Tutorial Paper

THE DESIGN OF DIGITAL FILTERS FOR BIOMEDICAL SIGNAL PROCESSING (in three parts)

PART 1: BASIC CONCEPTS

R.E. Challis* and R.I. Kitney†

ABSTRACT

The four main techniques used to describe signals and systems are reviewed: time domain impulse response, and Fourier, Laplace and z-transform methods. The design of simple digital filters is discussed and it is shown how z-transform methods lead to the construction of convenient recurrence relationships and the estimation of frequency response and mid-band gain.

Keywords: Electric filters, digital filters, signal processing, transform calculus.

INTRODUCTION

Digital filtering operations form an important part of the signal analysis procedures which are used in the study of many biological phenomena. This has been brought about to some extent by the increasing use of computers of various degrees of complexity in biological research. In medicine and physiology, digital filters have been used during the course of studies on all of the major systems of the body. The signals involved include blood pressure and flow, electrograms from the heart, brain and smooth and striated muscle, nerve action potentials and pressure signals from the lumina of the gut, the urinary tract and the uterus. The tasks required of the filters have been manifold, ranging from trend detection and the extraction or accentuation of those special features in a signal which are required to test a hypothesis down to the minimization or removal of interfering noise and artefact components in raw data. The personnel involved in biomedical signal processing activity derive from a wide range of academic backgrounds, ranging from engineering to psychology; it would appear that many of these workers experience difficulty in understanding the theoretical bases of digital filter theory and its relationship to continuous time filter theory as it applies to non-sampled signals or systems. The outcome is that a few simple digital filters which are known to work (the moving average, for example) get used over and over again, often in situations where the best choice of filter has not been established in any formal way. The purpose of these three papers is to explain the theory of digital filters in such a way that the reader gains a firm grasp of the relationships between discrete and continuous time filter systems and a clear understanding of the many practical aspects of digital filter design.

In the first paper, the Fourier, Laplace and time domain descriptions of signals and systems are summarized and the z-transform is introduced in the context of its relationship to those of Fourier and Laplace. The design of simple digital filters is discussed in detail and we show how z-transform methods lead to the construction of convenient recurrence relationships and the estimation of frequency response and mid-band gain. The second paper deals with more advanced design procedures and some of the many compromises which must be made in practice. These include the trade-off between filter impulse response shape and duration and side lobe generation in the frequency response, and the problems involved with

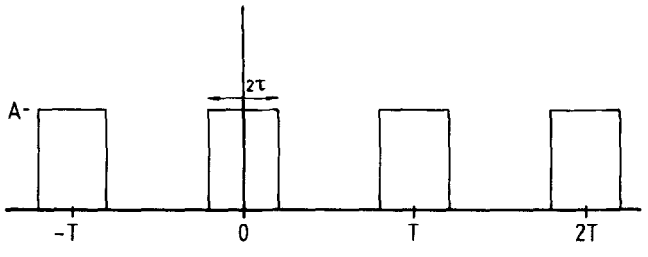


Figure 1 The periodic gate function – a train of rectangular pulses of height A, width 2τ repeating at intervals of T in the range $-\infty < t < +\infty$.

*Departments of Biophysics and Physiology, Chelsea College, University of London, UK
†Engineering in Medicine Laboratory, Department of Electrical Engineering, Imperial College, London, UK

0141-5426/82/040267-12 $03.00

the design of high pass digital filters. The bilinear transform is introduced as a technique by which the digital equivalent of analogue filter designs can be derived. In the third and final paper the methods developed in the previous two will be used to develop digital versions of Butterworth and Chebychev filters of various orders.

It is hoped that our readers will be convinced that a formal approach to the design of digital filters is worthwhile and indeed quite simple.

The exponential Fourier series, the Fourier transform and the sampled signal

A large number of text books[1,2,3] deal with the exponential Fourier series, the Fourier transform and the sampled signal and we do not propose to do more than summarize the essential concepts here. The basis of the Fourier series representation is that a given function can be represented by a summation of a set of orthogonal functions, and that the representation improves as the number of functions in the orthogonal set increases. Complex exponential functions are most commonly used because they not only form an orthogonal set, but also, any pair multiplied together forms another member of the set; they also have the useful property that their real and imaginary parts consist of sine and cosine functions.

A signal $f(t)$, expressed as its exponential Fourier series becomes

$$f(t) = \sum_{n=-\infty}^{\infty} F_n e^{jn\omega_0 t} \tag{1a}$$

where

$$F_n = \frac{1}{T} \int_{t_0}^{t_0+T} f(t) e^{-jn\omega_0 t}\, dt \tag{1b}$$

and

$$T = 2\pi/\omega_0.$$

1a and 1b are a transform pair. T is the time period over which $f(t)$ is assumed to repeat. The exponential Fourier series representation clearly applies to signals which repeat indefinitely with period T, or can be assumed to repeat for analytical convenience. A

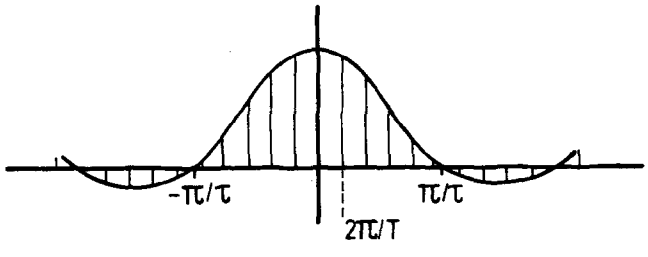


Figure 2 Fourier series representation of periodic gate function. The sin x/x envelope decays at a rate depending on the gate width and the frequency interval between spectral lines corresponds to the repetition frequency.

once-for-all signal, such as a single pulse does not repeat periodically and is described by the Fourier transform instead of the Fourier series representation. The transform is the limiting case of the series as period T tends towards infinity.

The Fourier transform pair has the form

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t}\, d\omega \tag{2a}$$

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t}\, dt \tag{2b}$$

In order to compare the series and transform representations we consider the sequence of rectangular pulses shown on *Figure 1*. Evaluating 1b we get

$$F_n = \frac{2A\tau}{T} \frac{\sin n\omega_0 \tau}{n\omega_0 \tau} \tag{3}$$

This function has a sinc (sin x/x) envelope but only exists for integer values of n; it therefore consists of vertical lines on an amplitude-frequency graph and is known as a line spectrum (*Figure 2*). The spectral lines occur at harmonics of the repetition frequency, $\omega_0 = 2\pi/T$. The distance between zeroes in the sinc envelope is given by $n\omega_0 = \pi/\tau$ and the number of spectral lines within the first lobe is $\pi/\omega_0\tau = T/2\tau$, truncated to the nearest integer. The amplitude at zero frequency increases proportionally with $2\tau/T$, as one might expect from the fact that an increase in pulse width contributes directly to the d.c. component in the signal. If τ remains fixed and we increase the period T such that it approaches infinity, the number of spectral lines on *Figure 2* increases without limit until the spectrum becomes the continuous function represented by the sinc envelope. This is the Fourier transform of a single rectangular pulse of width 2τ. The first zero still occurs at π/τ and the d.c. amplitude becomes $2A\tau$ instead of $2A\ \tau/T$.

If the period T remains finite but τ is reduced to zero in a limiting process such that each pulse has unit area, then the signal on *Figure 1* forms a train of equispaced unit impulses or Dirac functions in time known as a sampling function. Multiplication of this new sequence with any arbitrary time function performs the operation of sampling that function. The result of the multiplication process is a sequence of equispaced impulses whose weights correspond to the instantaneous values of $f(t)$ at the instants it was sampled (*Figure 3*). The interval between the impulses is known as the sampling interval and its reciprocal the sampling frequency.

The Fourier transform or spectrum of the sampling function is the limiting case of *Figure 2* with the first zero at infinite frequency. *Figure 4a* shows the spectrum; the interval between spectral lines is the sampling frequency. The spectrum of the sampled signal can be derived using the fact that multiplication in the time domain is equivalent to a convolution process in the frequency domain. If the spec-

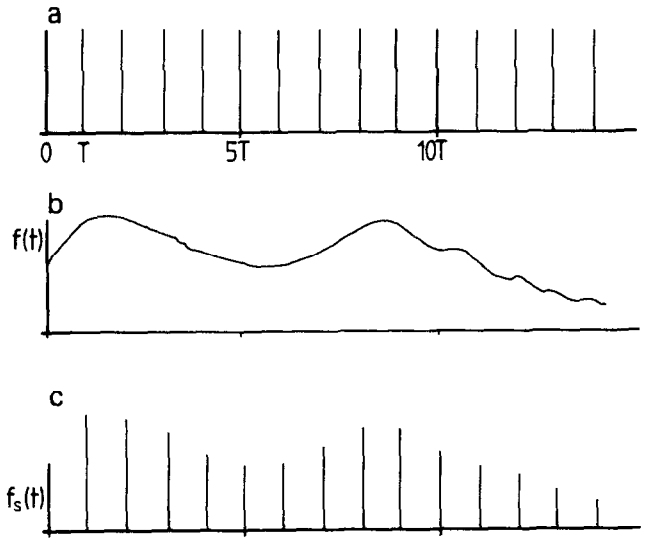


Figure 3 (a) The unit impulse sequence can be regarded as a limiting case of the periodic gate function. It has the sampling property that multiplication of an arbitrary signal (b) by the impulse sequence (a) results in the samples version of the signal shown in (c).

trum of the original time function is that shown in *Figure 4b* then the spectrum of its sampled version is that shown in *Figure 4c*. It consists of a series of identical repeats of the original spectrum, centred about multiples of the sampling frequency. The sampling process is only valid if the unsampled signal is bandlimited such that its highest frequency component, ω_N, is less than half the sampling frequency; that is $2\omega_N \leqslant \omega_s$. This is the condition for adjacent sidebands on *Figure 4c* to remain separate and is known as the Nyquist criterion. If this condition is not met 'aliassing' is said to have occurred. The term derives from the fact that a spectral component in the lower sideband centred on ω_s, will appear 'alias' a component in the upper sideband centred on zero frequency. If the Nyquist criterion is upheld then all the information contained in the unsampled waveform still exists in the sampled waveform. The original signal can be recovered from its sampled representation by low pass filtering to remove all frequency components outside the range $-\omega_s/2 > \omega > \omega_s/2$.

The sampled system and the digital representation of its responses

The behaviour of a system can be described by the relationships between its outputs and signals applied to its inputs. Inputs most frequently used in system characterization are sinusoids, unit step functions, unit impulse functions, pseudo-random binary sequences, and broadband noise. The unit impulse is particularly useful because the system response to it, the impulse response, can be used to build the response to more complicated input waveforms by convolution, if the system is linear. *Figure 5* shows a generalized linear system with a unit impulse function applied to input 3. The response at output 4 is the impulse response for this input-output pair h_{34} say. h_{34} defines the signal transfer properties of the system between input 3 and output 4. When a signal, $f_3(t)$, is applied to input 3 the output at port 4 can be determined by the convolution of the input signal with the system impulse response.

$$y_4(t) = \int_{-\infty}^{\infty} f_3(\tau)\, h_{34}(t-\tau)\, d\tau \qquad (3a)$$

In order to visualize this operation more clearly, imagine that $f_3(t)$ itself consists of a train of impulses (*Figure 5b*); when each of these passes through the system it will elicit the system impulse response at the output, weighted according to the amplitude of the input impulse. The whole output will consist of the superposition of these separate impulse responses. A continuous time input signal can be considered as a train of suitably weighted impulses spaced apart by an infinitesimally small time interval. The output will be built up of the corresponding impulse responses superimposed upon each other and spaced apart by the same time interval. Equation 3a expresses this process of superposition mathematically.

The Fourier transform of the impulse response yields the system frequency response between input 3 and output 4. The convolution operation, 3a, can be considered as a multiplication process in the frequency domain.

Symbolically,

$$Y_4(\omega) = F_3(\omega) \cdot H_{34}(\omega) \qquad (3b)$$

where

$$Y_4(\omega) \rightleftharpoons y_4(t)$$

$$F_3(\omega) \rightleftharpoons f_3(t)$$

and

$$H_{34}(\omega) \rightleftharpoons h_{34}(t)$$

are Fourier transform pairs.

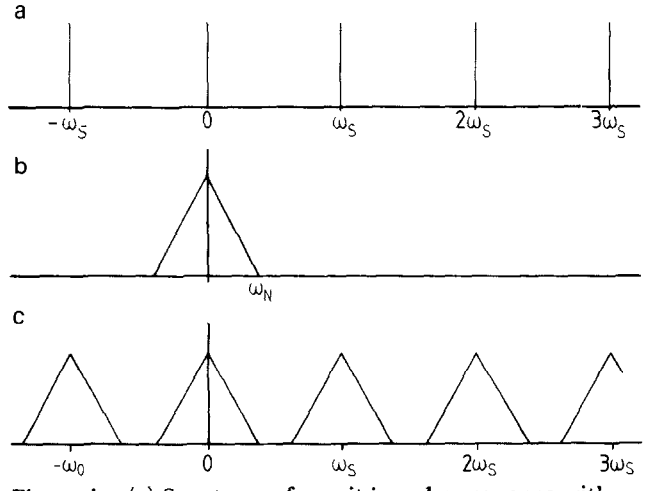


Figure 4 (a) Spectrum of a unit impulse sequence with sample interval $T = 2\pi/\omega_s$.
(b) Spectrum of unsampled signal.
(c) Spectrum of sampled version of *b*.

A simple two port system has only one input and one output and the subscripts are no longer required. The use of a digital computer as part of a system or to simulate a system, requires that the system itself be considered as a sampled entity. A sampled system impulse response has a spectrum which has all the properties of a sampled signal spectrum; in particular, the spectrum repeats identically at intervals of the sampling frequency.

A digital filter is a sampled two-port system and its sampling interval must be such that all the frequency components required for its complete description are contained in the interval $0 < \omega < \omega_s/2$. This condition may not be satisfied where a digital filter is designed by simply digitising the impulse response of a hardware filter. We shall see in paper II how this problem can be overcome in practice.

The Laplace transform, system transfer function and the *s*-plane

There are some continuous time functions for which the Fourier Transform does not converge (for example a ramp or step function) and cannot therefore be used for analysis. This prevents the application of the Fourier transform to a range of problems associated with the transient response of systems and filters. The Laplace transform is a generalisation of the Fourier transform which contains a convergence factor $e^{-\sigma t}$, thus

$$L[f(t)] = \int_{-\infty}^{\infty} f(t)e^{-\sigma t}e^{-j\omega t}dt \qquad (4)$$

where $L[f(t)]$ denotes the Laplace transform of $f(t)$. The transform is equivalent to describing $f(t)$ by an orthogonal set of exponentially growing or decaying sinusoids each of which is defined by the decay constant σ and radian frequency, ω, in one complex number, $s = \sigma + j\omega$. The Laplace transform is a function of s and equation 4 is frequently written

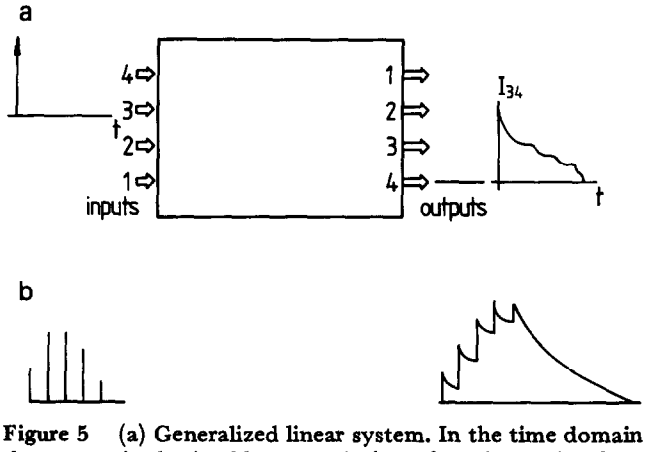


Figure 5 (a) Generalized linear system. In the time domain the output is obtained by convolution of the input signal with the system impulse response I_{34}.
(b) A signal consisting of a five equispaced impulses is applied to input 3. The output at port 4 consists of a corresponding set of five impulse responses, I_{34}, separated by the same interval, weighted according to the input impulse weights, and superimposed upon each other.

$$F(s) = \int_{-\infty}^{\infty} f(t)e^{-st}dt \qquad (5)$$

If $f(t)$ is causal, i.e., it only exists for $t > 0$, the lower limit of integration is 0 and the result is the unilateral or one sided Laplace transform. The Laplace transform of the impulse response of a system is known as the system transfer function, $G(s)$, which in the most general terms consists of the ratio of two polynomials in s.

$$G(s) = \frac{a_0 + a_1 s + a_2 s^2 \ldots\ldots + a_n s^n}{b_0 + b_1 s + b_2 s^2 \ldots\ldots + b_k s^k} \qquad (6)$$

Factorizing the numerator and denominator we get

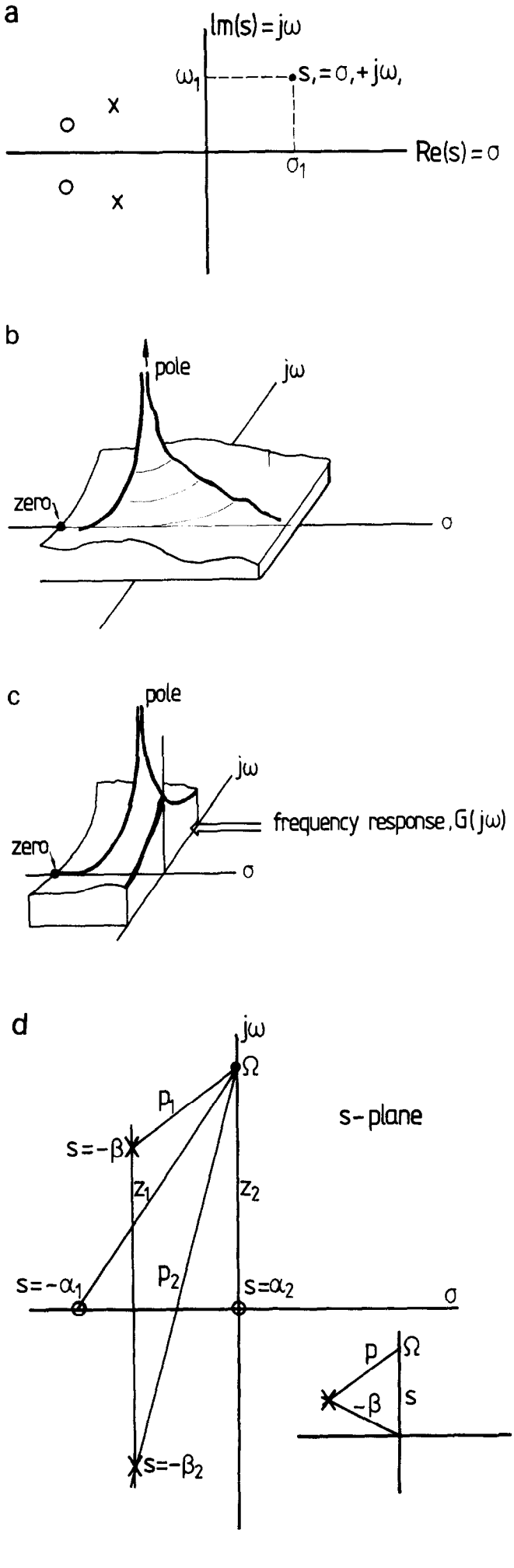
$$G(s) = \frac{(s+\alpha_1)(s+\alpha_2)\ldots\ldots(s+\alpha_n)}{(s+\beta_1)(s+\beta_2)\ldots\ldots(s+\beta_k)} \qquad (7)$$

Since s is a complex number it can be plotted on an Argand diagram and this is normally called the s-plane (*Figure 6a*). The function $G(s)$ can be represented by a surface on the s-plane which would come vertically out of the paper on *Figure 6a;* the height of the surface above the s-plane is the modulus of $G(s)$. (*Figure 6b*). The roots of the numerator of equation 7 ($s = -\alpha$) give values of s for which $G(s) = 0$; they are known as the zeros of $G(s)$. Similarly, the roots of the denominator yield the values of s for which $G(s)$ is infinite and these are known as the poles of $G(s)$. Poles and zeros are usually plotted on the s-plane as X's and O's respectively. *Figure 6b* shows $G(s)$ plotted as a surface on the s-plane; a pole shows as a physical 'pole' (a vertical upshoot of infinite height) and zeros are the points where the surface touches the s-plane. The frequency response of a system is merely $G(s)$ evaluated for $s = j\omega$, that is along the $j\omega$ axis on the s-plane. If we take a slice through the surface representing $G(s)$ on 6b corresponding to a cut along the $j\omega$ axis then the exposed edge of our surface corresponds to the frequency response of the system, it being no more than a two dimensional graph of $G(j\omega)$ versus ω (*Figure 6c*). The frequency response can be calculated by means of a graphical construction on the s-plane. *Figure 6d* shows poles and zeros marked on the s-plane and a point on the $j\omega$ axis marked as Ω. The system response at frequency Ω is simply the product of all the zero vector magnitudes, Z_i, divided by the product of all the pole vector magnitudes, P_i. Hence, for the pole-zero pattern on 6d we get

$$G(j\omega) = \frac{z_1 z_2}{p_1 p_2} \qquad (8)$$

The phase response, $\angle G(j\omega)$ can be obtained from the same diagram, it is merely the sum of all the zero vector angles minus the sum of all the pole vector angles. The vectors are assumed to be drawn from the zeros or poles to $j\Omega$ and their angles are the anticlockwise angular displacements from the horizontal axis.

The s-plane pole-zero description of systems has



proved to be an invaluable aid to the design of continuous time filters and in the analysis of continuous time systems. It is limited in its application to sampled data or systems because it cannot conveniently deal with the infinitely repeating sampled signal spectra. We shall see in the following sections how the z-transform overcomes this problem and how s-plane designs of filters can be translated into their z-transform equivalent.

The z-transform, the s-plane and the z-plane

We have seen that a sampled signal or impulse response consists of a sequence of equispaced samples, T seconds apart, whose weights are equal to the amplitude of the original function at the instants at which it was sampled. Such a sequence may be represented in the time domain as

$$f(t) = x(0)\,\delta(t) + x(1)\,\delta(t-T) + \ldots + x(n)\,\delta(t-nT)$$

or

$$f(t) = \sum_{n=-\infty}^{\infty} x(n)\,\delta(t-nT) \qquad 9(a)$$

The Laplace Transform of this sequence is

$$F(s) = x(0) + x(1)\,e^{-sT} + x(2)\,e^{-2sT} + \ldots . x(n)\,e^{-nsT} \qquad (10)$$

since the time delays nT are equivalent to multiplying by e^{-nsT} under the Laplace Transformation.

It is convenient to use the shorthand notation, $z = e^{sT}$, whence equation 10 becomes

$$X(z) = x(0) + x(1)\,z^{-1} + x(2)z^{-2} + \ldots . x(n)z^{-n}$$

or

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n} \qquad (11)$$

$X(z)$ is the z-transform of the sequence $x(n)$.

For $s = j\omega$ equation 11 becomes the discrete Fourier transform, which is the Fourier transform of a signal which only exists at intervals of T. The real and imaginary parts of the variable z can be mapped onto an Argand diagram known as the z-plane. The special relevance of the z-transform to sampled data systems can be illustrated by considering the relationship between quantities expressed

Figure 6 The nature and properties of the s-plane.
(a) The s-plane is an Argand diagram representation of the complex number s. The real (horizontal) axis represents σ whilst the imaginary axis (vertical) represents $j\omega$. Poles and zeros in the function $G(s)$ are represented by a surface on the s-plane. (b) At zeros the surface is coincident with the plane and at poles its height above the plane is infinite.
(c) A section through the surface taken along the $j\omega$ axis yields the frequency response $G(j\omega)$ versus $j\omega$.
(d) The s-plane showing a complex pole pair with zeros on the σ axis. The insert shows that a vector from $s = j\Omega$ to a pole at $-\beta$ has a length $p = s + \beta$ (since $-\beta + p = s$).

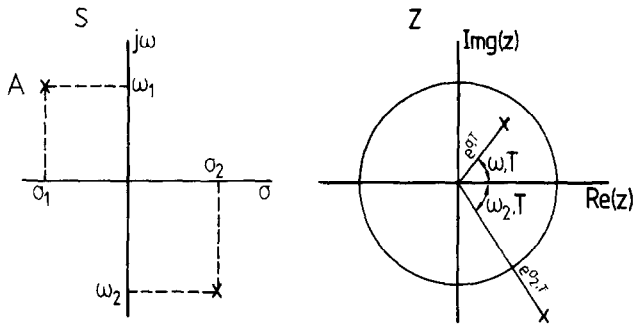


Figure 7 Mapping the s-plane on to the z-plane under the transformation $z = e^{sT}$. Points to the left of the $j\omega$ axis map inside the unit circle on the z-plane.

on the s-plane and the way in which they map onto the z-plane.

Figure 7 shows the s and z-planes and we consider the mapping of point A on the s-plane onto the z-plane. Now

$$z = e^{\sigma T} e^{j\omega T}$$

On the z-plane point A is described by the vector $z_1 = e^{j\omega_1 T}$. It has magnitude $r_1 = e^{\sigma_1 T}$ and id phase $\omega_1 T$ radians. r_1 will be <1 for $\sigma_1 < 0$ and >1 for $\sigma_1 > 0$. The $j\omega$ axis corresponds to $z = 1.e^{j\omega T}$; it therefore maps onto a unit circle on the z-plane. The s-plane origin ('d.c.') maps to $z = +1$. At points on the $j\omega$ axis corresponding to multiples of the sampling frequency we have $\omega = N\omega_s = N2\pi/T$. All these points map to $z = +1$ on the z-plane; the z-transform thus conveniently compresses the repeats of sampled signal spectra into a single trajectory round the z-plane unit circle. To further illustrate this we consider the s and z-plane plots shown on *Figure 8*. The pole-zero configurations on the s-plane consist of zeros at multiples of the sampling frequency flanked by pole pairs marked α and β. The z-plane map shows that all these sets of a single zero and two poles map into a single zero and pole pair.

Under the relationship $z = e^{sT}$ points on the s-plane which lie to the left of the $j\omega$ axis map to points inside the unit circle on the z-plane. Some authors define $z = e^{-sT}$ and under this relationship the mapping of the two halves of the s-plane onto the z-plane is reversed.

Simple digital filters

A digital filtering operation consists of an operation on an input sequence, $x(n)$, by a filtering sequence, $h(n)$, to produce an output sequence, $y(n)$. *Figure 9* illustrates an arbitrary sequence $x(n)$ and a simple rectangular window for $h(n)$ which consists of five unit impulses. The operation of $h(n)$ on $x(n)$ consists of a convolution and we consider the process stepwise:

We begin calculating $y(n)$ at $n = 3$ and perform the following operations:

1st pass let $y(3) = \{x(1) + x(2) + x(3) + x(4) + x(5)\}/5$

2nd pass shift $h(n)$ forward one interval,
let $y(4) = \{x(2) + x(3) + x(4) + x(5) + x(6)\}/5$

3rd pass shift $h(n)$ forward one more interval,
let $y(5) = \{x(3) + x(4) + x(5) + x(6) + x(7)\}/5$

and so on.

The operation, shift, add and divide by 5 is continued until $h(n)$ has been passed over all the input samples. Such a filter is known as a *moving average* filter and $h(n)$ is called its *weighting function*. The filter has the effect of smoothing out undulations in $x(n)$ and is thus a low pass filter. An $h(n)$ with more impulses would smooth more strongly and would thus have a sharper cut off. Other weighting functions are common and *Figure 10* illustrates the triangular window, which has the same effect as a rectangular window of 5 unit impulses passed over the data twice. In the sections which follow, we consider how the properties of the z-transform may be used to reduce the computational effort involved in the filtering operations and how they lead to a simple technique for establishing the frequency response of the filter.

The z-transform of digital filters, the auto regressive form and frequency response estimation

The simple description of the digital filtering operation outlined in the last section can be formalised using the z-transform method and the number of computing operations required for its execution can be greatly reduced when the filter is put into its autoregressive form. In this technique the nth output term of the filter is calculated from both

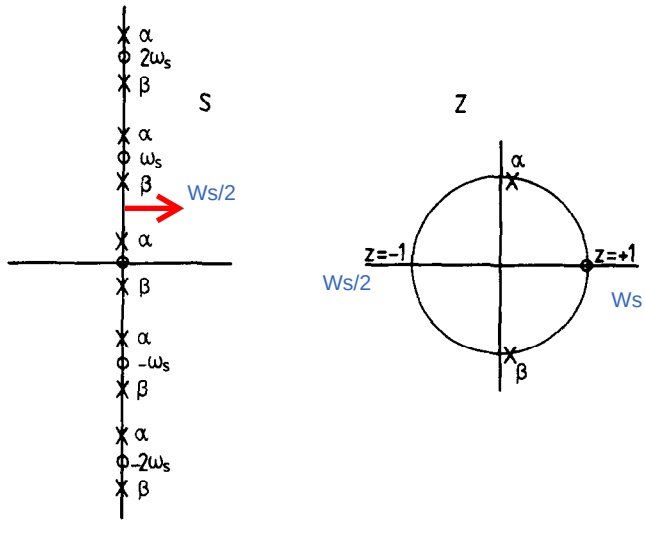


Figure 8 Mapping the $j\omega$ axis from the s-plane to the z-plane. Pole-zero combinations which repeat at intervals ω_s on the s-plane are compressed into one revolution on the z-plane.

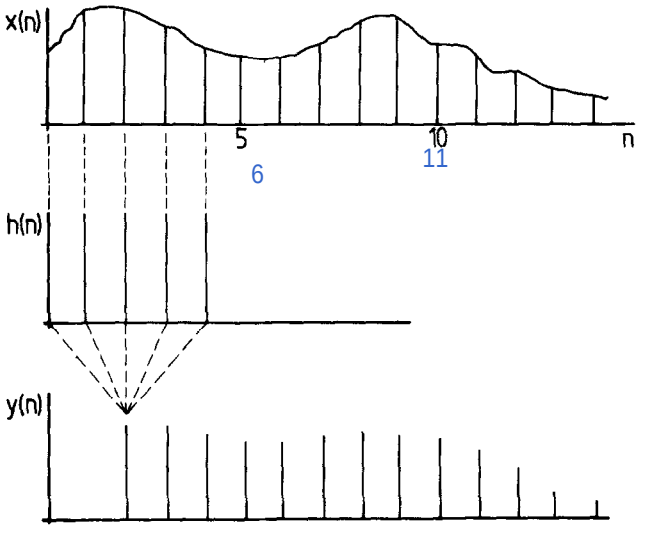


Figure 9 A sampled signal, $x(n)$ and 5 point moving average filter $h(n)$; each output sample, $y(n)$, is formed by taking the average of five successive input samples, the effect is to low pass filter the signal.

previous output terms and input terms, the total number of terms generally being less than required for non-autoregressive operation.

The operation of convolution can be described as:

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)\, h(n-k) \qquad (12)$$

It can be shown that[2] equation 12 is equivalent to a multiplication of z-transforms:

$$Y(z) = H(z)\, X(z) \qquad (13)$$

where

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

and

$$H(z) = \sum_{n=-\infty}^{\infty} h(n)\, z^{-n} \qquad (14)$$

For the simple 5 unit impulse moving average filter considered in the last section

$$H(z) = \sum_{n=0}^{4} z^{-n} = \frac{1-z^{-5}}{1-z^{-1}} \qquad (15)$$

Combining 13 and 15 we get

$$Y(z) = z^{-1}\, Y(z) + X(z) - z^{-5}\, X(z) \qquad (16)$$

Remembering that multiplication by the operator z is equivalent to a shift of minus one sampling interval, we can easily derive the inverse transform of equation 16. It is

$$y(n) = y(n-1) + x(n) - x(n-5) \qquad (17)$$

Equation 17 represents the autoregressive form of the filter and is known as the recurrence relationship.

[Note that multiplication by Z^m corresponds to shifting $y(n)$ to $y(n+m)$ in the untransformed sequence. Multiplication by z^1 shifts the sequence $y(n)$ backwards one sample so that we operate on a sample one interval later, i.e. $y(n+1)$. Multiplication by z^{-m} shifts the sequence $y(n)$ forwards so that we operate on an earlier sample, $y(n-m)$.]

Each output term, $y(n)$ is evaluated as a function of the previous output term and two input terms. The advantage here is that only three terms are required to calculate each output term, instead of five input terms described previously. As the number of samples in $h(n)$ increases autoregressive operation becomes relatively more advantageous since, for a simple moving average filter of N samples duration equation 17 becomes

$$y(n) = y(n-1) + x(n) - x(n-N) \qquad (18)$$

the number of samples required for the evaluation of each output sample is still only three. Similar economies can be obtained with filters of much greater complexity.

The frequency response of the digital filter is simply obtained by evaluating $H(z)$ around the unit circle on the z-plane. This corresponds to the substitution $z = e^{j\omega t}$. For the N-impulse moving average filter we get

$$H(j\omega) = \frac{1-e^{-jN\omega T}}{1-e^{-j\omega T}} \qquad (19)$$

rearranging and taking the modulus we get

$$|H(j\omega)| = \frac{\sin \frac{N\omega T}{2}}{\sin \frac{\omega T}{2}} \qquad (20)$$

The response for $N = 5$ is shown in *Figure 12b*. It is unity for $\omega T = 0$ and $\omega T = 2\pi$ (the sampling frequency) and contains N lobes between these limits.

This method for frequency response estimation is convenient for fairly simple filters of the type described here, but is not well suited to the rapid estimate of frequency response of more complex filters. Further, synthesis of a filter to a given frequency response specification would be difficult using this technique. The z-plane description of a digital filter greatly facilitates these tasks and is described in the next section.

The z-plane description of a digital filter

In the most general terms, the z-transform of a digital filter consists of the ratio of two polynomials in z.

$$H(z) = \frac{c_0 + c_1 z + c_2 z^2 \ldots\ldots + c_n z^n}{d_0 + d_1 z + d_2 z^2 \ldots\ldots + d_k z^k} \qquad (21)$$

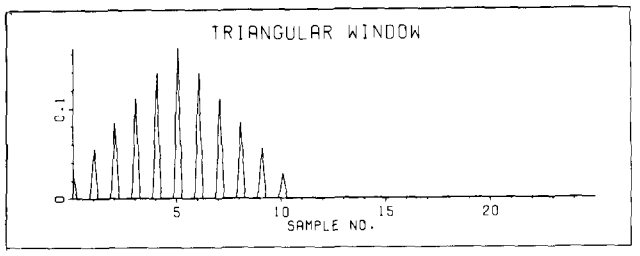


Figure 10 Triangular window weighting function.

This is analogous to the way in which the system transfer function has previously been described by the ratio of polynomials in the Laplace variable. Factorizing equation 21, $H(z)$ becomes

$$H(z) = \frac{(z+\gamma_1)(z+\gamma_2)\ldots\ldots(z+\gamma_n)}{(z+\Delta_1)(z+\Delta_2)\ldots\ldots(z+\Delta_k)} \quad (22)$$

The roots of the numerator and denominator yield the poles and zeros of $H(z)$ and these can be plotted on the z-plane diagram by setting $z = -\gamma$ for zeros and $z = -\Delta$ for poles. The z-plane pole-zero plot can be used to build up the filter frequency response. *Figure 11* shows a z-plane diagram with three poles and two zeros. The filter response at any frequency, Ω, is obtained by plotting the vectors from the poles and zeros to the point on the unit circle corresponding to $\omega = \Omega$. These are shown on *Figure 11* as γ' and Δ'. Since $\gamma' = z + \gamma$ for a zero at $-\gamma$ and $\Delta' = z + \Delta$ for a pole at $-\Delta$ we get

$$H(j\Omega) = \frac{\gamma'_1\,\gamma'_2}{\Delta'_1\,\Delta'_2\,\Delta'_3} \quad (23)$$

This construction can be used to estimate the frequency response by eye. To obtain the estimate one 'moves' round the unit circle, starting at $z = 1$, in order to follow the $j\omega$ axis. The effect of the proximity of poles and zeros can be considered separately: if one moves (a) close to a zero, the response will approach zero; (b) close to a pole the response will tend towards infinity; (c) close to a coincident pole-zero pair the response will approach unity; (d) near a pole-zero pair in close proximity, but not coincident, the response tends towards infinity if the pole is closer and towards zero if the zero is closer. Points distant from the pair will be relatively unaffected, since the two tend to cancel when their vectors approach equality. This effect is illustrated by γ_1 and Δ_1 on *Figure 11*. They have little effect over most of the frequency band, but produce a notch in the response at $\omega T = \pi$. The frequency response is shown on *Figure 11*.

The pole-zero pattern for simple moving average filters is very easily obtained from the z-transform. Considering

$$H(z) = \frac{1-z^{-N}}{1-z^{-1}} = \frac{1-z^N}{z^{N-1}(1-z)} \quad (24)$$

we see that $H(z)$ has a single pole at $z = 1$ and $N-1$ poles at the origin, $z = 0$. The factors in the numerator correspond to the N roots of $z^N = 1$. Recalling that $e^{\pm j2\pi} = 1$ we get

$$z^N = e^{j2\pi m} \quad (25)$$

Whence the z-plane zeros correspond to values of z given by

$$z = e^{j2\pi m/N} \quad (26)$$

In other words, an N impulse moving average filter has N equispaced zeros on the unit circle.

Figure 12a illustrates the pole-zero pattern of the 5 point moving average filter considered earlier. The frequency response, from zero up to the sampling frequency is shown in *Figure 12b*.

Digital filter mid band gain

The successive summations involved in a digital filtering operation result in a pass band gain which is not necessarily unity. It is therefore necessary to determine the mid band gain of the filter so that the filter output sequence can be scaled down to yield unity gain at the pass band centre. For low pass filters we need to establish the gain at zero frequency, which is given by

$$K_0 = \lim_{z \to 1} [H(z)] \quad (27)$$

for our N-point moving average filter we get

$$K_0 = \lim_{z \to 1} \left[\frac{1-z^{-N}}{1-z^{-1}}\right] \quad (28)$$

Setting $z = 1 + \delta$

$$K_0 = \lim_{\delta \to 0} \left[\frac{1-(1+\delta)^{-N}}{1-(1+\delta)^{-1}}\right]$$

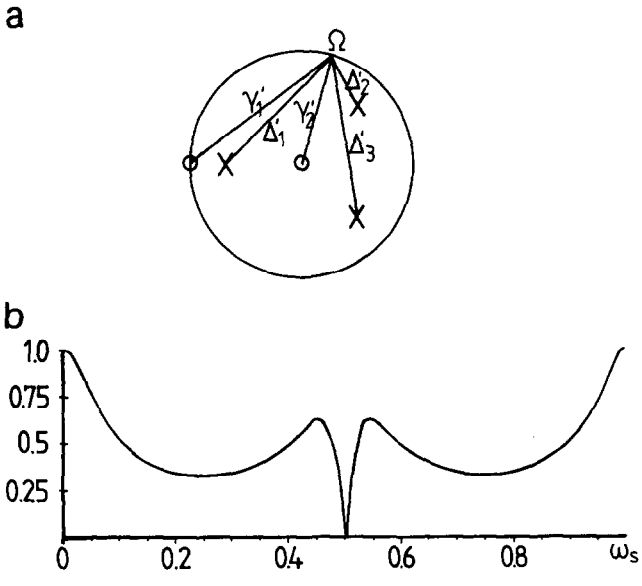


Figure 11 (a) Pole-zero plot on the z-plane. The filter response at a given frequency, Ω, is the modulus of the product of zero vectors, divided by the pole vectors. The corresponding frequency response is shown in (b). The notch at $\omega_s/2$ is formed by the close proximity of the zero γ_1, and the pole Δ_1.

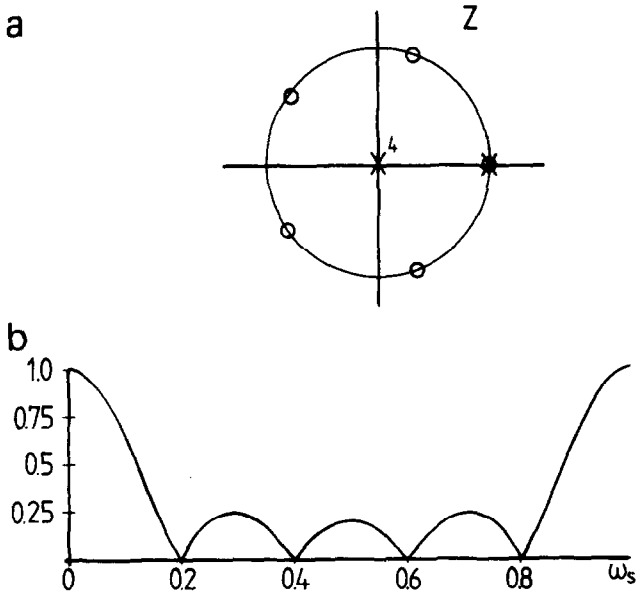


Figure 12 (a) Pole-zero pattern of 5 point moving average filter together with its frequency response (b).

whence

$$K_0 = N \tag{29}$$

For other types of filter, which may not be low pass, K is obtained by letting z approach its mid band value in a similar limit process.

Digital filter phase response

Considering again our elementary moving average filter we investigate the effect of dividing or multiplying $H(z)$ by integer powers of z. On the z-plane this is equivalent to placing poles or zeros at the origin, $z = 0$. They have no effect on the shape of the filter impulse response except that the response is shifted in time, forwards one sample interval for each pole at $z = 0$ and backwards for each zero at $z = 0$. For example, the two filters

$$H_0(z) = \frac{1 - z^{-5}}{1 - z^{-1}} \tag{30a}$$

$$H_{-3}(z) = \frac{(1 - z^{-5})\, z^{-3}}{1 - z^{-1}} \tag{30b}$$

have the following recurrence relationships:

$$H_0(z): y(n) = y(n-1) + x(n) - x(n-5) \tag{31a}$$

$$H_{-3}(z): y(n) = y(n-1) + x(n-3) - x(n-8) \tag{31b}$$

The operations involved are identical except that the nth output of the $H_{-3}(z)$ is derived from input samples which occur three intervals earlier than those used by $H_0(z)$; the output is thus delayed in the case of $H_{-3}(z)$. In a similar fashion, multiplying $H_0(z)$ by z^{+3} would produce a filter whose nth output corresponded to input samples $x(n+3)$ and $x(n-2)$; it is advanced in time such that it depends upon input samples which occur later than its own position in time. This is a feasible operation in situations where data is processed off-line, since all 'future inputs' can be made available to the filter simultaneously.

Time shifts of this nature are equivalent to phase shifts in the frequency domain. Multiplication by z^m has the effect of multiplying the frequency response by $e^{jm\omega T}$. That is to say, it advances the phase spectrum by $m\omega T$ radians. Digital filters whose poles and zeros lie either at $z = 0$ or on the unit circle have a linear phase characteristic; this means that the phase shifts imposed by the filter are proportional to frequency. Linear phase shift is a very useful characteristic since it implies that the temporal relationships between frequency components in a signal are preserved in a filtering operation and only the amplitude is affected by the filter.

A zero phase shift filter does not advance or delay any frequency components. This characteristic is particularly useful when one wishes to study phase spectra on a digital computer, since it avoids the $\pm \pi$ range limitation associated with software ARCTAN functions. For a digital filter to impose zero phase shift, it must have an impulse response which is symmetrical about $t = 0$. For our 5 point moving average filter the z-transform becomes

$$H(z) = \frac{(1 - z^{-5})\, z^{+2}}{1 - z^{-1}} \tag{32}$$

which gives the recurrence relationship

$$y(n) = y(n-1) + x(n+2) - x(n-3) \tag{33}$$

It should be clear that one of the conditions for such symmetry is that the filter impulse response consists of an odd number of samples; this implies that the z-plane will be divided by an odd number to get the positions of the filter zeros.

Synthesis of simple digital filters

We will now describe the steps involved in the design of simple filters, which have all their zeros either on the unit circle or at the origin of the z-plane.

Step 1. Remembering that the unit circle corresponds to the $j\omega$ axis between zero frequency and ω_s, mark out band centres and the approximate position of the zeros on the unit circle.

Step 2. Reorganise all these points to lie on an integer dividend of the unit circle.

Step 3. Superimpose poles and zeros at band centres.

Step 4. Build the factors in the z-transform, $H(z)$. Each factor will consist of pairs of poles or zeros, symmetrical about the real axis. Consider a zero pair situated at radius r, frequency ω. These form a factor in $H(z)$ thus

$$H(z) = (z - re^{j\theta})\,(z - re^{-j\theta}) \tag{34a}$$

where $\theta = \omega T$

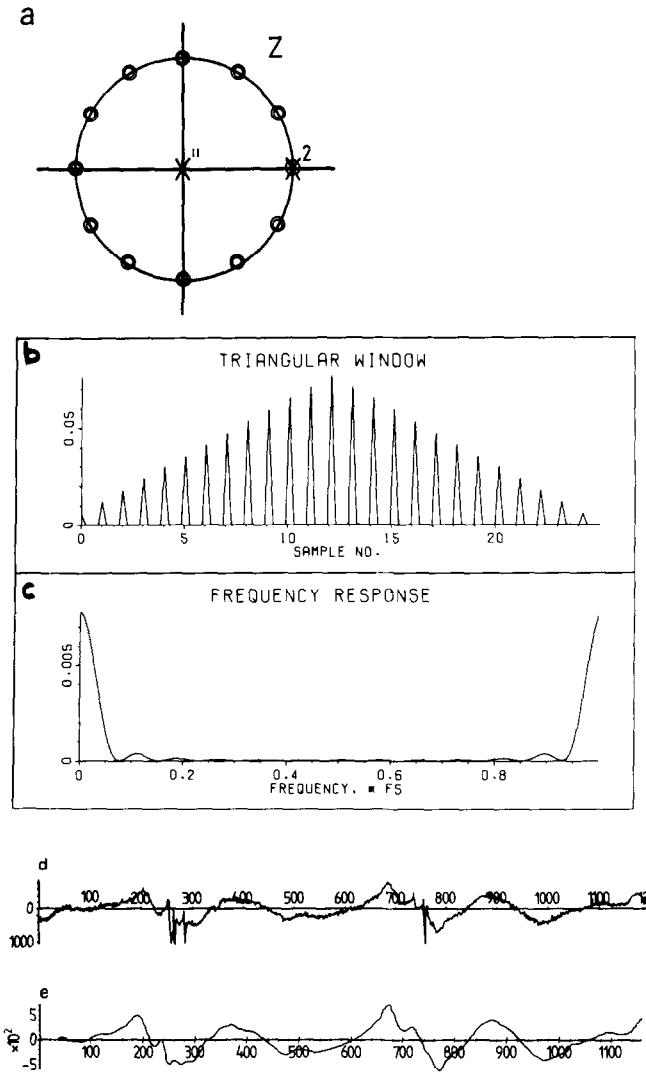


Figure 13 The z-plane (a), impulse response (b) and frequency response (c) of a simple zero phase low pass filter. (d) shows a twelve second record of e.m.g. taken from the serosal surface of the canine duodenum. It consists of high frequency components (spikes) superimposed on low frequency activity (the pace-setter potential). (e) shows the e.m.g. record after filtering with this low pass design.

multiplying out we get

$$H(z) = z^2 + r^2 - 2r\,z\cos\theta \qquad (34b)$$

on the unit circle $r = 1$ and

$$H(z) = z^2 + 1 - 2z\cos\theta \qquad (34c)$$

If $\cos\theta$ can be chosen such that it is the quotient of two integers then the coefficients in the recurrence relationship become integer and the filter can be implemented using integer arithmetic operations, with subsequent savings in speed and data array space.

Step 5. Multiply out the relationship $Y(z)/X(z) = H(z)$ and derive the recurrence relationship.

Step 6. Test the filter on a specially constructed or chosen data set. If low side bands are required, the filter can be applied to the data twice by squaring $H(z)$ before deriving the recurrence relationship.

We now consider three design examples.

Example 1: Lowpass filter. First zero at about 8 Hz. Sampling rate 100 Hz. Low side band levels.

The z-plane of *Figure 13* has the first zero at 8.3 Hz. Poles and zeros have been squared to give low side band levels. For $1 + N$ divisions of the unit circle the z-transform is

$$H(z) = \frac{(1 - z^{N+1})^2}{z^N(1 - z)^2} \qquad (35a)$$

We have chosen $N = 11$, dividing the unit circle by 12. The recurrence relationship is

$$y(n) = 2y(n-1) - y(n-2) + x(n-N-2) - 2x(n-1) + x(n+N) \qquad (35b)$$

with midband gain (i.e. at d.c.)

$$K_0 = (N+1)^2 \quad (\mathrm{H(z))2}$$

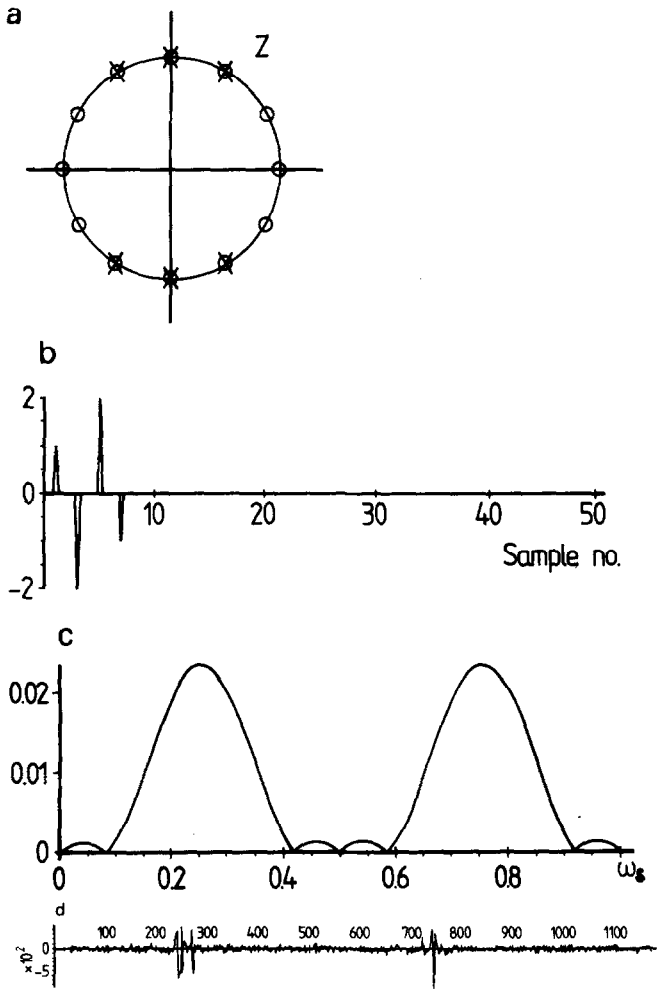
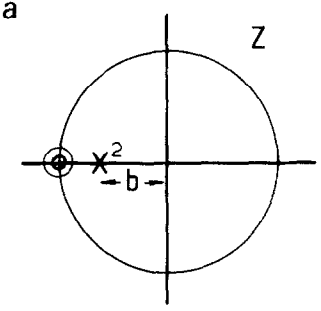


Figure 14 The z-plane (a), impulse response (b), and frequency response (c) of a simple band pass filter, band centre at $\omega = \omega_s/4$. (d) shows the result when this filter is applied to the e.m.g. data of the last example.

The impulse response of this filter is shown on *Figure 13b* and the frequency response on *13c*. *Figure 13d* shows an e.m.g. signal obtained from the serosal surface of the canine duodenum, digitized at 100 Hz. The signal consists of two main components, both of which are of interest to the physiologist; these are action potential 'spikes' and slow rhythmic oscillations in transmembrane potential, the so-called pacesetter potential or Basic Electrical Rhythm. It is convenient to separate these two components for independent analysis and digital filters can be used to advantage in the process. *Figure 13e* shows the e.m.g. after processing by this simple triangular window filter. Clearly the pacesetter potential has been successfully extracted and the higher frequency components suppressed.

a



Example 2: Band pass filter, sampling rate 100 Hz, mid band at 25 Hz, zeros at 25 ± 8 Hz, side band levels not critical. Referring to the z-plane diagram of *Figure 14a* we see that a filter with six pole zero pairs at $z = e^{\pm j/3}$, $e^{\pm j/2}$ and $e^{\pm 2/3}$ should be suitable.

j pi/3 j pi/2 j 2pi/3

The numerator of $H(z)$ is $(1-z^{12})$, yielding twelve equispaced zeros. The denominator is built up using equation 34c and $H(z)$ becomes

$$H(z) = \frac{1 - z^{12}}{(1 + z^2)(1 + z^2 - z)(1 + z^2 + z)} \qquad (36)$$

Multiplying out the denominator we get the relation ship

$$(1 - z^{12})X(z) = (z^6 + 2z^4 + 2z^2 + 1)Y(z) \qquad (37)$$

which yields the recurrence relationship

$$y(n) = x(n-6) - x(n+6) - 2y(n-2) - 2y(n-4) - y(n-6) \qquad (38)$$

the gain of the filter evaluated at mid band ($z = \pm j$) is 6. *Figures 14b* and *c* show the impulse and frequency responses of this filter. Its effect on the e.m.g. signal of the last example is shown on *Figure 14d;* the low frequency component in the signal has been suppressed and the high frequency 'spiking' activity has been extracted by the filter.

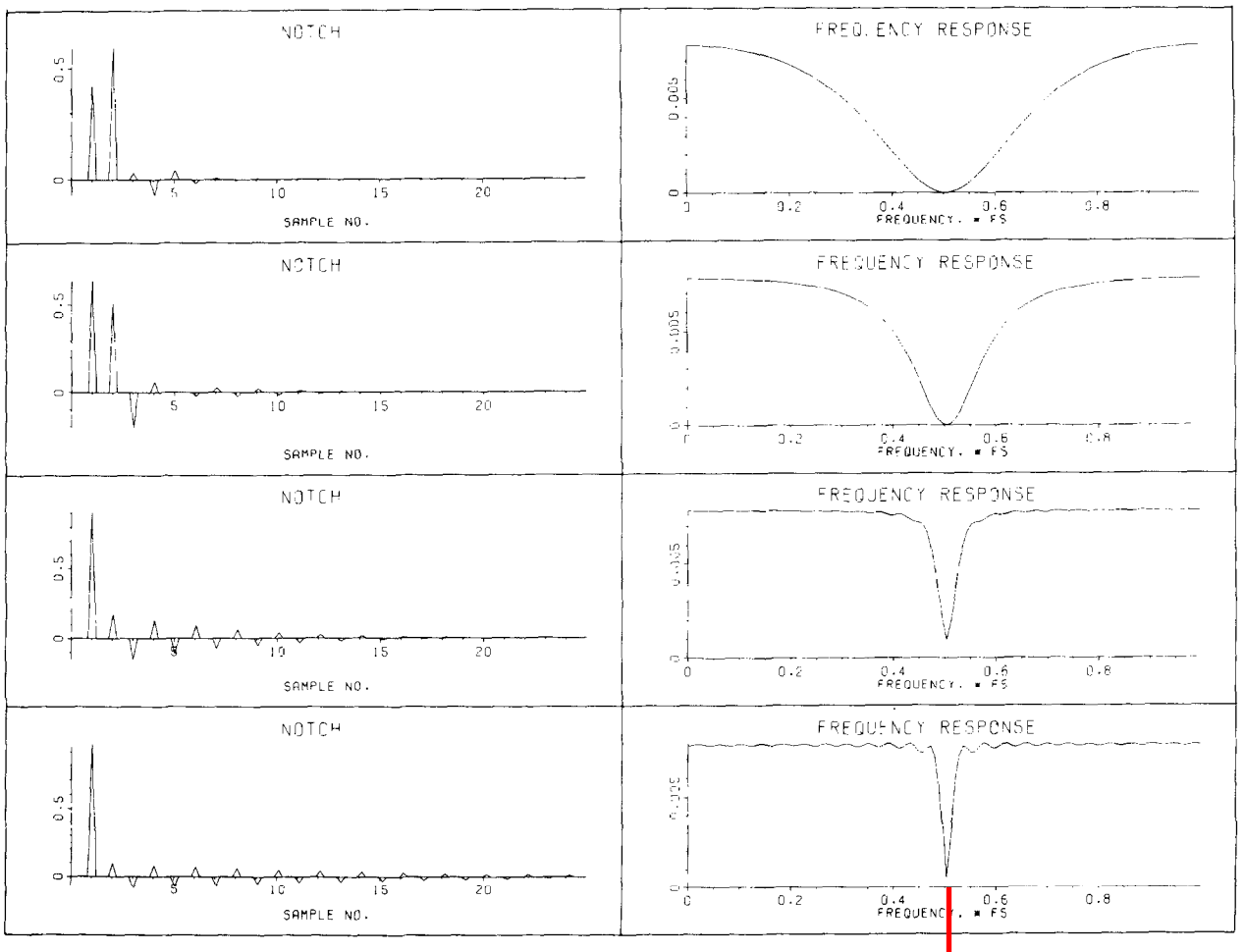


Figure 15 (a) z-plane plot of a notch filter at $\omega_s/2$ with cut off rate determined by parameter b. (b) shows the impulse responses for values of b equal to 0.3 (top), 0.6, 0.9, and 0.95 (bottom). (c) The frequency responses corresponding to the four values of b.

Ws/2 fim da banda

resposta simétrica

Example 3: In many physiological experiments mains frequency artefact often appears as an unwanted component in the signal under investigation. In this example a notch filter is designed to reject 50 Hz noise in data sampled at 100 Hz. The cut off rate is adjustable, by varying the parameter b in the range $0 < b < 1$.

The z-plane diagram of *Figure 15* shows a double zero at $z = -1$, which corresponds to 50Hz. The double pole at $z = -b$ controls the sharpness of cut off. The z-transform is

$$H(z) = \frac{(1+z)^2}{(b+z)^2} \tag{39}$$

which yields the recurrence relationship

$$y(n) = x(n) + 2x(n-1) + x(n-2) - 2by(n-1) - b^2y(n-2) \tag{40a}$$

and zero frequency gain

$$K_0 = \frac{4}{(1+b)^2} \tag{40b}$$

Figure 15 shows the frequency response of this filter for four values of b. It is clear that as the poles approach the zeros the influence of the zeros becomes more strongly localized at the notch frequency.

CONCLUSION

This paper has reviewed briefly the four main techniques by which signals and systems may be described. These were the time domain impulse response and the Fourier, Laplace and z-transform methods. The relationships between these techniques provide a powerful conceptual framework within which analytical procedures can be investigated. z-plane methods enable digital filters to be designed for specific applications and provide a convenient way by which suitable recurrence relationships can be worked out. The midband gain and frequency response of digital filters can be obtained directly from their z-transform.

Examples have been given of fairly basic digital filters which have simple z-plane patterns, many of which have their poles and zeros concentrated on the unit circle and at $z = 0$. Although such filters suffice for many biomedical signal processing exercises there are, however, some applications where the properties required cannot be met by these simple designs and more advanced techniques are indicated. In the two papers which follow, more advanced design methods will be introduced and the design compromises which are frequently found to be necessary in practice will be discussed. The relationships between the filter impulse response shape and side lobe generation in the frequency response will be studied in the context of a set of low pass filters. The problems associated with the design of high pass digital filters will be discussed and techniques will be given which provide for the derivation of high pass and band-pass filters from low pass designs. The Bilinear Transformation will be introduced as a technique which provides for the design of the digital equivalents of analogue filters avoiding serious aliasing phenomena in the frequency response. In the final paper these techniques will be used to develop digital Butterworth and Chebychev filters.

REFERENCES

1 Lynn, P.A. *An Introduction to the Analysis and Processing of Signals.* Macmillan London 1973.
2 Gabel, R.A. and Roberts, R.A. *Signals and Linear Systems.* Wiley New York 1973.
3 Oppenheim, A.V., and Schafer, R.W. *Digital Signal Processing.* Prentice-Hall New Jersey 1975.