65

NASA TECHNICAL MEMORANDUM

Report No. 53950

A GRAPHICAL APPROACH TO CONVOLUTION

By James C. Taylor
Astrionics Laboratory

September 26, 1969

NASA

George C. Marshall Space Flight Center
Marshall Space Flight Center, Alabama

NASA — GEORGE C. MARSHALL SPACE FLIGHT CENTER

INTERNAL NOTE
IN-ASTR-R-69-1

Changed to TM X-53950, September 26, 1969

A GRAPHICAL APPROACH TO CONVOLUTION

By

James C. Taylor

ASTRIONICS LABORATORY
RESEARCH AND DEVELOPMENT OPERATIONS

A GRAPHICAL APPROACH TO CONVOLUTION

By

James C. Taylor

George C. Marshall Space Flight Center
Marshall Space Flight Center, Alabama 35812

ABSTRACT

Many pages have been devoted to the development of the theory and application of convolution. In fact one might ask, "Why write any additional pages unless some fundamental contribution or additional degree of sophistication is added?" A reasonable answer to this question is that most of these pages are devoted to a rigorous analytical treatment or a discussion to support a specialized subject. The purpose of this paper is to pursue a graphical approach with sufficient analytical support to clearly illustrate the mathematical operations involved, the physical significance, and some important areas of application of convolution in a manner as simple and direct as possible.

TABLE OF CONTENTS

A GRAPHICAL APPROACH TO CONVOLUTION

GENERAL COMMENTS

Convolution is a powerful tool in communication theory and system analysis, both from a theoretical and a physical applications point of view. It provides a convenient method to relate the convolution of two functions in the time domain to the product of their spectra in the frequency domain, thereby offering, in many cases, a more convenient means of finding the inverse transform of certain functions. It offers a convenient method of evaluating the response of a linear system to an arbitrary excitation function in terms of its response to an impulse function. Furthermore, it provides a means of graphical solution when the functions involved cannot be expressed by analytical means, and it is most helpful in visually displaying relatively abstract concepts. Convolution is not a Fourier or Laplace transformation, although it is used extensively with transformation methods. It is an operation in its own right and obeys the distributive, associative, and commutative laws of algebra. Convolution of two functions resembles cross-correlation in that both include displacement, multiplication, and integration. However, convolution includes a folding or reflection which cross-correlation does not. The convolution of two functions, $f_1(t)$ and $f_2(t)$, is generally written as $f_1(t) * f_2(t)$ and is accomplished by folding one of the functions, shifting or displacing it by a given amount, multiplying it by the other function, and integrating the product curve over its domain of definition. Another solution would be to say the convolution of two functions can be obtained by scanning one of the functions with each element of the other function from $-\infty$ to $+\infty$ and summing the products at each point of the scan.

Because of the product or the overlap of the two functions, the convolution function is extended beyond the bounds of either function. Mathematically, the convolution of $f_1(t)$ with $f_2(t)$ is written

$$f_1(t) * f_2(t) = \int_{-\infty}^{+\infty} f_1(\tau) f_2(t-\tau)\, d\tau \quad . \tag{1}$$

The expression on the right-hand side of equation (1) is the convolution integral. Sometimes it is called the Faltung integral. The notation in equation (1) is quite popular and it should be noted that a dummy integration variable τ has been introduced so that $f_1(t)$ becomes $f_1(\tau)$ and $f_2(t)$ becomes $f_2(\tau)$. The function $f_2(\tau)$ is folded by the substitution of $-\tau$ for τ and is displaced along the τ-axis by an amount t. Either function could have been chosen to fold. However, once the choice is made, the folded function is the one that must be shifted. As a general rule, the simplest function is folded. Sometimes equation (1) is written as

$$f_1(\tau) \underset{t}{\otimes} f_2(\tau) = \int_{-\infty}^{+\infty} f_1(\tau)\, f_2(t-\tau)\, d\tau$$

where it is understood that the variable of integration is the dummy variable τ but the integral is indeed a function of the offset t. For each offset value of t, the product curve $f_1(\tau)f_2(t - \tau)$ is integrated from $-\infty$ to $+\infty$.

The convolution of two functions can often be evaluated graphically when the functions are so complicated that they defy analytical integration. Also they may be curves arrived at by experiment which cannot be expressed analytically. Therefore it is most important to understand the folding, displacement, product, and summation steps that are embodied in equation (1).

GRAPHICAL INTERPRETATION OF CONVOLUTION

Remembering the preceding general remarks, we will depart from the usual analytical approach to convolution and discuss the graphical procedures that lead to solution in considerable detail. It is believed that this approach will supply a better understanding of the physical significance than the more abstract analytical analysis. First, folding and displacement will be discussed, and then the product of the two functions as one is successively displaced will follow. Final'$_{\nu}$, the area under the product curve will be plotted as a function of the displacement or the offset. This is the convolution function.

Consider a function $f_2(t)$, shown in Figure 1a, in which a variable τ has been introduced in Figure 1b. It is possible to fold $f_2(\tau)$ around $\tau = 0$ (and this fact is important in convolution) by substituting $-\tau$ as illustrated in Figure 2a. This folding operation is the essential difference between convolution and cross-correlation. Figure 2b illustrates the folded function $f_2(-\tau)$ shifted to the right in the positive direction along the τ-axis by an amount t. The folded and

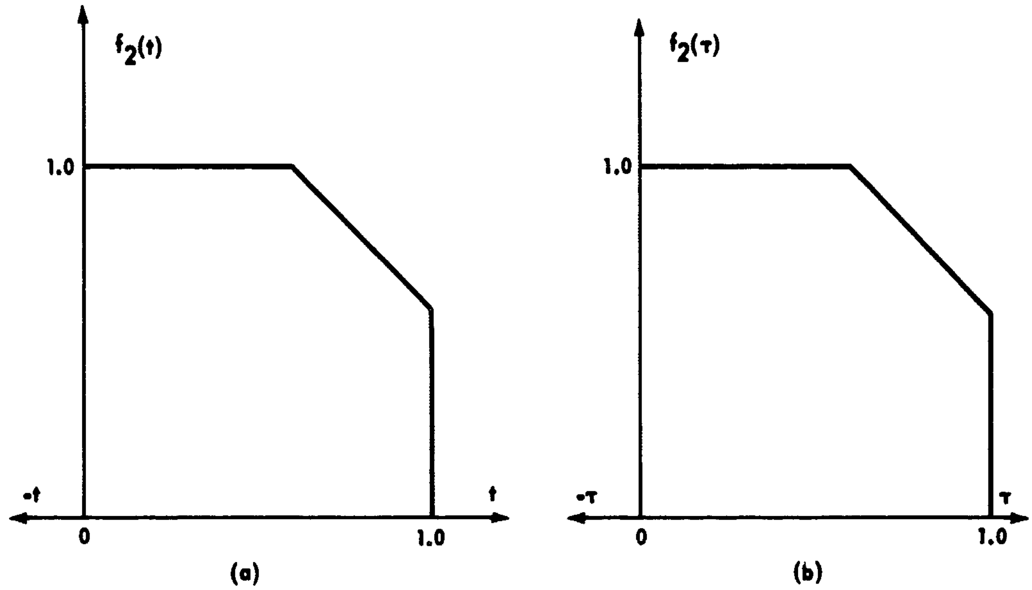


FIGURE 1

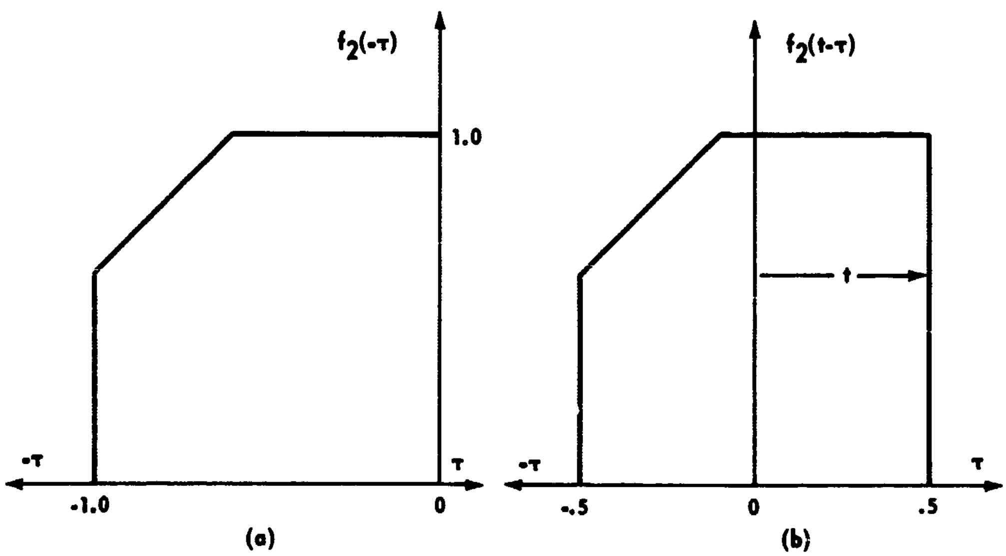


FIGURE 2

shifted function is now expressed as $f_2(t - \tau)$. If the function had been shifted to the left in the direction of the negative τ-axis, it would be expressed as $f_2(-t - \tau)$. Clearly for each value of t, there is a different position for the single pulse $f_2(t - \tau)$ along the τ-axis. Therefore by selecting successively larger positive values for t (t=1, t=2, t=3, . . .), the pulse is shifted along the τ-axis in the positive direction. If another function $f_1(\tau)$ is plotted and $f_2(t - \tau)$ is shifted across $f_1(\tau)$, the area under the product curve (area under the curve obtained by multiplying the two functions in the region of overlap) is a function of the offset t. If the area under the product curve is plotted as a function of the offset t, the result is the locus of the convolution function $f_1(t) * f_2(t)$.

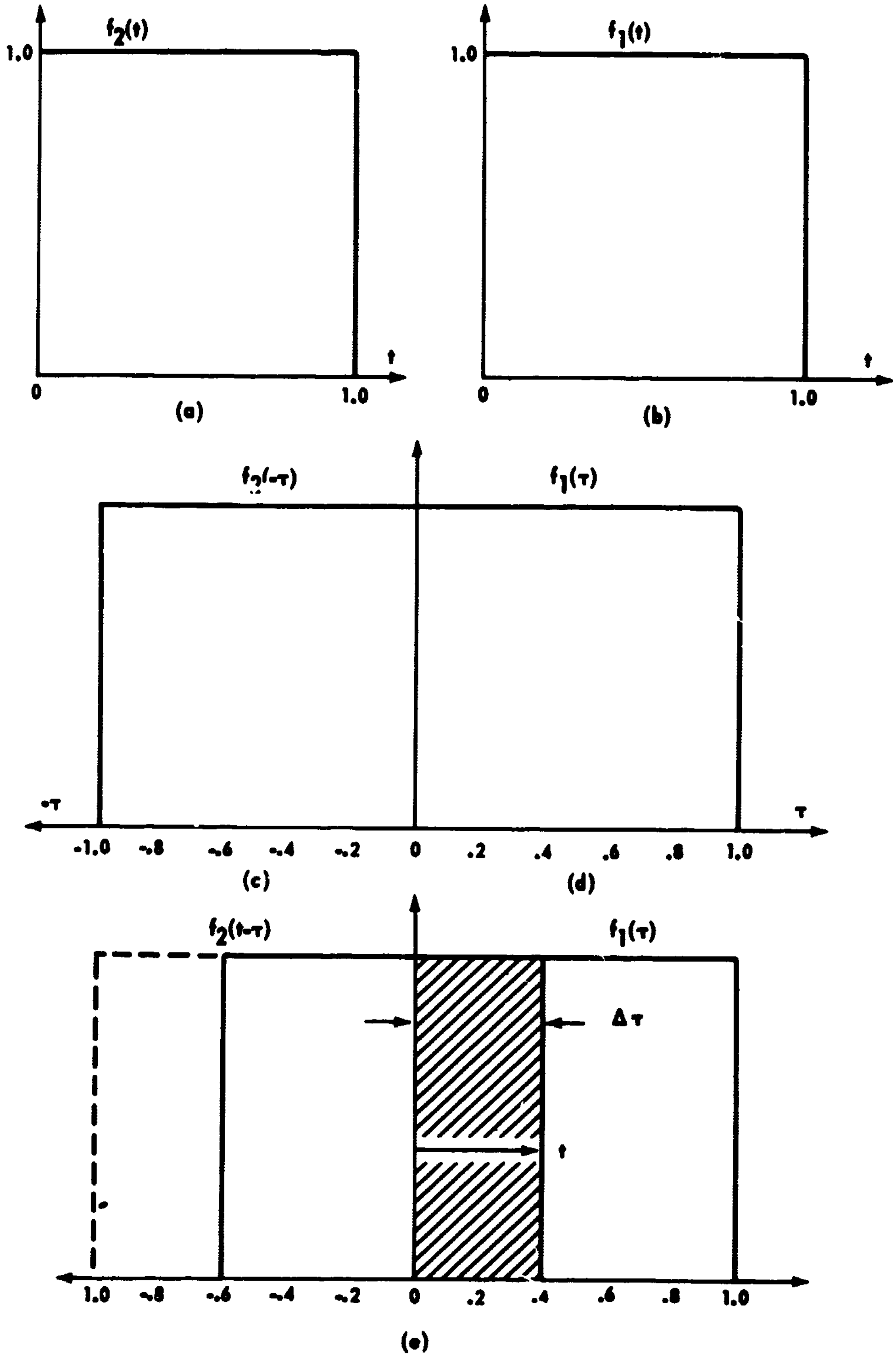
CONVOLUTION OF TWO SIMILAR PULSES $f_1(t) = f_2(t)$

Two similar pulses, $f_1(t) = f_2(t)$, are shown in Figures 3a and 3b. The dummy variable τ is introduced directly in $f_1(t)$. Since $f_2(t)$ is selected to be folded, $-\tau$ is directly substituted for t, as shown in Figures 3c and 3d. For convenience, both functions are plotted on the same set of coordinates and the folded function is shifted by an arbitrary amount in the positive direction along the τ-axis in Figure 3e. Since the folded function has been shifted by an amount t, it is expressed as $f_2(t - \tau)$. Figure 4 shows six successive values of t as the function $f_2(t - \tau)$ is shifted across $f_1(\tau)$. The "hashed" area is the area in which the two functions overlap. It is clear that the amplitude of the product of the two functions in the region of overlap is unity since the respective amplitudes $a_1 = a_2 = 1$. The area under this product curve for any offset t is simply

$$A = a_1 a_2 \Delta\tau = 1 \times 1 \times \Delta\tau \qquad (2)$$

where $\Delta\tau$ is the amount of overlap and is clearly a function of t. Table I tabulates the pertinent graphical information shown in Figures 3 and 4. Now if we plot the area under the product curve versus t, we have the locus of the convolution function, as shown in Figure 5. The convolution of two similar rectangular pulses results in a triangular spectrum. The range of the convolution is 2.0 while the range of $f_1(t) = f_2(t) = 1.0$.

Expansion of the foregoing graphical exercise is most convenient to illustrate the significance of equation (1). This equation states that a product (curve) is formed by multiplying $f_1(\tau)$ by $f_2(t - \tau)$ which has been folded and shifted by some value t. Furthermore, for each value of t the integral has a value which is obtained by integrating the product curve over the limits $-\infty$ to $+\infty$. Figure 6 is the result of redrawing Figure 4a. The region of overlap $\Delta\tau$



f2(t)
1.0
0
1.0
t
(a)
f1(t)
1.0
0
1.0
t
(b)
f2(-τ)
f1(τ)
-τ
τ
-1.0
-.8
-.6
-.4
-.2
0
.2
.4
.6
.8
1.0
(c)
(d)
f2(t-τ)
f1(τ)
Δτ
t
1.0
-.8
-.6
-.4
-.2
0
.2
.4
.6
.8
1.0
(e)

FIGURE 3

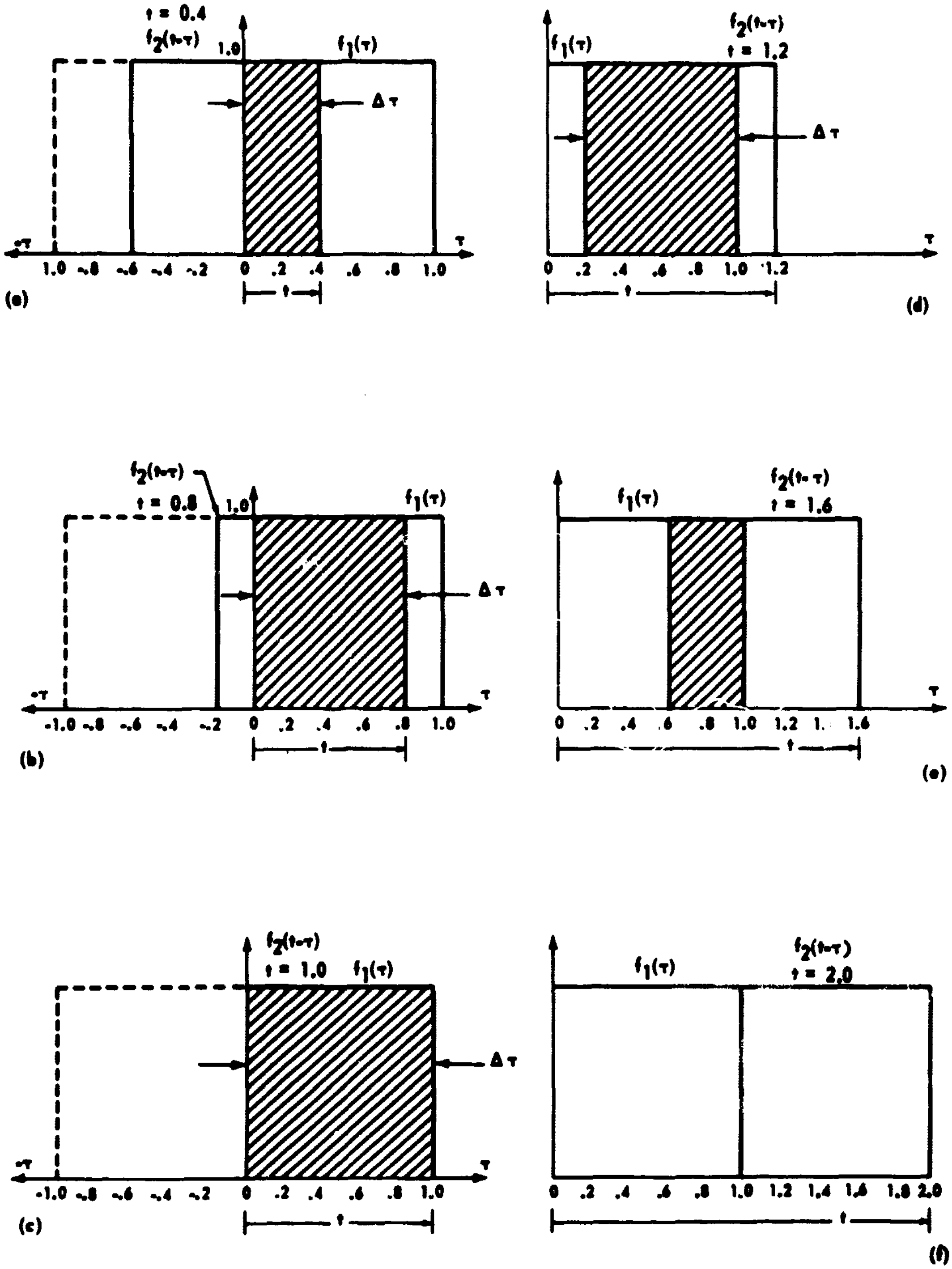

t = 0.4
f2(t-τ)
1.0
f1(τ)
Δτ
-τ
τ
1.0 -.8 -.6 -.4 -.2 0 .2 .4 .6 .8 1.0
t
(a)
f2(t-τ)
t = 1.2
f1(τ)
Δτ
τ
0 .2 .4 .6 .8 1.0 1.2
t
(d)
f2(t-τ)
t = 0.8
1.0
f1(τ)
Δτ
-τ
τ
-1.0 -.8 -.6 -.4 -.2 0 .2 .4 .6 .8 1.0
t
(b)
f1(τ)
f2(t-τ)
t = 1.6
τ
0 .2 .4 .6 .8 1.0 1.2 1.6
t
(e)
f2(t-τ)
t = 1.0
f1(τ)
Δτ
-τ
τ
-1.0 -.8 -.6 -.4 -.2 0 .2 .4 .6 .8 1.0
t
(c)
f1(τ)
f2(t-τ)
t = 2.0
0 .2 .4 .6 .8 1.0 1.2 1.4 1.6 1.8 2.0
t
(f)

FIGURE 4

TABLE I

Figure	Value of Offset t	$f_1(\tau) f_2(t - \tau) \Delta \tau$ Area Under the Product Curve for Offset t
3c, 3d	0	0 No Overlap
4a	0.4	0.4
4b	0.8	0.8
4c	1.0	1.0
4d	1.2	0.8
4e	1.6	0.4
4f	2.0	0 No Overlap

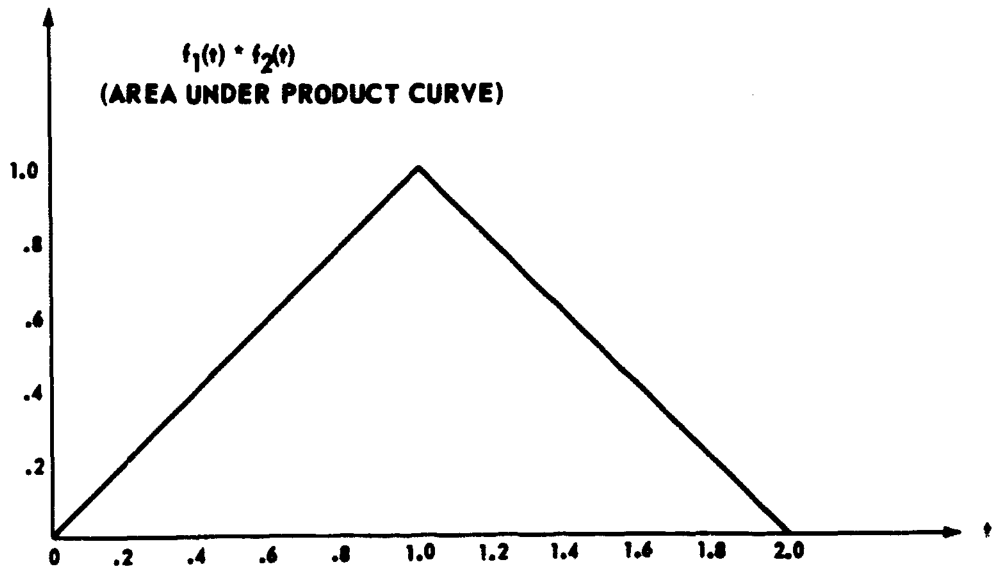


FIGURE 5

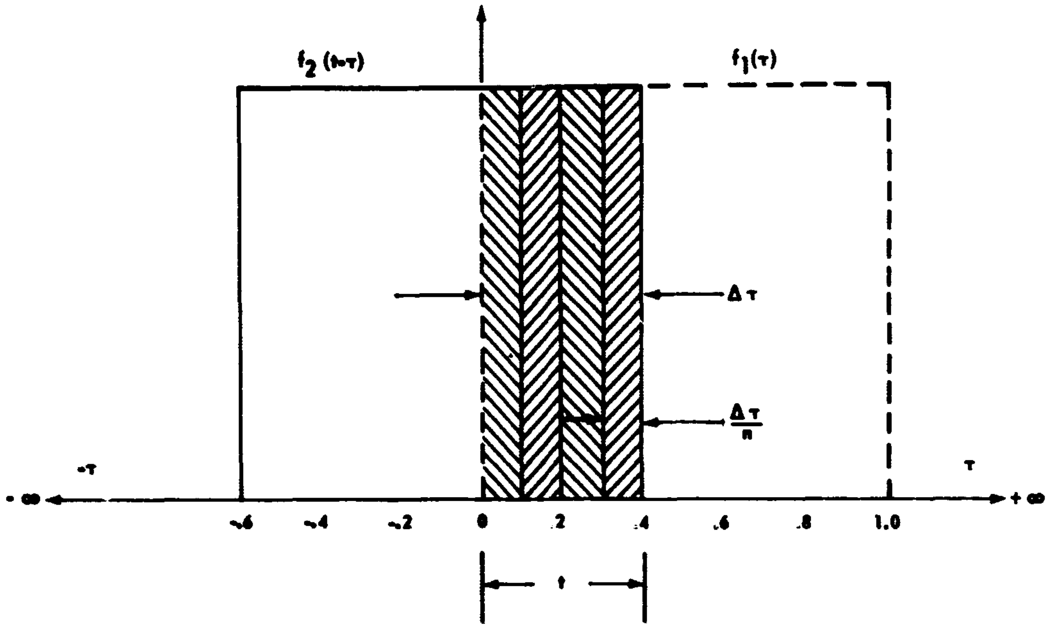


FIGURE 6

in Figure 4a has been divided into n increments of width $\Delta\tau/n$. Each incremental area is

$$\Delta A = f_1(\tau) f_2(t-\tau) \frac{\Delta\tau}{n}$$

where the area is a function of the offset t. Then the area under the product curve in the interval t=0 to t=0.4 is the sum of all incremental areas within the interval

$$A = \sum_{t=0}^{t=0.4} f_1(\tau) f_2(t-\tau) \frac{\Delta\tau}{n} \quad .$$

From the fundamental theorem of integral calculus

$$\lim_{\substack{\Delta\tau \to 0 \\ n \to \infty}} \sum_{0}^{t} f_1(\tau) f_2(t-\tau) \frac{\Delta\tau}{n} = \int_0^t f_1(\tau) f_2(t-\tau)\, d\tau \quad .$$

Since the product curve is zero outside the interval of overlap, nothing is added to the integral to replace the limits of integration by $-\infty$, $+\infty$

$$\int_{0}^{t} f_1(\tau) f_2(t - \tau)\, d\tau = \int_{-\infty}^{+\infty} f_1(\tau) f_2(t - \tau)\, d\tau \quad . \tag{3}$$

This is the convolution integral given in equation (1).

GRAPHICAL CONVOLUTION OF TWO DISSIMILAR PULSES $f_1(t) \neq f_2(t)$

The convolution of two dissimilar rectangular pulses results in a truncated isoceles triangle; the sides become steeper as the width of the narrow pulse becomes more narrow. The understanding of this convolution leads conveniently to an important property of convolution; that is, if a function $f_1(t)$ is convolved with a delta function or unit impulse $\delta(t)$, the result is the function $f_1(t)$ itself or

$$f_1(t) * \delta(t) = f_1(t) \quad . \tag{4}$$

This concept will be extended to show that if a function $f_1(t)$ is convolved with a unit impulse function $\delta(t - t_o)$ positioned at $t = t_o$, the function $f_1(t)$ is translated along the t-axis in the positive direction by an amount t_o which is written

$$f_1(t) * \delta(t - t_o) = f_1(t - t_o) \quad . \tag{5}$$

The procedure will be identical with that of the graphical convolution of two similar rectangular pulses. However, $f_2(t)$ is chosen to be a narrow pulse compared to $f_1(t)$. Upon comparison of the two examples, it will be clear that as the width of f_2 becomes very narrow, the convolution function becomes more rectangular. Furthermore as the width becomes more narrow, the height of the convolution function decreases. If the height of f_2 is allowed to increase as the width decreases in such a manner that the area becomes and remains unity, f_2 becomes the delta function δ (by definition) and equation (4) will hold.

The two functions are illustrated in Figures 7a and 7b. A dummy variable τ has been introduced in place of t in Figures 7c and 7d, and f_2 has

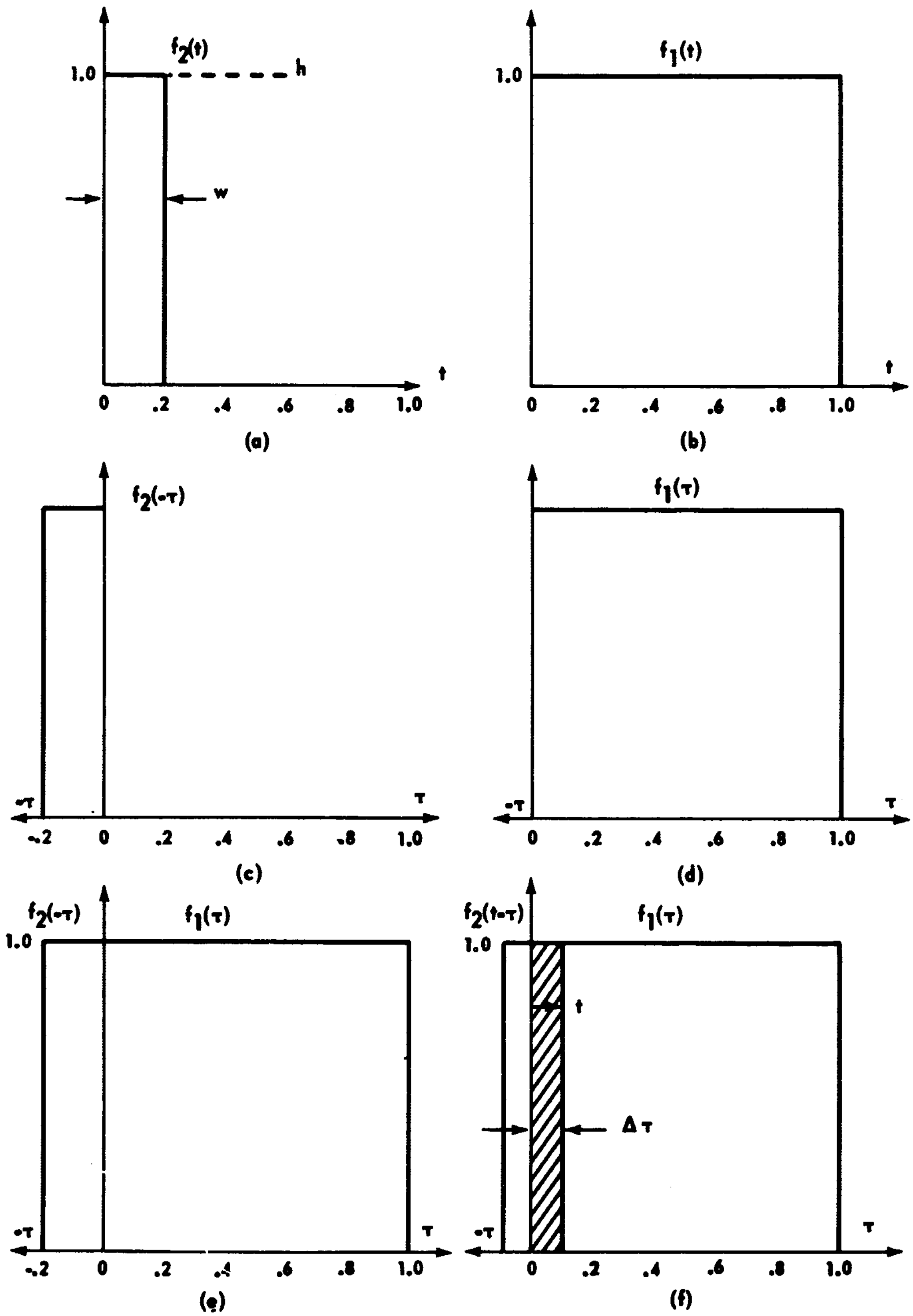


FIGURE 7

been rotated about the origin by substitution of $-\tau$ for τ. Both functions are drawn on the same coordinate system in Figure 7e. The function f_2 has been shifted along the positive direction of the τ-axis by an amount t which causes an overlap $\Delta\tau$ of $f_1(\tau)$ and the folded and shifted function $f_2(t-\tau)$ in Figure 7f. A product curve exists as long as the two functions overlap, and the area under this curve for each t is easily determined for this case by the simple product of the amplitude of each curve and the amount of overlap $\Delta\tau$ which is

$$A = f_1(\tau) f_2(t-\tau)\Delta\tau = 1 \times 1 \times \Delta\tau \quad . \tag{6}$$

Table II tabulates the area under the product curve for the corresponding offset t along with its functional overlap $\Delta\tau$. The area under the product curve is plotted as a function of the offset in Figure 8. Comparison of Figure 8 with Figure 5 illustrates graphically that as the width of one rectangular pulse becomes more narrow, the slopes of the sides of the product curve increase and the peak becomes truncated and decreases in magnitude. In fact, the shape of the curve in Figure 8 begins to approach the shape of the function $f_1(t)$, except the height is greatly reduced.

TABLE II

Offset t	Overlap $\Delta\tau$	Area Under Product Curve $f_1(\tau)f_2(t-\tau)\Delta\tau$ for t
0	0	0
0.1	0.1	0.1
0.2	0.2	0.2
0.4	0.2	0.2
0.6	0.2	0.2
0.8	0.2	0.2
1.0	0.2	0.2
1.1	0.1	0.1
1.2	0	0

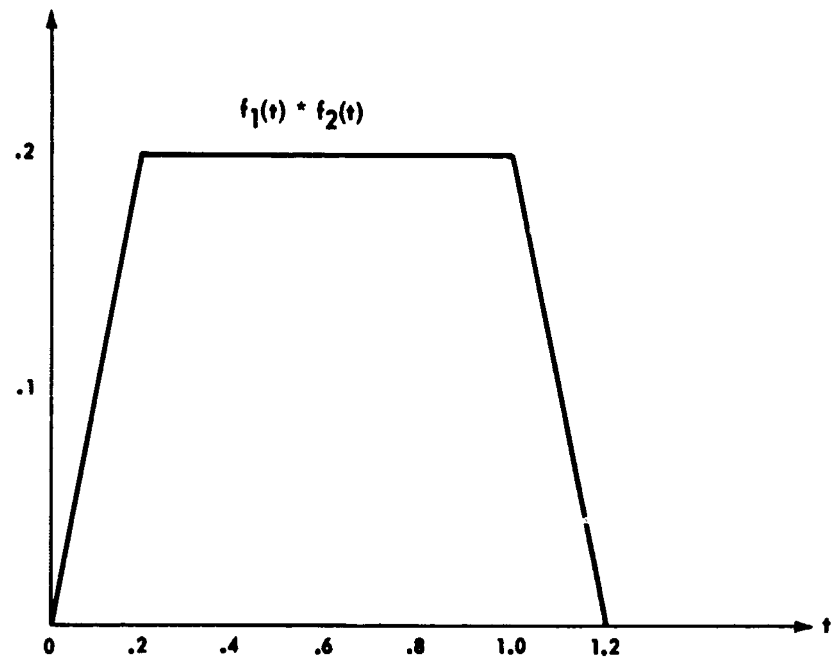


FIGURE 8

To more closely reproduce $f_1(t)$ we must increase the height of $f_2(t)$ as its width w is decreased. If we decrease the width and increase the height so the product of the width and the height equal unity, the maximum value of the area under the product curve will equal the amplitude of $f_1(t)$ throughout the region in which the overlap is constant (as the narrow pulse is shifted through the broad one). This case is illustrated in Figure 9 where the function $f_2(t)$ of Figure 3a is chosen with an amplitude five times that of $f_1(t)$ and with a width w which is one-fifth that of $f_1(t)$. The area of the pulse is unity. Data for the graphical convolution of the functions of Figure 9 are tabulated in Table III and the convolution function is plotted in Figure 10. If the convolution $f_1(t) * f_2(t)$ of Figure 10 is compared with $f_1(t)$ in Figure 3, it shows that

$$f_1(t) * f_2(t) = f_1(t) \tag{7}$$

for $0.2 \leq t \leq 1.0$. As the width of $f_2(t)$ becomes very small and its amplitude is increased so the area remains unity, the equality of equation (7) holds over almost the entire range of $f_1(t)$. Finally in the limit as the width approaches zero and the height approaches infinity, $f_2(t)$ becomes the Dirac delta function or unit impulse $\delta(t)$, equation (7) becomes equation (4), and equation (1) may be rewritten as

$$f_1(t) * \delta(t) = \int_{-\infty}^{+\infty} f_1(\tau)\, \delta(t-\tau)\, d\tau = f_1(t) \quad . \tag{8}$$

TABLE III

t	$\Delta\tau$	$f_1(\tau) f_2(t-\tau)\Delta\tau$
0	0	0
0.1	0.1	0.5
0.2	0.2	1.0
0.4	0.2	1.0
0.6	0.2	1.0
0.8	0.2	1.0
1.0	0.2	1.0
1.1	0.1	0.5
1.2	0	0

CONVOLUTION OF A FUNCTION WITH AN IMPULSE FUNCTION

It is now convenient to consider the important case of convolving a function $f_1(t)$ with an impulse function $\delta(t - t_o)$ positioned at $t = t_o$. First, as in the preceding example, we will convolve $f_1(t)$ with a narrow finite pulse $f_2(t - t_o)$ that is positioned at $t = t_o$. Both functions are shown in Figure 11a. A dummy variable is introduced in Figure 11b. The function $f_2(\tau - t_o)$ is rotated around the origin ($\tau = 0$) by substitution of $-\tau$ in Figure 11c. Finally the rotated function is shifted in the positive direction along the τ-axis by an amount t. The overlap $\Delta\tau$ and the area under the product curve $f_1(\tau) f_2(t - t_o - \tau)\Delta\tau$ are tabu-

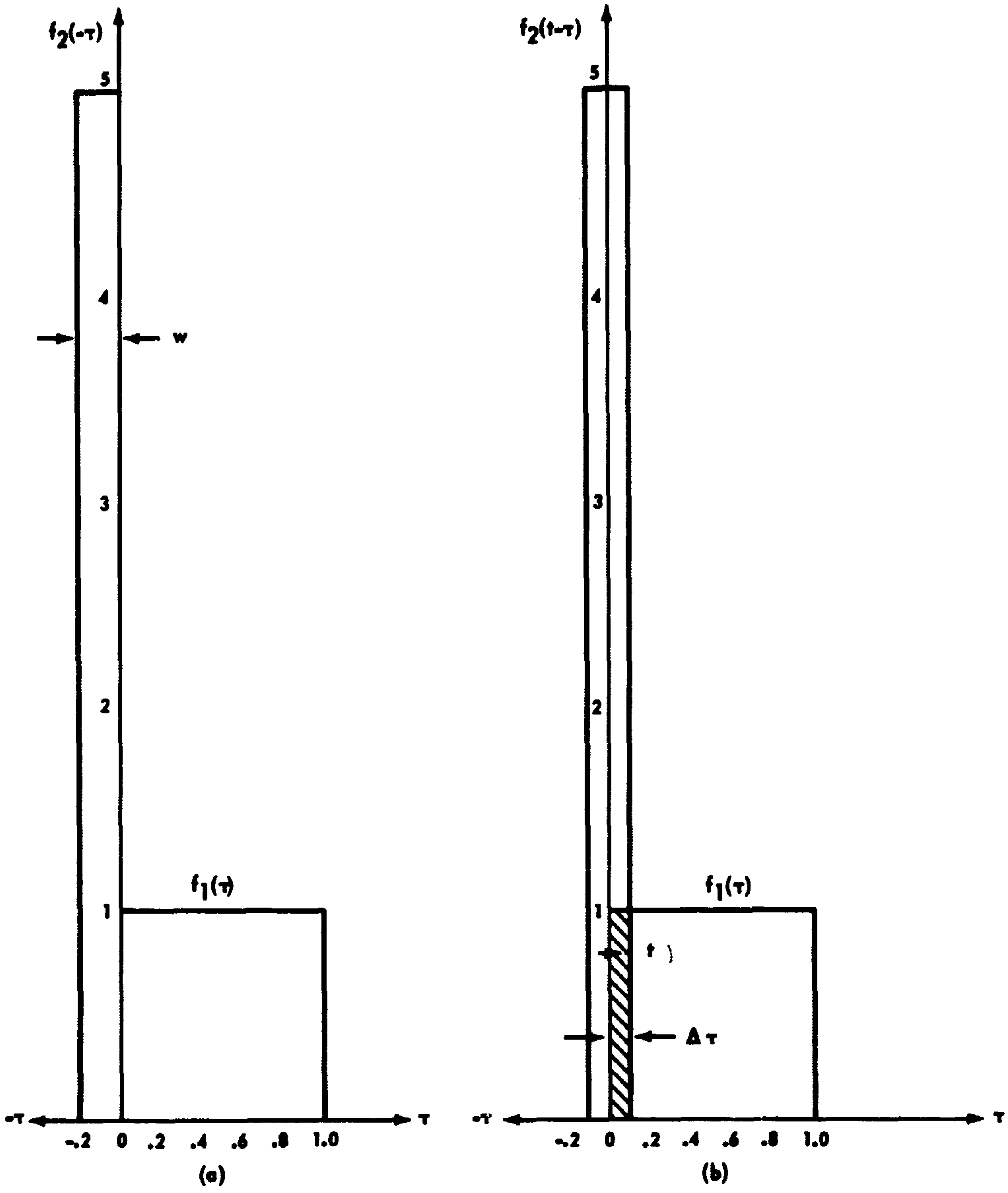


FIGURE 9

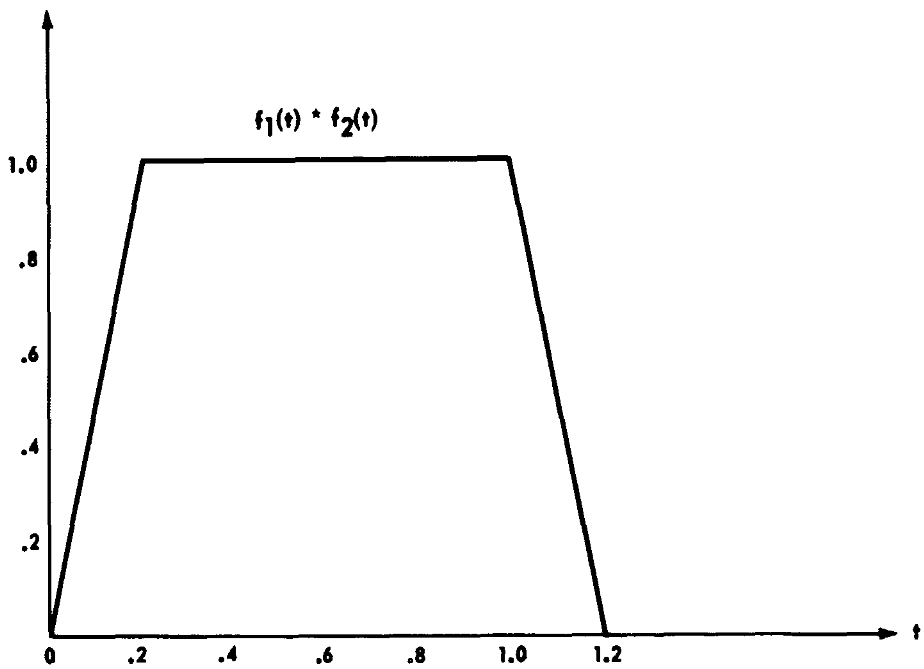
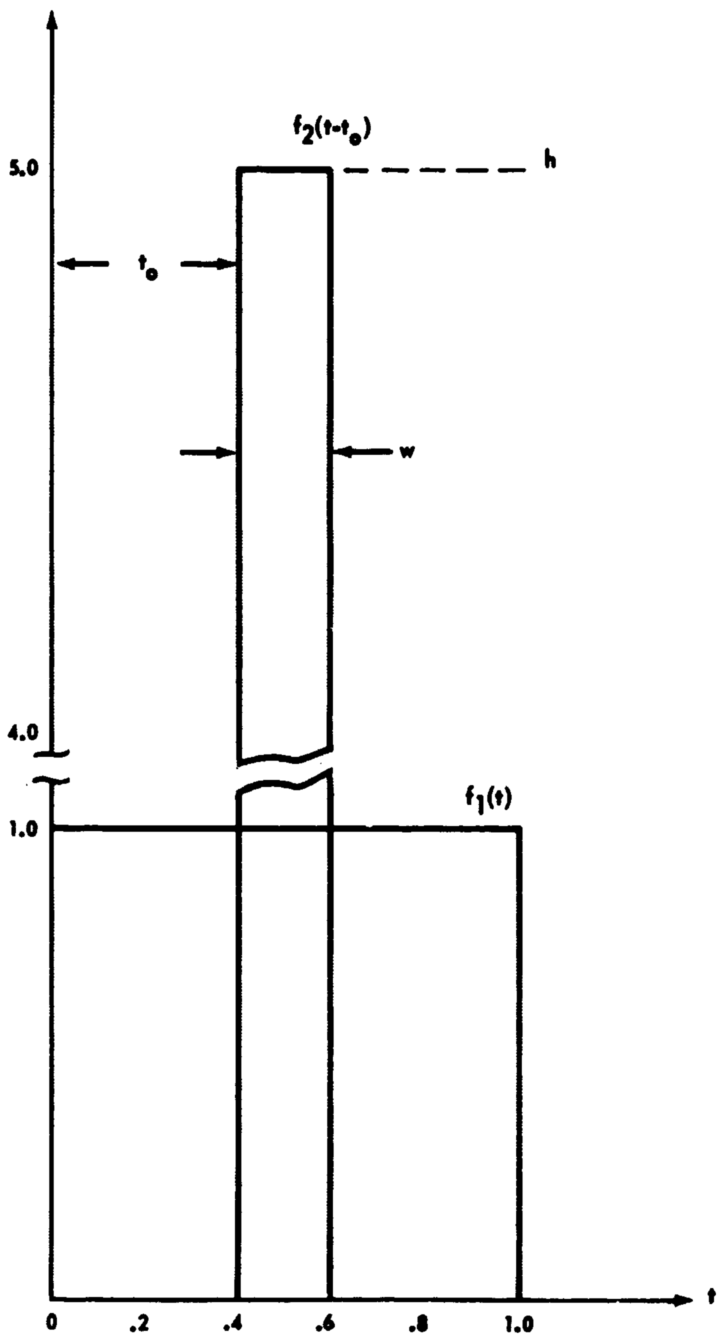


FIGURE 10

lated for several corresponding values of t in Table IV. The convolution function is plotted in Figure 12. By comparing Figure 12 with Figure 10, the convolution function of Figure 12 is identical except it has been displaced by an amount $t = t_o$. Furthermore, as the width w of the function f_2 is allowed to become smaller and smaller as the height is allowed to become greater and greater and the area is held constant at unity, the height of the plateau of Figure 12 remains equal to the amplitude of f_1 (which is 1.0 in this case) over a larger and larger range of the offset t. Finally in the limit as $f_2(t-t_o)$ becomes the delta function $\delta(t - t_o)$ (which is a function with zero width, infinite height, and unit area) the convolution function of Figure 12 becomes that of Figure 13, which is exactly that of $f_1(t)$ of Figure 11a shifted to the right by an amount $t = t_o$.

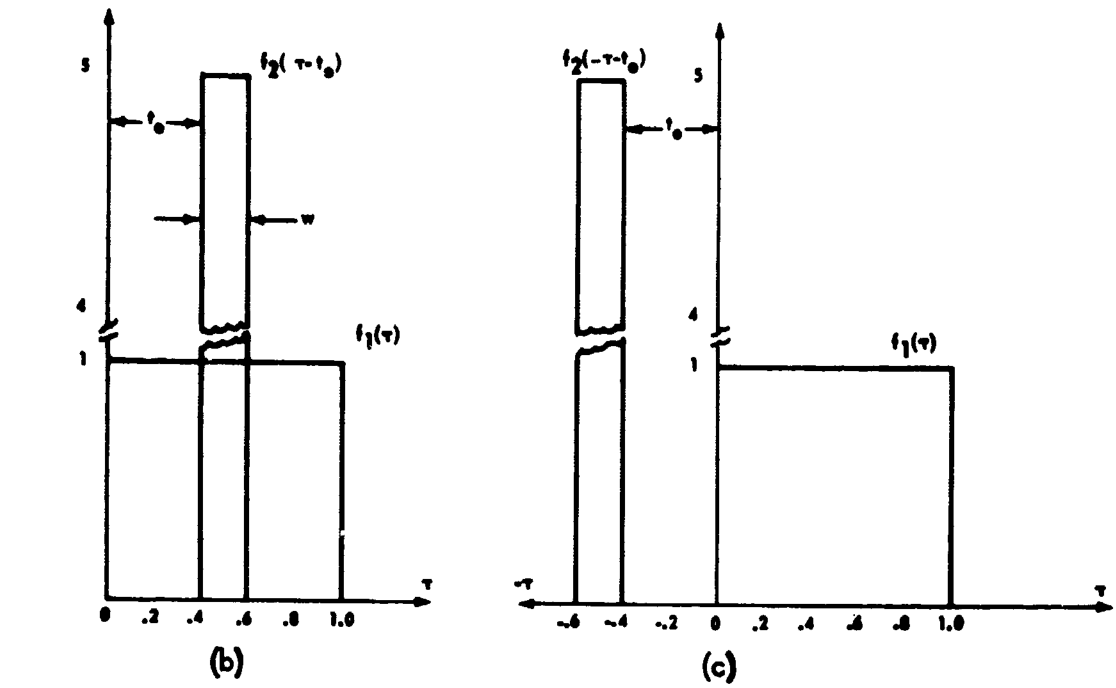
From an analytical point of view, the same argument used to arrive at equation (3) yields

$$f_1(t) * f_2(t - t_o) = \int_{-\infty}^{+\infty} f_1(\tau) f_2(t - t_o - \tau)\, d\tau$$



(a)

FIGURE 11



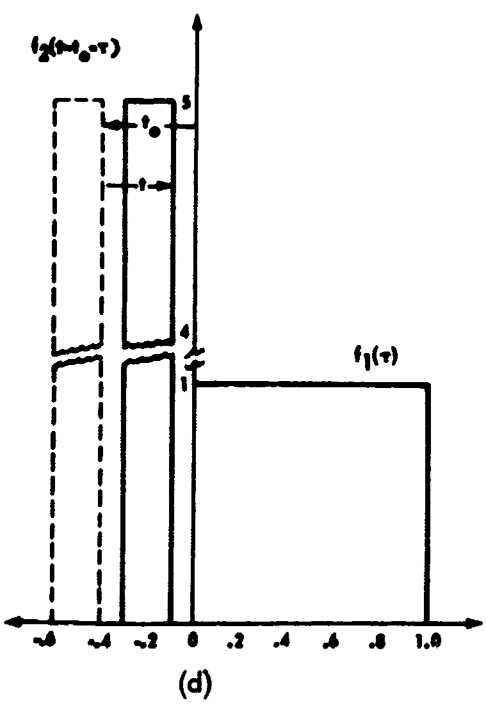


FIGURE 11 (Concluded)

TABLE IV

t	$\Delta\tau$	$f_1(\tau)\, f_2(t-t_0-\tau)\, \Delta\tau$
0	0	0
0.2	0	0
0.4	0	0
0.5	0.1	0.5
0.6	0.2	1.0
0.8	0.2	1.0
1.0	0.2	1.0
1.2	0.2	1.0
1.4	0.2	1.0
1.5	0.1	0.5
1.6	0	0

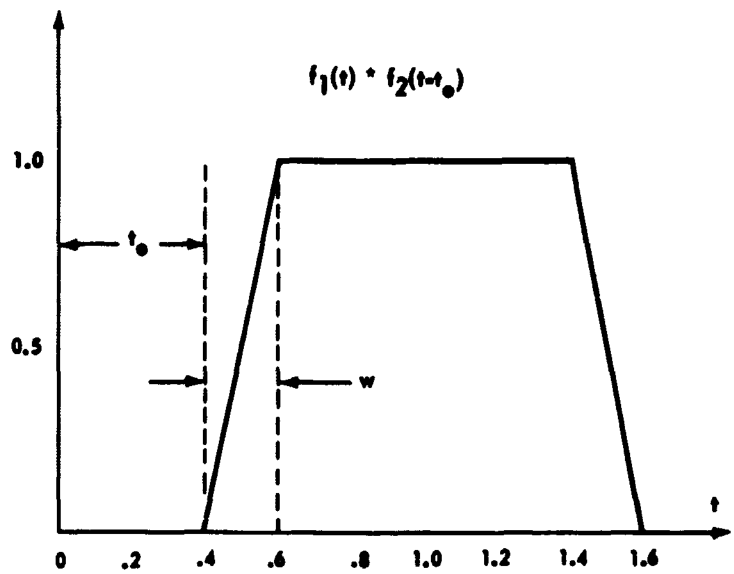


FIGURE 12

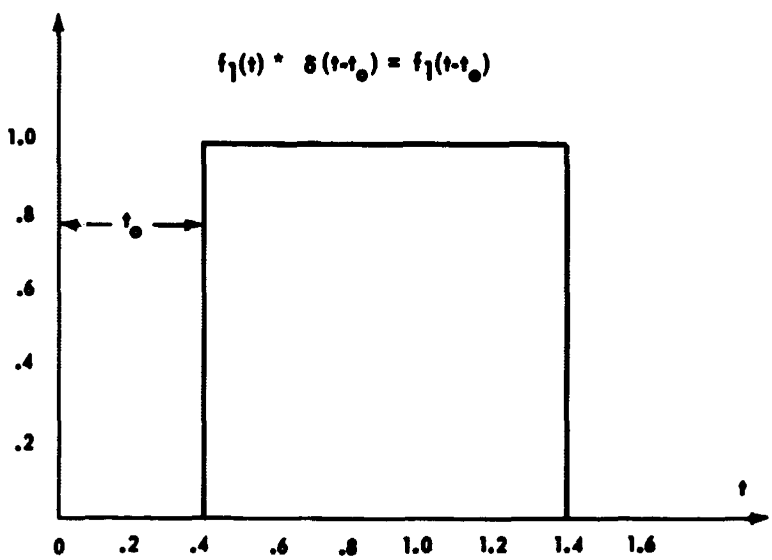


FIGURE 13

and as

$$f_2(t - t_o) \rightarrow \delta(t - t_o)$$

$$f_1(t) * \delta(t - t_o) = \int_{-\infty}^{+\infty} f_1(\tau)\, \delta(t - t_o - \tau)\, d\tau = f_1(t - t_o) \quad . \tag{9}$$

DIRAC DELTA FUNCTION

Expansion on the Dirac delta function should add substance to the previous discussion as well as enhance physical interpretation of subsequent convolution techniques. The delta function is a powerful tool in analyzing transformation problems. It may be regarded by the civil engineer as a concentrated load, by the electronic engineer as a unit impulse function, and by the physicist analyzing an optical system as a point (or line) source. In many applications the analyst may determine the response of a linear system to an arbitrary excitation function from its response to the delta function. This function is not an ordinary function in that it has a definite value for each point within the domain of definition and must be used with discretion to avoid inconsistencies. A rigorous treatment of the properties of the delta function must be based upon distribution theory rather than conventional mathematics. The delta function at the origin is defined to be zero everywhere except for $t = 0$ or

$$\delta(t) = 0 \qquad t \neq 0 \tag{10a}$$

such that the area is unity

$$\int_{-\infty}^{+\infty} \delta(t)\, dt = 1. \tag{10b}$$

This function is represented by Figure 14a, in which the function exists only at the point $t = 0$ and the one in parenthesis reminds us that the area of the function is unity. If the delta function is translated along the t-axis in the positive direction by an amount $t = t_o$,

$$\delta(t = t_o) = 0 \qquad t \neq t_o \tag{11a}$$

and the area is

$$\int_{-\infty}^{+\infty} \delta(t - t_o)\,dt = 1 \,. \tag{11b}$$

In this case the function only exists at the point $t = t_o$, as shown in Figure 14b.

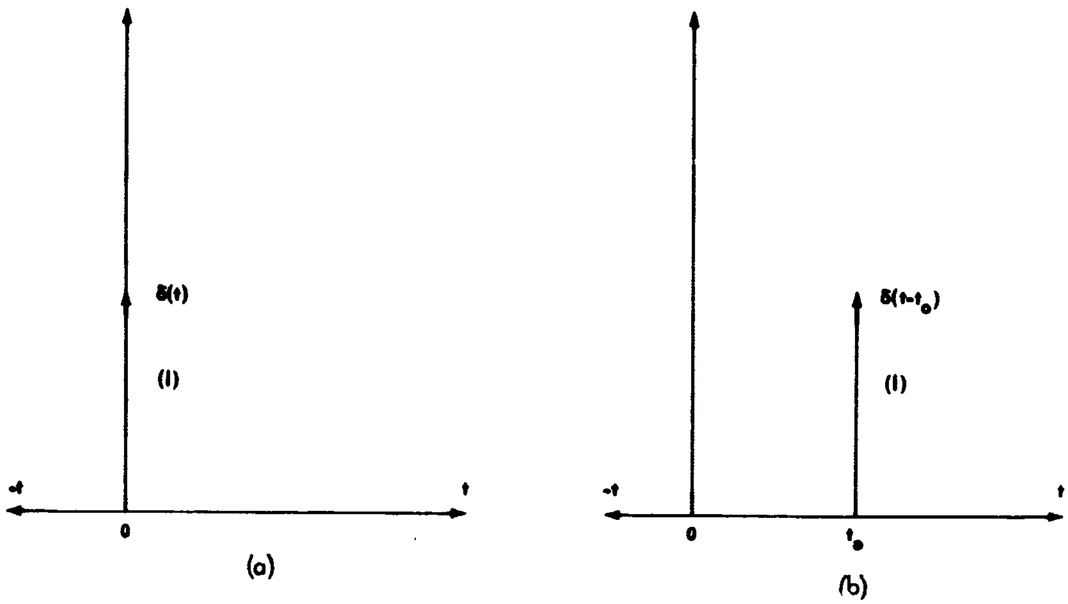


FIGURE 14

A very important property of the delta function will be shown and subsequently utilized to analytically arrive at the convolution result given in equation (9). Figure 15 illustrates a function of a real variable t plotted along with a delta function positioned at $t = \tau$, which by definition does not exist at any point except $t = \tau$. The two functions intersect at $f(t) = f(\tau)$; therefore

$$f(t)\,\delta(t - \tau) = f(\tau)\,\delta(t - \tau) \,. \tag{12a}$$

Let $t = \sigma - t_o$ and we have

$$f(\sigma - t_o)\,\delta(\sigma - t_o - \tau) = f(\tau)\,\delta(\sigma - t_o - \tau) \,. \tag{12b}$$

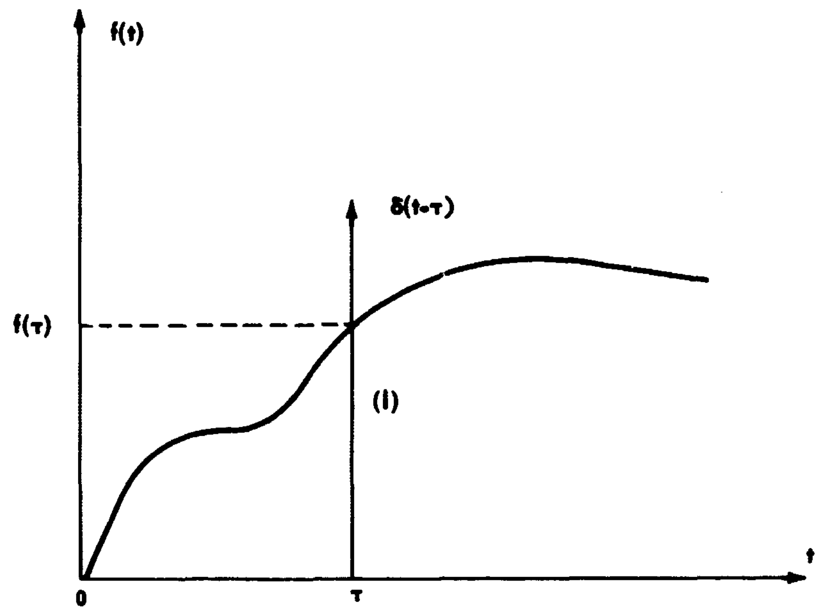


FIGURE 15

Integrating equation (12a) over the range of t($-\infty$ to $+\infty$) gives

$$\int_{-\infty}^{+\infty} f(t)\ \delta(t-\tau)\,dt = \int_{-\infty}^{+\infty} f(\tau)\ \delta(t-\tau)\,dt \ .$$

The factors in the integrands on both sides are equivalent from equation (12a). However, the factor $f(\tau)$ in the integrand on the right (equation 12a) is not a function of the variable of integration t and the right-hand side may be written as

$$\int_{-\infty}^{+\infty} f(\tau)\ \delta(t-\tau)\,dt = f(\tau) \int_{-\infty}^{+\infty} \delta(t-\tau)\,dt \tag{13}$$

and since by definition of the delta function the integral on the right is unity,

$$f(\tau) \int_{-\infty}^{+\infty} \delta(t-\tau)\,dt = f(\tau) \tag{14}$$

which states that multiplying a delta function $\delta(t - \tau)$ by the function of a real variable $f(t)$ and integrating over a range of t from $-\infty$ to $+\infty$ is equivalent to substituting τ for t.

With the above principles in mind and with the aid of Figure 16, we have direct analytical proof of the important property that if a function $f(t)$ is convolved with the unit impulse function $\delta(t - t_0)$, the function $f(t)$ is translated in the positive direction along the t-axis by an amount t_0. Figure 16 shows that

$$f(t) * \delta(t - t_0) = \int_{-\infty}^{+\infty} f(\tau)\, \delta(t - t_0 - \tau)\, d\tau \quad .$$

The factors in the integrand are of the same form as equation (12b); therefore

$$f(t) * \delta(t - t_0) = \int_{-\infty}^{+\infty} f(t - t_0)\, \delta(t - t_0 - \tau)\, d\tau \quad .$$

The factor $f(t - t_0)$ does not contain the variable of integration and is brought outside the integral

$$f(t) * \delta(t - t_0) = f(t - t_0) \int_{-\infty}^{+\infty} \delta(t - t_0 - \tau)\, d\tau \quad .$$

By definition of a delta function the integral is unity and the convolution is

$$f(t) * \delta(t - t_0) = f(t - t_0) \tag{15}$$

which is $f(t)$ of Figure 16a displaced by an amount $t = t_0$. This is plotted in Figure 17.

PRACTICAL IMPULSE FUNCTION

From a practical point of view, the unit impulse function is generally regarded as an impulse of infinitesimal duration $d\tau$ and an amplitude h which is

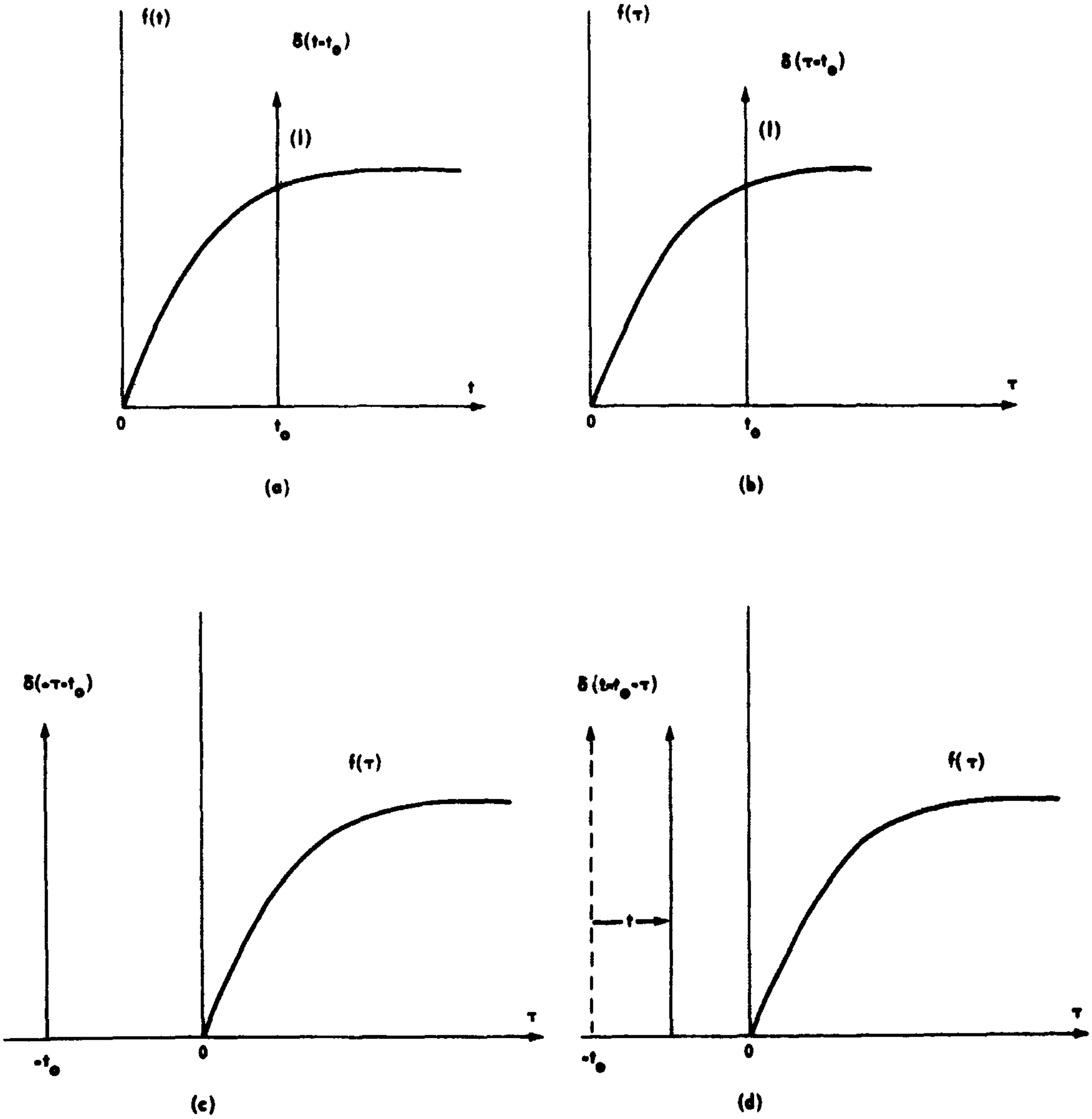


FIGURE 16

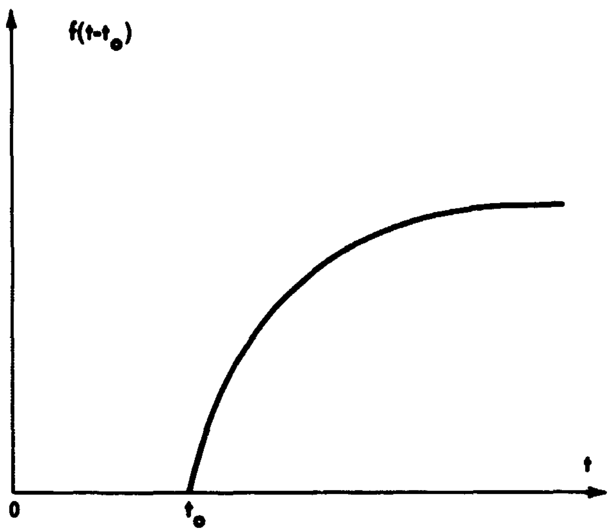


FIGURE 17

large (tends to infinity). This very narrow and very long rectangular pulse has an area equal to unity which is an even function. Since the area is finite, the pulse can be integrated over its interval. This type of pulse is approximated by the application of a voltage for a very short duration or the instantaneous charge transfer. Quite often it is convenient and very useful to work with an impulse function rather than a unit impulse function. In this case, the impulse has an area other than unity. The area k of the impulse is regarded as its strength and is written as a coefficient in the form $k\,\delta(t - t_o)$. This type of impulse is similar to the delta function except that the amplitude is controllable. In truth, it is what one is most likely to encounter in practice. Subsequently it will be shown how a convolution can be built up from a succession of impulses of proper amplitude within a chosen interval. Figure 18a illustrates an impulse of width $d\tau$ and height h with a strength $k = h\,d\tau$ which is positioned symmetrically around $t = \tau$. The factor k gives the strength, and the factor $\delta(t - \tau)$ gives the position. Therefore the impulse may be shifted along the t-axis by selecting values of τ. Suppose the height h of an impulse of constant but infinitesimal width $d\tau$ follows some function f(t). Then any i^{th} impulse is written $k_i\,\delta(t - \tau_i)$, as shown in Figure 18b, where $k_i = h_i d\tau = f(\tau_i) d\tau$.

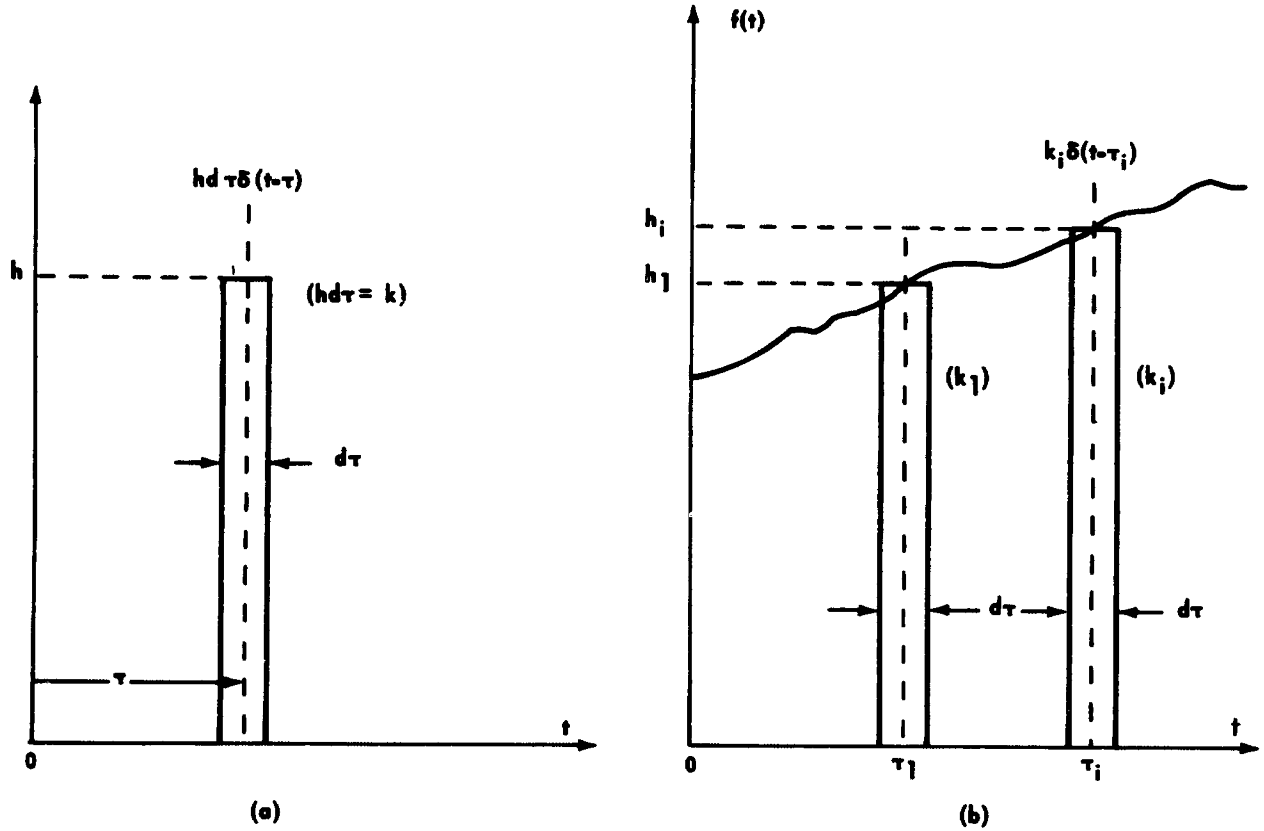


FIGURE 18

SCANNING AND SUPERPOSITION PROPERTIES OF CONVOLUTION

It is important to note that when two functions are convolved, the result is the same as if one of the functions is scanned by each element of the other and the products are summed. Recognition of this scanning property leads directly to the realization that the convolution integral is in truth a superposition integral. Thus, if $f_1(t)$ is convolved with $f_2(t)$, it is possible to partition $f_2(t)$ into small increments of width and convolve $f_1(t)$ with each increment. A superposition of the incremental convolutions results in the total convolution $f_1(t) * f_2(t)$. Strictly speaking, the superposition of the incremental convolutions approaches $f_1(t) * f_2(t)$ as the incremental width of the partitions approaches an infinitesimal. In this case, the function $f_2(t)$ is thought to be divided into adjacent impulses such that any i^{th} impulse is $k_i\,\delta(t - t_i)$ in which the impulse has a strength k_i and is positioned at t_i along the t-axis. Each of these impulses within the domain of definition of the function $f_2(t)$ is then convolved with $f_1(t)$. Superposition of all these convolutions yields the total convolution. This is easily shown by graphical means. Figures 3a and 3b are redrawn in Figure 19 where $f_2(t)$ is partitioned into five increments (a, b, c, d, and e) of width $d\tau$ and height h ($h = 1$ in this case). The increments are respectively positioned at t_o, t_1, t_2, t_3, and t_4. These partitions are rather large to consider as impulse functions (a good approximation of the unit impulse is to choose a width $d\tau$ so the impulse height will be $1/d\tau$ and the area $d\tau \times 1/d\tau = 1$). They are convenient for graphical illustration and, in this simple case, give ideal results. Impulse notation will be used to emphasize the convolution of partitions of a function $f_2(t)$ with $f_1(t)$. The partition a of $f_2(t)$ in Figure 19 is clearly the function $f_2(t)$ in Figure 7a, whereas the function $f_1(t)$ of Figure 19 is the function $f_1(t)$ of Figure 7b. The convolution function of the two dissimilar pulses shown in Figure 7 is plotted in Figure 8. Therefore the convolution function of the partition a of $f_2(t)$ with $f_1(t)$ in Figure 19 must be identical with the convolution function of Figure 8. This convolution $f_1(t) * k\,\delta(t - t_o)$ is plotted in Figure 20a. The other partitions are identical in width and height and only differ in successive position along the t-axis. Therefore their convolution functions are identical with that of partition a except they are successively displaced along the t-axis by an amount $d\tau$. These convolution functions are b, c, d, and e of Figure 20. Addition of the functions a, b, c, d, and e results in the convolution function $f_1(t) * f_2(t)$ of Figure 20f which is identical with that of Figure 5. Therefore we may convolve $f_1(t)$ and $f_2(t)$ directly or we may partition $f_2(t)$, convolve each partition with $f_1(t)$, superimpose the convolutions, and arrive at the same result.

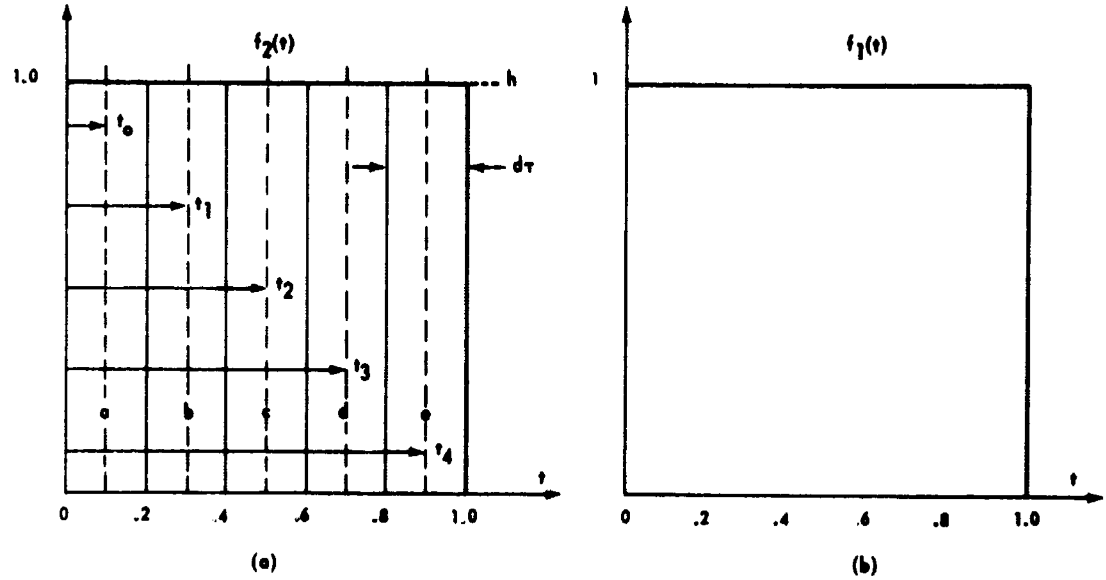
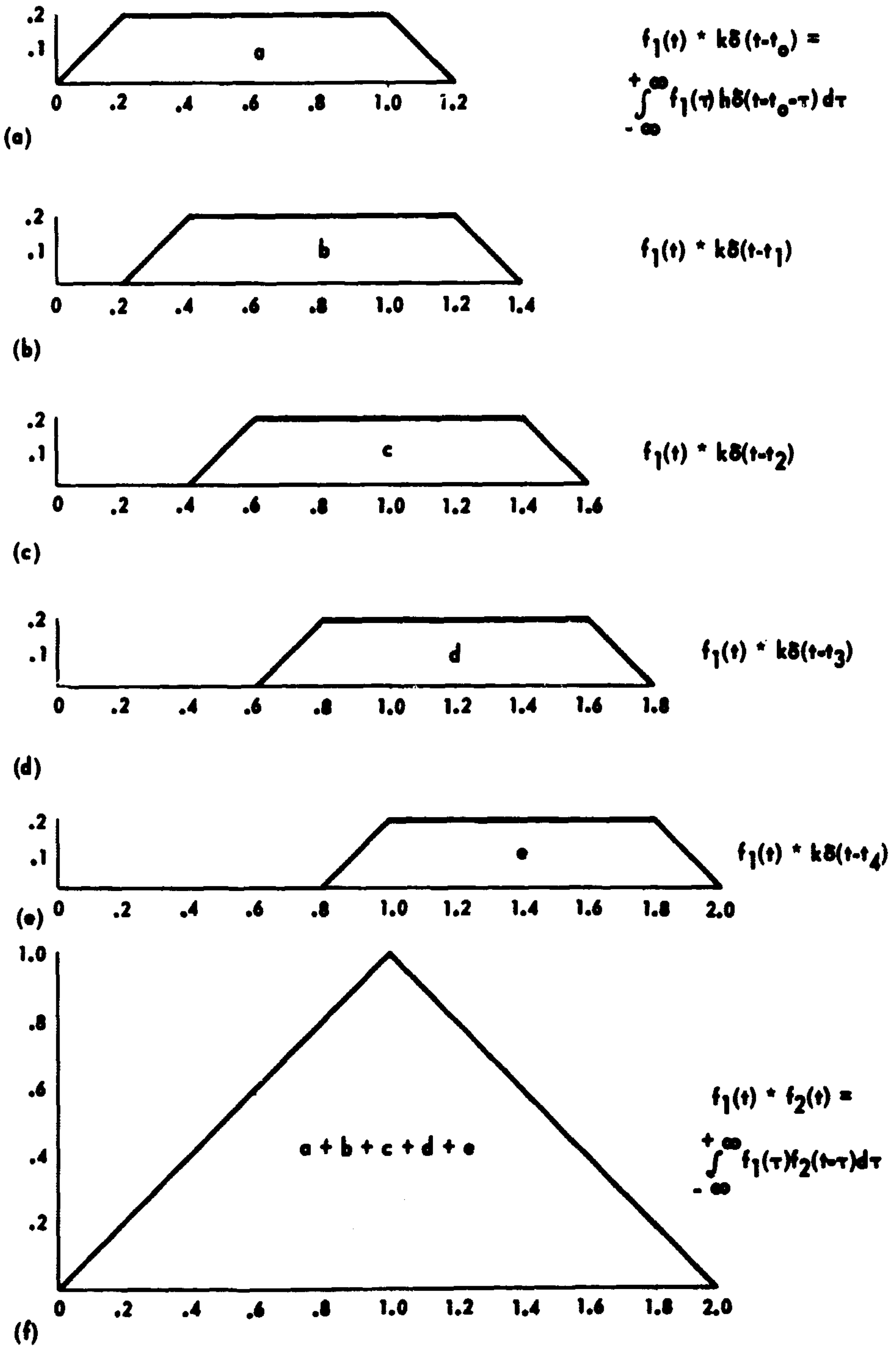


FIGURE 19

CONVOLUTION AND THE RESPONSE OF A LINEAR SYSTEM

The superposition properties of the convolution integral make it a most useful and convenient means of specifying the response of a linear system in terms of its impulse response and excitation function. In fact, a very important theorem sometimes referred to as Borel's theorem states, "The response of a linear system to an arbitrary excitation is the convolution of the excitation function and the system impulse response." Therefore if the impulse response of the system is once determined, it may be convolved with any arbitrary excitation function to result in an expression for the system response. The importance of this statement is magnified by the fact that if there is no analytical expression for the excitation function, the integral can be evaluated graphically. Furthermore it is often convenient to partition the excitation function into a succession of impulses similar to the delta function except that the amplitudes are controlled according to the amplitude of the partitioned function. Convolving the system impulse function with each successive excitation impulse and superimposing the results gives the desired system response. The impulse response is most convenient and widely used to characterize the input-output relations of a system and is defined in the following way. If the input excitation of a system is the delta function $\delta(t)$, the output response is the impulse response $h(t)$. Sometimes the impulse response is called Green's function (by the field theorist) or the spread function (by the optical physicist).



$f_1(t) * k\delta(t-t_o) = \int_{-\infty}^{+\infty} f_1(\tau)\, h\delta(t-t_o-\tau)\, d\tau$
(a)
$f_1(t) * k\delta(t-t_1)$
(b)
$f_1(t) * k\delta(t-t_2)$
(c)
$f_1(t) * k\delta(t-t_3)$
(d)
$f_1(t) * k\delta(t-t_4)$
(e)
$f_1(t) * f_2(t) = \int_{-\infty}^{+\infty} f_1(\tau) f_2(t-\tau)\, d\tau$
a + b + c + d + e
(f)

FIGURE 20

Borel's theorem will be illustrated but not proved, although a rather general proof may be given without undue difficulty. (The proof is based on the classical theory of second-order linear differential equations, a modicum of operational calculus, and the definition of Green's function.) In so doing, consider a linear system such that an input delta function $\delta(t)$ causes an output $h(t)$ as shown in Figure 21a. Since the system is said to be linear, the input and output could be represented by a differential equation with constant coefficients, and the invariance of the input and output relations would not be disturbed by a translation of time. Therefore if the input excitation is delayed by a time $t = \tau$, the output response is likewise delayed by a time $t = \tau$ as shown in Figure 21b (this function is that of $f_1(t)$ in Figure 19), and by definition the impulse response of the system is $h(t - \tau)$. Let an arbitrary input excitation $f_i(t)$ (identical with $f_2(t)$ in Figure 19) be applied to the input as shown in Figure 21c. According to Borel's theorem the output response $f_o(t)$ is the convolution of the excitation function $f_i(t)$ and the impulse response $h(t - \tau)$, and we may write

$$f_o(t) = f_i(t) * h(t) \tag{16}$$

which is

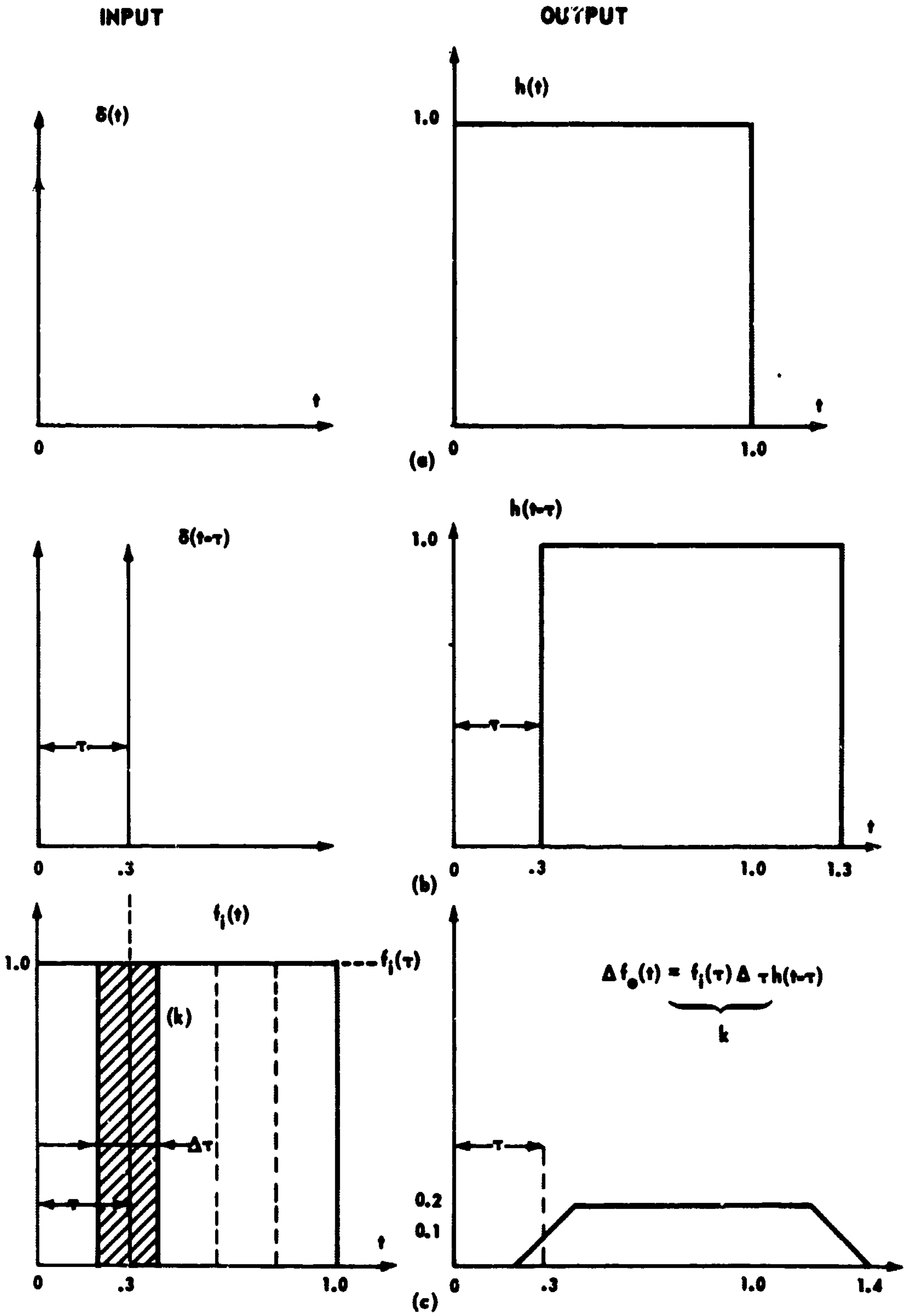
$$f_o(t) = \int_{-\infty}^{+\infty} f_i(\tau)\, h(t - \tau)\, d\tau \ . \tag{17}$$

A plot of the system response $f_o(t)$ is the same as that in Figure 20f.

To emphasize the principle of linear superposition embodied in the integral of equation (17), it will be developed with the aid of Figure 21. The input function $f_i(t)$ is partitioned in increments of width $\Delta\tau$ and height $f_i(\tau)$ where the height is $f_i(t)$ evaluated at the position τ of the given increment. As $\Delta\tau$ becomes smaller, the function $f_i(t)$ becomes smaller, the function $f_i(t)$ becomes more closely approximated by a series of adjacent impulses of strength $k = f_i(\tau)\Delta\tau$ which are positioned at $\delta(t - \tau)$; therefore a typical input pulse is

$$f_i(\tau)\, \Delta\tau\, \delta(t - \tau) \ .$$

If the increments are sufficiently small, the pulses approximate the unit impulse function (strength unity) so that, if the unit impulse function $\delta(t - \tau)$ results in a response $h(t - \tau)$, the incremental input impulse with strength $f_i(\tau)\Delta\tau$ located at $\delta(t - \tau)$ results in an incremental response (Figure 21c).



INPUT
OUTPUT
δ(t)
h(t)
1.0
t
0
1.0
(a)
δ(t-τ)
h(t-τ)
τ
.3
1.3
(b)
$f_i(t)$
$f_i(\tau)$
(k)
Δτ
$\Delta f_o(t) = \underbrace{f_i(\tau)\,\Delta\tau}_{k}\, h(t-\tau)$
0.2
0.1
1.4
(c)

FIGURE 21

$$\Delta f_o(t) = f_i(\tau)\Delta\tau\, h(t-\tau) \quad . \tag{18}$$

According to the principle of linear superposition, the response of a linear system to an excitation function $f_i(t)$ up to a given time t is the sum of its responses to all of the incoming impulses up to the time t. In other words, we may add up all of the contributions (response) to the present time to arrive at the total.

Therefore the sum of all the responses to the individual input impulses from t = 0 to the final time t is

$$f_o(t) = \sum_{t=0}^{t} f_i(\tau)\Delta\tau\, h(t-\tau) \quad . \tag{19}$$

It has been emphasized that as $\Delta\tau$ becomes smaller (amplitude is now controlled to correspond with the amplitude of the input function) $f_i(t)$ is more closely approximated by the series of adjacent impulses; therefore equation (19) more closely approximates the system response. In the limit as $\Delta\tau \rightarrow 0$ the expression

$$f_o(t) = \lim_{\Delta\tau \rightarrow 0} \sum_{t=0}^{t} f_i(\tau)\, h(t-\tau)\Delta\tau$$

becomes the integral

$$f_o(t) = \int_0^t f_i(\tau)\, h(t-\tau)\, d\tau \tag{20}$$

which is a convolution integral that defines the system response in terms of the impulse response and the input. Equation (20) is sometimes referred to as the superposition theorem.

Since the integral is zero for any value of t for which the product of the factors in the integrand is zero, nothing would be added to change the limits of the integral from $t = -\infty$ to $+\infty$ and equation (20) becomes

$$f_o(t) = \int_{-\infty}^{+\infty} f_i(\tau)\, h(t-\tau)\, d\tau \quad . \tag{21}$$

Equation (21) holds if $f_i(t)$ and $f_o(t)$ are periodic, aperiodic, or random.

CONVOLUTION THEOREM

From the standpoint of solving linear differential equations, the transfer function is the reciprocal of the characteristic function in the transform solution of all linear differential equations. From a systems point of view, it is simply the ratio of the response output transform to the excitation input transform and, in addition to the impulse response, is widely used to characterize the input-output relations of a linear system. The transfer function is discussed here in the context of arriving at the very important convolution theorem.

If the input to a linear system $f_i(t)$ Fourier transforms into $F_i(\omega)$ and the output response $f_o(t)$ Fourier transforms into $F_o(\omega)$; that is

$$\mathfrak{F}[f_i(t)] = F_i(\omega)$$

$$\mathfrak{F}[f_o(t)] = F_o(\omega) \quad ,$$

the transfer function $H(\omega)$ is by definition

$$H(\omega) = \frac{F_o(\omega)}{F_i(\omega)}$$

and the Fourier transform of the output response is

$$F_o(\omega) = H(\omega)\, F_i(\omega) \quad . \tag{22}$$

The inverse Fourier transform of equation (22) is the response of the system in terms of the transfer function and the input excitation

$$f_o(t) = \mathfrak{F}^{-1}[H(\omega) F_i(\omega)] \quad . \tag{23}$$

By comparing equations (23) and (21) we can write

$$\int_{-\infty}^{+\infty} f_i(\tau)\, h(t-\tau)\, d\tau = \mathfrak{F}^{-1}[H(\omega) F_i(\omega)] \tag{24a}$$

or equally

$$\mathfrak{F}\left[\int_{-\infty}^{+\infty} f_i(\tau)\, h(t-\tau)\, d\tau\right] = H(\omega) F_i(\omega) \quad . \tag{24b}$$

Equations (24a) and (24b) are a form of the convolution theorem which shows the equivalence of two powerful methods of arriving at the response of a linear system. Whether the system response is determined by the convolution of the input and the impulse response or by the inverse of the product of the input and the transfer function transforms is a matter of choice.

The convolution theorem is one of the most widely used tools in frequency analysis and is worthy of a more general statement and proof. Before doing so, however, the shifting theorem will be reviewed because it will be used in a very straightforward proof. According to the shifting theorem, if the Fourier transform of $f_2(t)$ is

$$\mathfrak{F}[f_2(t)] = \int_{-\infty}^{+\infty} f_2(t)\, e^{-i\omega t} dt = F_2(\omega) \tag{25}$$

then the Fourier transform of $f_2(t-\tau)$ is

$$\mathfrak{F}[f_2(t-\tau)] = e^{-i\omega\tau} F_2(\omega) \; . \tag{26}$$

Equation (26) states that a shift of τ in the time domain is equivalent to the multiplication by the factor $e^{-i\omega\tau}$ in the frequency domain. To prove equation (26), it is only necessary to perform the operation indicated on the left-hand side of the equation which is

$$\mathfrak{F}[f_2(t-\tau)] = \int_{-\infty}^{+\infty} f_2(t-\tau)\, e^{-i\omega t}\, dt \tag{27}$$

and make the substitution

$$\eta = t - \tau$$

so that

$$\mathfrak{F}[f_2(t-\tau)] = \int_{-\infty}^{+\infty} f_2(\eta)\, e^{-i\omega(\eta+\tau)}\, d\eta \quad .$$

The exponential factor in the integrand contains a factor $e^{-i\omega\tau}$ which does not contain the variable of integration and may be moved outside the integral,

$$\mathfrak{F}[f_2(t-\tau)] = e^{-i\omega\tau} \int_{-\infty}^{+\infty} f_2(\eta)\, e^{-i\omega\eta}\, d\eta \quad .$$

The integral is clearly the Fourier transform of $f_2(\eta)$ and is identical in form with the Fourier transform of $f_2(t)$ in equation (25); therefore we may write

$$\mathfrak{F}[f_2(t-\tau)] = e^{-i\omega\tau}\, F_2(\omega) \tag{28}$$

which is proof of the shifting theorem.

Now it is convenient to state and prove the convolution theorem. If $f_1(t)$ transforms into $F_1(\omega)$ and $f_2(t)$ transforms into $F_2(\omega)$; that is, if

$$\mathfrak{F}[f_1(t)] = F_1(\omega)$$

and

$$\mathfrak{F}[f_2(t)] = F_2(\omega)$$

then

$$\mathfrak{F}\left[\int_{-\infty}^{+\infty} f_1(\tau)\, f_2(t-\tau)\, d\tau\right] = F_1(\omega)\, F_2(\omega) \quad . \tag{29}$$

To prove that equation (29) is true, it is necessary to perform the operation indicated on the left, and equation (29) becomes

$$\int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} f_1(\tau)\, f_2(t - \tau)\, d\tau \right] e^{-i\omega t}\, dt = F_1(\omega)\, F_2(\omega) \ .$$

To use the shifting theorem, the expression within the brackets of the above equation may be rearranged so that

$$\int_{-\infty}^{+\infty} f_1(\tau) \left[\int_{-\infty}^{+\infty} f_2(t - \tau)\, e^{-i\omega t}\, dt \right] d\tau = F_1(\omega)\, F_2(\omega) \ .$$

The expression within the brackets in the integral on the left-hand side of the above equation is the Fourier transform of the function $f_2(t - \tau)$ of equations (26) and (27); therefore

$$\int_{-\infty}^{+\infty} f_1(\tau)\, F_2(\omega)\, e^{-i\omega\tau}\, d\tau = F_1(\omega)\, F_2(\omega) \ .$$

The factor $F_2(\omega)$ in the integrand of the above integral does not contain the variable of integration and may be moved outside.

$$F_2(\omega) \int_{-\infty}^{+\infty} f_1(\tau)\, e^{-i\omega\tau}\, d\tau = F_1(\omega)\, F_2(\omega) \ .$$

The integral is now the Fourier transform of $f_1(\tau)$, which is of the form $f_1(t)$, and the result is

$$F_1(\omega)\, F_2(\omega) = F_1(\omega)\, F_2(\omega)$$

which proves the convolution theorem.

BIBLIOGRAPHY

Cheng, David K.: Analysis of Linear Systems. Addison-Wesley, 1959.

Jennison, R. C.: Fourier Transforms and Convolutions for the Experimentalist. Pergamon, 1961.

Lathi, B. P.: Communication Systems. John Wiley & Sons, 1968.

Lee, Y. W.: Statistical Theory of Communications. John Wiley & Sons, 1960.

O'Neill, Edward L.: Introduction to Statistical Optics. Addison-Wesley, 1963.

Sneddon, Ian N.: Fourier Transforms. McGraw-Hill, 1951.

APPROVAL

A GRAPHICAL APPROACH TO CONVOLUTION

By James C. Taylor

The information in this report has been reviewed for security classification. Review of any information concerning Department of Defense or Atomic Energy Commission programs has been made by the MSFC Security Classification Officer. This report, in its entirety, has been determined to be unclassified.

This document has also been reviewed and approved for technical accuracy.

W. Haeussermann

W. HAEUSSERMANN
Director, Astrionics Laboratory