TECHNICAL REPORT 400-56

# THE APPROXIMATION PROBLEM FOR DIGITAL FILTERS

K. Steiglitz

March 1962

NEW YORK UNIVERSITY
COLLEGE OF ENGINEERING
DEPARTMENT OF ELECTRICAL ENGINEERING

University Heights
New York 53, New York

## ACKNOWLEDGMENT

The author wishes to thank Dr. S. S. L. Chang of New York University for many stimulating discussions, for valuable suggestions, and for encouragement. A similar debt is owed to Dean Ragazzini and to the automatic control group at New York University, which provided time on the IBM 1620 computer.

This research was supported in part by the Air Force Office of Scientific Research under contract no. AF49-(638)-586.

# THE APPROXIMATION PROBLEM FOR DIGITAL FILTERS

## I. INTRODUCTION

The problem of finding a realizable transfer function with a prescribed magnitude or phase angle on the $j\omega$ axis has been an important one to the designers of continuous wave filters and compensators. The Butterworth and Tchebycheff filters, together with their high-pass and bandpass transformations, are well known solutions to common approximation problems. The possibility of designing digital systems in terms of frequency response, however, has remained relatively unexplored since the early work of Salzer[1], despite the advent of sampled-data systems and the increasingly wide use of digital computers in control and measurement. It is the purpose of this paper to explore some approximation techniques for digital filters, and finally, to show the equivalence of the approximation problems for digital filters and continuous filters. An immediate application is the design of spectral windows for spectrum analysers when the input data is discrete[2,3]. Also, when some time delay is tolerable and a digital computer is available, it might be practical to use a digital filter in tandem with a hold circuit to filter a continuous wave. Hopefully, other applications will be found.

The advantages of using a digital computer as a filter ~~is~~ are the flexibility, accuracy, and stability which can be readily obtained, and which ~~is~~ are practically impossible to achieve with analogue hardware. The constants of a digital filter can be set to a high degree of accuracy, can be changed very fast, and are not subject to unwanted variation with temperature or age. Furthermore, with the use of pulse-code modulation for low noise transmission of data over large distances, the availability of signals already in digital form can make the use of digital filters very practical.

First, let us clarify the notion of digital filter. The term digital filter will be applied to any linear computation scheme producing a discrete output time series $y(nT)$ from a discrete input time series $x(nT)$. Thus, any digital filter can be realized with an appropriately programmed digital computer. Theoretically, a pencil and paper will serve as well. In our context, the synthesis problem is trivial, for we need only carry out the indicated computations. Lewis[4] has considered the synthesis of digital filters with networks of open- and short-circuited transmission lines.

A word about our nomenclature: Our definition is independent of any physical, continuous wave filter. Hence the use of the term <u>digital filter</u>, as opposed to a filter which must always be the z-transform of some continuous filter.

## II. THE TRANSFER FUNCTION OF A DIGITAL FILTER

If a digital filter must operate in real time, only past and present inputs are available. We call such filters <u>real time</u> filters. They can be characterized by the following computational scheme:

$$y(nT) = \sum_{k=0}^{N} a_k x(nT-kT) - \sum_{k=1}^{M} b_k y(nT-kT) \quad , \tag{1}$$

where T is the sampling period. If we use the two-sided z-transform notation[5]

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

$$Y(z) = \sum_{n=-\infty}^{\infty} y(nT)z^{-n} \quad , \tag{2}$$

$$z = e^{sT} \quad ,$$

(1) becomes

$$Y(z) = \sum_{k=0}^{N} a_k z^{-k} \; X(z) - \sum_{k=1}^{M} b_k z^{-k} \; Y(z) \quad . \tag{3}$$

Thus, the transfer function is

$$D(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^{N} a_k z^{-k}}{1 + \sum_{k=1}^{M} b_k z^{-k}} \quad , \tag{4}$$

which is a rational function of z, whose magnitude is a periodic function of $\omega$. The phase angle of D(z) is not really periodic in $\omega$, but increases by a multiple of $2\pi$ every time $\omega$ increases by $\frac{2\pi}{T}$. To see this, suppose that inside the unit circle in the z plane the number of zeros of D(z) minus the number of poles of D(z) is m. Then when $\omega$ increases by $\frac{2\pi}{T}$, the point z traverses the unit circle once in the counterclockwise direction. Therefore, the phase angle of D(z) increases by $m2\pi$.

If a digital filter D(z) requires P future inputs, $z^{-P}D(z)$ will be a real time filter. On the $j\omega$ axis, the delay operator $z^{-P}$ does not affect the magnitude characteristic and contributes only linear phase.

When all the poles of D(z) are inside the unit circle in the z plane, we call D(z) a stable filter. The impulse response of a filter that is not stable will not approach zero as time increases. Since unstable filters will generally be of no practical importance, we shall not accept them as solutions to an approximation problem.

## III. FOURIER SERIES TECHNIQUES

Guillemin[6] has suggested the use of Fourier series in the approximation of magnitude characteristics of continuous wave filters. His approximation procedure consists of employing a bilinear transformation of the frequency variable to make the desired characteristic a periodic function of frequency, using a truncated Fourier series in the new frequency variable, and then reversing the transformation to give a rational function of $\omega$. Since our desired characteristic is already periodic, we can use Fourier series directly.

Suppose, then, that we are given the desired magnitude characteristic $M(\omega)$ of some digital filter. Since this is an even function of $\omega$ with period $2\pi/T$, we can approximate it in a least mean square error sense with the truncated Fourier series

$$M(\omega) \cong \sum_{m=-K}^{K} c_m e^{-jm\omega T} \quad , \tag{5}$$

where

$$c_m = c_{-m} = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} M(\omega) e^{jm\omega T} \, d\omega \quad . \tag{6}$$

The digital filter

$$D(z) = \sum_{m=0}^{2K} c_{m-K} z^{-m} = z^{-K} \sum_{m=-K}^{K} c_m z^{-m} \tag{7}$$

will then approximate $M(\omega)$, because when $z=e^{j\omega T}$

$$|D(z)| = \left| \sum_{m=-K}^{K} c_m z^{-m} \right| \cong M(\omega) \quad . \tag{8}$$

The series (5) is a cosine series, so that the only phase shift of $D(z)$ is that caused by the delay factor $z^{-K}$. Thus, if a delay of KT is tolerable, these filters can be considered to introduce no phase distortion. The filters obtained by this method will always be stable since they are polynomials in $z^{-1}$ and will have poles only at the origin. In fact, the impulse response of these filters will be zero after a finite number of sampling periods.

Example 1.

As an example, suppose we wish to approximate the ideal low-pass characteristic shown as a dashed line in Fig. 1. For convenience, take $T=\pi/2$, the cutoff frequency at $\omega=1$, and the Nyquist frequency at $\omega=2$. From (6) we get

$$c_m = c_{-m} = \begin{cases} 1/2 & m=0 \\ \dfrac{(-1)^{(m-1)/2}}{m\pi} & m=1,3,5,\ldots \\ 0 & m=2,4,6,\ldots \end{cases}$$

The normalized magnitudes of the first three of these filters are plotted vs. $\omega$ in Fig. 1:

Curve A: $D(z) = \frac{1}{\pi} + \frac{1}{2}z^{-1} + \frac{1}{\pi}z^{-2}$

Curve B: $D(z) = -\frac{1}{3\pi} + \frac{1}{\pi}z^{-2} + \frac{1}{2}z^{-3} + \frac{1}{\pi}z^{-4} - \frac{1}{3\pi}z^{-6}$

Curve C: $D(z) = \frac{1}{5\pi} - \frac{1}{3\pi}z^{-2} + \frac{1}{\pi}z^{-4} + \frac{1}{2}z^{-5} + \frac{1}{\pi}z^{-6} - \frac{1}{3\pi}z^{-8} + \frac{1}{5\pi}z^{-10}$ .

One disadvantage of this method is that the resulting digital filters are always polynomials in $z^{-1}$ and hence will not utilize past outputs. While this might be desirable when there are storage problems, the class of filters is very restrictive and in general will not use computing facilities efficiently.

As can be seen from the figure, another objectionable feature of this method is the ripple and overshoot that is characteristic of Fourier approximations. One way of alleviating this difficulty might be to use Fejér means[7] for the coefficients $c_k$. Very high order filters will be required, however, to achieve sharp cutoff and good rejection in the stop band. Because of the above disadvantages, we shall not pursue the subject further.

## IV. Z-TRANSFORMS OF CONTINUOUS FILTERS

Suppose that we design a continuous filter $F(s)$ which approximates the desired digital filter characteristics for $|\omega| \leq \pi/T$, and is small outside this range. If $F(s)$ has all its poles in the left-half plane, the z-transform of $F(s)$,

$$\mathcal{Z}[F(s)]\big|_{s=j\omega} = \frac{1}{T}\sum_{n=-\infty}^{\infty} F(j\omega - jn\frac{2\pi}{T}) , \qquad (9)$$

will be a stable digital filter with approximately the desired characteristics. The main difficulty with this is the addition of unwanted terms in (9) due to the aliasing of the filter function. If we use this idea, we must use high order filters with relatively low cutoff frequencies, so that the characteristic decreases fast in $|\omega| > \pi/T$.

This method requires that we z-transform our F(s). If F(s) is of a high order, this is a laborious task. Furthermore, once $\mathcal{Z}[F(s)]$ is determined, the calculation of its magnitude and phase for $s=j\omega$ will be a still more difficult job. This means that the error introduced by the aliasing will not be easy to assess.

Example 2.

To illustrate this problem, consider the third-order maximally flat Butterworth low-pass filter with unit cutoff frequency:

$$F(s) = \frac{1}{s^3 + 2s^2 + 2s + 1} .$$

The z-transform is

$$D(z) = \frac{0.3703z^{-1} + 0.1346z^{-2}}{1 - 0.3981z^{-1} + 0.2474z^{-2} - 0.04321z^{-3}} ,$$

where again we have taken $T=\pi/2$. The normalized magnitude is plotted as curve A in Fig. 2. Because of the high cutoff frequency and the low order of the filter, the effects of aliasing are quite pronounced - the cutoff is not sharp and the rejection is poor.

## V. CONVERSION TO THE APPROXIMATION PROBLEM FOR CONTINUOUS FILTERS

The difficulties associated with the methods so far presented can be eliminated by just reversing Guillemin's[6] procedure. More explicitly, our procedure will be as follows: We start with a desired digital filter frequency characteristic. (This may be a magnitude, phase angle, real part, or imaginary part.) By the inverse of the transformation used by Guillemin, we convert this to a continuous wave filter characteristic. This can then be approximated with a rational function, using any one of the many procedures available for continuous filters. By then reversing the transformation, we get a rational function of $z=e^{sT}$ whose frequency characteristic approximates the one given. This method allows us to apply all the known techniques for continuous filters to the approximation problem for digital filters.

The desired transformation is the familiar bilinear transformation

$$z = \frac{1 + \bar{s}}{1 - \bar{s}} ,$$

$$\bar{s} = \frac{z - 1}{z + 1} \quad , \tag{10}$$

where $\bar{s} = \bar{\sigma} + j\bar{\omega}$ is the frequency variable for continuous wave filters, and $z=e^{sT} = e^{(\sigma + j\omega)T}$ is the frequency variable for digital filters. As shown in Fig. 3, the entire $j\bar{\omega}$ axis is mapped onto the unit circle $z=e^{j\omega T}$ in the z plane. When $\bar{s}=j\bar{\omega}$, we have

$$z = e^{j\omega T} = \frac{1 + j\bar{\omega}}{1 - j\bar{\omega}} \quad , \tag{11}$$

or

$$\omega = \frac{2}{T} \arctan \bar{\omega} \quad . \tag{12}$$

The entire $\bar{\omega}$ axis is thus mapped into ~~strips~~ intervals on the $\omega$ axis $2\pi/T$ wide. Since the left-half $\bar{s}$ plane is mapped into the unit circle in the z plane, any continuous wave filter that has all its poles in the left-half $\bar{s}$ plane will become a stable digital filter under the transformation (10).

Suppose now that we are given some periodic function of $\omega$, $C(\omega)$, that is to be the desired characteristic (magnitude, phase, real or imaginary part) of a digital filter. $C(\frac{2}{T}\arctan \bar{\omega})$ will then be the corresponding characteristic for a continuous filter. We then approximate $C(\frac{2}{T}\arctan \bar{\omega})$ and arrive at a rational function of $\bar{s}$, say $F(\bar{s})$. Then $F(\frac{z-1}{z+1})$ will be a digital filter that approximates the desired characteristic.

Loosely speaking, we have taken the strip $|\omega| \leq \pi/T$ and stretched it out; done our approximation for a continuous filter; and then squeezed the $\bar{\omega}$ axis back into the original strip. Notice that there is no aliasing of the filter function. Although the $\bar{\omega}$ axis is compressed, many of the widely used approximation criterea, such as equal ripple, maximal flatness, etc., carry over directly to the digital filter case. If a continuous filter has magnitude $M(\bar{\omega})$, phase $\Phi(\bar{\omega})$, real part $R(\bar{\omega})$, and imaginary part $I(\bar{\omega})$; then the bilinear-transformed digital filter will have magnitude $M(\tan \omega T/2)$, phase $\Phi(\tan \omega T/2)$ in $|\omega| \leq \pi/T$, real part $R(\tan \omega T/2)$, and imaginary part $I(\tan \omega T/2)$.

<u>Example 3</u>.

Suppose, then, that we wish to approximate the same ideal low-pass filter as in examples 1 and 2. The desired filter magnitude characteristic has a perfect cutoff at $\bar{\omega} = 1$. We can therefore use the same Butterworth filter that we used in example 2:

$$F(\bar{s}) = \frac{1}{\bar{s}^3 + 2\bar{s}^2 + 2\bar{s} + 1} \quad .$$

When we let $\bar{s} = \frac{z-1}{z+1}$, this becomes the digital filter

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

whose normalized magnitude is plotted as curve B in Fig. 2, along with the z-transformed filter. Note that this is now a maximally flat digital filter. The point $\omega=2$ corresponds to $\bar{\omega} = \pm\infty$, and the filter magnitude at this point is zero. The phase characteristic is that of the continuous Butterworth filter (except for multiples of $2\pi$), compressed repetitively into strips, the same way as the magnitude characteristic.

Example 4.

As another illustration of this method, suppose we want a low-pass digital filter with equal ripple in the pass band. Take $T=\pi/2$, the cutoff at $\omega=\frac{1}{2}$, and the Nyquist frequency at $\omega=2$. The continuous filter must have its cutoff frequency at

$$\bar{\omega} = \tan \omega T/2 = \tan \pi/8 = 0.4142 .$$

Suppose, then, that we start with the $4^{th}$-order Tchebycheff filter with about 10% ripple ($\epsilon^2 = 1/5$), and a cutoff frequency at $\bar{\omega}=1$:

$$F(\bar{s}) = \frac{1}{\bar{s}^4 + 1.034\bar{s}^3 + 1.535\bar{s}^2 + 0.8306\bar{s} + 0.3062} .$$

If we substitute $\bar{s}/(0.4142)$ for $\bar{s}$, we get

$$F_1(\bar{s}) = \frac{1}{\bar{s}^4 + 0.4284\bar{s}^3 + 0.2633\bar{s}^2 + 0.05903\bar{s} + 0.009011},$$

which has a cutoff frequency at $\bar{\omega} = 0.4142$. We then substitute $\bar{s} = \dfrac{z-1}{z+1}$ to obtain the desired digital filter:

$$D(z) = \frac{1 + 4z^{-1} + 6z^{-2} + 4z^{-3} + z^{-4}}{1.760 - 4.703z^{-1} + 5.527z^{-2} - 3.225z^{-3} + 0.7849z^{-4}}.$$

Fig. 4 shows the normalized magnitude of this Tchebycheff equal ripple digital filter.

## VI. USING DIGITAL FILTERS TO MAKE CONTINUOUS WAVE FILTERS

We conclude with some remarks about constructing a continuous wave filter with a sampler, a digital filter $D(z)$, and a data reconstruction circuit $H(s)$. With reference to Fig. 5, the overall transfer function is

$$\frac{Y(s)}{X^*(s)} = D^*(s)H(s) \quad ,$$

where $D^*(s) = D(e^{sT})$ ,

and $X^*(s)$ is the Laplace transform of the sampled input. Note that the transfer function is with respect to the sampled (aliased) input signal. Thus, we must sample at a frequency at least twice as great as the bandwidth of $x(t)$ in order that the filter characteristic represent the action of the filter on the original signal in the range $|\omega| \leq \pi/T$. Any transmission outside this range does not represent transmission at all, but represents spurious harmonics of the input signal.

Let us assume for the purposes of discussion that $H(s)$ is a simple hold circuit (a zero-order hold), and that we want to convert a digital filter to a continuous filter whose magnitude characteristic is small for $|\omega| > \pi/T$. Ideally, then, we would want $|H(j\omega)|$ to be 1 in $|\omega| \leq \pi/T$, and zero elsewhere. Actually, we have

$$H(s) = \frac{1 - e^{-sT}}{sT} \quad , \tag{13}$$

and

$$|H(j\omega)| = \left| \frac{\sin \omega T/2}{\omega T/2} \right| .$$

$|H(j\omega)|$ has its first zero at $\omega = 2\pi/T$, which is twice the Nyquist frequency, and has lobes of appreciable magnitude well outside the range $|\omega| \leq \pi/T$. The overall transfer function therefore has spurious responses at high frequencies. This is an unavoidable consequence of an imperfect reconstruction device. It is evident that a fairly crude low-pass continuous filter placed at the output, with a cutoff frequency near the Nyquist frequency, would be effective in reducing the harmonics in the output. It is also desirable to compensate for the falling off of the magnitude of the hold circuit in the pass band by appropriately shaping the digital filter characteristic. This technique, plus the use of a low-pass continuous wave postfilter, will probably make it uneconomical to use hold circuits more complicated than a simple clamp.

Example 5.

As an example, we use the hold circuit of (13) with the D(z) of example 4. The resulting low-pass magnitude characteristic is shown in Fig. 6. As mentioned before, the high frequency pass bands are not truly in the transfer characteristic, but represent harmonics of the aliased input signal.

It is interesting to note that the filtering characteristic of our final system can be changed as fast as the coefficients in the digital computer program can be changed. If we used band-pass digital filters, for instance, we might then be able to use our computer system to replace a bank of fixed filters or a frequency sweeping system.

## REFERENCES


1. J. M. Salzer. "Frequency Analysis of Digital Computers Operating in Real Time." Proc. IRE, vol. 42 (February 1954), pp. 457-66.

2. R. B. Blackman and J. W. Tukey. The Measurement of Power Spectra. New York: Dover Publications, 1959. Reprinted from Bell System Technical Journal , vol. 37 (January and March, 1958).

3. U. Grenander and M. Rosenblatt. Statistical Analysis of Stationary Time Series. New York: John Wiley and Sons, 1957.

4. P. M. Lewis II. "Synthesis of Sampled-Signal Networks." IRE Trans. on Circuit Theory, vol. CT-5 (March 1958) pp. 74-7.

5. J. R. Ragazzini and G. F. Franklin. Sampled-Data Control Systems. New York: McGraw-Hill Book Company, 1958.

6. E. A. Guillemin. Synthesis of Passive Networks. New York: John Wiley and Sons, 1957; pp. 614-19.

7. B. J. Ley, S. G. Lutz and C. F. Rehberg. Linear Circuit Analysis. New York: McGraw-Hill Book Company, 1959; pp. 373-75.

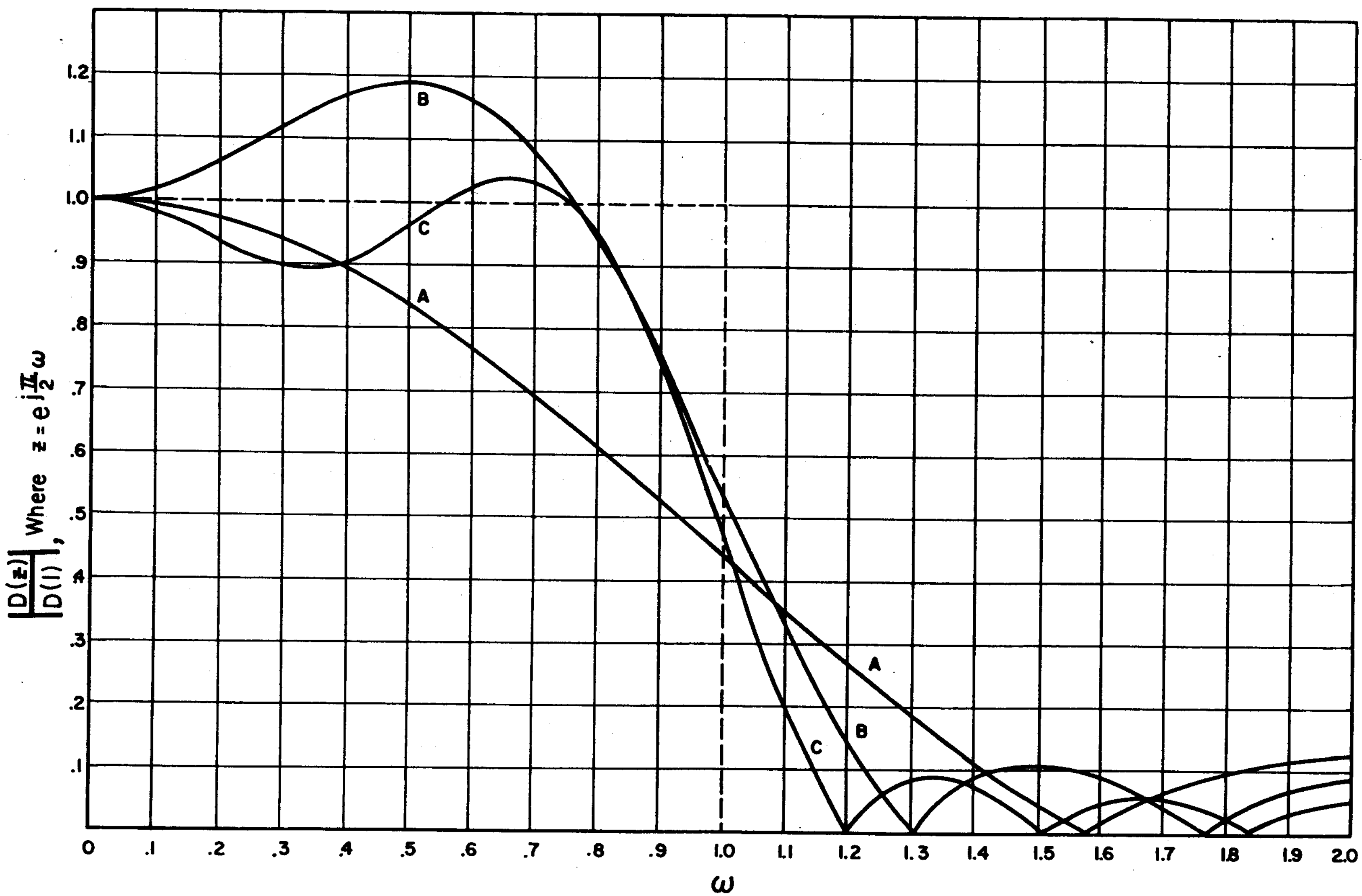


*Fig. 1*

Fourier Series Approximations To An Ideal Low-pass Digital Filter Magnitude Characteristic

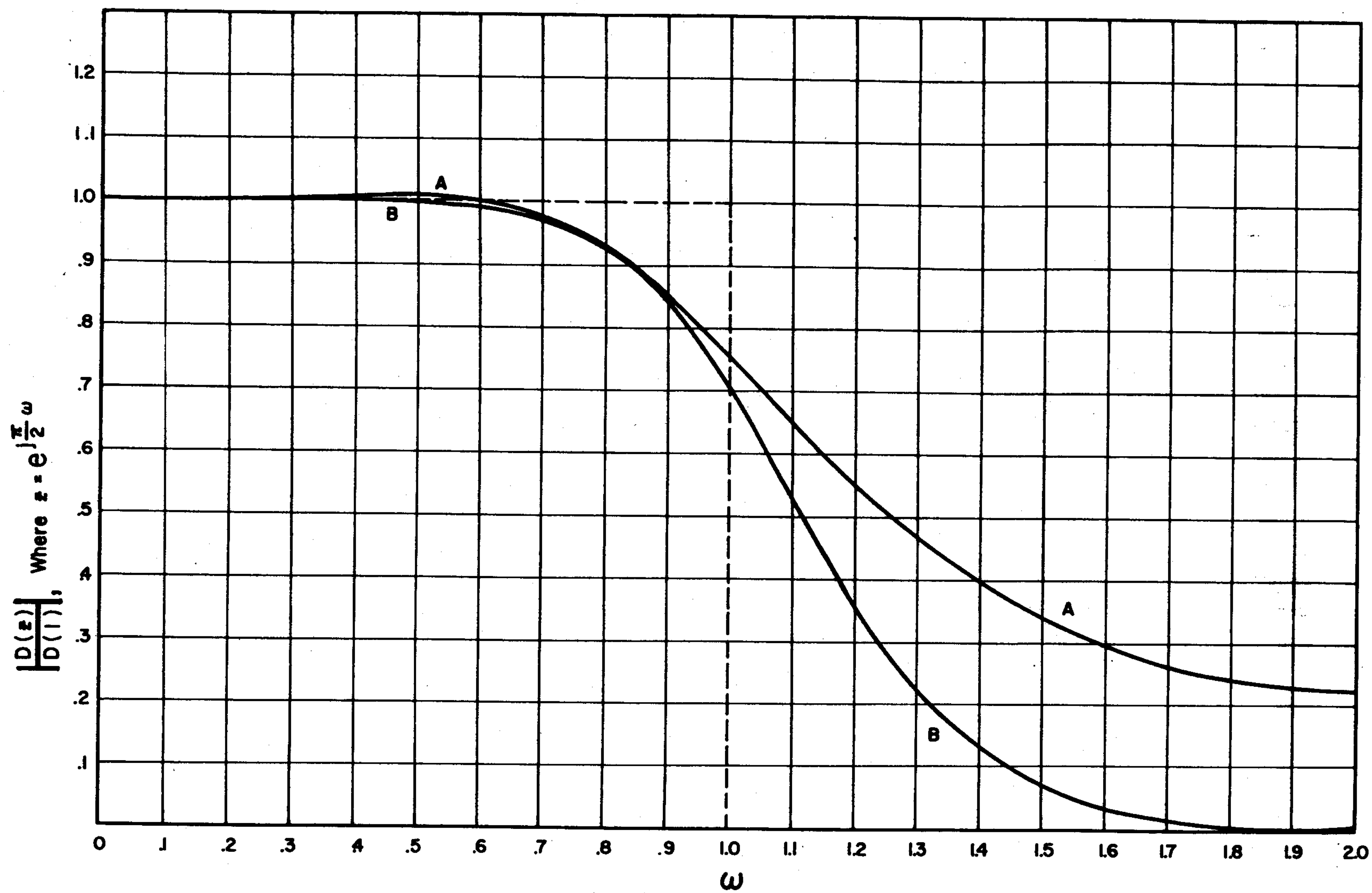


*Fig. 2*

Curve A: The Digital Filter Magnitude Characteristic Corresponding To The $z$ Transform Of A Third Order Butterworth Filter. Curve B: The Bilinear – Transformed Digital Filter Corresponding To The Same Butterworth Filter.

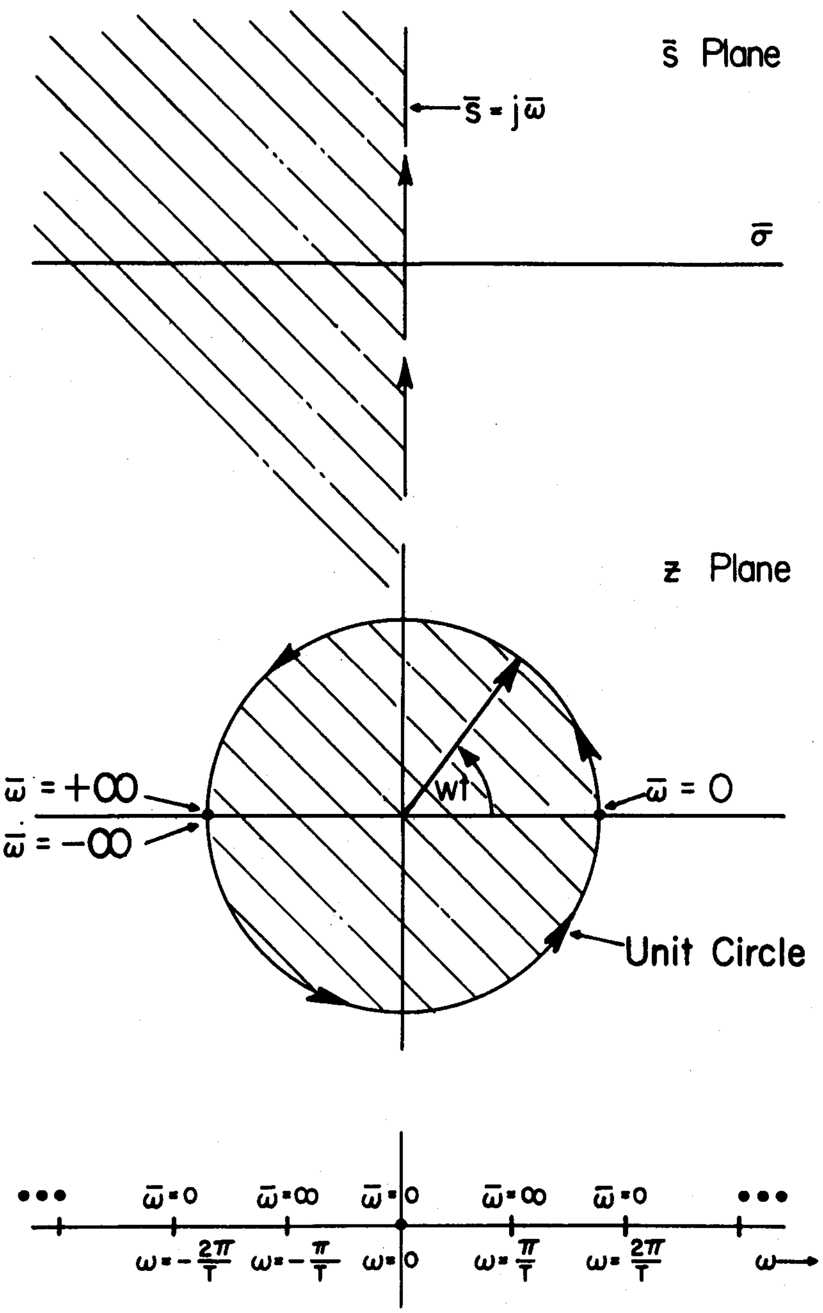


*Fig. 3*

The $\bar{s}$ Plane, The $z = e^{sT}$ Plane, And The $j\omega$ Axis;
When $z = \dfrac{1+\bar{s}}{1-\bar{s}}$ .

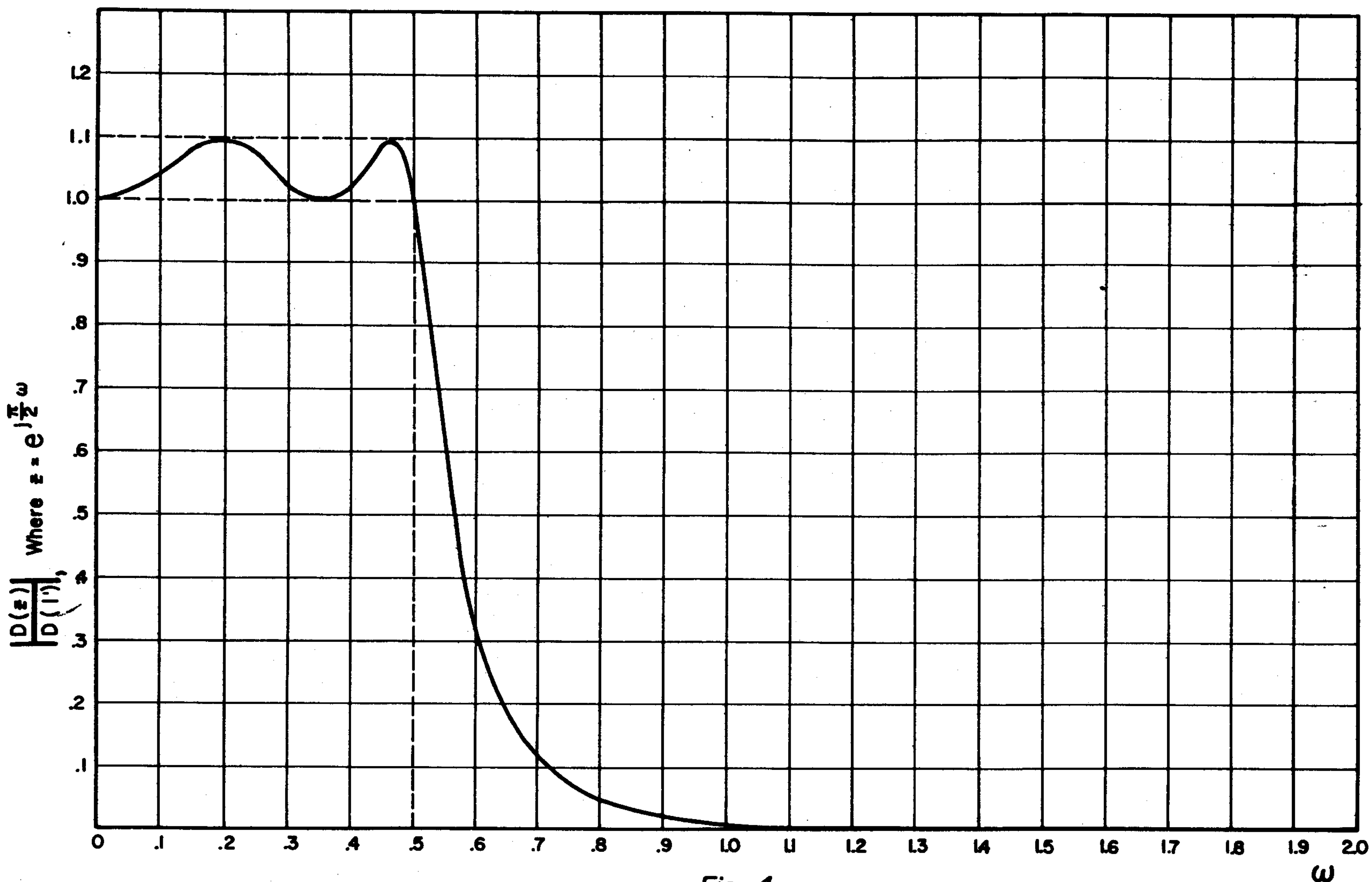


*Fig. 4*

The Normalized Magnitude Characteristic Of The Bilinear - Transformed Digital Filter Corresponding To A 4th. — Order Tchebycheff Filter With 10% Ripple.

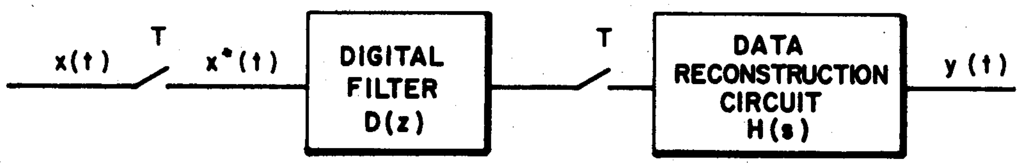


*Fig. 5*

A Continuous Wave Filter Constructed From A Digital. Filter.

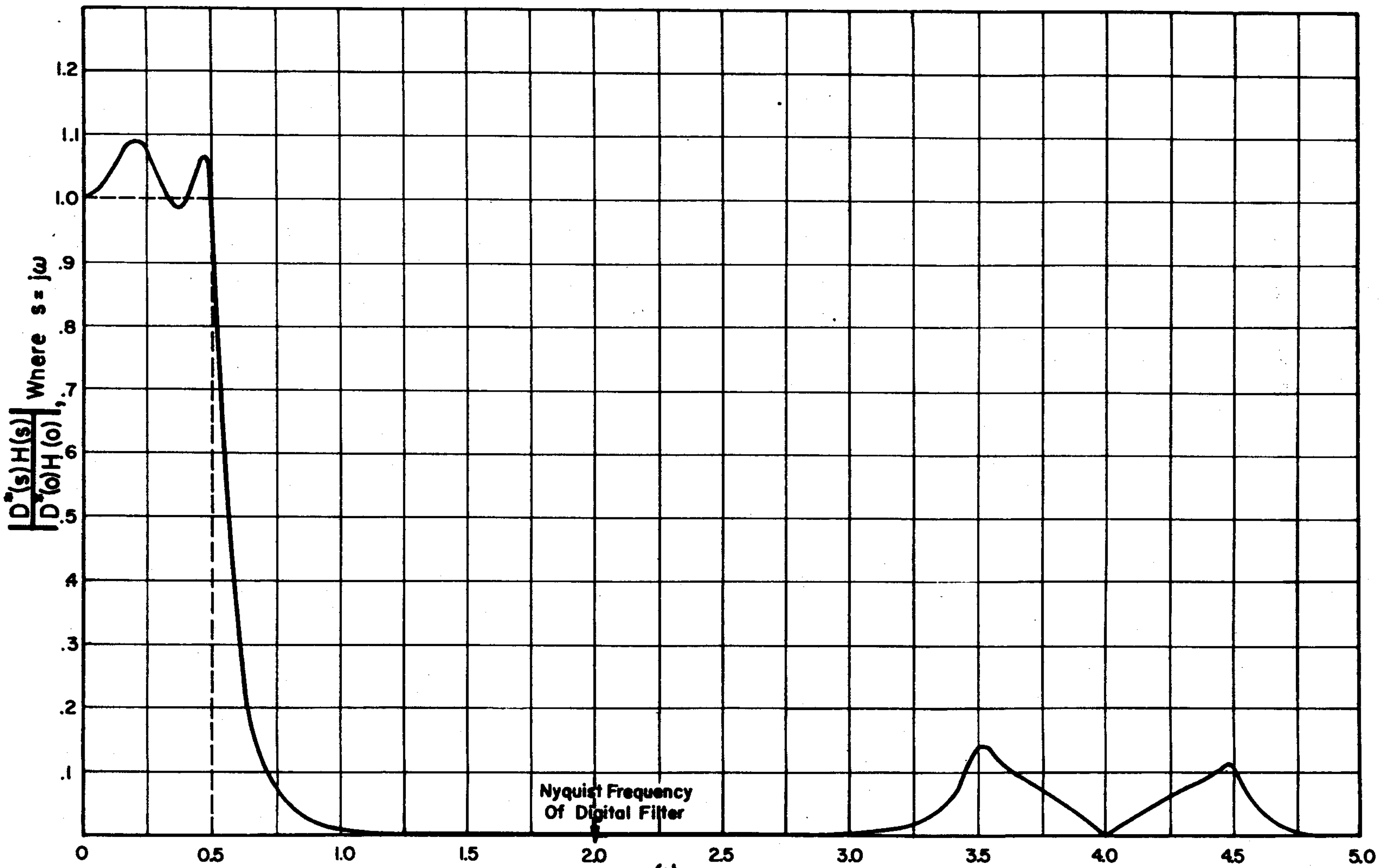


*Fig. 6*

The Normalized Magnitude Characteristic Of A Continuous Wave Low-pass Filter Constructed With A Digital Filter And A Zero-Order Hold Circuit.