

CONVOLUTION: Digital Signal Processing

National Semiconductor
Application Note 237
January 1980



Introduction

As digital signal processing continues to emerge as a major discipline in the field of electrical engineering, an even greater demand has evolved to understand the basic theoretical concepts involved in the development of varied and diverse signal processing systems. The most fundamental concepts employed are (not necessarily listed in the order of importance) the sampling theorem^[1], Fourier transforms^[2] ^[3], convolution, covariance, etc.

The intent of this article will be to address the concept of convolution and to present it in an introductory manner hopefully easily understood by those entering the field of digital signal processing.

It may be appropriate to note that this article is Part II (Part I is titled "An Introduction to the Sampling Theorem") of a series of articles to be written that deal with the fundamental concepts of digital signal processing.

Let us proceed . . .

Part II Convolution

Perhaps the easiest way to understand the concept of convolution would be an approach that initially clarifies a subject relating to the frequency spectrum of linear networks.

Determining the frequency spectrum or frequency transfer function of a linear network provides one with the knowledge of how a network will respond to or alter an input signal. Conventional methods used to determine this entail the use of spectrum analyzers which use either sweep generators or variable-frequency oscillators to impress upon a network all possible frequencies of equal amplitude and equal phase.

The response of a network to all frequencies can thus be determined. Any amplitude and phase variations at the output of a network are due to the network itself and as a result define the frequency transfer function.

Another means of obtaining this same information would be to apply an impulse function to the input of a network and then analyze the network impulse-response for its spectral-frequency content. Comparison of the network-frequency transfer function obtained by the two techniques would yield the same information.

This is found to be easily understood (without elaborate experimentation) if the implications of the impulse function are initially clarified.

If the pulse of *Figure 1a* is examined, using the Fourier integral, its frequency spectrum is found to be

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \quad (1)$$

$$= \int_{-T/2}^{T/2} A e^{-j\omega t} dt \quad (2)$$

$$F(\omega) = AT \left[\frac{\sin\left(\frac{\omega T}{2}\right)}{\left(\frac{\omega T}{2}\right)} \right]$$

as shown in *Figure 1b*.

Decreasing the pulse width while increasing the pulse height to allow the area under the pulse to remain constant, *Figure 1c*, shows from eq(1) and eq(2) the bandwidth or spectral-frequency content of the pulse to have increased, *Figure 1d*.

Further altering the pulse to that of *Figure 1e* provides for an even broader bandwidth, *Figure 1f*. If the pulse is finally altered to the limit, i.e., the pulsewidth being infinitely narrow and its amplitude adjusted to still maintain an area of unity under the pulse, it is found in 1g and 1h the unit impulse produces a constant, or "flat" spectrum equal to 1 at all frequencies. Note that if $AT = 1$ (unit area), we get, by definition, the unit impulse function in time.

Since this time function contains equal frequency components at all frequencies, applying it or a good approximation of it to the input of a linear network would be the equivalent of simultaneously impressing upon the system an array of oscillators inclusive of all possible frequencies, all of equal amplitude and phase. The frequencies could thus be determined from this one input time function. Again, variations in amplitude and phase at the system output would be due to the system itself.

Empirically speaking the frequency spectrum or the network frequency transfer function can thus be determined by applying an impulse at the input and using, for example, a spectrum analyzer at the network output. At this point, it is important to emphasize that the above discussion holds true for only linear networks or systems since the superposition principle (The response to a sum of excitations is equal to the sum of the responses to the excitations acting separately), and its analytical techniques break down in non-linear networks.

Since an impulse response provides information of a network frequency spectrum or transfer function, it additionally provides a means of determining the network response to any other time function input. This will become evident in the following development.

If the input to a network, *Figure 2*, having a transfer function $H(\omega)$ is an impulse function $\delta(t)$ at $t=0$, its Fourier transform using eq(1) can be found to be $F(\omega) = 1$.

The output of the network $G(\omega)$ is therefore

$$G(\omega) = H(\omega) \cdot F(\omega)$$

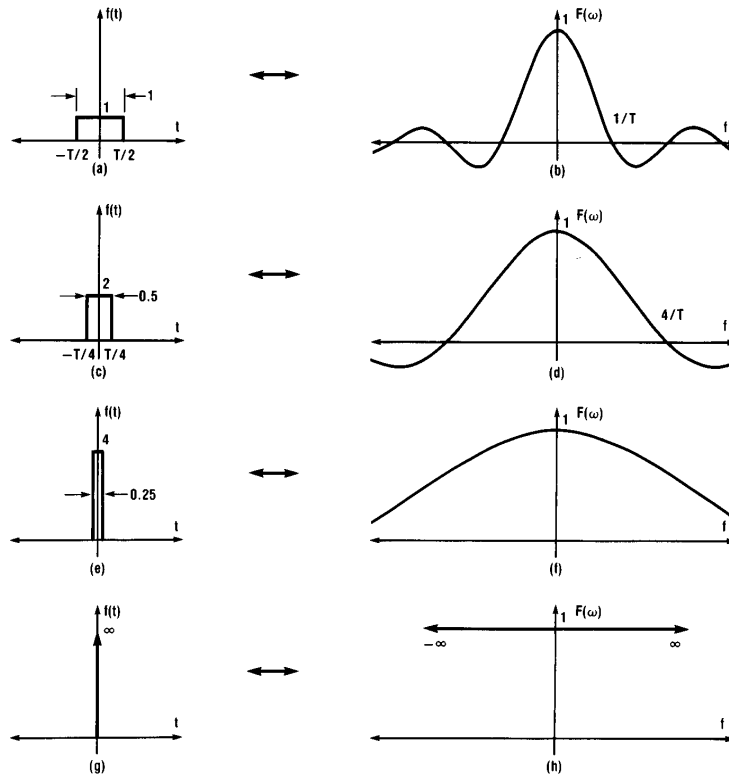
$$G(\omega) = H(\omega)$$

The inverse transform is

$$g(t) = h(t)$$

and $h(T)$ is defined as the impulse response of the network as a result of being excited by a unit impulse time function at $t=0$.

Extending this train of thought further, the response of a network to any input excitation can be determined using the same technique.



TL/H/5621-1

**FIGURE 1. Development of a unit impulse;
(a) (c) (e) (g) its time function
(b) (d) (f) (h) its frequency spectrum**

Hence, finding the Fourier transform of the input excitation, $F(\omega)$, multiplying it by the transfer function transform $H(\omega)$ (or the transform of time domain network impulse response) and inverse transforming to find the output $g(t)$ as a function of time.

By definition the convolution integral¹

$$f(t) * h(t) = \int_0^t f(\tau) h(t - \tau) d\tau \quad (3)$$

(where * denotes the convolution operation, $h(t)$ denotes the impulse response function described above and both $f(t)$ and $h(t)$ are zero for $t < 0$. Note that the meaning of the variables t and τ will be clarified, later in the article) makes the same claim but in the realm of the time domain alone.

If this is true then the Fourier transform of the convolution integral eq(3) should have the following equivalence –

$$F \left[\int_0^t f(\tau) h(t - \tau) d\tau \right] = F(\omega) \bullet H(\omega) \quad (4)$$

As a proof using eq(1) let

$$F \left[f(t) * h(t) \right] = \int_0^\infty e^{-j\omega t} \left[\int_0^t f(\tau) h(t - \tau) d\tau \right] dt \quad (5)$$

Defined by the shifted step function

$$u(t - \tau) = 1 \text{ for } \tau \leq t \quad (6a)$$

and

$$u(t - \tau) = 0 \text{ for } \tau > t \quad (6b)$$

Footnote:

1. It is important to note that the convolution integral is commutative. This implies the reversability of the $f(t)$ and $h(t)$ terms in the definition.

$$g(t) = \int_{-\infty}^\infty f(\tau) h(t - \tau) d\tau = F^{-1}[F(\omega) \bullet H(\omega)]$$

$$\int_{-\infty}^\infty f(t - \tau) h(\tau) d\tau = F^{-1}[H(\omega) \bullet F(\omega)]$$

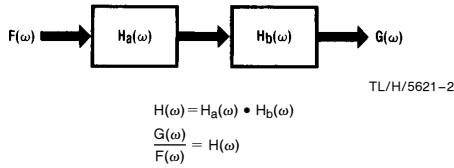


FIGURE 2. Block diagram of a network transfer function

the following identity can be made

$$\int_0^t f(\tau) h(t - \tau) d\tau = \int_0^\infty f(\tau) h(t - \tau) u(t - \tau) d\tau \quad (7)$$

Rewriting eq(5) as

$$F[f(t) * h(t)] = \int_0^\infty \epsilon^{-j\omega t} \int_0^\infty f(\tau) h(t - \tau) u(t - \tau) d\tau dt \quad (8)$$

and letting $x = t - \tau$ so that

$$\epsilon^{-j\omega t} = \epsilon^{-j\omega(x + \tau)} \quad (9)$$

eq(8) finally becomes

$$\begin{aligned} F[f(t) * h(t)] &= \int_0^\infty \int_0^\infty f(\tau) h(x) u(x) \epsilon^{-j\omega t} \epsilon^{-j\omega x} d\tau dx \\ &= \int_0^\infty h(x) u(x) \epsilon^{-j\omega x} dx \int_0^\infty f(\tau) \epsilon^{-j\omega \tau} d\tau \end{aligned}$$

$$F[f(t) * h(t)] = H(\omega) \bullet F(\omega) \quad (10)$$

which is the equivalent of eq(4).

In essence the above proof describes one of the most important and powerful tools used in signal processing . . . the convolution theorem. In words,

Convolution Theorem:

The convolution theorem allows one to mathematically convolve in the time domain by simply multiplying in the frequency domain. That is, if $f(t)$ has the Fourier transform $F(\omega)$ and $x(t)$ has the Fourier transform $X(\omega)$, then the convolution $f(t) * x(t)$ has the Fourier transform $F(\omega) \bullet X(\omega)$.

For the time convolution

$$f(t) * x(t) \longleftrightarrow F(\omega) \bullet X(\omega) \quad (11)$$

and the dual frequency convolution is

$$f(t) \bullet x(t) \longleftrightarrow F(\omega) * X(\omega) \quad (12)$$

Convolutions are fundamental to time series sampled data analysis. First of all, as described earlier all linear networks can be completely characterized by their impulse response functions and furthermore the response to any input is given by its (the input function) convolution with the network impulse response function. Digit filters being linear systems accomplish the filtering task using convolutions. A network or filter transfer function for example can be represented by its impulse response in the form of a Fourier series. A filtered input excitation response can then be found by convolving the input time function with the network Fourier series or impulse response. With the aid of a high speed computer the same result could be obtained by storing the FFT (Fast Fourier Transform) of the network impulse response into memory, performing an FFT on the sampled continuous in-

put excitation function, multiplying the two transforms and finally computing the inverse FFT of the product.

Moving averages and smoothing operations can further be characterized as lowpass filtering functions and can additionally be implemented using convolution. The above are just a few of the many operations convolution performs and the remainder of this discussion will focus on how convolution is realized.

To start with, an illustrative analysis will be performed assuming continuous functions followed by one performed in discrete form similar to that realized in computer aided sampled-data systems techniques.

As an example, if it were desired to determine the response of a network to the excitation pulse $f(t)$ shown in *Figure 3a*, knowing the network impulse $h_\delta(t)$, *Figure 3b*, the impulse response of an RC network, would allow one to determine the output $g(t)$ using the convolution integral, eq(3).

The convolution of $f(t)$ and $h_\delta(t)$

$$f(t) = 10[u(t) - (u(t - T_o))] \quad (13)$$

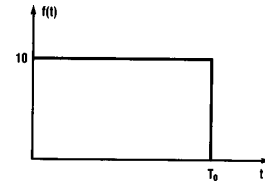
$$h_\delta(t) = \epsilon^{-at}$$

could be obtained by first substituting the dummy variable $t - \tau$ for t in $h_\delta(t)$ so that

$$h_\delta(t - \tau) = \epsilon^{-a(t - \tau)} \quad (15)$$

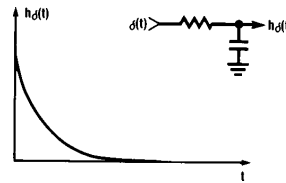
By definition $g(t) = f(t) * h_\delta(t)$ thus becomes

$$\int_0^t f(\tau) h_\delta(t - \tau) d\tau = \int_0^t 10[u(t) - u(t - T_o)] \epsilon^{-a(t - \tau)} d\tau \quad (16)$$



$$f(t) = 10[u(t) - u(t - T_o)]$$

(a)



$$h_\delta(t) = \epsilon^{-at}$$

(b)

**FIGURE 3. (a) rectangular pulse excitation
(b) impulse response of a single RC network**

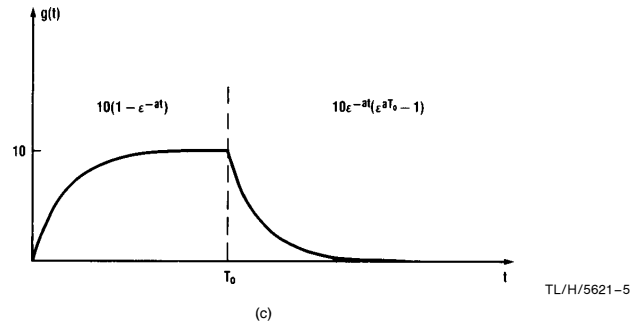


FIGURE 3. (c) output or convolution of the network (b) excited by (a)

Since the piecewise nature of the excitation makes it convenient to calculate the response in corresponding pieces the output is found to be

$$\begin{aligned} 0 < t \leq T_0 \\ g(t) &= f(t) h_\delta(t) = \int_0^t 10e^{-a(t-\tau)} d\tau \\ &= 10(1 - e^{-at}) \end{aligned} \quad (17)$$

$$\begin{aligned} t \geq T_0 \\ g(t) &= f(t) * h_\delta(t) = \int_0^{T_0} 10e^{-a(t-\tau)} d\tau \\ &= 10e^{-at} (e^{aT_0} - 1) \end{aligned} \quad (18)$$

The output response $g(t)$ is plotted in *Figure 3c* and is clearly what might be expected from a simple RC network excited by a rectangular pulse.

Though simplistic in its nature, the analysis of the above example quickly becomes unrealistically cumbersome when complex excitation and impulse response functions are used. Turning to a numerical evaluation of the convolution integral may perhaps be the most desirable method of realization. Prior to a numerical development however, an intuitive graphical illustration of convolution will be presented which should make discrete numeric convolution easily understood.

The convolution integral —

$$\int_0^t f(\tau) h_\delta(t - \tau) d\tau$$

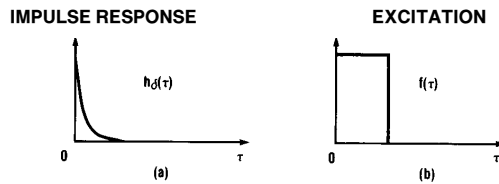


FIGURE 4. a) $h_\delta(\tau)$: network impulse response
b) $f(\tau)$: excitation function

defines the graphical procedure. Using the same example depicted in *Figure 3* the excitation and impulse response functions replaced with the dummy variable is defined as past data or historical information to be used in a convolution process. Thus

$$f(\tau) = 10[u(\tau) - u(\tau - T_0)] \quad (19)$$

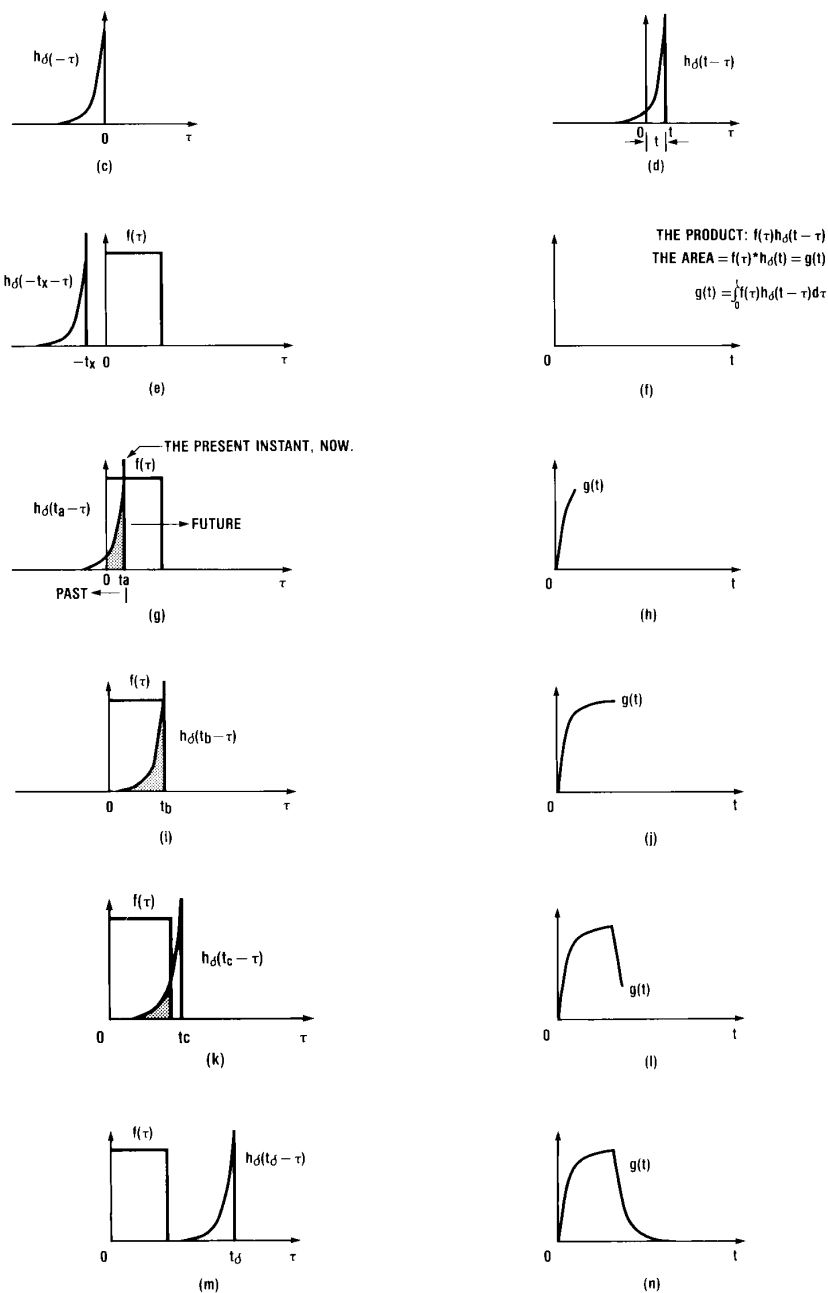
and

$$h_\delta(\tau) = e^{-a\tau} \quad (20)$$

are shown in *Figure 4a* and *b*. *Figure 4c*, $h_\delta(-\tau)$, represents the impulse response folded over [mirror image of $h_\delta(\tau)$] about the ordinate and *Figure 4d*, $h_\delta(t - \tau)$, is simply the function $h_\delta(-\tau)$ time shifted by the quantity t .

Evaluation of the convolution integral is performed by multiplying $f(\tau)$ by each incremental shift in $h_\delta(t - \tau)$. It is understood in *Figure 4e* that a negative value of $-t$ produces no output. For $t > 0$ however as the present time t varies, the impulse response $h_\delta(t - \tau)$ scans the excitation function $f(\tau)$, always producing a weighted sum of past inputs and weighing most heavily those values of $f(\tau)$ closest to the present. As seen in *Figures 4e* through *4n*, the response or output of the network at anytime t is the integral of the functions or calculated shaded area under the curves. In terms of the superposition principle the filter response $g(t)$ may be interpreted as being the weighted superposition of past input $f(\tau)$ values weighted or multiplied by $h_\delta(t - \tau)$.

An extension of the continuous convolution to its numerical discrete form is made and shown in *Figure 5*. Again the excitation and impulse response of *Figure 3* are used and are further represented as two finite duration sequences $f(n)$



**FIGURE 4. cont'd c) $h_d(-\tau)$: $h_d(\tau)$ folded about the ordinate
 d) $h_d(t-\tau)$: $h_d(\tau)$ folded and shifted
 e) through n) the output response $g(t)$ of the network whose
 impulse response $h_d(\tau)$ is excited by a function $f(\tau)$.
 Or the convolution, $f(\tau) * h_d(t)$, of $f(t)$ with $h_d(t)$.**

TL/H/5621-7

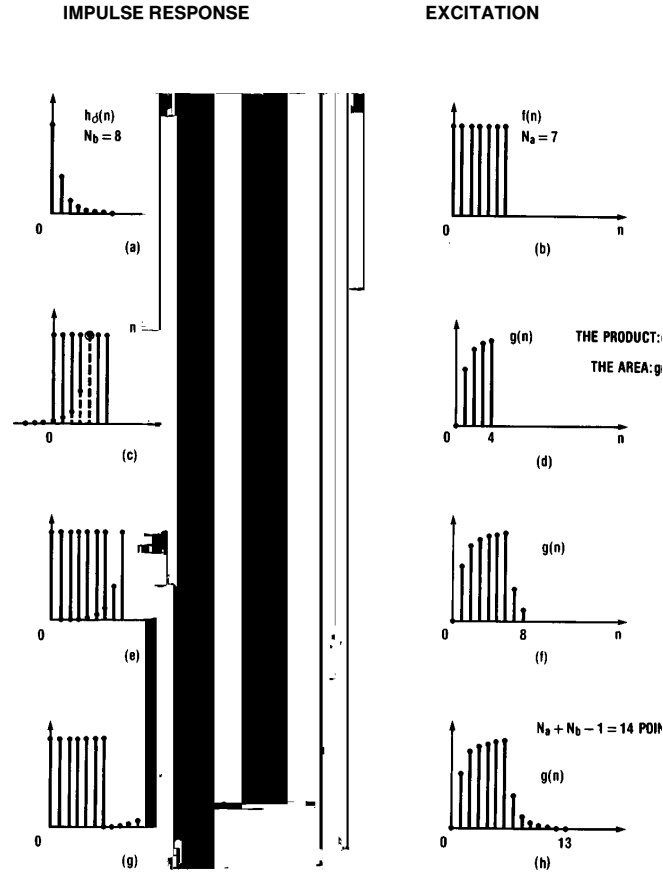


FIGURE 5. Illustrative description of discrete convolution

TL/H/5621-8

and $h_d(n)$ respectively, Figures 5a and b.

It is observed additionally that the duration of $f(n)$ is $N_a = 7$ samples [$f(n)$ is nonzero for the interval $0 \leq n \leq N_a - 1$ and the duration of $h_d(n)$ is $N_b = 8$ samples [$h_d(n)$ is nonzero for the interval $0 \leq n \leq N_b - 1$]. The sequence $g(n)$, a discrete convolution, can thus be defined as

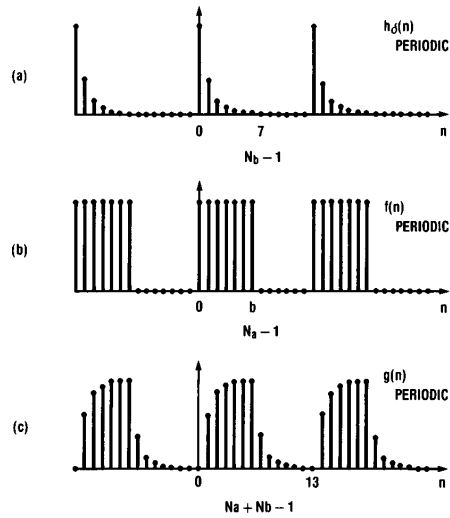
$$g(n) = \sum_{x=0}^n f(x) h_d(n-x) \quad (21)$$

having a finite duration sequence of $N_a + N_b - 1$ samples, Figure 5h. The convolution using numerical integration (area under the curve) can be defined as

$$g(n)T = T \sum_{x=0}^n f(x) h_d(n-x) \quad (22)$$

where T is the sampling interval used to obtain the sampled data sequences.

If $f(n)$ and $h_d(n)$ were next considered to be periodic sequences and a convolution was desired using either shifting techniques or performing an FFT on the excitation and impulse response sequences and finally inverse FFT transforming to achieve the output response, some care must be taken when preparing the convolving sequences. From Figure 5h it is observed that the convolution is completed in a $N_a + N_b - 1$ point sequence. To acquire the nonoverlapping or nondistorted periodic sequence of Figure 6c the convolution thus requires $f(n)$ and $h_d(n)$ to be $N_a + N_b - 1$ point sequences. This is achieved by appending the appropriate number of zero valued samples, also known as zero filling, to $f(n)$ and $h_d(n)$ to make them both $N_a + N_b - 1$ point sequences. The undistorted and correct convolution can now be performed using the zero filled sequences Figure 6a and 6b to achieve 6c.



TL/H/5621-9

FIGURE 6. Linear periodic discrete convolution of $f(n)$ and $h_d(n)$, $f(n) * h_d(n)$.

A Final Note

This article attempted to simplify the not-so-obvious concept of convolution by first developing the readers knowledge and feel for the implications of the impulse function and its effect upon linear networks. This was followed by a short discussion of network transfer functions and their relative spectrum. Having set the stage, the convolution integral and theorem were introduced and supported with an analytical and illustrative example. This example showed how the response of a simple RC network excited by a rectangular pulse could be determined using the convolution integral.

Finally, two examples of discrete convolution were presented. The first example dealt with finite duration sequences and the second dealt with periodic sequences. Additionally, precautions in the selection on n -point sequences was discussed in the second example to alleviate distorting or spectrally overlapping the excitation and impulse response functions during the convolution process.

Appendix A

1. Carson Chen, *An Introduction to the Sampling Theorem*, National Semiconductor Corporation.
2. A. Papoulis, *The Fourier Integral and Its Applications*, McGraw-Hill, 1962.
3. E. O. Brigham, *The Fast Fourier Transform*, Prentice-Hall, 1979.
4. L. Enochson, R. K. Otnes, *Applied Time Series Analysis*, John Wiley, 1978.
5. B. Gold, C. M. Rader, *Digital Processing of Signals*, McGraw-Hill, 1969.
6. F. F. Kuo, *Network Analysis and Synthesis*, John Wiley, 1966.
7. R. W. Hamming, *Digital Filters*, Prentice-Hall, 1977.
8. B. Gold, L. R. Rabiner, *Theory and Application of Digital Signal Processing*, Prentice-Hall, 1975.
9. M. Schwartz, *Information Transmission, Modulation, and Noise*, McGraw-Hill, 1970.
10. S. D. Stearns, *Digital Signal Analysis*, Haysen, 1975.
11. E. Kreyszig, *Advanced Engineering Mathematics*, John-Wiley, 1979.
12. S. A. Tretter, *Introduction to Discrete-Time Signal Processing*, John-Wiley, 1976.
13. A. Papoulis, *Signal Analysis*, McGraw-Hill, 1977.
14. M. Schwartz, L. Shaw, *Signal Processing: Discrete, Spectral Analysis, Detection, and Estimation*, McGraw-Hill, 1975.
15. W. D. Stanley, *Digital Signal Processing*, 1975.
16. D. F. Tuttle, Jr., *Circuits*, McGraw-Hill, 1977.
17. R. C. Agawal, C. S. Burrus, *Number Theoretical Transforms to Implement Fast Digital Convolution*, Proc. IEEE, Vol. 63, pp. 550–560, April 1975.
18. J. W. Cooley, P. A. W. Lewis, P. D. Welch, *Application of the Fast Fourier Transform to Computation of Fourier Integrals, Fourier Series, and Convolution Integrals*, IEEE Trans, Audio Electroacoustics, Vo. Au-15, pp. 79–84, June 1967.

Acknowledgements

The author wishes to thank James Moyer and Barry Siegel for their guidance and constructive criticisms.

LIFE SUPPORT POLICY

NATIONAL'S PRODUCTS ARE NOT AUTHORIZED FOR USE AS CRITICAL COMPONENTS IN LIFE SUPPORT DEVICES OR SYSTEMS WITHOUT THE EXPRESS WRITTEN APPROVAL OF THE PRESIDENT OF NATIONAL SEMICONDUCTOR CORPORATION. As used herein:

1. Life support devices or systems are devices or systems which, (a) are intended for surgical implant into the body, or (b) support or sustain life, and whose failure to perform, when properly used in accordance with instructions for use provided in the labeling, can be reasonably expected to result in a significant injury to the user.
2. A critical component is any component of a life support device or system whose failure to perform can be reasonably expected to cause the failure of the life support device or system, or to affect its safety or effectiveness.



National Semiconductor Corporation
1111 West Bardin Road
Arlington, TX 76017
Tel: 1(800) 272-9959
Fax: 1(800) 737-7018

National Semiconductor Europe
Fax: (+49) 0-180-530 85 86
Email: cnjwge@tevm2.nsc.com
Deutsch Tel: (+49) 0-180-530 85 85
English Tel: (+49) 0-180-532 78 32
Français Tel: (+49) 0-180-532 93 58
Italiano Tel: (+49) 0-180-534 16 80

National Semiconductor Hong Kong Ltd.
19th Floor, Straight Block,
Ocean Centre, 5 Canton Rd.
Tsimshatsui, Kowloon
Hong Kong
Tel: (852) 2737-1600
Fax: (852) 2736-9960

National Semiconductor Japan Ltd.
Tel: 81-043-299-2309
Fax: 81-043-299-2408

National does not assume any responsibility for use of any circuitry described, no circuit patent licenses are implied and National reserves the right at any time without notice to change said circuitry and specifications.