensitive mismatch.

where long duration square waves should be used are where sharp irregularities exist in the frequency characteristic. Typical cases of sharp irregularities are low end cutoffs or sharp resonances anywhere in the pass band.

If high-end cutoffs are being observed, a relatively high repetition frequency with its widely-spaced spectral lines is usually permissible. The reason for this is that high-end cutoffs are relatively broad frequency effects (consume only a short time in the time domain).

Sharp mid-band effects may be explored even with high repetition frequencies if the repetition fre-

quency is variable. Sweeping the repetition frequency will always cause some harmonic of the square wave to coincide with a sharp mid-band effect and thus produce an observable transient. In fact, long duration resonances that are scarcely visible in the step function response because of their low amplitude will become quite prominent when the proper frequency square wave is used. The reason for this is that the proper frequency square wave will cause the successive excitations of the resonance to reinforce preceding excitations with the result that a much larger amplitude oscillation is produced.

The complete step function response of a system having a low frequency cutoff is only displayed if the repetition frequency is much less than the frequency of the cutoff. Since it is often inconvenient to use such a low frequency, it is customary to use a frequency such that the transients from the successive positive and negative steps do not vanish completely but rather overlap considerably. In such cases the square wave response needs to be interpreted. Table II shows some typical overlapped low frequency responses.

—B. M. Oliver

TABLE II
EFFECTS OF TYPICAL LOW END DISTORTIONS ON SQUARE WAVE

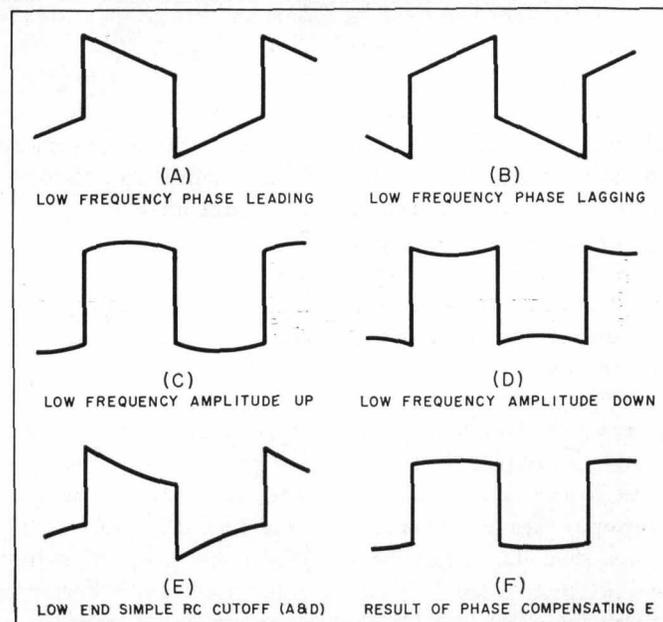


TABLE I

STEP FUNCTION RESPONSE OF TYPICAL NETWORKS

NO.	STEP RESPONSE	SYSTEM FREQ. CHARACTERISTIC	TYPICAL NETWORK	NO.	STEP RESPONSE	SYSTEM FREQ. CHARACTERISTIC	TYPICAL NETWORK
1				4			
2				5			
3				6			

CASE 4. Typical simple high-frequency cutoff transmission. Commonly encountered in simple (not "peaked") interstages, and wherever shunt capacity (as from connecting cables) loads down a resistive source.

CASE 1. This is the typical simple low-frequency cutoff such as might be produced by a series condenser-shunt resistor combination. The step response shows an abrupt rise to unity followed by an exponential decay. Usually encountered in amplifier interstages and so-called "differentiating networks." In interstages, f_0 is typically a few cycles; in differentiating networks, f_0 may be as high as several megacycles in which case the step response is very nearly an impulse.

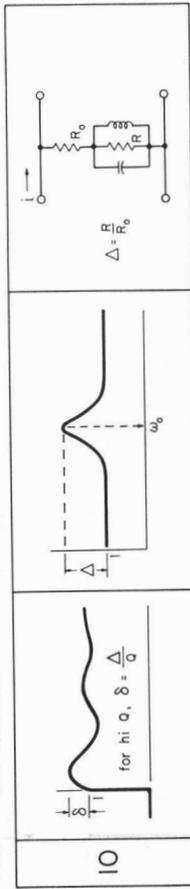
CASE 5. Two simple RC high-frequency cutoffs in tandem. Typical rise characteristic of two-stage resistance coupled amplifier without "peaking." Principal differences compared with case 4: (1) longer rise time for same ω_0 , (2) zero slope at $t = 0$.

CASE 2. Rising simple step in the frequency characteristic. Step response rises initially to amplitude determined by high frequency transmission, falls exponentially to level determined by low frequency (or d-c) transmission. This is commonly encountered in improperly compensated resistance-capacity dividers, such as scope probes, d-c amplifier interstages.

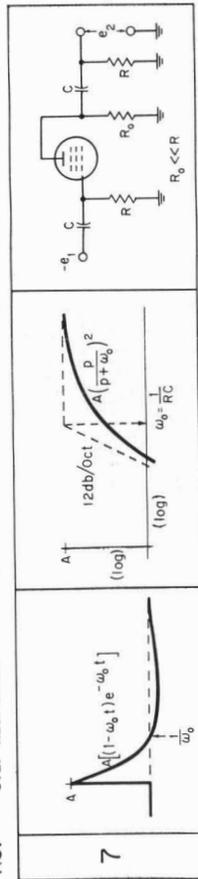
CASE 6. Phase-compensated low end cutoff. Step function response falls to zero eventually, but initial slope is zero. As a result square wave response shows little or no tilt. May be produced in a single network, or by two networks (cases 1 and 3) in tandem. Often found in video amplifiers.

CASE 3. The counterpart of case 2. Here it is the high frequency transmission which is deficient.

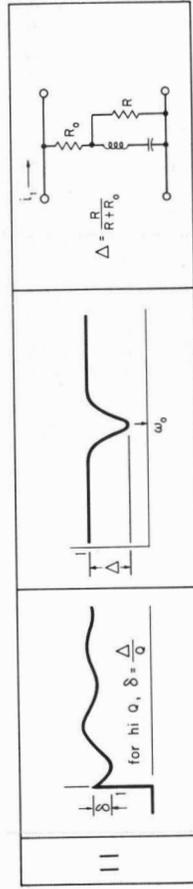
CASE 5. For each additional high frequency cutoff one more derivative of step response vanishes at $t = 0$. Thus, if high frequency transmission falls (ultimately) at $6n$ db/octave, all derivatives of step response up to the n th are zero at $t = 0$.



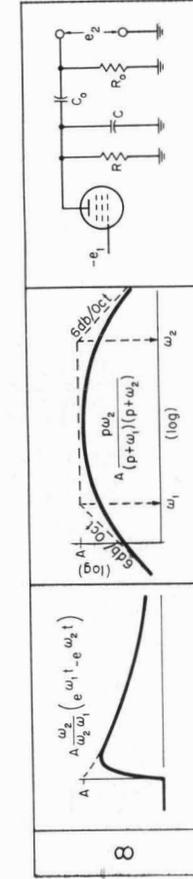
CASE 10. Small resonance in an otherwise flat characteristic. Response consists of unit step due to flat transmission plus damped oscillation due to resonance. Initial amplitude of oscillation is related to amplitude of hump in frequency characteristic as indicated in figure. For the same amplitude of hump, increasing the Q decreases amplitude of oscillation but oscillation persists longer. If hump is near top of band, time scale will be



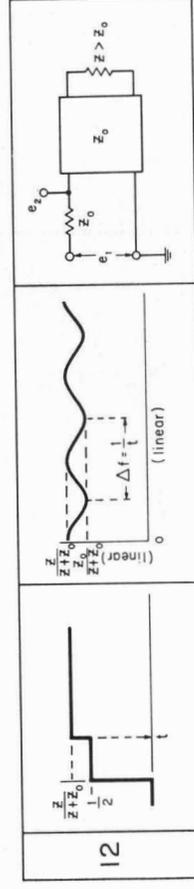
CASE 7. Two simple low frequency cutoffs (case 1) in tandem. Typical low frequency transient response of single-stage resistance-coupled amplifier with input blocking condenser or two-stage amplifier with no input blocking condenser. Principal differences compared with case 1: (1) faster initial rate of fall for same ω_0 ; (2) response goes negative crossing axis at $t = 1/\omega_0$. With each additional low-end cutoff one additional axis crossing is produced. Thus, if the low end response falls off (ultimately) at δ ndb/octave, there will be $n-1$ axis crossings. They do not occur at regular intervals—each successive half cycle takes longer.



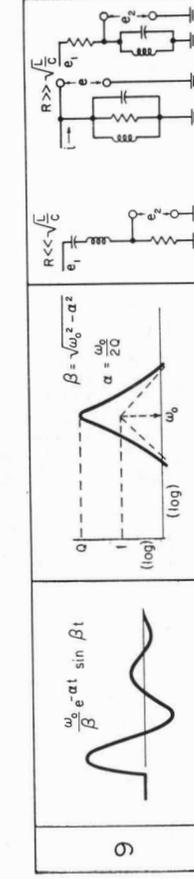
CASE 11. Similar to case 10 but here there is a resonant dip. Note that the effect of a complete null ($\Delta = 1$) is no worse than that of a 6 db hump. The pilot separation filters used in the coaxial television system produce this type of dip



CASE 8. Simple high- and low-frequency cutoff. The step response rises exponentially at a rate determined by high frequency cutoff, then falls exponentially at a rate determined by low frequency cutoff. Typical complete resistance-coupled interstage response. If $\omega_2/\omega_1 \gg 1$, then on a

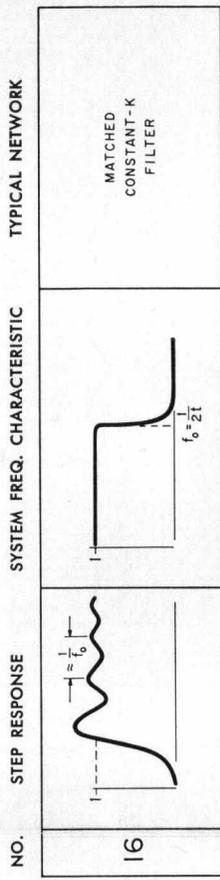


CASE 12. Positive echo. Associated frequency characteristic has nearly sinusoidal ripple in amplitude and phase. Frequency interval between successive maxima or minima is reciprocal of echo delay. The longer the delay, the closer the ripples.

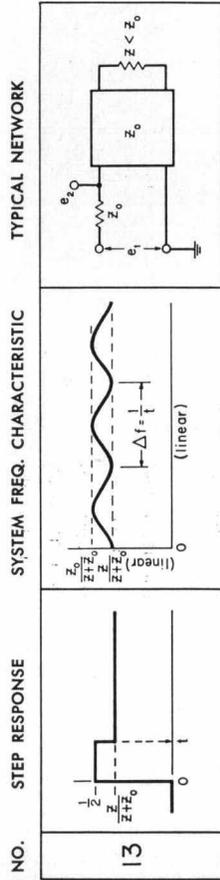


CASE 9. Typical damped oscillation. The dotted lines in the frequency characteristic are the asymptotes which the actual characteristic approaches for $\omega/\omega_0 \ll 1$ and $\omega/\omega_0 \gg 1$. The peak of the resonance curve is Q times as high as the

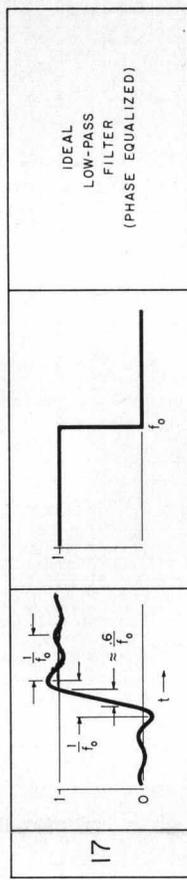
Commonly encountered in systems having faulty or misterminated delay lines. Also in measurements where multipath transmissions can exist such as acoustic measurements.



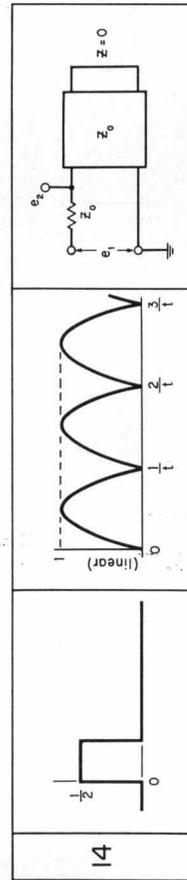
CASE 16. Rise characteristic (qualitative only) of a low pass filter without phase correction. Following the initial smooth rise there is a ripple whose apparent frequency approaches the cutoff frequency after several cycles. With an increasing number of sections this ripple increases in amplitude and duration. The "ringing sound" so often attributed to sharp cutoff filters is not due to exaggeration of frequencies near cutoff nor to the sharp cutoff *per se*, but rather to the delay distortion which exists near cutoff causing those upper frequencies which are passed to arrive too late and thus be separately audible. The effect is noticeable only in extreme cases and with proper delay equalization the effect disappears.



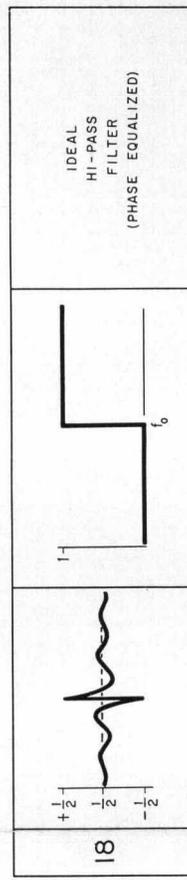
CASE 13. Negative echo. Same frequency ripples as in case 11 but reversed 180°.



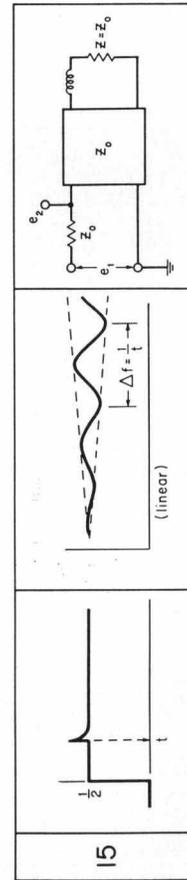
CASE 17. The "ideal" low pass filter passes all frequencies below f_0 with the same amplitude and delay while attenuating completely those above f_0 . Its step response is the sine integral. This function differs from zero (except at discrete points) for all $t > -\infty$. Hence the ideal filter cannot be realized without infinite delay. A practical approximation will have a finite delay and its step response therefore will execute only a finite number of wiggles before the main rise. Here again, the ripples in the step response do not indicate high frequency enhancement, but are the "Gibb's effect" encountered in Fourier series, and are properly called *band elimination ripples*.



CASE 14. Rectangular pulse response. Can be considered to be a 100% negative echo. Minima of frequency ripples have now become nulls. Shape of amplitude characteristic is that of rectified sine wave. Phase characteristic is sawtooth end shorted.



CASE 18. The ideal high pass filter. By superposition the response of this filter is obtained by subtracting the response of the ideal low pass filter from an equally delayed unit step.



CASE 15. "Differentiated Echo." This is the sort of disturbance produced when a delay line is terminated in such a way that the reflection coefficient increases with frequency. Typical causes are (1) series inductance or shunt capacitance in the termination of a smooth line, (2) termination of a constant-k filter in simple resistances. With both ends matched at low frequencies the transmitted echo involves two reflections both of which increase with frequency and so tends to be "doubly differentiated" and smaller.