DIGITAL FILTERS

Sanjit K. Mitra*
Visiting Scientist
STANFORD LINEAR ACCELERATOR CENTER
Stanford University
Stanford, California

PREPARED FOR THE U. S. ATOMIC ENERGY
COMMISSION UNDER CONTRACT NO.

* Department of Electrical Engineering, University of California, Davis, California 95616

1. INTRODUCTION

The term *filtering* is used to denote the process of spectrum shaping either in the frequency domain or in the time domain. There are basically two different ways the filtering can be achieved. The more commonly used means is to use analog components like inductors, capacitors, amplifiers as the basic elements of the filter. Such filters are known as *analog* or continuous filters. Alternately, the signal processing can be done by using digital components like gates, flip-flops as the basic elements. The latter type of filters are called the *digital filters*. In a continuous filter the input and output waveforms are continuous functions of time, whereas in a digital filter the input and output waveforms are sampled signals, i.e., discrete functions of time (see Figure 1). Strictly speaking, the term digital filter represents the "computational process or algorithm by which a sampled signal or a sequence of numbers is transformed into a desired sequence of numbers"[1]. This transformation is assumed to be a linear operation. Applications of digital filtering techniques include computer simulation of linear dynamic and continuous systems like speech communication systems, processing of data signals like geophysical data in a computer. It can be pointed out here that recent developments in monolithic integrated digital circuits indicate the possibility of eventual replacement of ananlog hardware by real time digital filtering systems of lower cost and size and of greater flexibility.

The purpose of this document is to present the fundamentals of digital filters. This report is based on the calssroom notes the author prepared for a series of lectures given at SLAC during August, 1969.

2. PRINCIPLE OF OPERATION

The principle of operation of a digital filter can be explained with reference to Figure 1, where $\hat{x}(t)$ denotes the set of input samples occurring at time intervals of T seconds, and $\hat{y}(t)$ denotes the set of output samples occurring at time intervals of T seconds:

$$\begin{aligned}\hat{x}(t) &= \{x(0),\ x(T),\ x(2T),\ x(3T),\ldots\} \\ \hat{y}(t) &= \{y(0),\ y(T),\ y(2T),\ y(3T),\ldots\}\end{aligned} \tag{1}$$

The output sample at any instant, in general is related to the input sample at that instant and also to the input and output samples at previous time intervals. Thus we can write for $t = nT$,

$$y(nT) = \sum_{i=0}^{m} A_i\ x(nT-iT) + \sum_{i=1}^{k} B_i\ y(nT-iT) \quad m < k \tag{2}$$

where A_i and B_i are constant coefficients. Hence, a kth order digital filter is described by a kth order difference equation. Equation (2) defines the digital filtering algorithm and as a result is suitable for computer implementation.

$\hat{x}(t)$ is physically obtained by sampling a continuous wave form $x(t)$ by means of a sampler (Figure 2). Real time digital filtering consists of performing the algorithm indicated by Equation (1) once for each arrival of a new input sample and completing the operation in less than T seconds. T is known as the sampling interval.

Even though the problem of designing a digital filter is not a very difficult one, there are many associated difficulties which may occur in an actual implementation of the design and as a result lead to errors. One source of error is the sampling process itself which in practice is not ideal. Thus the actual samples will have a finite width. Another source of error arises because of quantization of the samples and the coefficients A_i and B_i. Equation (2) also indicates that a multiplication process is involved with which are associated inherent round off errors caused by finite register lengths. Examination of these errors is beyond the scope of this report.

3. SAMPLING PROCESS

Since sampling is an essneital part of a digital filter design involving continuous input signals, let us examine in some details the sampling process. For our purpose we assume the sampler to be an ideal switch. This implies that it makes and breaks contact instantly, duration of sample is negligible and the sampler samples periodically every T seconds. These assumptions make the mathematics very simple without any loss of insight. The relation between a signal g(t) fed into a sampler and the output $\hat{g}(t)$ of the sampler is illustrated in Figure 3. Note that the height of sampled pulses is equal to the value of g(t) at that instant, i.e.,

$$\hat{g}(nT) = g(nT) \tag{3}$$

A convenient representation of $\hat{g}(t)$ is by means of a train of delta functions which is described next.

3.1 Delta Function

A delta function (or unit impulse function) $\delta(t)$ is defined as follows:

$$\delta(t) = 0 \qquad t \neq 0 \tag{4}$$

$$\int_{-a}^{b} \delta(t)\, dt = 1 \qquad a > 0 \qquad b > 0 \tag{5}$$

$\delta(t)$ is said to occur at $t = 0$. An impulse function occurring at $t = t_0$ is denoted by $\delta(t-t_0)$. Thus

$$\delta(t-t_0) = 0 \qquad t \neq t_0$$

$$\int_{t_0-\varepsilon}^{t_0+\varepsilon} \delta(t-t_0)\, dt = 1 \qquad \varepsilon > 0 \tag{6}$$

An important property of a delta function is that when modulated by another function $g(t)$, it in essence samples the function. More precisely

$$\int_{-\infty}^{\infty} \delta(t-t_0) g(t) dt = g(t_0) \tag{7}$$

Often the above property is loosely defined as follows:

$$\delta(t-t_0) g(t) = g(t_0) \tag{8}$$

In using Equation (8), it is to be understood that the sampling property of the delta function holds only under integration.

An extension of the above idea leads to a sequence of impulses $\delta_T(t)$ occurring at $t = nT$ where $n = -\infty, \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots +\infty$. The function $\delta_T(t)$ can be represented as:

$$\delta_T(t) = \sum_{n=-\infty}^{\infty} \delta(t-nT) \tag{9}$$

We can thus consider the output of a sampler as a sequence of pulses in which the input function g(t) modulates the impulse train:

$$g^*(t) = g(t)\, \delta_T(t) = g(t) \left[\sum_{n=-\infty}^{\infty} \delta(t-nT) \right]$$

$$= \sum_{n=-\infty}^{\infty} g(nT)\, \delta(t-nT) \tag{10}$$

3.2 Spectral Properties of a Sampled Signal

It is interesting to examine the spectral properties of $g^*(t)$. Now $\delta_T(t)$ being a periodic function of period T (sampling interval), it can be represented by a Fourier series:

$$\delta_T(t) = \sum_{k=-\infty}^{\infty} c_k \, e^{j2\pi kt/T} \tag{11}$$

where c_k is given by

$$c_k = \frac{1}{T} \int_{-T/2}^{T/2} \delta_T(t) \, e^{-j2\pi kt/T} \, dt$$

$$= \frac{1}{T} \int_{-T/2}^{T/2} \delta(t) \, e^{-j2\pi kt/T} \, dt = \frac{1}{T} \tag{12}$$

Hence an alternate way to write $g^*(t)$ would be to express it as:

$$g^*(t) = g(t) \, \delta_T(t)$$

$$= g(t) \left[\frac{1}{T} \sum_{k=-\infty}^{\infty} e^{j2\pi kt/T} \right]$$

$$= \frac{1}{T} \sum_{k=-\infty}^{\infty} g(t) \, e^{j2\pi kt/T} \tag{13}$$

Let us denote the Fourier transform of $g(t)$ by $G(j\omega)$:

$$G(j\omega) = \mathcal{F}\{g(t)\} = \int_{-\infty}^{\infty} g(t) e^{-j\omega t} dt \tag{14}$$

We now take the Fourier transform of both sides of Equation (13):

$$G^*(j\omega) = \mathcal{F}\{g^*(t)\} = \frac{1}{T} \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} g(t) e^{j2\pi kt/T} \, e^{-j\omega t} \, dt \tag{15}$$

Recall that by shifting theorem:

$$\mathcal{F}\{g(t)\, e^{jmt}\} = G[j(\omega-m)] \tag{16}$$

Therefore

$$G^*(j\omega) = \frac{1}{T} \sum_{k=-\infty}^{\omega} G(j\omega - j\frac{2\pi k}{T})$$

$$= \frac{1}{T} \sum_{k=-\infty}^{\infty} G[j(\omega - k\omega_o)] \tag{17}$$

where we have used the notation

$$\omega_o = \frac{2\pi}{T} \tag{18}$$

$f_o = 1/T$ is known as the <u>sampling frequency</u>. Equation (17) says that the effect of impulse sampling is to produce a sampled signal whose spectrum is given by a succession of spectra which are proportional to the original spectrum $G(j\omega)$ but shifted periodically by a frequency separation $k\omega_o$. This is illustrated in Figure 4 for a typical $G(j\omega)$.

An important conclusion can be made from above discussion: If $G(j\omega)$ is band-limited, i.e.

$$G(j\omega) = 0 \quad \text{for} \quad \omega > \omega_\alpha$$

then the original signal can be recovered (within a multiplicative constant) from $g^*(t)$ by passing $g^*(t)$ through an ideal low-pass filter of cutoff frequency $\frac{\omega_o}{2}$ provided $\omega_\alpha \leq \frac{\omega_o}{2}$.

Summarizing we can state:

<u>Sampling Theorem</u>

In order to recover a band limited signal, the sampling frequency ω_o should be equal to or larger than the twice the highest-frequency component of the input signal:

$$\omega_o \geq 2\omega_\alpha \tag{19}$$

i.e., sampling interval T should be less than or equal to $1/2f_\alpha$.

If sampling frequency ω_o does not satisfy Equation (19) then it is evident from Figure 4 that it would not be possible to recover the input signal because of overlaps. For simulation work this bandlimiting of input signal is most essential in order to minimize errors due to overlap of frequency spectrums. One way to achieve this is to insert a band limiter (low-pass filter) before the sampler. Figure 5 shows the general scheme for system simulation using digital filter representation for continuous system dynamics.

4. Z-TRANSFORM

In the analysis and design of continuous systems, the Laplace Transform is a valuable tool which enables the conversion of differential equations into algebraic form for easier manipulation. In a similar manner, the Z-Transform is found to be useful in the analysis and design of digital filters.

The development of Z-Transform [2] is described next. We have shown before in Equation (10) a sampled signal g*(t) can be represented by means of an amplitude modulated impulse train. Now if g(t) is a causal signal i.e., g(t) = 0 for t < 0, then Equation (10) can be rewritten as:

$$g^*(t) = \sum_{n=0}^{\infty} g(nT)\delta(t-nT) \tag{20}$$

Taking the Laplace transforms of both sides yields:

$$\begin{aligned} G^*(s) &= \mathcal{L}\{g^*(t)\} = \int_0^\infty g^*(t)e^{-st}dt \\ &= \sum_{n=0}^{\infty} g(nT)\mathcal{L}\{\delta(t-nT)\} \end{aligned} \tag{21}$$

But

$$\mathcal{L}\{\delta(t-nT)\} = \int_0^\infty \delta(t-nT)e^{-st}\,dt = e^{-nsT} \tag{22}$$

which when used in Equation (21) yields:

$$G^*(s) = \sum_{n=0}^{\infty} g(nT)\,e^{-nsT} \tag{23}$$

If we use the notation

$$z = e^{sT} \tag{24}$$

then Equation (23) becomes

$$\mathcal{Z}\{g(t)\} = G(z) = G^*(s)\Big|_{z=e^{sT}} = \sum_{n=0}^{\infty} g(nT)\,z^{-n} \tag{25}$$

$G(z)$ as defined above is known as the "Z-transform of $g^*(t)$." Note that the Laplace transforms of sampled signals are transcendental functions of the complex frequency variable s. Introduction of a new variable z defined by Equation (24) makes the Z-transform a rational function in z which simplifies mathematical manipulation.

Example 1: Let us determine the Z-transform of the following function:

$$g(t) = \begin{cases} e^{-at} & t > 0 \\ 0 & t < 0 \end{cases} \tag{26}$$

Note that the above function reduces to the unit step function $u(t)$ for $a = 0$. A convenient representation of the function of (26) is $e^{-at}\,u(t)$. Now

$$\mathcal{Z}\{e^{-at}u(t)\} = \sum_{n=0}^{\infty} e^{-anT}\,z^{-n}$$

$$= \frac{1}{1 - e^{-aT}\,z^{-1}} = \frac{z}{z - e^{-aT}} \tag{27}$$

From Equation (27) it can be seen

$$\mathcal{Z}\{u(t)\} = \frac{z}{z-1} \tag{28}$$

Since in a digital filter, input and output are sequence of numbers, it is convenient to use the Z-transform techniques for analysis and design of such filters.

4.1 Relation between s-plane and z-plane - Stability Conditions

It is profitable to examine the relationship between the s-plane and the z-plane. The transformation $z = e^{sT}$ transforms a strip in the s-plane into the entire z-plane. The left half-plane portion of the strip bounded by $+j\frac{\pi}{T}$ and $-j\frac{\pi}{T}$ (shown shaded in Figure 6a) is mapped into the interior of the unit circle in the z-plane (shaded portion of Figure 6b). The right half plane portion of the strip is mapped to the exterior of the unit circle. Since $G^*(s)$ is a periodic function, each successive strip as shown is mapped in a similar fashion. This can be easily seen from the inverse transformation,

$$s = \frac{1}{T}\ln z \tag{29}$$

which is multiple-valued with period $j\pi$. Thus a periodic function $G^*(s)$ is transformed into a non-periodic function.

Relating the stability results in the s-plane to that in the z-plane, we conclude:

For strict stability the poles of the system function in the z-plane must be situated inside the unit circle.

4.2 Evaluation of $G(z)$ from $G(s)$

Often it is necessary to determine the Z-transform from a prescribed $G(s)$. The desired technique is derived next. We recall

$$g^*(t) = g(t)\,\delta_T(t) \tag{10}$$

In practice, $g(t)$ is a "causal" function, i.e. $g(t) = 0$ for $t < 0$.

Then the function $\delta_T(t)$ in Equation (10) can be replaced by an impulse train $\delta_T'(t)$ which is also a causal function:

$$\delta_T'(t) = \sum_{n=0}^{\infty} \delta(t-nT) \tag{30}$$

Now

$$\mathcal{L}\{\delta_T'(t)\} = \Delta(s) = \sum_{n=0}^{\infty} e^{-nTs}$$

$$= \frac{1}{1-e^{-sT}} \tag{31}$$

From Equation (10),

$$\mathcal{L}\{g^*(t)\} = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} G(\lambda)\ \Delta(s-\lambda)d\lambda \tag{32}$$

by the convolution theorem in the time-domain. Using the expression for $\Delta(s)$ in Equation (32), we obtain

$$\mathcal{L}\{g^*(t)\} = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{G(\lambda)}{1-e^{-T(s-\lambda)}}\, d\lambda \tag{33}$$

Evaluation of the above integral is accomplished through contour integration by closing the path of integration either in the right half s-plane (which encloses the poles of $\Delta(s-\lambda)$) or in the left-half s-plane (which encloses the poles of $G(s)$). If $G(s)$ is a rational function, the number of poles is finite. Thus, it is convenient to close the path of integration to the left. Then,

$$\int_{c-j\infty}^{c+j\infty} \frac{G(\lambda)}{1-e^{-T(s-\lambda)}}\, d\lambda = 2\pi j \sum_{\text{all } p_k} \text{Residue} \left[\frac{G(\lambda)}{1-e^{-T(s-\lambda)}}\right]\Bigg|_{\lambda=p_k} \tag{34}$$

where $\lambda = p_k$ is a pole of $G(\lambda)$ and the summation above is carried out for all such poles (assuming simple poles only). From Equations (33) and (34), we easily obtain

$$G^*(s) = \sum_{\text{all } p_k} \text{Residue} \left[\frac{G(\lambda)}{1-e^{-T(s-\lambda)}}\right]\Bigg|_{\lambda=p_k} \tag{35}$$

Example 2. Suppose

$$G(s) = \frac{1}{(s+a)(s+b)} \tag{36}$$

Note that the poles of $G(s)$ are at $p_1 = -a$ and $p_2 = -b$.

$$\text{Residue} \left[\frac{G(\lambda)}{1-e^{-T(s-\lambda)}}\right]\Bigg|_{\lambda=-a} = \frac{(\lambda+a)}{(\lambda+a)(\lambda+b)[1-e^{-T(s-\lambda)}]}\Bigg|_{\lambda=-a}$$

$$= \frac{1}{(b-a)[1-e^{-sT}e^{-aT}]}$$

$$\text{Residue} \left[\frac{G(\lambda)}{1-e^{-T(s-\lambda)}}\right]\Bigg|_{\lambda=-b} = \frac{\lambda+b}{(\lambda+a)(\lambda+b)[1-e^{-T(s-\lambda)}]}\Bigg|_{\lambda=-b}$$

$$= \frac{1}{(a-b)[1-e^{-sT}\ e^{-bT}]}$$

Hence

$$G(z) = \frac{1}{b-a}\left[\frac{1}{1-e^{-aT}\ z^{-1}} - \frac{1}{1-e^{-bT}\ z^{-1}}\right]$$

$$= \frac{z^{-1}\ (e^{-aT} - e^{-bT})}{(b-a)(1-e^{-aT}\ z^{-1})(1-e^{-bT}\ z^{-1})} \tag{37}$$

For multiple poles, the above procedure is modified slightly[2].

4.3 Inverse Z-Transform

From a prescribed $G(z)$, the corresponding sampled time function $g^*(t)$ can be determined by using the inverse Z-transform:

$$g(nT) = \frac{1}{2\pi j} \oint G(z)\, z^{n-1}\, dz \tag{38}$$

where the path of integration encloses the origin.

Example 3. Let $g(t) = k^{t/T}$

This implies

$$g(nT) = k^n$$

Hence

$$G(z) = \sum_{n=0}^{\infty} k^n z^{-n} = \frac{z}{z-k}$$

Inverse transform of $G(z)$ yields

$$g(nT) = \frac{1}{2\pi j} \int \frac{z^{n-1}\, z}{z-k}\, dz$$

The contour integral can be solved by Cauchy's Residue theorem by taking a contour which encloses the pole at $z = k$:

$$\oint \frac{z^n}{z-k}\, dz = 2\pi j \quad \text{Residue} \left[\frac{z^n}{z-k}\right]\Bigg|_{z=k}$$

$$= 2\pi j \quad k^n$$

As a result

$$g(nT) = k^n$$

as expected.

If the interest is only on the first few terms of the pulse sequence, an alternate procedure in obtaining $g^*(t)$ from a rational $G(z)$ is as illustrated in the next example.

<u>Example</u> 4. Let

$$G(z) = \frac{z^2+2z+3}{2z^2+z+4}$$

Rewrite $G(z)$ as

$$G(z) = \frac{N(z^{-1})}{D(z^{-1})} = \frac{1+2z^{-1}+3z^{-2}}{2+z^{-1}+4z^{-2}}$$

Divide the numerator $N(z^{-1})$ of $G(z)$ by its denominator $D(z^{-1})$ in a long division:

$$
\begin{array}{r|l|l}
2 + z^{-1} + 4\,z^{-2} & 1 + 2\,z^{-1} + 3\,z^{-2} & \frac{1}{2} + \frac{3}{4}\,z^{-1} + \frac{1}{8}\,z^{-2} - \frac{25}{16}\,z^{-3} \\
 & 1 + \frac{1}{2}\,z^{-1} + 2\,z^{-2} & \qquad + \frac{17}{32}\,z^{-4} + \dots \\
\hline
\end{array}
$$

$$
\begin{array}{l}
\frac{3}{2}\,z^{-1} + z^{-2} \\
\frac{3}{2}\,z^{-1} + \frac{3}{4}\,z^{-2} + 3\,z^{-3} \\
\hline
\frac{1}{4}\,z^{-2} - 3\,z^{3} \\
\frac{1}{4}\,z^{-2} + \frac{1}{8}\,z^{-3} + \frac{1}{2}\,z^{-4} \\
\hline
-\frac{25}{8}\,z^{-3} - \frac{1}{2}\,z^{-4} \\
-\frac{25}{8}\,z^{-3} - \frac{25}{16}\,z^{-4} - \frac{25}{4}\,z^{-5} \\
\hline
\frac{17}{16}\,z^{-4} + \frac{25}{4}\,z^{-5}
\end{array}
$$

Hence

$$G(z) = \frac{1}{2} + \frac{3}{4} z^{-1} + \frac{1}{8} z^{-2} - \frac{25}{16} z^{-3} + \frac{17}{32} z^{-4} + ..$$

As a result

$$g^*(t) = \frac{1}{2} \delta(t) + \frac{3}{4} \delta(t-T) + \frac{1}{8} \delta(t-2T)$$

$$- \frac{25}{16} \delta(t-3T) + \frac{17}{32} \delta(t-4T) + \ldots$$

4.4 Properties of Z-Transform

Based on the definition as given by Equation (25), several useful properties of the Z-transform can be derived. A few of these are given next.

Linearity. Z-transform is a linear operation. Thus, if we define $\mathcal{Z}[g_1^*(t)] = G_1(z)$ and $[g_2^*(t)] = G_2(z)$, then

$$Z\{a\, g_1^*(t) + b\, g_2^*(t)\} = a\, G_1(z) + b\, G_2(z) \qquad (39)$$

Delayed Sequence. If we denote by $G(z)$, the Z-transform of $g^*(t)$, then the Z-transform of the delayed sequence $g^*(t-kT)$ is easily derived as shown below.

$$\mathcal{Z}\{g^*(t-kT)\} = \sum_{n=o}^{\infty} g(nT-kT)\, z^{-n}$$

$$= \sum_{n=k}^{\infty} g(nT-kT)\, z^{-n}$$

$$= z^{-k} \sum_{n=k}^{\infty} g(nT-kT)\, z^{-(n-k)}$$

$$= z^{-k} \sum_{m=o}^{\infty} g(mT)\, z^{-m} = z^{-k}\ G(z) \qquad (40)$$

Product of Z-Transforms: Convolution Theorem

If

$$Y(z) = G(z)\, X(z) \tag{41}$$

where

$$Y(z) = \sum_{n=o}^{\infty} y(nT)\, z^{-n} \tag{42}$$

$$G(z) = \sum_{n=o}^{\infty} g(nT)\, z^{-n} \tag{43}$$

$$X(z) = \sum_{n=o}^{\infty} x(nT)\, z^{-n} \tag{44}$$

then

$$y(nT) = \sum_{m=o}^{n} g(mT)\, x(nT-mT)$$

$$= \sum_{m=o}^{n} g(nT-mT)\, x(mT) \tag{45}$$

Expression (45) can easily be proved by substituting Equations (42)-(44) in Equation (41), and equating coefficient of z^{-n} on both sides.

5. FIRST-ORDER DIGITAL FILTERS[3]

The kth order difference equation given in Equation (2) describes a kth order digital filter. It thus follows that a first order digital filter will be characterized by a difference equation of the form:

$$y(nT) = A_0\, x(nT) + A_1\, x(nT-T) + B_1\, y(nT-T) \tag{46}$$

Equation (46) can be solved sequentially to express y(nT) in terms of the initial state y(-T) for n = 0, 1, 2, ... If x*(t) is a causal function, i.e., x(nT) = 0 for n < 0, then we have

$$y(0) = B_1\, y(-T) + A_o\, x(0)$$

$$y(T) = B_1\, y(0) + A_o\, x(T) + A_1\, x(0)$$

$$= B_1^2\, y(-T) + A_o\, x(T) + [B_1 A_o + A_1]\, x(0)$$

$$y(nT) = B_1^{\,n+1}\, y(-T) + A_o\, x(nT) + \sum_{m=o}^{n-1} B_1^{\,n-1-m}[B_1 A_o + A_1] x(mT) \tag{47}$$

If the input $x(t)$ is a unit step function at $t = 0$, i.e., $x(t) = u(t)$

$$x(nT) = 1 \qquad \text{for all values of } n$$

This implies

$$y(nT) = B_1^{\,n+1} y(-T) + A_o + [B_1 A_o + A_1] \sum_{m=o}^{n-1} B_1^{\,n-1-m}$$

$$= B_1^{\,n+1}\, y(-T) + A_o + \frac{1-B_1^{\,n}}{1-B_1}[B_1 A_o + A_1] \tag{48}$$

provided $|B_1| < 1$. If $|B_1| < 1$ then $B_1^n \approx 0$ for large values of n so that the steady-state response for large values of n becomes

$$y(nT)\Big|_{\text{steady state}} \doteq A_o + \frac{B_1 A_o + A_1}{1-B_1}$$

$$= \frac{A_o + A_1}{1-B_1} \tag{49}$$

It is evident from Equation (48), if $|B_1|$ is not less than 1, then the first order digital filter is unstable, i.e., $y(nT) \to \infty$ as $n \to \infty$.

A plot of $y(nT)$ for some typical values of the parameters is shown in Figure 7 with $y(-T) = 0$. Note the similarity of the response for $A_o = 0$ with the step response of an RC integrating circuit.

A convenient representation of the first order digital filter described by Equation (46) is shown in Figure 8.

Several comments are here in order with regard to either digital computer simulation or digital hardware realization of Equation (48). Note that even for the simple case of $A_1 = 0$ and $A_o = 1$, three registers are needed to store B_1, hold $y(nT)$ and hold $x(nT)$. In addition, facilities for multiplication and addition must be available. This type of implementation appears to be expensive for realizing a simple first order filter. But the real gain is achieved if these digital components were utilized to realize large number of first order filters by making use of time multiplexing [3].

6. SYSTEM TRANSFER FUNCTION

As in the case of the continuous system, we can define a system transfer function for a digital filter by expressing the Z-transform of the response as

$$Y(z) = H(z)\, X(z) \tag{50}$$

where $X(z)$ is the Z-transform of the input. In Equation (50) $H(z)$ is the system transfer function. Using Equation (50) in Equation (2) , we readily obtain for a kth order filter:

$$H(z) = \frac{Y(z)}{X(z)} = z^{k-m} \cdot \frac{A_m + A_{m-1}z + \ldots + A_1 z^{m-1} + A_o z^m}{-(B_k + B_{k-1}z + \ldots + B_1 z^{k-1}) + z^k} \tag{51}$$

If we let

$$z = e^{j\omega T} \tag{52}$$

then we can write

$$H(e^{j\omega T}) = |H| e^{j\psi} \tag{53}$$

where

$$\psi = \arg H\,(e^{j\omega T}) \tag{54}$$

$H(e^{j\omega T})$ thus determines the frequency response of the digital filter.

For the first order digital filter of Equation (46), the system transfer function is

$$H(z) = \frac{A_o z + A_1}{z - B_1} \tag{55}$$

Consequently

$$H(e^{j\omega T}) = \frac{A_o e^{j\omega T} + A_1}{e^{j\omega T} - B_1} \tag{56}$$

Thus

$$|H| = \left[\frac{(A_1 + A_o \cos\omega T)^2 + (A_o \sin\omega T)^2}{(\cos\omega T - B_1)^2 + (\sin\omega T)^2}\right]^{1/2}$$

$$= \left[\frac{(A_1^{\,2} + A_o^{\,2}) + 2A_1 A_o \cos\omega T}{(1 + B_1^{\,2}) - 2B_1 \cos\omega T}\right]^{1/2} \tag{57}$$

and

$$\psi = \tan^{-1} \frac{A_o \sin\omega T}{A_1 + A_o \cos\omega T} - \tan^{-1} \frac{\sin\omega T}{\cos\omega T - B_1} \tag{58}$$

The magnitude function $|H|$ has been plotted for in Figure 9 for typical values of various parameters. The possibility of obtaining frequency selectivity with digital filters is illustrated by this figure.

7. SECOND-ORDER DIGITAL FILTERS

Let us now examine several realizations of the second-order digital filter. The methods discussed can readily be extended for higher order filters.

The general form of a second order filter is obtained from Equation (2) as:

$$y(nT) = A_o x(nT) + A_1 \ (nT-T) + A_2 x(nT-2T)$$

$$+ B_1 y(nT-T) + B_2 y(nT-2T) \qquad (59)$$

Taking the Z-transforms of both sides, assuming $y(-T) = y(-2T) = 0$ and re-arranging terms, we obtain

$$Y(z) = \frac{A_o + A_1 z^{-1} + A_2 z^{-2}}{1 - B_1 z^{-1} - B_2 z^{-2}} \cdot X(z) \qquad (60)$$

The system transfer function is then

$$H(z) = \frac{A_o + A_1 z^{-1} + A_2 z^{-2}}{1 - B_1 z^{-1} - B_2 z^{-2}} \qquad (61)$$

Direct realization of this filter is sketched in Figure 10.

An alternate realization is obtained by writing Equation (60) as two equations:

$$G(z) = \frac{X(z)}{1 - B_1 z^{-1} - B_2 z^{-2}} \qquad (62)$$

or

$$Y(z) = (A_o + A_1 z^{-1} + A_2 z^{-2}) G(z) \qquad (63)$$

We rewrite Equation (62) as

$$G(z) = X(z) + B_1 z^{-1} G(z) + B_2 z^{-2} G(z) \qquad (64)$$

Realization of the filter described by Equations (63) and (64) is illustrated in Figure 11 a. Note that the signal at points a' is g(nT-T), and so is the signal at point a''. Thus, these two points can be combined. This would eliminate one delay unit. Similarly, points b' and b'' can be combined to eliminate another delay unit. The modified realization is shown in Figure 11(b). For a second order digital filter, the minimum number of delay units necessary is two. Hence the realization of Figure 11(b) is "optimal" with respect to the number of delay units and is said to be "canonic" in form. Another "canonic" realization is indicated in Figure 12.

The second-order digital filter can be considered as the basic building block for the realization of higher order digital filters. It is generally known that such method leads to lowest coefficient accuracy requirements. The kth order digital filter transfer function as given by Equation (51) can be expressed either as a product or as a sum of second-order transfer functions if k is even[1]. If k is odd, then a first order building block will be needed in either case. To illustrate this approach, consider the case of $k = 4$. Then we have

$$H(z) = c \left[\frac{A_{10}+A_{11}z^{-1}+A_{12}z^{-2}}{1-B_{11}z^{-1}-B_{12}z^{-2}}\right] \left[\frac{A_{20}+A_{21}z^{-1}+A_{22}z^{-2}}{1-B_{21}z^{-1}-B_{12}z^{-2}}\right] \tag{65}$$

or

$$H(z) = D + \frac{\alpha_{10}+\alpha_{11}z^{-1}}{1-B_{11}z^{-1}-B_{12}z^{-2}} + \frac{\alpha_{20}+\alpha_{21}z^{-1}}{1-B_{21}z^{-1}-B_{22}z^{-2}} \tag{66}$$

The product-form of (65) is obtained by factoring roots of the denominator and numerator of $H(z)$. The sum-form of Equation (66) is obtained by a partial fraction expansion and assuming simple poles only. For the multiple pole case the form is slightly modified. The two corresponding realizations are shown in Figures 13 and 14.

8. DESIGN TECHNIQUES

There are basically two types of digital filters. If the filter is described by Equation (2) with all B_i being zeros, it is called a non-recursive or transversal type digital filter. In this case, the output at any time depends only on the present and past samples of the input, and does not depend on the previous samples of the output. If on the other hand, at least one B_i and one A_j are non-zero, then the filter is said to be recursive type digital filter.

We now present four methods of digital filter design, the first three lead to recursive type filters and the last one for the non-recursive type filter.

8.1 Impulse Response Invariance Technique

By this method, one can design a digital filter whose impulse response is identical to the sampled impulse response of a given continuous filter. From the prescribed impulse response $y(t)$, we first obtain the system function $Y(s) = \mathcal{L}\{y(t)\}$. Assuming simple poles, $H(s)$ can be expressed in a partial fraction expansion as:

$$Y(s) = \sum_{i=1}^{m} \frac{a_i}{s+b_i} \tag{67}$$

which implies

$$y(t) = \sum_{i=1}^{m} a_i \, e^{-b_i t}$$

Hence

$$y(nT) = \sum_{i=1}^{m} a_i \, e^{-b_i nT} \tag{69}$$

We want $y(nT) \equiv h(nT)$ where $h^*(t)$ is the impulse response of the digital filter. Thus

$$H(z) = \sum_{n=0}^{\infty} h(nT) z^{-n}$$

$$= \sum_{i=1}^{m} a_i \left[\sum_{n=0}^{\infty} e^{-b_i nT} \, z^{-n} \right]$$

$$= \sum_{i=1}^{m} \frac{a_i}{1-e^{-b_i T} z^{-1}} \qquad (70)$$

Note that the above results could have been obtained directly by making use of the relation between Z-transform and the Laplace transform outlined in Section 4.2.

If b_i is complex, then we can combine the terms corresponding to b_i and its conjugate b_i^* in Equation (67) to yield either of the following forms:

$$Y_1(s) = \frac{s+\alpha}{(s+\alpha)^2+\beta^2} \qquad (71)$$

$$Y_2(s) = \frac{\beta}{(s+\alpha)^2+\beta^2} \qquad (72)$$

The corresponding expressions in $H(z)$ will be

$$H_1(z) = \frac{1-e^{-\alpha T}(\cos\beta T)z^{-1}}{1-2e^{-\alpha T}(\cos\beta T)z^{-1}+e^{-2\alpha T} z^{-2}} \qquad (73)$$

$$H_2(z) = \frac{e^{-\alpha T}(\sin\beta T)z^{-1}}{1-2e^{-\alpha T}(\cos\beta T)z^{-1}+e^{-2\alpha T} z^{-2}} \qquad (74)$$

The method can be routinely extended for the case of multiple poles.

8.2 Design based on Magnitude Function Specification

An elegant technique based on magnitude function approximation is as follows. We first note that $H(z)$ is a rational function in z^{-1} and for $z = e^{j\omega T}$ (i.e. on the unit circle), $H(e^{j\omega t})$ is a rational function of $e^{j\omega T}$. Consequently $|H(e^{j\omega T})|^2$ can always be expressed as a ratio of two trigonometric functions of ωT. To illustrate now the basic idea behind the second method, suppose we have approximated a specified magnitude function in the form:

$$|H(e^{j\omega T})|^2 = \frac{1}{1 + \dfrac{\tan^{2n}(\omega T/2)}{\tan^{2n}(\omega_c T/2)}} \qquad (75)$$

If we let $z = e^{j\omega T}$, then

$$\frac{z-1}{z+1} = j \tan \frac{\omega T}{2} \tag{76}$$

Using Equation (76) in Equation (75), we obtain

$$|H(z)|^2 = \frac{\tan^{2n}(\omega_c T/2)}{\tan^{2n}(\omega_c T/2) + (-1)^n \left[\frac{z-1}{z+1}\right]^{2n}} \tag{77}$$

It can be seen that $|H(z)|^2$ is a rational function in z, which has a zero of order $2n$ at $z = -1$. Determination of the poles is readily achieved by using a new transformation

$$p = \frac{z-1}{z+1} \tag{78}$$

Thus the 2n poles of $|H(p)|^2$ are uniformly spaced around a circle of radius $\tan(\omega_c T/2)$ in the p-plane. From the knowledge of the p-plane roots, z-plane roots are found by using the inverse transformation

$$z = \frac{1+p}{1-p} \tag{79}$$

and the roots inside the unit circle are chosen as the poles of $H(z)$.

<u>Example 5</u>. Consider $n = 2$

$$|H(e^{j\omega T})|^2 = \frac{\tan^4\left(\frac{\omega_c T}{2}\right)}{\tan^4\left(\frac{\omega_c T}{2}\right) + \tan^4\left(\frac{\omega T}{2}\right)}$$

Let $\omega_c T = \frac{2\Pi}{3}$ then $\tan\left(\frac{\omega_c T}{2}\right) = \sqrt{3}$ thus, Equation (77) becomes

$$|H(z)|^2 = \frac{9}{\left(\frac{z-1}{z+1}\right)^4 + 9}$$

Making use of the transformation indicated by Equation (78), we have

$$|H(p)|^2 = \frac{9}{p^4+9} = \frac{9}{(p^2+\sqrt{6}p+3)(p^2-\sqrt{6}p+3)}$$

Using inverse transformation of Equation (78) we thus get

$$|H(z)|^2 = \frac{9(z+1)^4}{\left[(4+\sqrt{6})z^2+4z+(4\sqrt{6})\right]\left[(4-\sqrt{6})z^2+4z+(4+\sqrt{6})\right]}$$

Considering only the poles inside the unit circle, we finally obtain the required system transfer function as

$$H(z) = \frac{3(z^2+2z+1)}{(4+\sqrt{6})z^2+4z+(4-\sqrt{6})}$$

8.3 <u>Technique Based on Bilinear Transformation</u>

Let $H(s)$ be a realizable analog filter transfer function. Its frequency response is found by evaluating $H(s)$ at points on the imaginary axis of s-plane. If in the function $H(s)$ s is replaced by rational function of z, $f(z)$, which maps the imaginary axis of the s-plane onto the unit circle of the z-plane, then the resulting function of variable z, $H'(z)$, where

$$H'(z) = H(s)\Big|_{s=f(z)} , \tag{80}$$

evaluated along the unit circle will take on the same set of values as $H(j\omega)$.

One such transformation is

$$s = \frac{z-1}{z+1} \tag{81}$$

Let ω_A be a particular analog frequency of interest, and the corresponding digital frequency variable be $\omega_D T$ such that the following holds

$$H(\omega_A) = H'(\omega_D T) \tag{82}$$

It can be shown easily that Equation (82) holds if

$$\omega_A = \tan \frac{\omega_D T}{2} \qquad (83)$$

Since the transformation indicated maps the left half s-plane onto the inside of the unit circle, $H'(z)$ is guaranteed stable, provided $H(s)$ was stable.

Example 6. A low-pass digital filter having a monotonic frequency response is to be designed for a 3KHz sampling rate. The 3-db cut-off is at 0.5 KHz and the response should be more than ~~18~~ db down at 1 KHz.

A Butterworth filter would satisfy these requirements in the analog domain. The pertinent frequencies of the digital filter are at:

$$\omega_D' T = 2\pi \cdot \left(\frac{1}{2} \times 10^3\right) \cdot \left(\frac{1}{3 \times 10^3}\right) = \frac{\pi}{3}$$

$$\omega_D'' T = 2\pi \left(10^3\right) \left(\frac{1}{3 \times 10^3}\right) = \frac{2\pi}{3}$$

Corresponding frequencies in the analog domain are at:

$$\omega_A' = \tan \frac{\omega_D' T}{2} = \tan \left(\frac{\pi}{6}\right) = \frac{1}{\sqrt{3}}$$

$$\omega_A'' = \tan \frac{\omega_D'' T}{2} = \tan \left(\frac{\pi}{3}\right) = \sqrt{3}$$

A normalized Butterworth filter has a 3-db cut-off at $\omega_o = 1$. This has to be de-normalized to $\omega_c = \omega_A' = \frac{1}{\sqrt{3}}$. Our next problem is to determine the order of the filter, we note at ω_A'' we must have

$$10 \log \left[1 + \left(\frac{\omega_A''}{\omega_A'}\right)^{2n}\right] \cancel{\geq 18\ \text{db}} \geq \text{[illegible] db}$$

Which is equivalent to

$$1 + \left(\frac{\omega''_A}{\omega'_A}\right)^{2n} = \cancel{63.1} \geq (63)$$

or

$$1 + (\overset{3}{\cancel{2}})^{2n} \geq \cancel{63.1}\ (63) \quad \text{or} \quad (\overset{3}{\cancel{2}})^{2n} \geq \cancel{62.1}\ (62)$$

Hence n must be 3. The resulting poles are at

$$s_1 = \frac{1}{\sqrt{3}}\left(-\frac{1}{2} - j\frac{\sqrt{3}}{2}\right)$$

$$s_2 = \frac{1}{\sqrt{3}}\left(-\frac{1}{2} + j\frac{\sqrt{3}}{2}\right)$$

$$s_3 = -\frac{1}{\sqrt{3}}$$

The corresponding analog transfer function is

$$H(s) = \frac{1}{(3\sqrt{3})s^3 + 6s^2 + (2\sqrt{3})s + 1}$$

$H'(z)$ is obtained from above by replacing s by $(z-1)/(z+1)$:

$$H'(z) = \frac{(z+1)^3}{(7+5\sqrt{3})z^3 - (7\sqrt{3}+3)z^2 + (7\sqrt{3}-3)z + (7-5\sqrt{3})}$$

8.4 The Fourier Series Approach

Let $H(s)$ be the desired analog transfer function which is to be approximately realized by a digital filter so that the magnitude characteristics match. We expand, $H(\omega)$ in a Fourier series over the band $|\omega| < \frac{\omega_s}{2}$:

$$H(\omega) = \sum_{n=0}^{\infty} a_n \cos n\omega T \,. \tag{84}$$

or

$$H(\omega) = \sum_{n=o}^{\infty} b_n \sin n\omega T \tag{85}$$

If $H(s) \cong Ks^m$ for small s, then the cosine-series of Equation (84) is used if m is even. If m is odd, then the sine-series expansion of Equation (85) is used. Since on the unit circle,

$$z \equiv e^{j\omega T} \tag{86}$$

Equations (84) and (85) lead to the corresponding system transfer functions for the non-recursive filters:

$$H(z) = a_o + \frac{1}{2} \sum_{n=o}^{\infty} a_n (z^n + z^{-n}) \text{ , m even} \tag{87}$$

$$= \frac{1}{2} \sum_{n=o}^{\infty} b_n (z^n + z^{-n}) \text{ , n odd} \tag{88}$$

In practice, the series is truncated which leads to errors in the approximation. Various approaches have been suggested to improve the approximation by modifying $\{a_n , b_n\}$ [1].

9. CONCLUSION

An introduction to basic concepts of digital filter along with the theory of analysis and design of such filters has been presented. More extensive information on the subject will be found in the references listed in the bibliography.

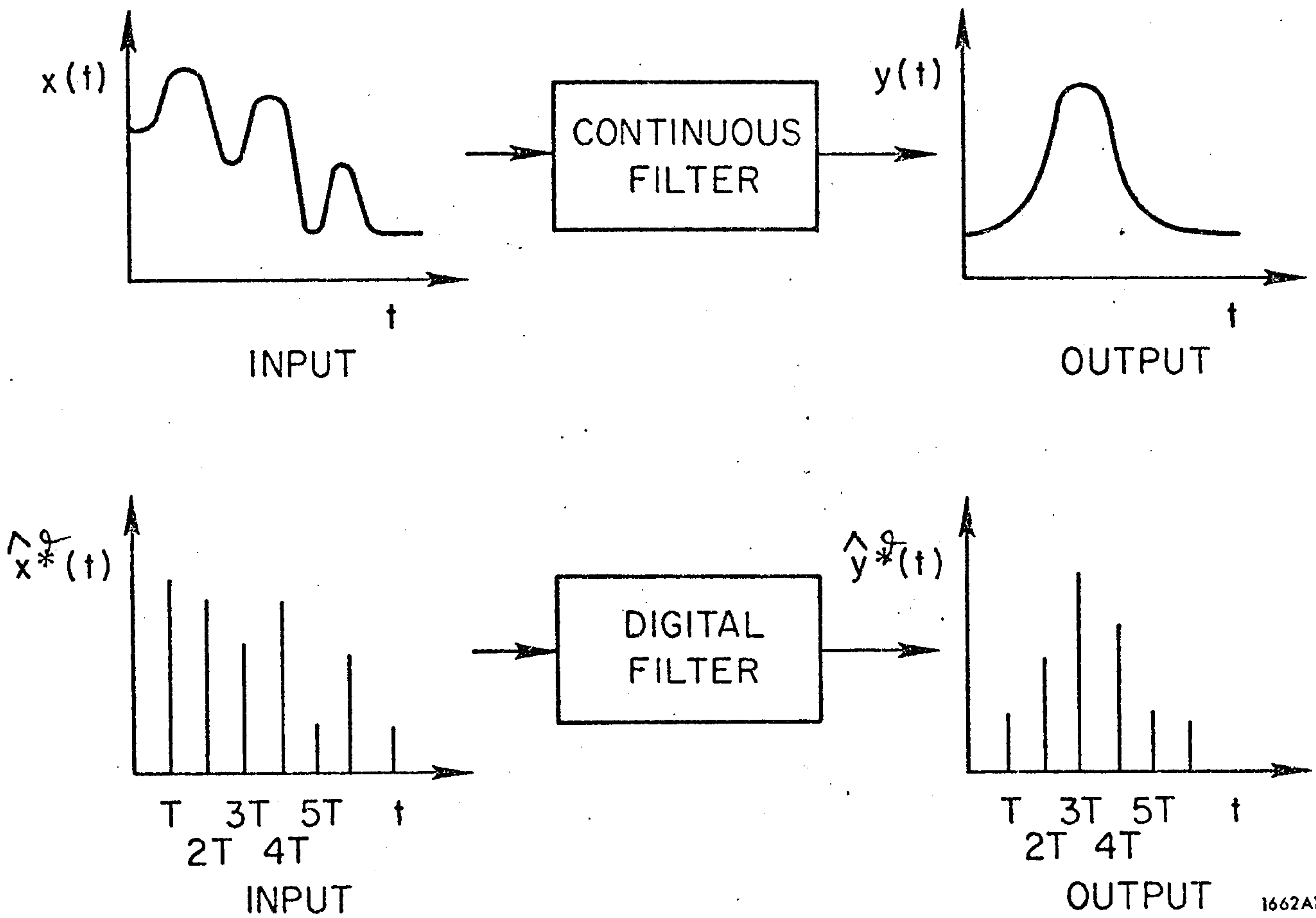
For conventional applications, a digital filter employs more components in comparison to an equivalent analog filter. This probably has restricted the use of such filters primarily to simulation work. With large scale integration of digital circuits implementation of digital filters would be cheaper with resultant decrease in size and an increase in reliability, making digital filters more attractive than their equivalent analog filters. An unique property of the digital filters is the ease with which they can be easily modified and time-shared. This particular property will be made more use of in future in designing filtering systems.

BIBLIOGRAPHY

1. J. F. Kaiser, "The Digital Filters," in "System Analysis by Digital Computer," F. F. Kuo and J. F. Kaiser, Eds., John Wiley, 1966, pp. 218-285.
2. E. I. Jury, "Theory and Application of the Z-Transform Method," John Wiley and Sons, 1964.
3. B. Gold and C. M. Rader, "Digital Processing of Signals," McGraw-Hill, 1969.
4. R. M. Golden, "Digital Filter Synthesis by Sampled-Data Transformation," IEEE Trans. on Audio and Electroaccoustics, vol. AU-16, No. 3, September 1968, pp. 321-329.
5. J. V. Wait, "Digital Filters," in "Active Filters," Ed. L. P. Huelsman, McGraw-Hill, 1970.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. J. F. Kaiser, Mr. R. S. Larsen and Mr. D. Porat for their help in the preparation of this report.



x(t)
CONTINUOUS
FILTER
y(t)
t
INPUT
t
OUTPUT
$\hat{x}^*(t)$
DIGITAL
FILTER
$\hat{y}^*(t)$
T 3T 5T t
2T 4T
INPUT
T 3T 5T t
2T 4T
OUTPUT
1662A1

Fig. 1

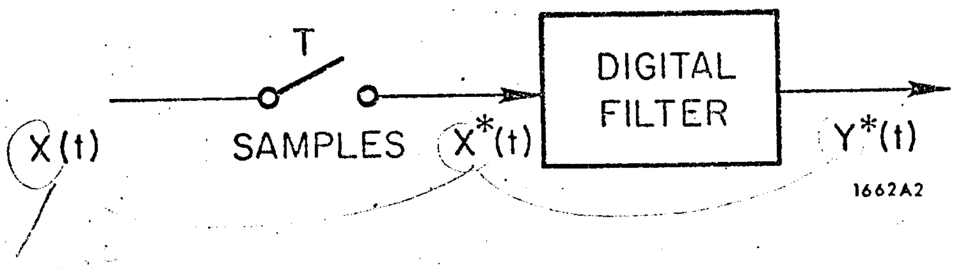


Fig. 2

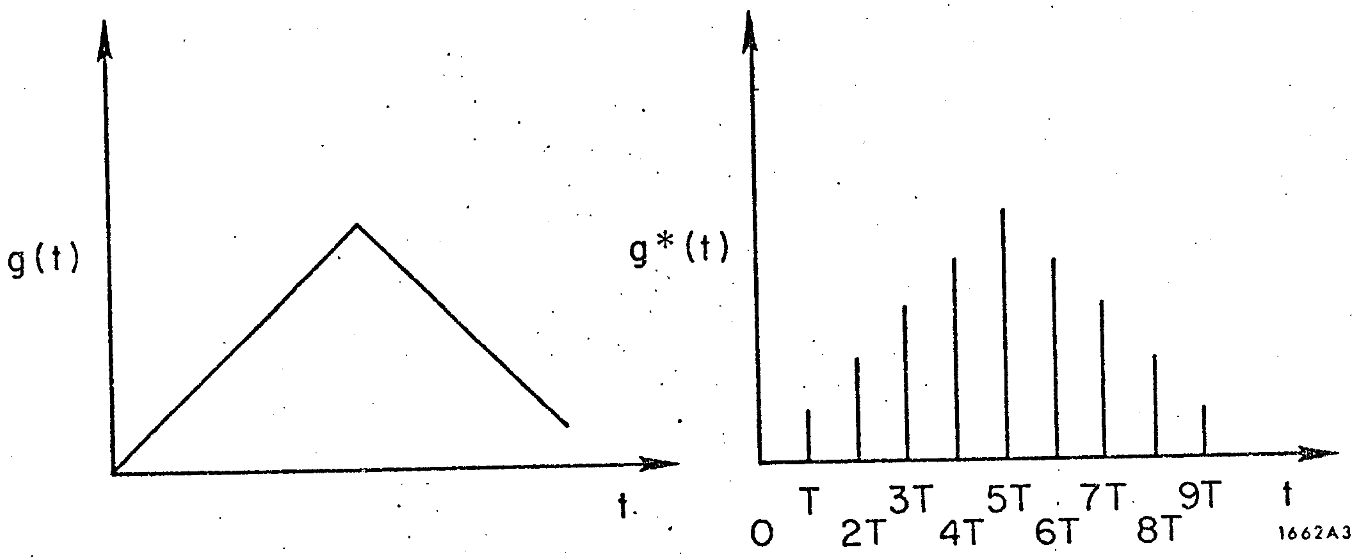


Fig 3

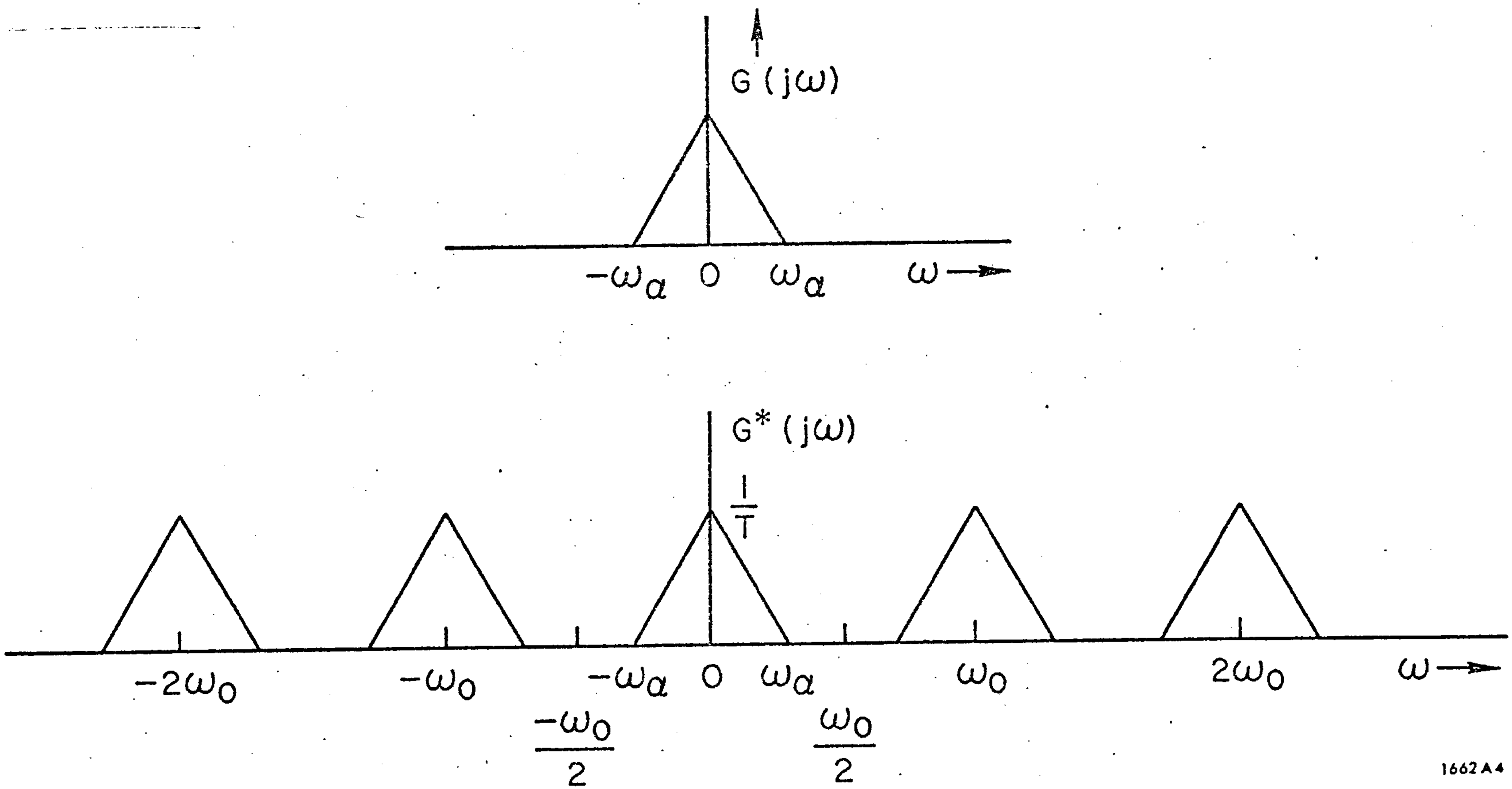


Fig. 4

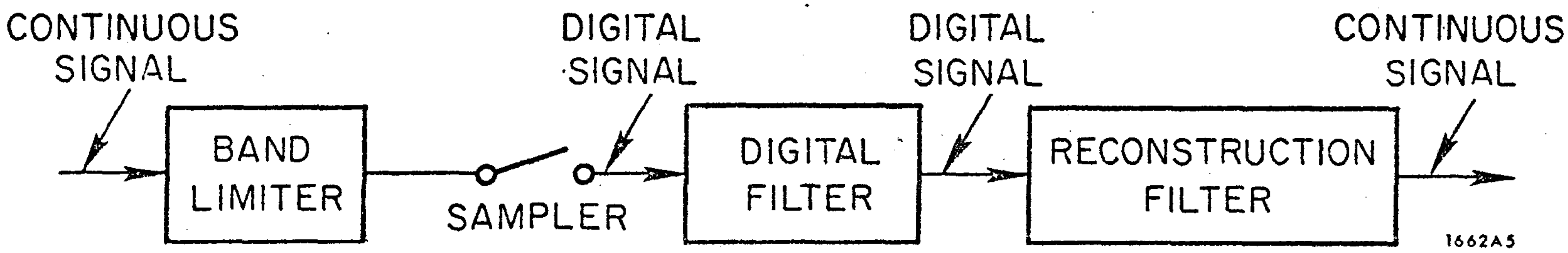
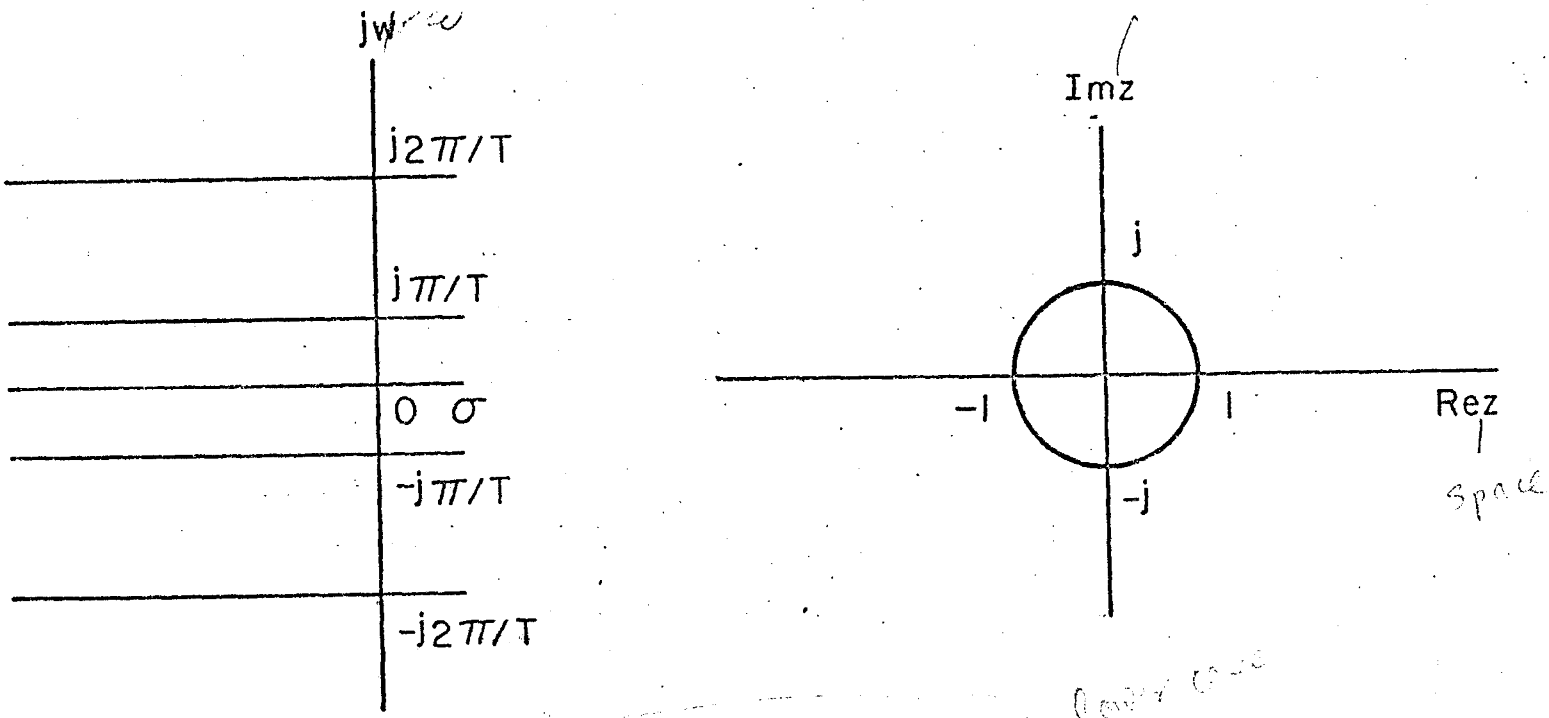

CONTINUOUS SIGNAL
BAND LIMITER
SAMPLER
DIGITAL SIGNAL
DIGITAL FILTER
DIGITAL SIGNAL
RECONSTRUCTION FILTER
CONTINUOUS SIGNAL
1662A5

Fig 5



(a) S-PLANE (b) Z-PLANE

1662A6

Fig. 6

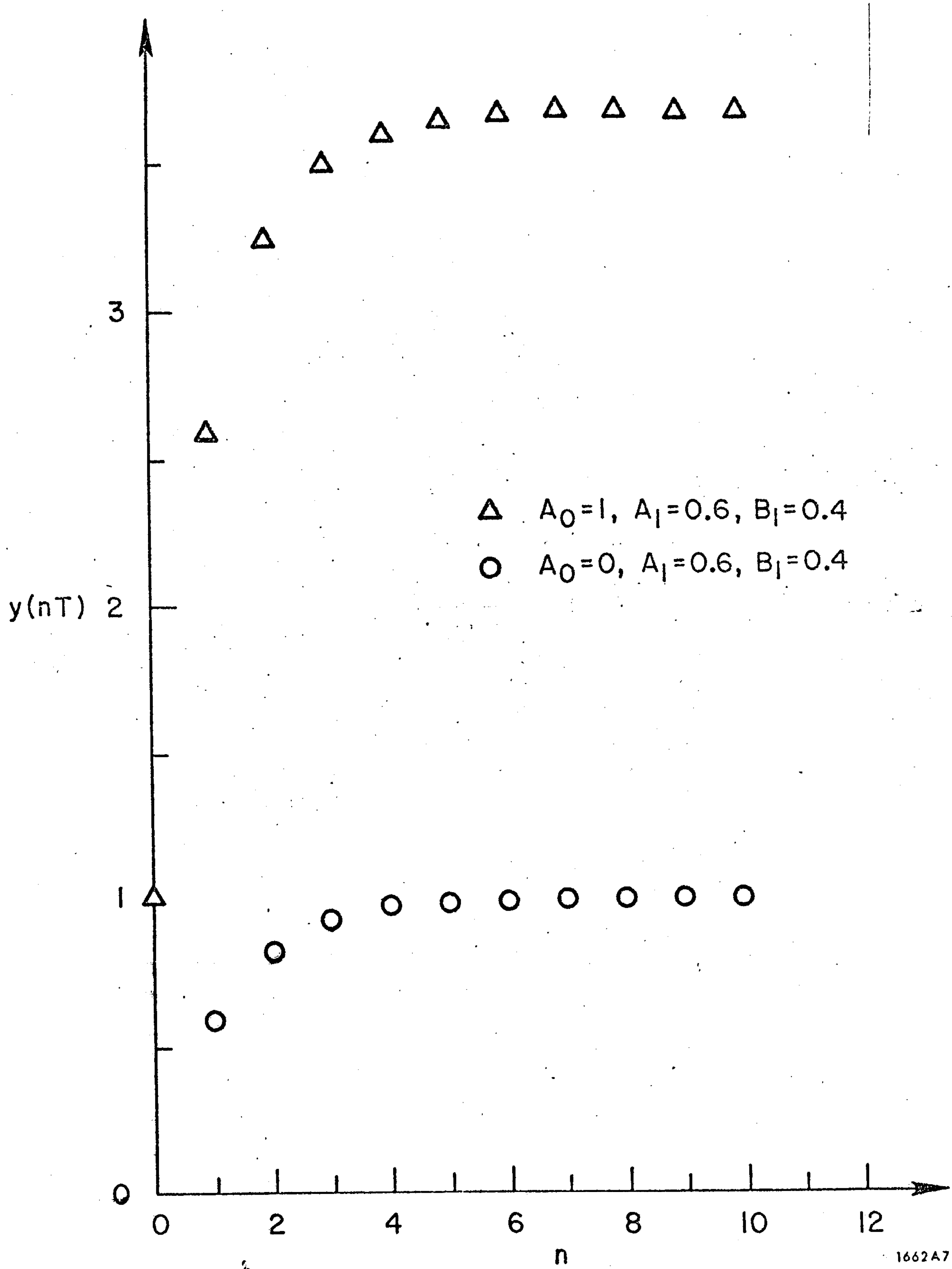
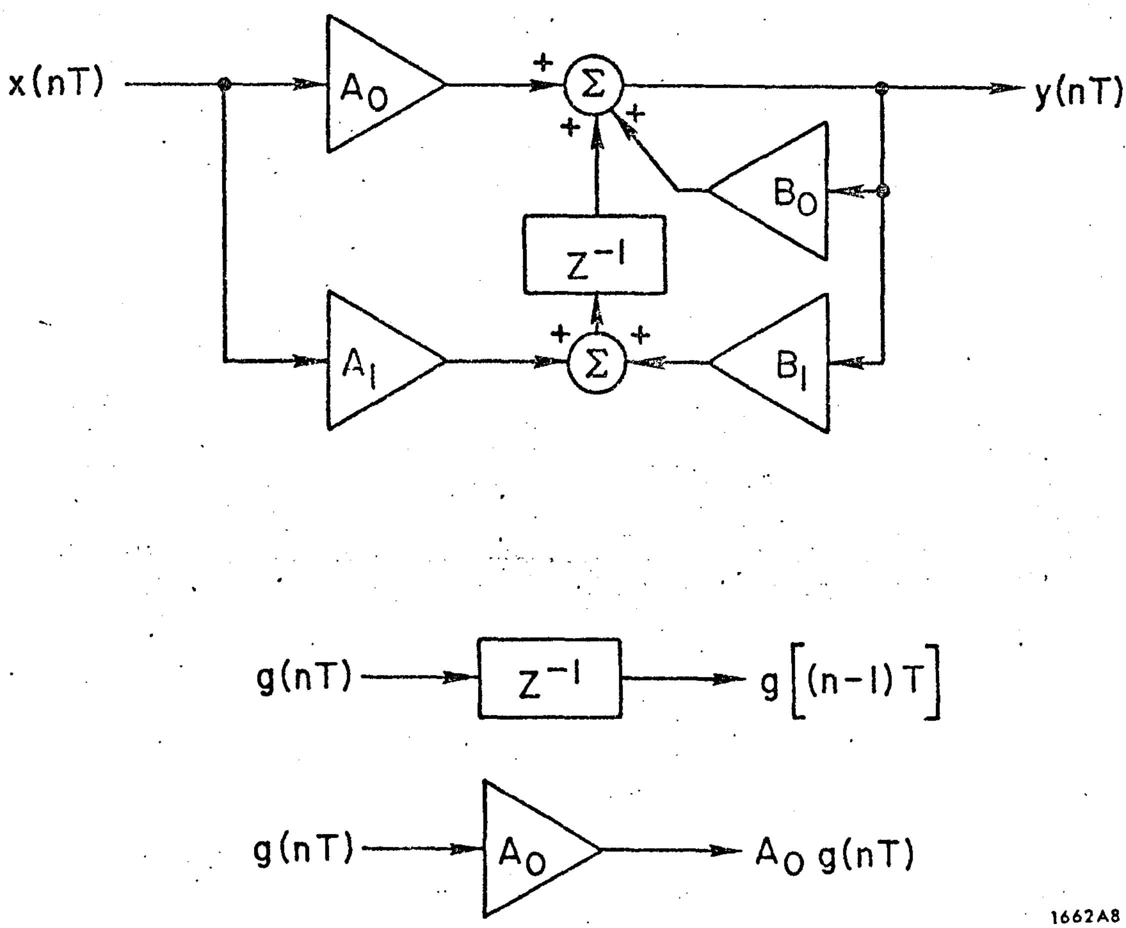
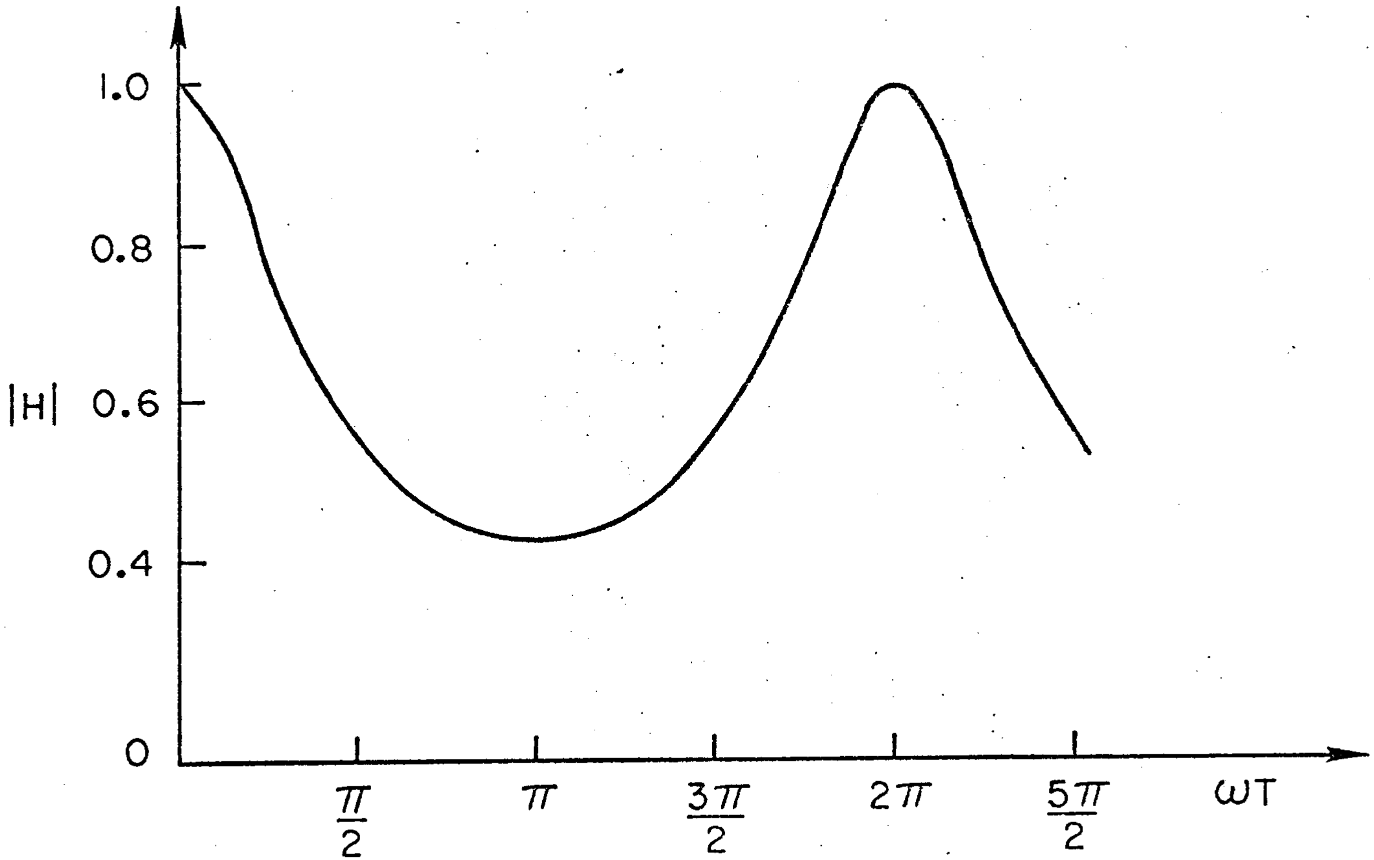


Fig. 7



1662A8

Fig. 8



$A_1 = 0.6 \quad B_1 = 0.4$

1662A9

Fig. 9

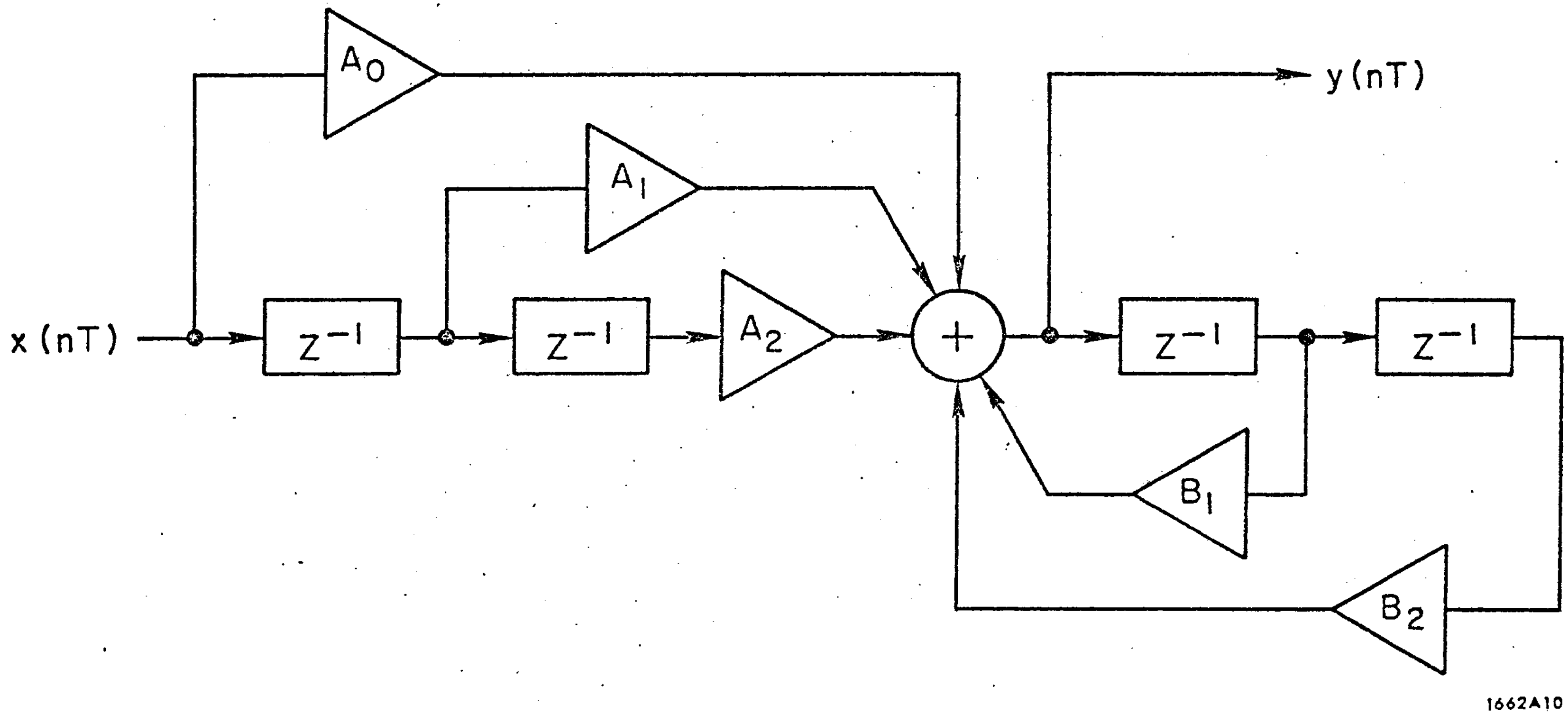
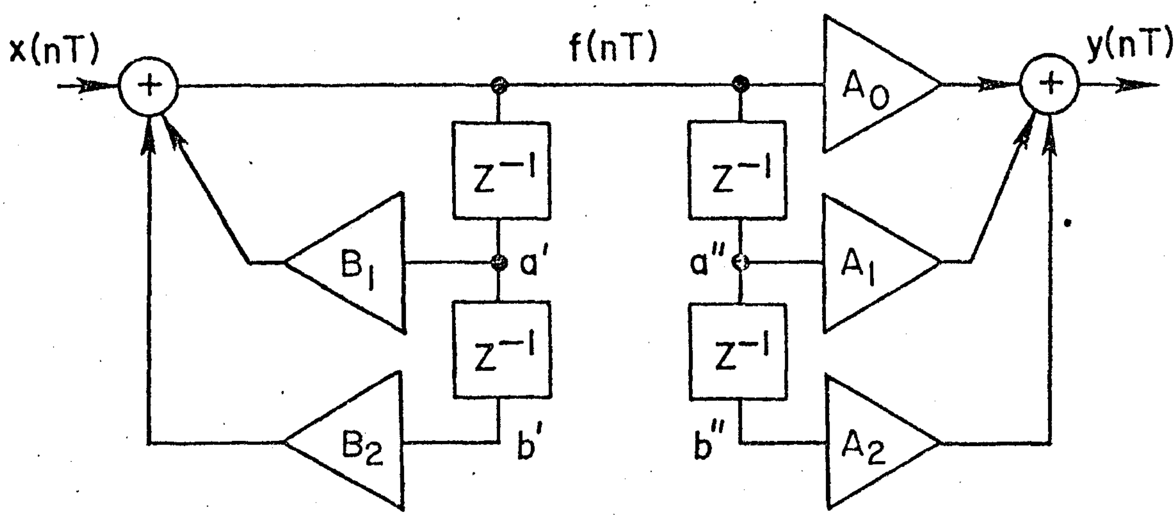
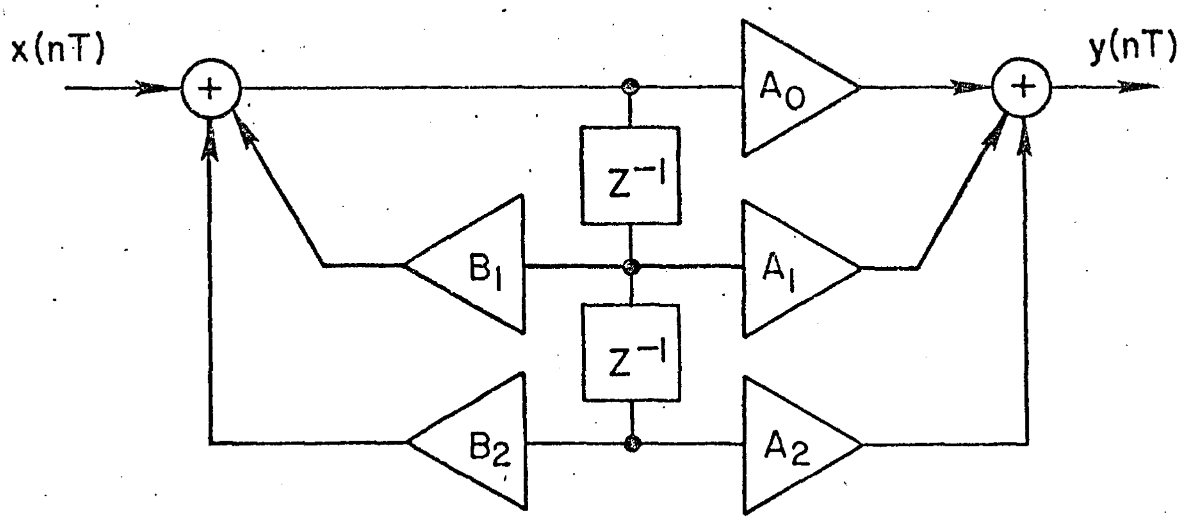


Fig. 10



(a)



(b)

1662A11

Fig. 11

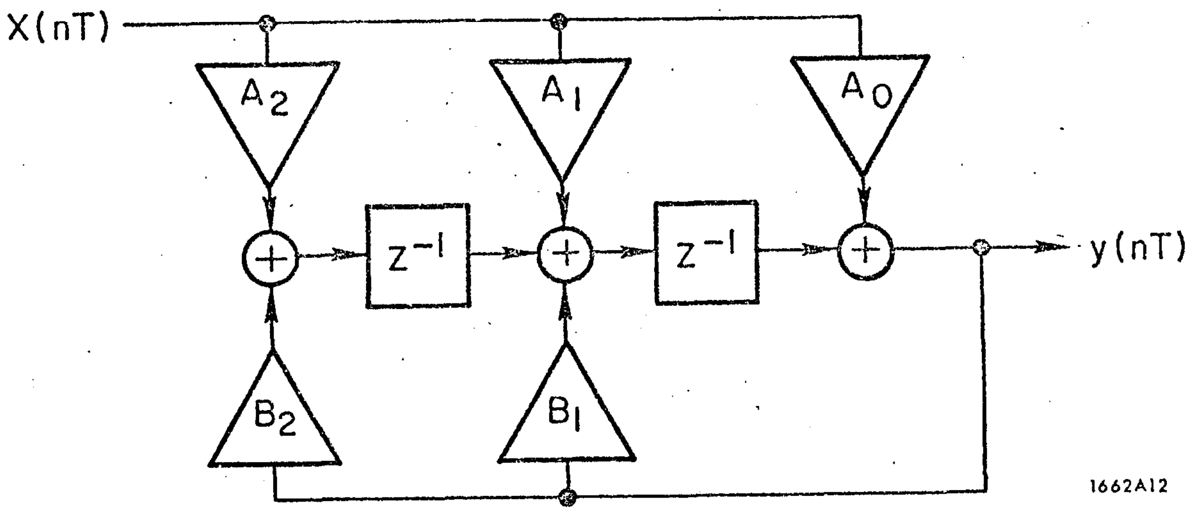


Fig. 12

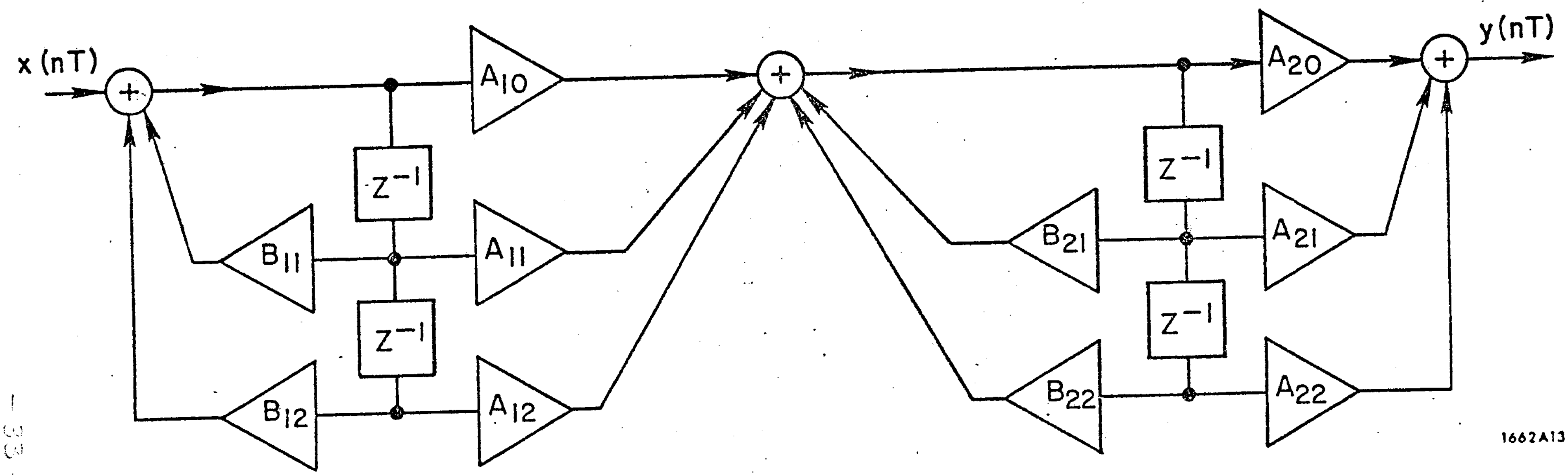

x (nT)
y (nT)
A10
A11
A12
B11
B12
Z−1
Z−1
A20
A21
A22
B21
B22
Z−1
Z−1
1662A13

Fig 13

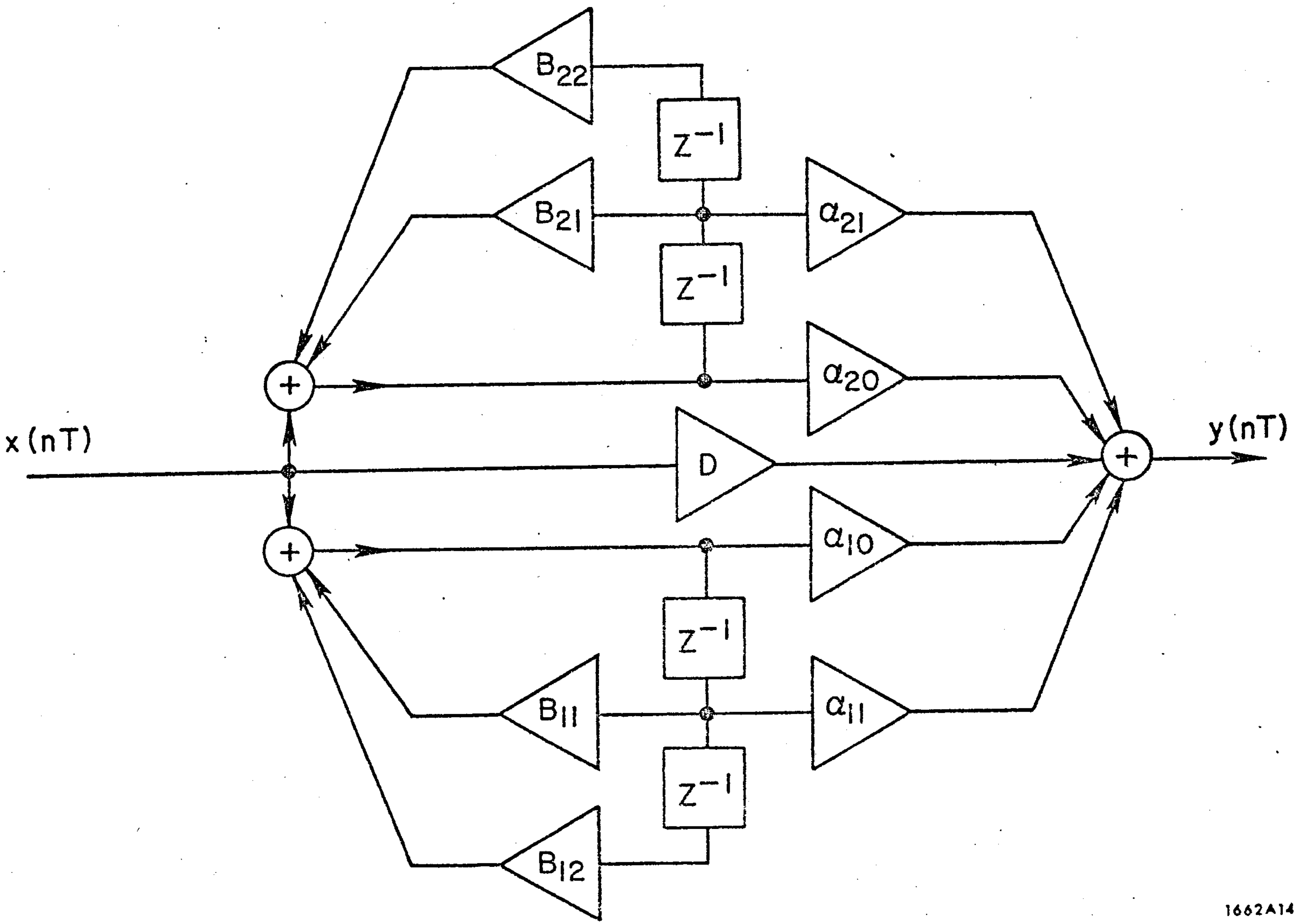


Fig 14.