GURUNANAK INSTITUTE OF TECHNOLOGY

157/F, Nilgunj Road, Panihati
Kolkata -700114
Website: www.gnit.ac.in
Email: info.gnit@jisgroup.org

Approved by A.I.C.T.E., New Delhi
Affiliated to MAKAUT, West Bengal

OCW

Digital Signal Processing

Course Level: Undergraduate

Credit: 3

Prepared by:

Ms. Anurima Majumdar
Assistant Professor (ECE)

Dr. Suparna Biswas
Associate Professor (ECE)

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

COURSE OBJECTIVES:

1. To analyze the Z-transform with the help of properties of ROC, Poles and Zeros , inverse z-transform using Contour integration - Residue Theorem, Power Series expansion and Partial fraction expansion.
2. To introduce students with transforms for analysis of discrete time signals and systems.
3. To understand the digital signal processing, sampling and aliasing.
4. To use and understand implementation of digital filters.
5. To study filter design techniques.
6. To study Discrete Fourier Transforms.
7. To study Fast Fourier Transforms.
8. To study fixed point and floating point digital signal processors.

COURSE OUTCOMES:

1. Able to analyze discrete time systems in frequency domain and their region of convergence using Z Transforms.
2. Able to define discrete systems in the Frequency domain using Fourier analysis tools like DFT, FFT.
3. Able to interpret the properties of discrete time signals in frequency domain.
4. Able to analyze discrete time signals and systems in frequency domain.
5. Able to describe the digital signal processing, sampling and aliasing.
6. Able to understand to implement digital filters.

PREREQUISITE:

Prerequisites for Digital Signal Processing are required a thorough understanding of various signals, systems, and the methods to process a digital signal and also the knowledge of arithmetic of complex numbers and a good grasp of elementary calculus. The questions reflect the kinds of calculations that routinely appear in Signals. The candidates are expected to have a basic understanding of discrete mathematical structures. The candidates will be required to do arithmetic only on integers, simple fractions, and simple radicals. In addition, you need to know the representation of sine, cosine, in exponential form.

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

Elementary Idea of DSP:

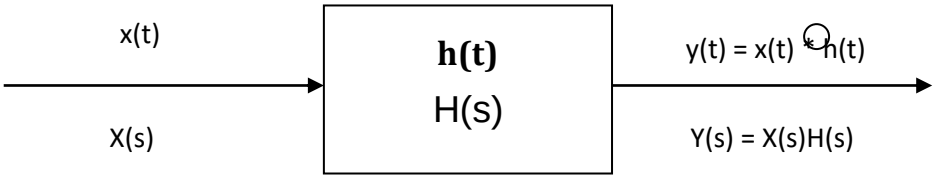
What is DSP?

DSP, or Digital Signal Processing, as the term suggests, is the processing of signals by digital means. A *signal* in this context can mean a number of different things. Historically the origins of signal processing are in electrical engineering, and a signal here means an electrical signal carried by a wire or telephone line, or perhaps by a radio wave. More generally, however, a signal is a stream of information representing anything from stock prices to data from a remote-sensing satellite.

Module I
Introduction to Z-Transforms

Introduction

A linear system can be represented in the complex frequency domain (s-domain where $s = \sigma + j\omega$) using the LaPlace Transform.



Where the direct transform is:

$$L\{x(t)\} = X(s) = \int_{t=0}^{\infty} x(t)\varepsilon^{-st}dt$$

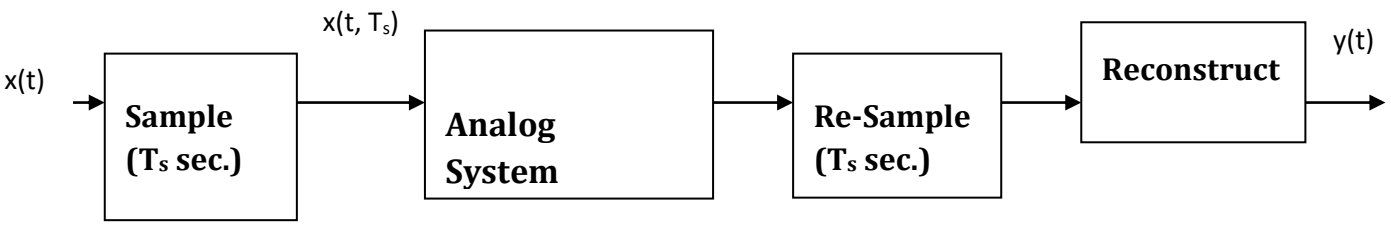
And x(t) is assumed zero for $t \leq 0$

The Inversion integral is a contour integral in the complex plane (seldom used, tables are used instead)

$$L^{-1}\{X(s)\} = x(t) = \frac{1}{2\pi j}\int_{s=\sigma-j\infty}^{\sigma+j\infty} X(s)\varepsilon^{st}ds$$

Where σ is chosen such that the contour integral converges.

If we now assume that x(t) is ideally sampled as in:



Where:

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

$$x_n = x(n*T_s) = x(t)\big|_{t=n*T_s}$$

and

$$y_n = y(n*T_s) = y(t)\big|_{t=n*T_s}$$

Analyzing this equivalent system using standard analog tools will establish the z-Transform.

Sampling

Substituting the Sampled version of x(t) into the definition of the LaPlace Transform we get

$$L\{x(t,T_s)\} = X_T(s) = \int_{t=0}^{\infty} x(t,T_s)\varepsilon^{-st}dt$$

But

$$x(t,T_s) = \sum_{n=0}^{\infty} x(t) * p(t - n*T_s)$$

(For x(t) = 0 when t < 0)

Therefore

$$X_T(s) = \int_{t=0}^{\infty}\left[\sum_{n=0}^{\infty} x(n*T_s) * \delta(t - n*T_s)\right]\varepsilon^{-st}dt$$

Now interchanging the order of integration and summation and using the sifting property of δ-functions

$$X_T(s) = \sum_{n=0}^{\infty} x(n*T_s)\int_{t=0}^{\infty}\delta(t - n*T_s)\varepsilon^{-st}dt$$

$$X_T(s) = \sum_{n=0}^{\infty} x(n*T_s)\varepsilon^{-nT_s s}$$

(We are assuming that the first sample occurs at t = 0+)

if we now adjust our nomenclature by letting:

$z = \varepsilon^{sT}$, $x(n*Ts) = x_n$, and $X(z) = X_T(s)\big|_{z=\varepsilon^{sT}}$

$$\boxed{X(z) = \sum_{n=0}^{\infty} x_n z^{-n}}$$

Which is the direct z-transform (one-sided; it assumes $x_n = 0$ for $n < 0$).

The inversion integral is:

$$\boxed{x_n = \frac{1}{2\pi j}\oint_c X(z)z^{n-1}dz}$$

(This is a contour integral in the complex z-plane)

(The use of this integral can be avoided as tables can be used to invert the transform.)

To prove that these form a transform pair we can substitute one into the other.

$$x_k = \frac{1}{2\pi j}\oint_c \left[\sum_{n=0}^{\infty} x_n z^{-n}\right] z^{k-1} dz$$

Now interchanging the order of summation and integration (valid if the contour followed stays in the region of convergence):

$$x_k = \frac{1}{2\pi j}\sum_{n=0}^{\infty} x_n \oint_c z^{k-n-1} dz$$

If "C" encloses the origin (that's where the pole is), the Cauchy Integral theorem says:

$$\oint_c z^{k-n-1} dz = \begin{matrix} o & for\ n \neq k \\ 2\pi j & for\ n = k \end{matrix}$$

And we get $x_k = x_k$

A Short Table of z-Transforms

f(t) (sampled)	F(z)	Region of Convergence
U(t)	$\frac{z}{z-1}$	|z| > 1
δ_{n-k}	z^{-k}	|z| > 1
t	$\frac{Tz}{(z-1)^2}$	|z| > 1
t^2	$\frac{T^2 z(z+1)}{(z-1)^3}$	|z| > 1
ε^{at}	$\frac{z}{z-\varepsilon^{aT}}$	|z| > ε^{at}
sin(βt)	$\frac{z*\sin(\beta T)}{z^2-2z*\cos(\beta T)+1}$	|z| > 1
cos(βt)	$\frac{z*[z-\cos(\beta T)]}{z^2-2z*\cos(\beta T)+1}$	|z| > 1

Properties of the z-Transform

The z-transform has properties that are analogous to those of the LaPlace Transform. The following table has some of the more useful ones listed.

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

$$X(z) \equiv \sum_{n=-\infty}^{\infty} x[n] z^{-n} \quad \square \quad x[n] \equiv \frac{1}{2\pi j} \oint_C X(z) z^{n-1} dz$$

where C is a closed contour that includes z=0

	Signal		z-Transform
	$x[n]$	□	$X(z)$
Superposition	$ax[n]+by[n]$	□	$aX(z)+bY(z)$
Time Shifting	$x[n-n_0]$	□	$z^{-n_0}X(z)$
	$e^{j\Omega_0 n}x[n]$	□	$X(e^{-j\Omega_0}z)$
	$z_0^n x[n]$	□	$X\left(\frac{z}{z_0}\right)$
	$a^n x[n]$	□	$X(a^{-1}z)$
Time inversion	$x[-n]$	□	$X(z^{-1})$
Time Convolution	$x[n]*y[n]$	(convolution)	$X(z)Y(z)$
Frequency Differentiation	$n\,x[n]$	□	$-z\frac{dX(z)}{dz}$
Summation	$\sum_{k=-\infty}^{n} x[k]$	□	$\frac{1}{1-z^{-1}}X(z)$

Example:

Find the z-transform of

$\delta_{n-k} = \begin{matrix} o & if\ n \neq k \\ i & if\ n=k \end{matrix}$ This is the "Unit Pulse" at n = k (assume k > 0)

$$\Delta(z) = \sum_{n=0}^{\infty} \delta_{n-k} z^{-n}$$

$\boxed{\Delta(z) = z^{-k}}$ **(Note: dividing by z is equivalent to a delay of one sample time)**

Example: $F(z) = \frac{2z}{(z-2)(z-1)^2}$ **, determine f_n**

A. By Infinite Series

$$F(z)=\frac{2z}{z^3-4z^2+5z-2}$$

Now divide (long division) with the polynomials written in descending powers of z

$2z^{-2}+8z^{-3}+22z^{-4}+52z^{-5}+114z^{-6}+\ldots$

$Z^3-4z^2+5z-2|2z$

$2z-8+10z^{-1}-4z^{-2}$

$8-10z^{-1}+04z^{-2}$

$8-32z^{-1}+40z^{-2}-16z^{-3}$

$22z^{-1}-36z^{-2}+016z^{-3}$

$22z^{-1}-88z^{-2}+110z^{-3}-44z^{-4}$

$52z^{-2}-094z^{-3}+044z^{-4}$

$52z^{-2}-208z^{-3}+260z^{-4}-104z^{-5}$

$114z^{-3}-216z^{-4}+104z^{-5}$

$$\therefore \quad F(z)=\sum_{n=0}^{\infty} f_n z^{-n} = 2z^{-2}+8z^{-3}+22z^{-4}+52z^{-5}+114z^{-6}+\ldots$$

And the time sequence for f_n is

n	0	1	2	3	4	5	6	…
f_n	0	0	2	8	22	52	114	…

Note that this method does NOT give a closed form for the answer, but it is a good method for finding the first few sample values or to check out that the closed form given by another method at least starts out correctly.

B. By Partial Fraction Expansion

$$F(z)=\frac{2z}{(z-2)(z-1)^2}=\frac{k_1 z}{z-2}+\frac{k_2 z}{z-1}+\frac{k_3 z}{(z-1)^2}$$

To find k1 multiply both sides of the equation by (z-2), divide by z, and let $z\to 2$

$$\frac{2z}{(z-1)^2}=k_1 z+\frac{k_2 z(z-2)}{z-1}+\frac{k_3 z(z-2)}{(z-1)^2}$$

$$\frac{2}{(z-1)^2}=k_1+\frac{k_2(z-2)}{z-1}+\frac{k_3(z-2)}{(z-1)^2}$$

$$\left.\frac{2}{(z-1)^2}\right|_{z=2} = k_1 + \left.\frac{k_2(z-2)}{z-1}\right|_{z=2} + \left.\frac{k_3(z-2)}{(z-1)^2}\right|_{z=2}$$

or

$k_1 = 2$

Similarly to find k_3 multiply both sides by $(z-1)^2$, divide by z, and let $z \to 1$

$$\frac{2}{(z-2)} = \frac{k_1(z-1)^2}{z-2} + k_2(z-1) + k_3 z \qquad \text{Equation A}$$

And

$k_3 = -2$

Finding k2 requires going back to Equation A above and taking the derivative of both sides

$$\frac{2}{(z-2)} = \frac{k_1(z-1)^2}{z-2} + k_2(z-1) + k_3 z$$

$$-\frac{2}{(z-2)^2} = k_1\left[\frac{2(z-1)}{z-2} - \frac{2(z-1)^2}{(z-2)^2}\right] + k_2$$

Now again let $z \to 1$

$k_2 = -2$

$$\therefore \quad F(z) = \frac{2z}{z-2} - \frac{2z}{z-1} - \frac{2z}{(z-1)^2}$$

Multiple choice questions:

1. Which of the following justifies the linearity property of z-transform?[x(n)↔X(z)] a) x(n)+y(n) ↔X(z)Y(z)
b) x(n)+y(n) ↔X(z)+Y(z)
c) x(n)y(n) ↔X(z)+Y(z)
d) x(n)y(n) ↔X(z)Y(z)

2. What is the z-transform of the signal x(n)=[3(2^n)-4(3^n)]u(n)?
a) $3/(1-2z^{-1})-4/(1-3z^{-1})$
b) $3/(1+2z^{-1})-4/(1+3z^{-1})$
c) 3/(1-2z)-4/(1-3z)
d) None of the mentioned

3. According to Time shifting property of z-transform, if X(z) is the z-transform of x(n) then what is the z-transform of x(n-k)?
a) $z^kX(z)$

b) $z^{-k}X(z)$
c) X(z-k)
d) X(z+k)

4. If X(z) is the z-transform of the signal x(n) then what is the z-transform of $a^n x(n)$?
   a) X(az)
   b) $X(az^{-1})$
   c) $X(a^{-1}z)$
   d) None of the mentioned

5. If the ROC of X(z) is r1<|z|<r2, then what is the ROC of $X(a^{-1}z)$?
   a) |a|r1<|z|<|a|r2
   b) |a|r1>|z|>|a|r2
   c) |a|r1<|z|>|a|r2
   d) |a|r1>|z|<|a|r2

6. If X(z) is the z-transform of the signal x(n), then what is the z-transform of the signal x(-n)?
   a) X(-z)
   b) $X(z^{-1})$
   c) $X^{-1}(z)$
   d) None of the mentioned

7. X(z) is the z-transform of the signal x(n), then what is the z-transform of the signal nx(n)?
   a) -z(dX(z))/dz
   b) zdX(z)/dz
   c) $-z^{-1}dX(z)/dz$
   d) $z^{-1}(dX(z))/dz$

8. What is the set of all values of z for which X(z) attains a finite value?
   a) Radius of convergence
   b) Radius of divergence
   c) Feasible solution
   d) None of the mentioned

9. What is the ROC of the signal x(n)=δ(n-k),k>0?
   a) z=0
   b) z=∞
   c) Entire z-plane, except at z=0
   d) Entire z-plane, except at z=∞

10. What is the ROC of the z-transform of the signal $x(n)= a^n u(n)+b^n u(-n-1)$?

    a) |a|<|z|<|b|
    b) |a|>|z|>|b|
    c) |a|>|z|<|b|
    d) |a|<|z|>|b|

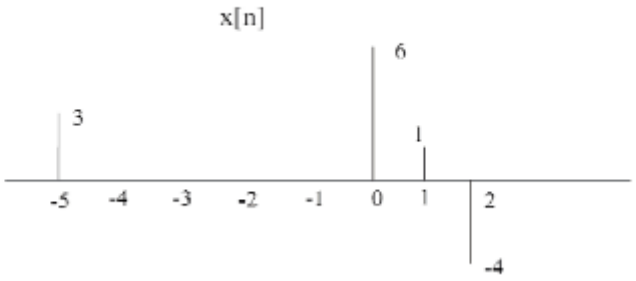
Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

Short Answer Type Questions:

Question no 1: Determine the Z-transform of the following signal:



Question no 2: Determine the Z-transform of the following signals. Also, sketch the corresponding zero-pole plot:

(a) $x[n] = (na^n \sin \omega_0 n)u[n]$

(b) $x[n] = (a^n + a^{-n})u[n]$, a real

3. Find the z transform of SinΩt u(t) , if sampling period is 1sec.

References:

1. A. Anand Kumar, 'Signals &Systems' -PHI
2. S. Haykin & B.V.Veen , 'Signals and Systems'-John Wiley
3. A. Nagoor Kani, Signals and Systems by - McGraw Hill
4. S Ghosh, 'Signals and Systems'- Pearson

Module II

Discrete Fourier Transform

It is required to know the frequency domain sampling i.e. X(ω). So, relationship between sampled Fourier transform and DFT is established in the following manner.

Similarly, periodic sequences can fit to this tool by extending the period N to infinity [3].

Let a non-periodic sequence be,

$$x(n) = \lim_{N\to\infty} x_N(n)$$

Defining its Fourier transform,

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n} X(k\delta\omega) \quad \ldots\ldots(1)$$

Here, X(ω) is sampled periodically, at every δω radian interval. As X(ω) is periodic in 2π radians, we require samples only in fundamental range. The samples are taken after equidistant intervals in the frequency range $0 \le \omega \le 2\pi$. Spacing between equivalent intervals is $\delta\omega = 2\pi/N$ radian [2].

Now evaluating, $\omega = 2\pi/N.k$

$$X\left(\frac{2\pi}{N}k\right) = \sum_{n=-\infty}^{\infty} x(n)e^{-j2\pi nk/N}, \quad \ldots\ldots(2)$$

where k=0,1,……N-1

After subdividing the above, and interchanging the order of summation

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

$$X(\frac{2\pi}{N}k) = \sum_{n=0}^{N-1}\left[\sum_{l=-\infty}^{\infty} x(n-Nl)\right]e^{-j2\pi nk/N}$$ .....(3)

$$\sum_{l=-\infty}^{\infty} x(n-Nl) = x_p(n) = \text{a periodic function of period N and its fourier series} = \sum_{k=0}^{N-1} C_k e^{j2\pi nk/N}$$

where, n=0,1,…..,N-1; 'p'- stands for periodic entity or function

The Fourier coefficients are,

$$C_k = \frac{1}{N}\sum_{n=0}^{N-1} x_p(n)e^{-j2\pi nk/N}$$ where k=0,1,2……N-1 …….(4)

Comparing equations 3 and 4, we get ;

$$NC_k = X(\frac{2\pi}{N}k)$$ where k=0,1,2……N-1 ……..(5)

$$NC_k = X\left(\frac{2\pi}{N}k\right) = X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x_p(n)e^{-j2\pi nk/N}$$ …… (6)

From Fourier series expansion,

$$x_p(n) = \frac{1}{N}\sum_{k=0}^{N-1} N\,C_k e^{j2\pi nk/N} = \frac{1}{N}\sum_{k=0}^{N-1} X\left(\frac{2\pi}{N}k\right)e^{j2\pi nk/N}$$ …… (7)

Where n=0,1,…,N-1

Here, we got the periodic signal from X(ω). $x(n)$ can be extracted from $x_p(n)$ only, if there is no aliasing in the time domain. $N \geq L$, N= period of $x_p(n)$ L= period of $x(n)$ [2]

$$x(n) = \begin{cases} x_p(n), & 0 \leq n \leq N-1 \\ 0, & Otherwise \end{cases}$$

So the DFT is given by:

$$X(k) = DFT[x(n)] = X\left(\frac{2\pi}{N}k\right) = \sum_{n=0}^{N-1} x(n)e^{-\frac{j2\pi nk}{N}}$$, where k=0,1,2……N-1

And IDFT is given by:

$$x(n) = IDFT[X(k)] = \frac{1}{N}\sum_{k=0}^{N-1} X(k)e^{\frac{j2\pi nk}{N}}$$ where n=0,1,2……N-1

Example1:

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

Compute the N-point DFT of $\boldsymbol{x(n) = 3\delta(n)}$

Solution: We know that,

$$X(K) = \sum_{n=0}^{N-1} x(n) e^{-\frac{j2\Pi kn}{N}}$$

$$= \sum_{n=0}^{N-1} 3\delta(n) e^{-\frac{j2\Pi kn}{N}}$$

$$= 3\delta(0) \times e^0 = 1$$

$$x(k) = 3, 0 \leq k \leq N-1$$

Example2:

Compute the N-point DFT of $\boldsymbol{x(n) = 7(n - n_0)}$

Solution: We know that,

$$X(K) = \sum_{n=0}^{N-1} x(n) e^{-\frac{j2\Pi kn}{N}}$$

Substituting the value of x(n),

$$\sum_{n=0}^{N-1} 7\delta(n - n_0) e^{-\frac{j2\Pi kn}{N}}$$

$$= e^{-j14\Pi kn_0/N}$$

Twiddle Factor

It is denoted as W_N and defined as $W_N = e^{-j2\pi/N}$. Its magnitude is always maintained at unity. Phase of W_N= -2π/ N. It is a vector on unit circle and is used for computational convenience. Mathematically, it can be shown as [1]

$$W_N^r = W_N^{r\pm N} = W_N^{r\pm 2N} = \cdots$$

It is function of r and period N. Consider N=8, r=0,1,2,3,….14,15,16…..

$$\Leftrightarrow W_8^0 = W_8^8 = W_8^{16} = \cdots = \cdots = W_8^{32} = \cdots = 1 = 1\angle 0$$

$$W_8^1 = W_8^9 = W_8^{17} = \cdots = \cdots = W_8^{33} = \cdots = \frac{1}{\sqrt{2}} - j\frac{1}{\sqrt{2}} = 1\angle -\frac{\pi}{4}$$

Computation of DFT can be performed using twiddle factor

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

$$x_N = \begin{bmatrix} x(0) \\ x(1) \\ . \\ . \\ x(N-1) \end{bmatrix} \quad X_N = \begin{bmatrix} X(0) \\ X(1) \\ . \\ . \\ X(N-1) \end{bmatrix}$$

$N-point\ DFT\ in\ matrix\ term\ is\ given\ by-\ X_N=W_N x_N$

$$\begin{bmatrix} 1 & 1 & 1 & \dots & \dots & 1 \\ 1 & W_N & W_N^2 & \dots & \dots & W_N^{N-1} \\ . & W_N^2 & W_N^4 & \dots & \dots & W_N^{2(N-1)} \\ . & & & & & . \\ 1 & W_N^{N-1} & W_N^{2(N-1)} & \dots & \dots & W_N^{(N-1)(N-1)} \end{bmatrix}$$

Therefore

$$\begin{bmatrix} X(0) \\ X(1) \\ . \\ . \\ X(N-1) \end{bmatrix} = \begin{bmatrix} x(0) \\ x(1) \\ . \\ . \\ x(N-1) \end{bmatrix} \text{X} \begin{bmatrix} 1 & 1 & 1 & \dots & \dots & 1 \\ 1 & W_N & W_N^2 & \dots & \dots & W_N^{N-1} \\ . & W_N^2 & W_N^4 & \dots & \dots & W_N^{2(N-1)} \\ . & & & & & . \\ 1 & W_N^{N-1} & W_N^{2(N-1)} & \dots & \dots & W_N^{(N-1)(N-1)} \end{bmatrix}$$

Properties of DFT

Linearity

It states that the DFT of a combination of signals is equal to the sum of DFT of individual signals. Let us take two signals $x_1(n)$ and $x_2(n)$, whose DFT s are $X_1(\omega)$ and $X_2(\omega)$ respectively [2]. So, if

$$x_1(n) \rightarrow X_1(\omega) \text{ and } x_2(n) \rightarrow X_2(\omega)$$

Then $ax_1(n)+bx_2(n) \rightarrow aX_1(\omega)+bX_2(\omega)$

where a and b are constants.

Symmetry

The symmetry properties of DFT can be derived in a similar way as we derived DTFT symmetry properties. We know that DFT of sequence x(n) is denoted by X(K). Now, if x(n) and X(K) are complex valued sequence, then it can be represented as under

$$x(n)=x_R(n)+jx_1(n), \qquad 0 \le n \le N-1$$

And $X(K)=X_R(K)+jX_1(K), \quad 0 \le K \le N-1$

Time Shift

$$x\left[((n-n_0))_N\right] \longleftrightarrow e^{-j\Omega_0 n_0 k} X\left[k\right]$$

Frequency Shift

$$e^{jin} x\left[n\right] \longleftrightarrow X\left[((k-i))_N\right]$$

Convolution

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

$$x_1[n] * x_2[n] = \sum_{m=0}^{N-1} x_1[((n-m))_N]\, x_2[m]$$

$$x_1[n] * x_2[n] \longleftrightarrow X_1[k]\, X_2[k]$$

Multiplication

$$X_1[k] * X_2[k] = \sum_{i=0}^{N-1} X_1[((k-i))_N]\, X_2[i]$$

$$x_1[n]\, x_2[n] \longleftrightarrow \frac{1}{N} X_1[k] * X_2[k]$$

Time Differencing

$$x[n] - x[((n-1))_N] \longleftrightarrow \left(1 - e^{-j\Omega_0 k}\right) X[k]$$

Conjugation

$$x^*[n] \longleftrightarrow X^*[((-k))_N]$$

Circular Operations

Mod N arithmetic may be used to de.ne operations on signals that are appropriate for the DFT. The circular reflection of x [n] is x [((-n)) $_N$] : The circular shift of x [n] by any integer n_0 is x [((n - n_0))$_N$] [1] : From circular shift, we de.ne the circular convolution of x_1 [n] and x_2 [n] is

$$x_1[n] * x_2[n] = \sum_{m=0}^{N-1} x_1[((n-m))_N]\, x_2[m]\,.$$

Periodic and circular convolution are similar, but apply to different circumstances, Periodic convolution requires periodic signals and produces a periodic signal; circular convolution requires finite duration signals and produces a finite duration signal.

$$x_1[n] *_{circ} x_2[n] = x_{1p}[n] *_{per} x_{2p}[n]$$

for 0 ≤n ≤N-1,Circular shift and periodic convolution may also be applied to X [k], it is important that circular shift and convolution appear in several of the DFT properties above [2].

The Fast Fourier Transform

The discrete (or digitized) version of the Fourier transform is called the Discrete Fourier Transform (DFT). This transform takes digitized time domain data and computes the frequency domain representation. While normal Fourier theory is useful for understanding how the time and frequency domain relate, the DFT allows us to compute the frequency domain representation of real-world time domain signals. This brings the power

of Fourier theory out of the world of mathematical analysis and into the realm of practical measurements. The Agilent 54600 scope with Measurement/Storage Module uses a particular algorithm, called the Fast Fourier Transform (FFT), for computing the DFT. The FFT and DFT produce the same result and the feature is commonly referred to as simply the FFT.

The FFT function uses 1000 of these points (every fourth point) to produce a 500 point frequency domain display. This frequency domain display extends in frequency from 0 to feff /2, where feff is the effective sample rate of the time record in Figure 1 (a) & (b)

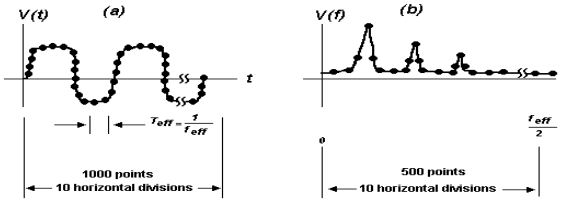


Figure 1

(a) The sampled time domain waveform.

(b) The resulting frequency domain plot using the FFT.

The effective sample rate is the reciprocal of the time between samples and depends on the time/div setting of the scope. For the Agilent 54600 series, the effective sample rate is given by:

$$f_{eff} = \frac{\text{record length}}{10^* \text{ time/div}} = \frac{1000}{10^* \text{ time/div}} = \frac{100}{\text{time/div}}$$

So for any particular time/div setting, the FFT produces a frequency domain representation that extends from 0 to feff /2 (Figure 3b). When the FFT function is active, the effective sample rate is displayed when the time/div knob is turned or the ± key is pressed. Note that the effective sample rate for the FFT can be much higher than the maximum sample rate of the scope. The maximum sample rate of the scope is 20 MHz, but the random-repetitive sampling technique places samples so precisely in time that the sample rate seen by the FFT can be as high as 20 GHz.

The default frequency domain display covers the normal frequency range of 0 to feff /2. The Center Frequency and Frequency Span controls can be used to zoom in on narrower frequency spans within the basic 0 to feff /2 range of the FFT. These controls do not affect the FFT computation, but instead cause the frequency domain points to be plotted in expanded form.

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

Aliasing

The frequency feff /2 is also known as the folding frequency. Frequencies that would normally appear above feff /2 (and, therefore, outside the useful range of the FFT) are folded back into the frequency domain display. These unwanted frequency components are called aliases, since they erroneously appear under the alias of another frequency. Aliasing is avoided if the effective sample rate is greater than twice the bandwidth of the signal being measured.

The frequency content of a triangle wave includes the fundamental frequency and a large number of odd harmonics with each harmonic smaller in amplitude than the previous one. In Figure 4a, a 26 kHz triangle wave is shown in the time domain and the frequency domain. Figure 2b shows only the frequency domain representation. The leftmost large spectral line is the fundamental. The next significant spectral line is the third harmonic. The next significant spectral line is the fifth harmonic and so forth. Note that the higher harmonics are small in amplitude with the 17th harmonic just visible above the FFT noise floor. The frequency of the 17th harmonic is 17 x 26 kHz = 442 kHz, which is within the folding frequency of feff /2, (500 kSa/sec) in Figure 2b. Therefore, no significant aliasing is occurring.

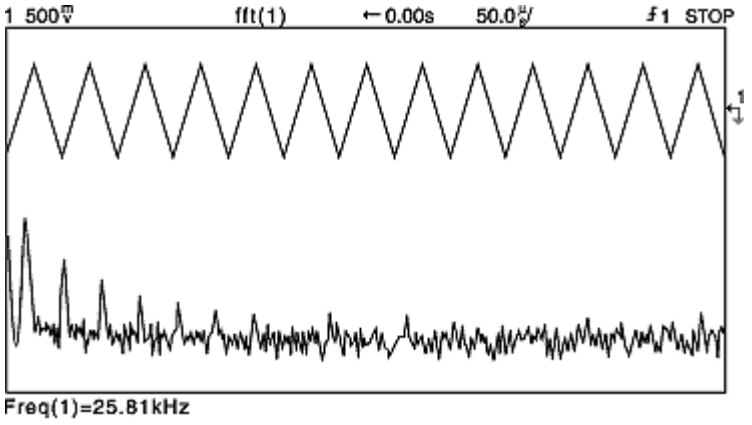


Figure 2a:

The time domain and frequency domain displays of a 26kHz triangle wave.

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

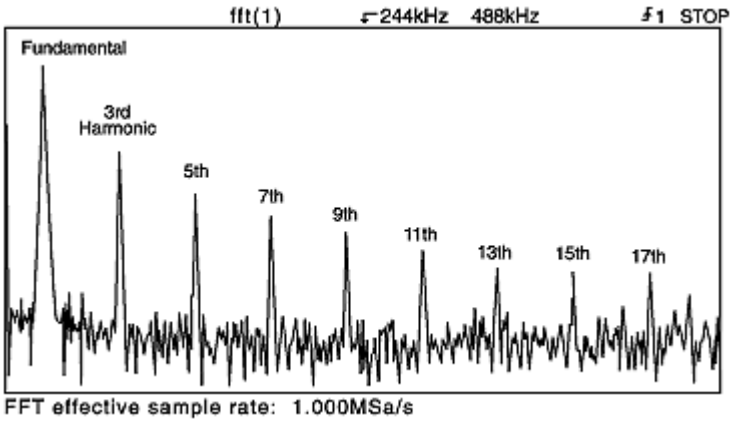


Figure 2b:

Frequency spectrum of a triangle wave.

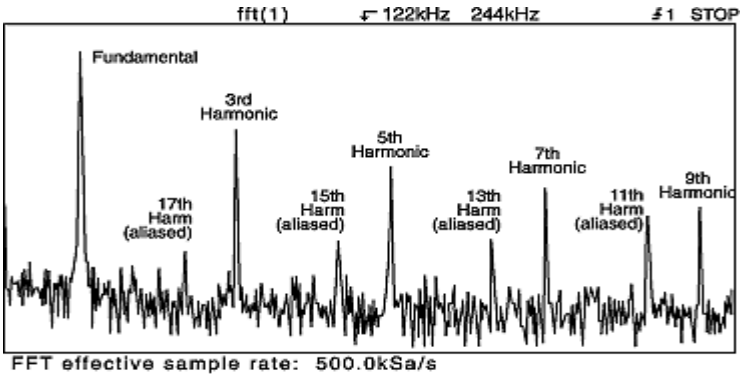


Figure 2c:

With a lower effective sample rate, the upper harmonics appear as aliases.

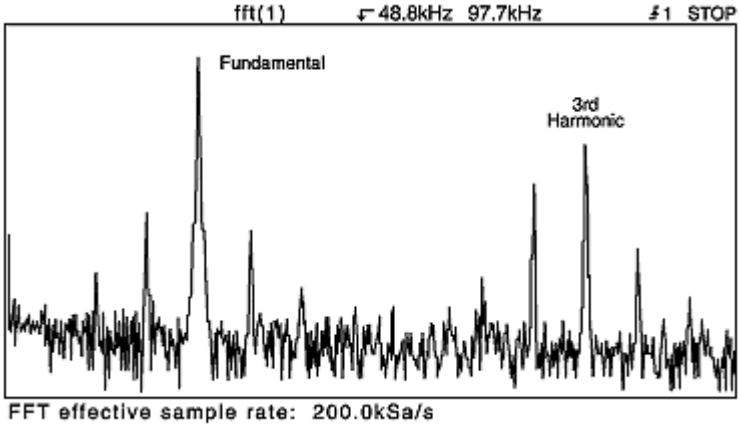


Figure 2d:

With an even lower effective sample rate, only the fundamental and third harmonic are not aliased.

Figure 4c shows the FFT of the same waveform with the time/div control turned one click to the left, resulting in an effective sample rate of 500 kSa/sec and a folding frequency of 250 kSa/sec. Now the upper harmonics of the triangle wave exceed the folding frequency and appear as aliases in the display. Figure 4d shows the FFT of the same triangle wave, but with an even lower effective sample rate (200 kSa/sec) and folding frequency (100 kSa/sec). This frequency plot is severely aliased.

Often the effects of aliasing are obvious, especially if the user has some idea as to the frequency content of the signal. Spectral lines may appear in places where no frequency components exist. A more subtle effect of aliasing occurs when low level aliased frequencies appear near the noise floor of the measurement. In this case the baseline can bounce around from acquisition to acquisition as the aliases fall slightly differently in the frequency domain.

Aliased frequency components can be misleading and are undesirable in a measurement. Signals that are bandlimited (that is, have no frequency components above a certain frequency) can be viewed alias-free by making sure that the effective sample rate is high enough. The effective sample rate is kept as high as possible by choosing a fast time/div setting. While fast time/div settings produce high effective sample rates, they also cause the frequency resolution of the FFT display to degrade.

If a signal is not inherently bandlimited, a lowpass filter can be applied to the signal to limit its frequency content (Figure 3). This is especially appropriate in situations where the same type of signal is measured often and a special, dedicated lowpass filter can be kept with the scope.

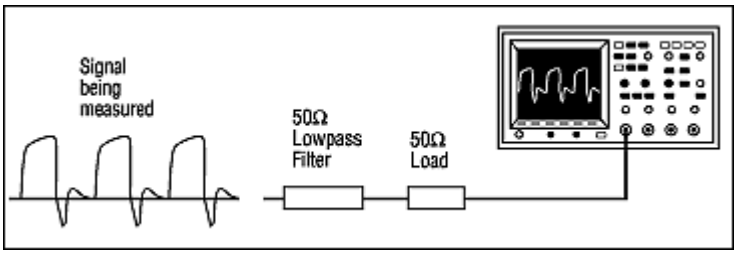


Figure 3: A lowpass filter can be used to band limit the signal, avoiding aliasing.

Leakage*

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

The FFT operates on a finite length time record in an attempt to estimate the Fourier Transform, which integrates over all time. The FFT operates on the finite length time record, but has the effect of replicating the finite length time record over all time (Figure 4). With the waveform shown in Figure 4a, the finite length time record represents the actual waveform quite well, so the FFT result will approximate the Fourier integral very well.

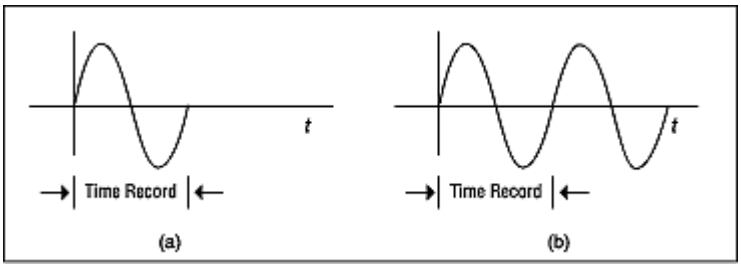


Figure 4

(a) A waveform that exactly fits one time record.

(b) When replicated, no transients are introduced.

However, the shape and phase of a waveform may be such that a transient is introduced when the waveform is replicated for all time, as shown in Figure 5. In this case, the FFT spectrum is not a good approximation for the Fourier Transform.

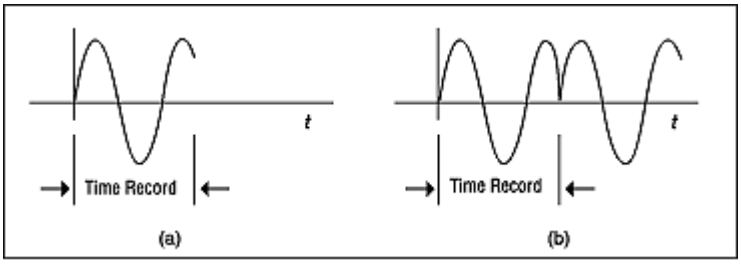


Figure 5

(a) A waveform that does not exactly fit into one time record.

(b) When replicated, severe transients are introduced, causing leakage in the frequency domain.

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

Since the scope user often does not have control over how the waveform fits into the time record, in general, it must be assumed that a discontinuity may exist. This effect, known as LEAKAGE, is very apparent in the frequency domain. The transient causes the spectral line (which should appear thin and slender) to spread out as shown in Figure 6.

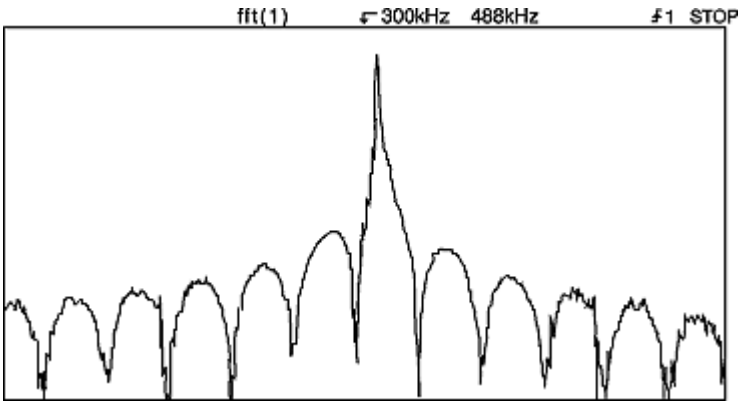


Figure 6

Leakage occurs when the normally thin spectral line spreads out in the frequency domain.

Multiple choice questions:

1. The selected frequency points in frequency sampling method are
   a)Fourier Transform coefficients
   b) DTFT coefficients
   c) DFT coefficients
   d) None

2. Let $x(n) = \left(\frac{1}{2}\right)^n u(n), y(n) = x^2(n)$ and $Y(e^{j\omega})$ be the Fourier Transform of $y(n)$. Then $y(e^{j0})$ is
   (a) $\frac{1}{4}$
   (b) 2
   (c) 4
   (d) $\frac{4}{3}$

3. The filtering is performed using DFT using

1) Limited size or blocks of data
2) Small memory size
3) Large memory size
4) Large segments of data

a. 1, 2 and 3 are correct

b. 3 and 4 are correct

c. 1 and 2 are correct

d. All the four are correct

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

4. $\{x(n)\}$ is real-valued periodic sequence with a period N. $x(n)$ and $X(k)$ form N-point Discrete Fourier Transform (DFT) pairs. The DFT Y(k) of the sequence

$$y(n) = \frac{1}{N}\sum_{r=0}^{N-1} x(r)\ X(n+r)\ is$$

(a) $|X(k)|^2$

(b) $\frac{1}{N}\sum_{r=0}^{N-1} X(r)X^*\ (k+r)$

(c) $\frac{1}{N}\sum_{r=0}^{N-1} X(r)X\ (k+r)$

(d) 0

5. The circular convolution of two sequences in time domain is equivalent to

a. Multiplication of DFTs of two sequences

b. Summation of DFTs of two sequences

c. Difference of DFTs of two sequences

d. Square of multiplication of DFTs of two sequences

6. For an N-point FFT algorithm with $N = 2^m$, which one of the following statements is TRUE?
   (a) It is not possible to construct is signal flow graph with both input and output in normal order
   (b) The number of butterflies in the m^{th} state is N/m
   (c) In-place computation requires storage of only 2N node data

7. Circular shift of an N point is equivalent to

a. Circular shift of its periodic extension and its vice versa

b. Linear shift of its periodic extension and its vice versa

c. Circular shift of its aperiodic extension and its vice versa

d. Linear shift of its aperiodic extension and its vice versa

8. Padding of zeros increases the frequency resolution.

a. True

b. False

9. Frequency selectivity characteristics of DFT refers to

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

a. Ability to resolve different frequency components from input signal

b. Ability to translate into frequency domain

c. Ability to convert into discrete signal

d. None of the above

10. DTFT is the representation of

a. Periodic Discrete time signals

b. Aperiodic Discrete time signals

c. Aperiodic continuous signals

d. Periodic continuous signals

References:

1. J.G.Proakis & D.G.Manolakis, Digital Signal Processing-Pearson Ed.
2. Prof. Dan Cobb, Digital Signal Processing Lecture Notes.
3. S.K.Mitra, Digital Signal processing – A Computer Based Approach-TMH.

Module III
Filter Design

The solution to the problem of leakage is to force the waveform to zero at the ends of the time record so that no transient will exist when the time record is replicated. This is accomplished by multiplying the time record by a WINDOW function. Of course, the window modifies the time record and will produce its own effect in the frequency domain. For a properly designed window, the effect in the frequency domain is a vast improvement over using no window at all.1 Four window functions are available in the Agilent 54600 scopes: Hanning, Flattop, Rectangular and Exponential.

The Hanning window provides a smooth transition to zero as either end of the time record is approached. Figure 1a shows a sinusoid in the time domain while Figure 1b shows the Hanning window which will be applied to the time domain data. The windowed time domain record is shown in Figure 9c. Even though the overall shape of the time domain signal has changed, the frequency content remains basically the same. The spectral line associated with the sinusoid spreads out a small amount in the frequency domain as shown in Figure 2

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

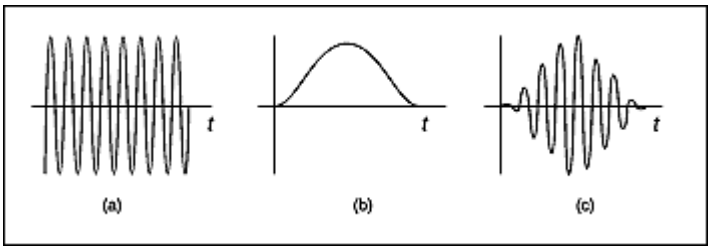


Figure 1

(a) The original time record.

(b) The Hanning Window.

(c) The windowed time record.

The shape of a window is a compromise between amplitude accuracy and frequency resolution. The Hanning window, compared to other common windows, provides good frequency resolution at the expense of somewhat less amplitude accuracy.

The FLATTOP window has fatter (and flatter) characteristic in the frequency domain, as shown in Figure 3. (Again, the figure is expanded in the frequency axis to show clearly the effect of the window.) The flatter top on the spectral line in the frequency domain produces improved amplitude accuracy, but at the expense of poorer frequency resolution (when compared with the Hanning window).

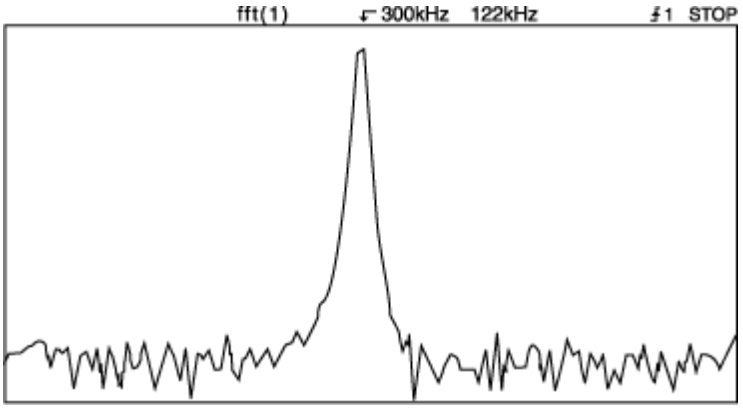


Figure 2

The Hanning Window has a relatively narrow shape in the frequency domain.

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

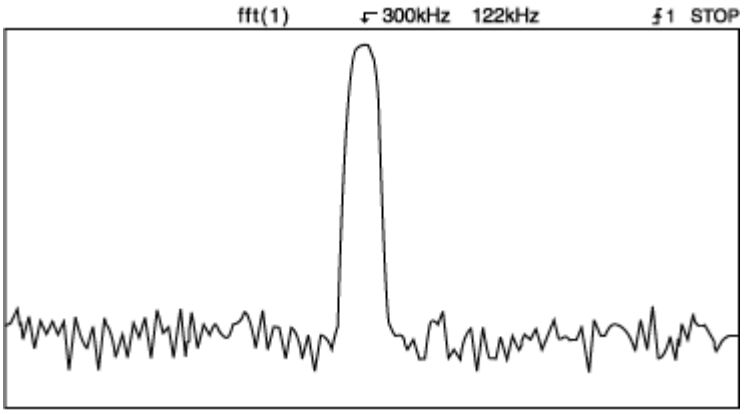


Fig. 3: The flattop window has a wider, flat-topped shape in the frequency domain.

The Rectangular window (also referred to as the Uniform window) is really no window at all; all of the samples are left unchanged. Although the uniform window has the potential for severe leakage problems, in some cases the waveform in the time record has the same value at both ends of the record, thereby eliminating the transient introduced by the FFT. Such waveforms are called SELF-WINDOWING. Waveforms such as sine bursts, impulses and decaying sinusoids can all be self-windowing.

A typical transient response is shown in Figure 4a. As shown, the waveform is self-windowing because it dies out within the length of the time record, reducing the leakage problem.

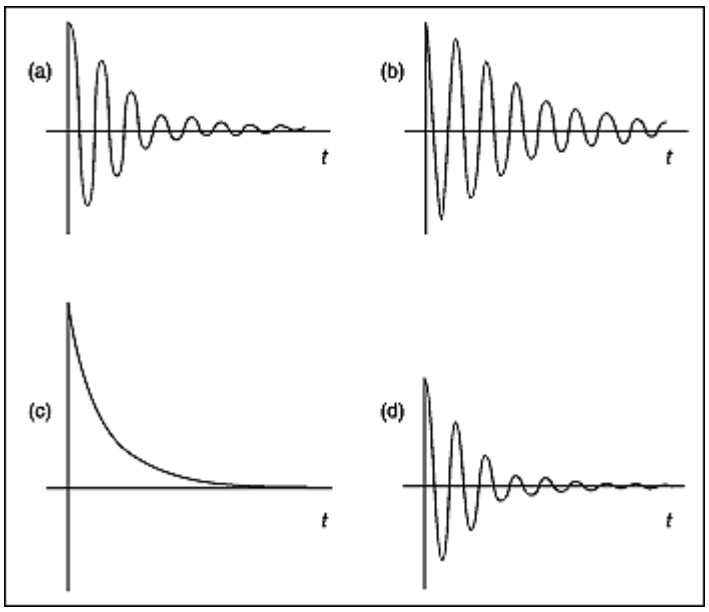


Figure 4

(a) A transient response that is self-windowing.

(b) A transient response which requires windowing.

(c) The exponential window.

(d) The windowed transient response.

If the waveform does not dissipate within the time record (as shown in Figure 4b), then some form of window should be used. If a window such as the Hanning window were applied to this waveform, the beginning portion of the time record would be forced to zero. This is precisely where most of the transient's energy is, so such a window would be inappropriate.

A window with a decaying exponential response is useful in such a situation. The beginning portion of the waveform is not disturbed, but the end of the time record is forced to zero. There still may be a transient at the beginning of the time record, but this transient is not introduced by the FFT. It is, in fact, the transient being measured. Figure 4c shows the exponential window and Figure 12d shows the resulting time domain function when the exponential window is applied to Figure 4b. The exponential window is inappropriate for measuring anything but transient waveforms.

Selecting a Window

Most measurements will require the use of a window such as the Hanning or Flattop windows. These are the appropriate windows for typical spectrum analysis measurements. Choosing between these two windows involves a tradeoff between frequency resolution and amplitude accuracy. Having used the time domain to explain why leakage occurs, now the user should switch into frequency domain thinking. The narrower the passband of the window's frequency domain filter, the better the analyzer can discern between two closely spaced spectral lines. At the same time, the amplitude of the spectral line will be less certain. Conversely, the wider and flatter the window's frequency domain filter is, the more accurate the amplitude measurement will be and, of course, the frequency resolution will be reduced. Choosing between two such window functions is really just choosing the filter shape in the frequency domain.

The rectangular and exponential windows should be considered windows for special situations. The rectangular window is used where it can be guaranteed that there will be no leakage effects. The exponential window is for use when the input signal is a transient.[3]

FIR Filter design using window function :

Windowing of a simple waveform like $\cos \omega t$ causes its Fourier transform to develop non-zero values (commonly called spectral leakage) at frequencies other than ω. The leakage tends to be worst (highest) near ω and least at frequencies farthest from ω.

If the waveform under analysis comprises two sinusoids of different frequencies, leakage can interfere with the ability to distinguish them spectrally. If their frequencies are dissimilar and one component is weaker, then

leakage from the stronger component can obscure the weaker one's presence. But if the frequencies are similar, leakage can render them *irresolvable* even when the sinusoids are of equal strength. The rectangular window has excellent resolution characteristics for sinusoids of comparable strength, but it is a poor choice for sinusoids of disparate amplitudes. This characteristic is sometimes described as *low dynamic range*.

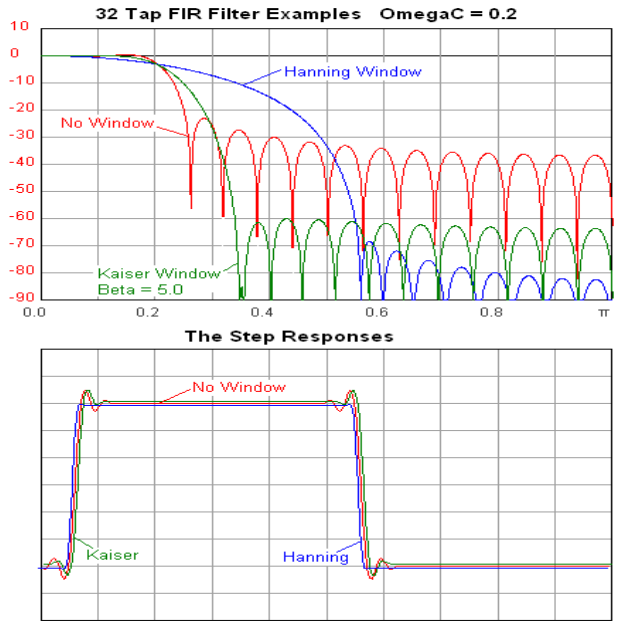
At the other extreme of dynamic range are the windows with the poorest resolution and ***sensitivity***, which is the ability to reveal relatively weak sinusoids in the presence of additive random noise. That is because the noise produces a stronger response with high-dynamic-range windows than with high-resolution windows. Therefore, high-dynamic-range windows are most often justified in *wideband applications*, where the spectrum being analyzed is expected to contain many different components of various amplitudes.

In between the extremes are moderate windows, such as Hamming and Hann. They are commonly used in *narrowband applications*, such as the spectrum of a telephone channel. In summary, spectral analysis involves a trade-off between resolving comparable strength components with similar frequencies and resolving disparate strength components with dissimilar frequencies. That trade-off occurs when the window function is chosen.

Need of Windowing in FIR Filter

the Kaiser or Hanning, are used to design FIR filters, not IIR. Windows are also used for spectral analysis, but I think you are only asking about them with regard to filter design.

The reason there are so many types of windows is that each generates a slightly different frequency response and time domain response, as shown here.



Direct Form I

The direct form I is derived by rearranging above equation for a0=1a0=1

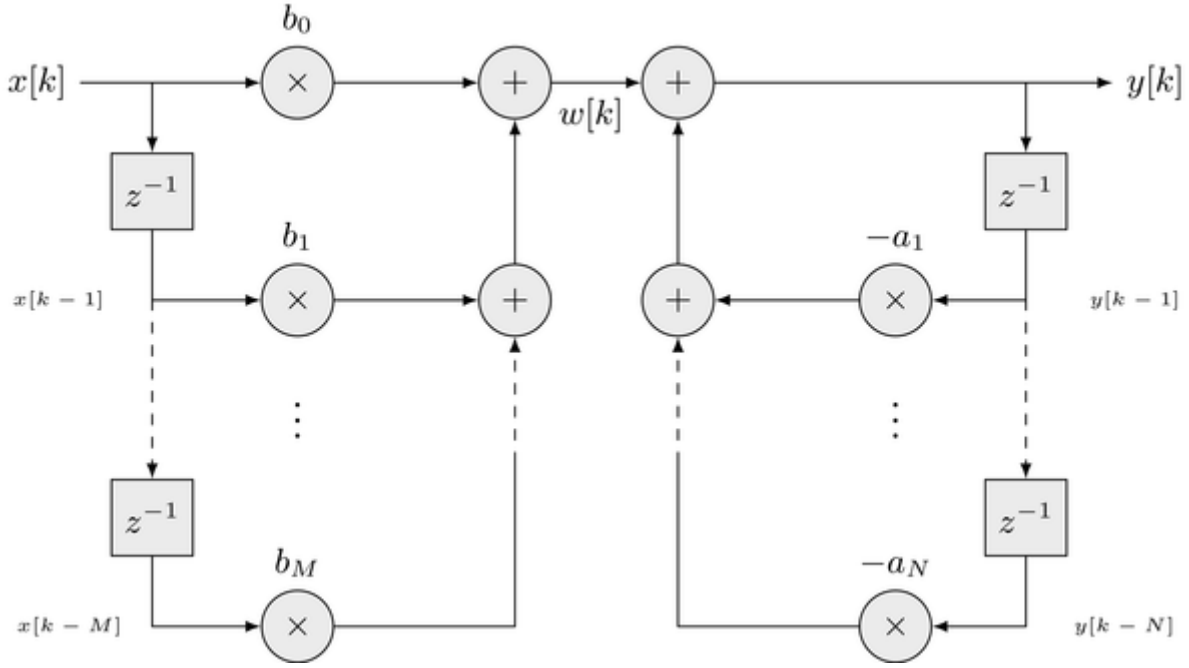
y[k]=∑m=0Mbmx[k−m]+∑n=1N−any[k−n]y[k]=∑m=0Mbmx[k−m]+∑n=1N−any[k−n]

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

It is now evident that we can realize the recursive filter by a superposition of a non-recursive and a recursive part. With the elements given above, this results in the following block-diagram



Direct form I filter

This representation is not canonical since N+MN+M unit delays are required to realize a system of order NN. A benefit of the direct form I is that there is essentially only one summation point which has to be taken care of when considering quantized variables and overflow. The output signal y[k]y[k] for the direct form I is computed by realizing above equation.

The block diagram of the direct form I can be interpreted as the cascade of two systems. Denoting the signal in between both as w[k]w[k] and discarding initial values we get

where h1[k]=[b0,b1,…,bM]h1[k]=[b0,b1,…,bM] denotes the impulse response of the non-recursive part and h2[k]=[1,−a1,…,−aN]h2[k]=[1,−a1,…,−aN] for the recursive part. From the last equality of the second equation and the commutativity of the convolution it becomes clear that the order of the cascade can be exchanged.

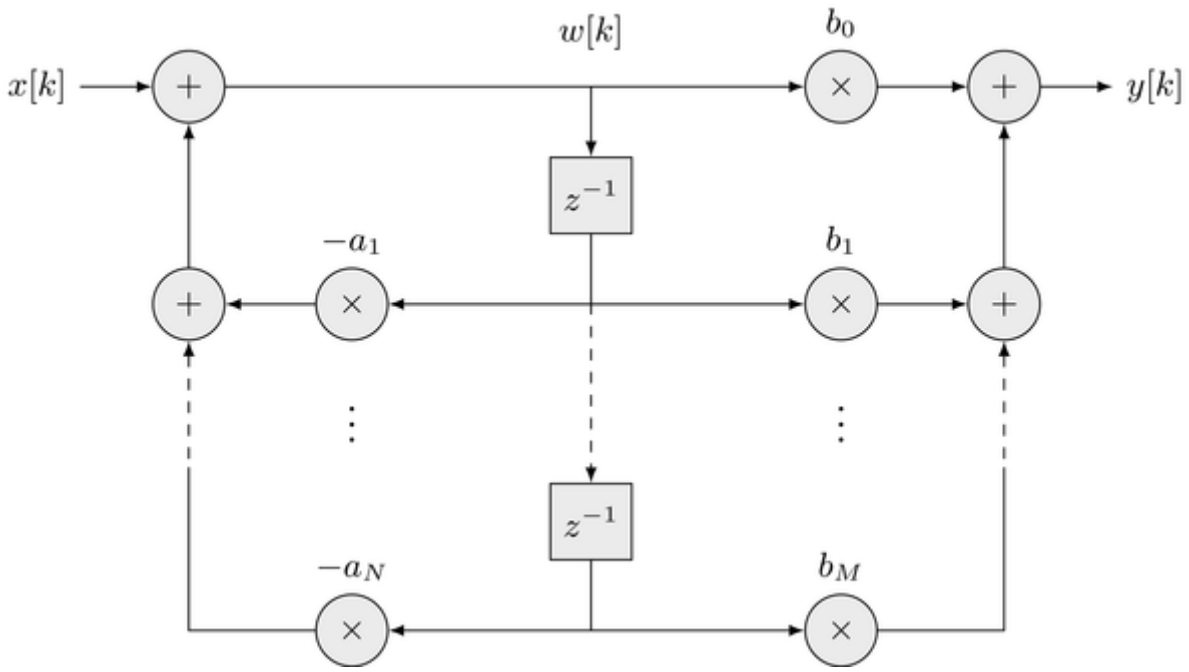
Direct Form II

The direct form II is yielded by exchanging the two systems in above block diagram and noticing that there are two parallel columns of delays which can be combined, since they are redundant. For N=MN=M it is given as

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th



Direct form II filter

Other cases with N≠MN≠M can be considered for by setting coefficients to zero. This form is a canonical structure since it only requires NN unit delays for a recursive filter of order NN. The output signal y[k]y[k] for the direct form II is computed by the following equations

The samples w[k−m]w[k−m] are termed state (variables) of a digital filter.

Transposed Direct Form II

The block diagrams above can be interpreted as linear signal flow graphs. The theory of these graphs provides useful transformations into different forms which preserve the overall transfer function. Of special interest is the transposition or reversal of a graph which can be achieved by

exchanging in- and output,

exchanging signal split and summation points, and

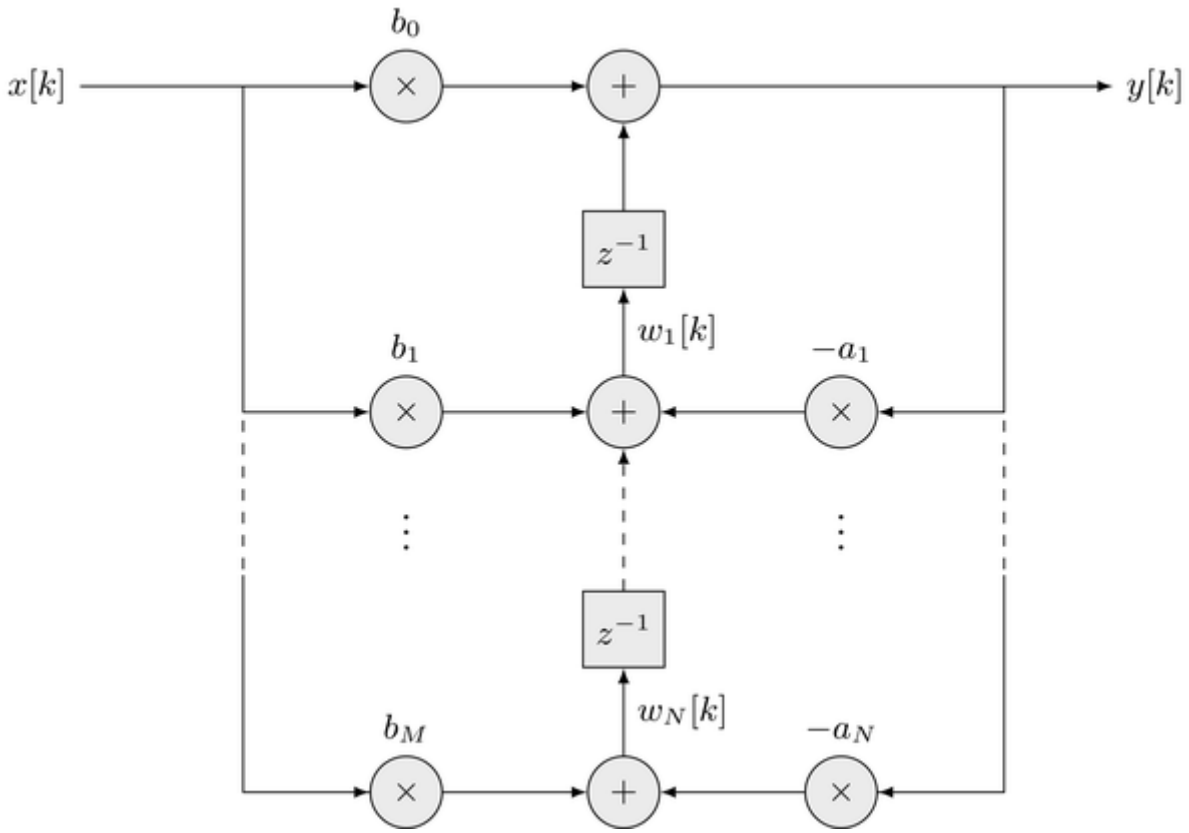
reversing the directions of the signal flows.

Applying this procedure to the direct form II shown above for N=MN=M yields the transposed direct form II

Subject Name: Digital Signal Processing

Subject Code: EC 503 **Contacts: 3L** **Credits: 3** **Lectures: 35 hrs.**

Semester: 5th



Transposed direct form II filter

The output signal of the transposed direct form II is given as

y[k]=b0x[k]+∑m=1Mbmx[k−n]−∑n=1Nany[k−n]

Structures of Digital Processors

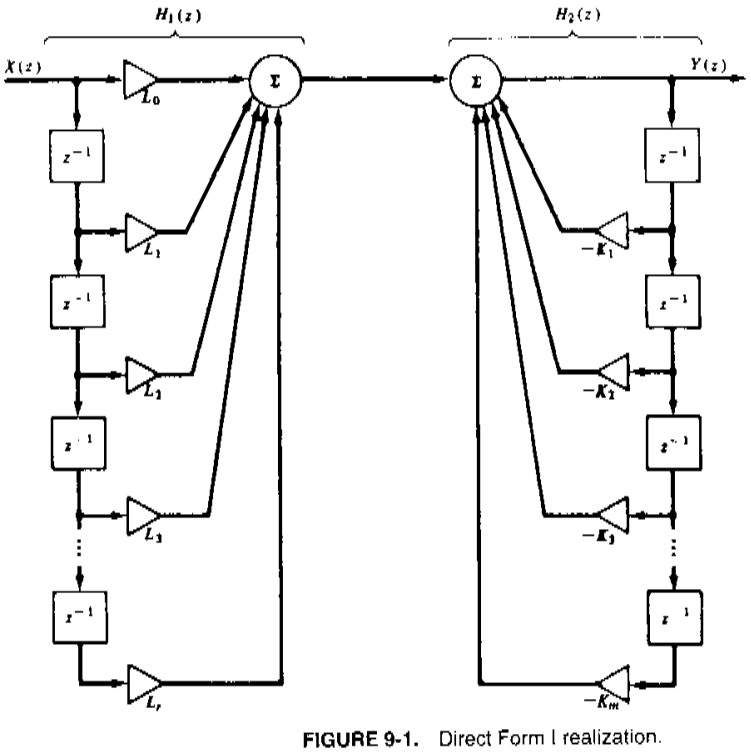


FIGURE 9-1. Direct Form I realization.

1. Direct-Form Realization

$$H(z)=\frac{Y(z)}{X(z)}=\frac{\sum_{i=0}^{r}L_i z^{-i}}{1+\sum_{j=1}^{m}k_j z^{-j}}$$

$$\Rightarrow y(nT)=\sum_{i=0}^{r}L_i x(nT-iT)-\sum_{j=1}^{m}k_j y(nT-jT)$$

The function is realized!

What's the issue here?

Count how many memory elements we need!

Can we reduce this number?

If we can, what is the concern?

$$H(z)=\frac{\sum_{i=0}^{r}L_i z^{-i}}{1+\sum_{j=1}^{m}k_j z^{-i}}=\underbrace{\left(\sum_{i=0}^{r}L_i z^{-i}\right)}_{H_1(z)}\underbrace{\frac{1}{1+\sum_{j=1}^{m}k_j z^{-j}}}_{H_2(z)}$$

$$\Rightarrow Y(z)=H(z)X(z)=H_1(z)H_2(z)X(z)$$

Denote $V(z)=H_2(z)X(z)\Rightarrow Y(z)=H_1(z)V(z)$

Implement H2(z) and then H1(z) ?

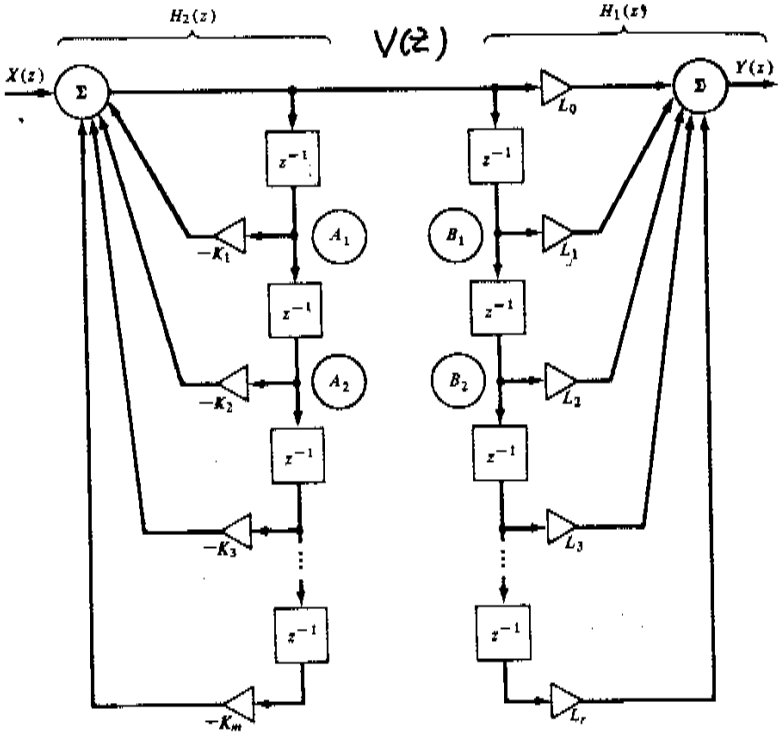


FIGURE 9-2. Rearrangement of Figure 9-1.

Why H2 is implemented?

(1) $V(z)=X(z)-k_1z^{-1}V(z)-\cdots-k_mz^{-m}V(z)$ (2)

$$(1+k_1z^{-1}+\cdots+k_mz^{-m})V(z)=X(z)$$

$$V(z)=\frac{1}{1+\sum_{j=1}^{m}k_jz^{-j}}X(z)$$

H2 is realized!

Can you tell why H1 is realized?

What can we see from this realization? Signals at A_j and B_j: always the same

➔ Direct Form II Realization

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

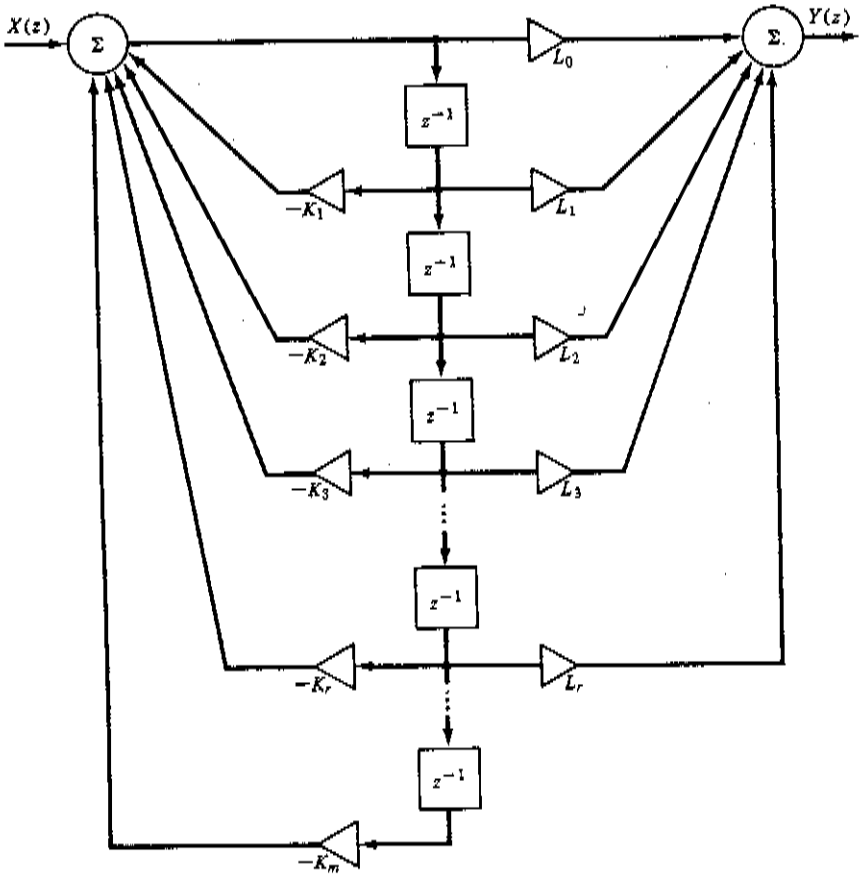
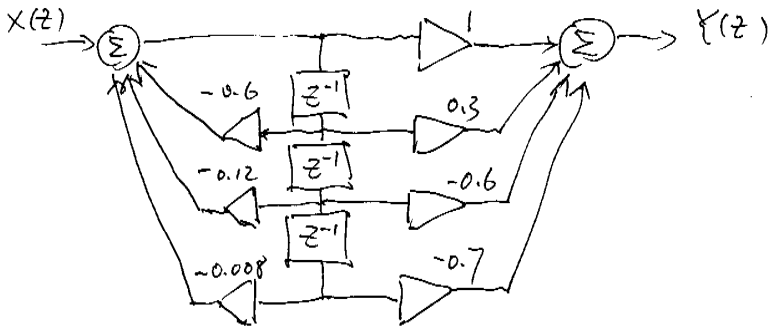


FIGURE 9-3. Direct Form II realization.

Solved Maths:

Example
$$H(z) = \frac{1 + 0.3z^{-1} - 0.6z^{-2} - 0.7z^{-3}}{(1 + 0.2z^{-1})^3}$$

Solution:
$$H(z) = \frac{1 + 0.3z^{-1} - 0.6z^{-2} - 0.7z^{-3}}{1 + 0.6z^{-1} + 0.12z^{-2} + 0.008z^{-3}}$$



Important: H(z)=B(z)/A(z) (1) A: 1+…. (2) Coefficients in A: in the feedback channel

2. Cascade Realization

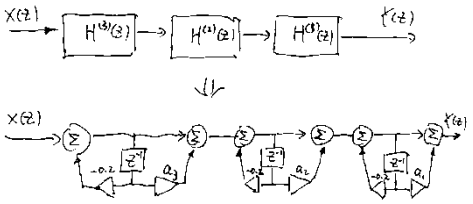
Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

Factorize $1+0.3z^{-1}-0.6z^{-2}-0.7z^{-3}=(1-a_1z^{-1})(1-a_2z^{-2})(1-a_3z^{-3})$

$$H(z)=\frac{1+0.3z^{-1}-0.6z^{-2}-0.7z^{-3}}{(1+0.2z^{-1})^3}=\underbrace{\frac{1-a_1z^{-1}}{1+0.2z^{-1}}}_{H^{(1)}(z)}\cdot\underbrace{\frac{1-a_2z^{-1}}{1+0.2z^{-1}}}_{H^{(2)}(z)}\cdot\underbrace{\frac{1-a_3z^{-1}}{1+0.2z^{-1}}}_{H^{(3)}(z)}$$



General Form

$$H(z)=kz^{-M}\underbrace{\frac{\prod_{i=1}^{N_1}(1-a_iz^{-1})\prod_{j=1}^{N_2}(1-b_j^{*}z^{-1})(1+b_j^{*}z^{-1})}{\prod_{k=1}^{D_1}(1-c_kz^{-1})\prod_{l=1}^{D_2}(1-d_l^{*}z^{-1})(1+d_l^{*}z^{-1})}}_{\Downarrow\qquad\qquad\Downarrow}$$

$$\frac{1-q_iz^{-1}}{1-c_kz^{-1}}\qquad\frac{1-(b_j+b_j^{*})z^{-1}+b_jb_j^{*}z^{-2}}{1-(d_l+d_l^{*})z^{-1}+d_ld_l^{*}z^{-2}}$$

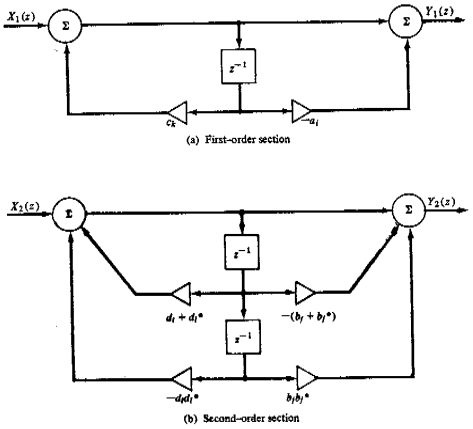


FIGURE 9-4. Basic first- and second-order filters.

Apply Direct II for each!

3. Parallel Realization (Simple Poles)

$$H(z)=\underbrace{\sum_{i=0}^{M}A_iz^{-i}}_{\text{If } r>m}+\underbrace{\sum_{k=1}^{D_1}B_k\frac{1}{1-C_kz^{-1}}}_{\text{realize the real poles}}+\underbrace{\sum_{l=1}^{D_2}C_l\frac{1-e_lz^{-1}}{(1-d_lz^{-1})(1-d_l^{*}z^{-1})}}_{\text{realize the complex conjugate poles}}$$

Example 9-1

$$H(z)=\frac{(1-z^{-1})^3}{(1-\frac{1}{2}z^{-1})(1-\frac{1}{8}z^{-1})}$$

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

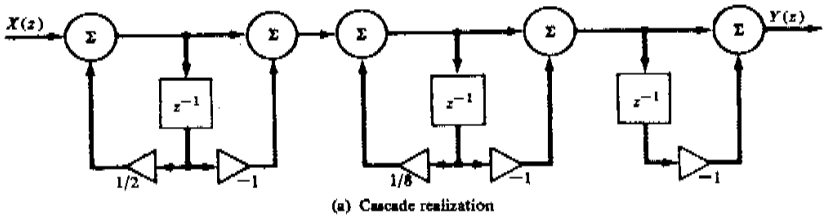
Semester: 5th

cascade and parallel realization!

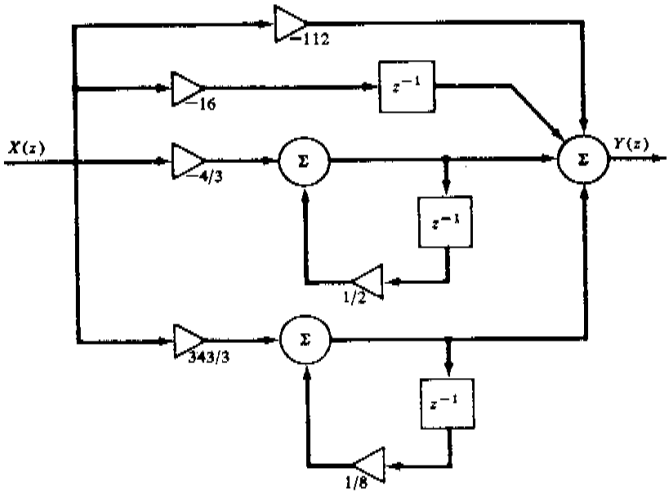
Solution:

Cascade:

$$H(z)=\frac{1-z^{-1}}{(1-\frac{1}{2}z^{-1})}\frac{1-z^{-1}}{(1-\frac{1}{8}z^{-1})}(1-z^{-1})$$



(a) Cascade realization



(b) Parallel realization

FIGURE 9-5. Cascade and parallel realizations for Example 9-1.

Parallel

$$H(z)=\frac{(1-z^{-1})^3}{(1-\frac{1}{2}z^{-1})(1-\frac{1}{8}z^{-1})}=\frac{(z-1)^3}{z(z-\frac{1}{2})(z-\frac{1}{8})}$$

In order to make deg(num)<deg(den)

$$\frac{H(z)}{z}=\frac{(z-1)^3}{z^2(z-\frac{1}{2})(z-\frac{1}{8})}$$

Partial-Fraction Expansion for *s*

$$=\frac{A}{z^2}+\frac{B}{z}+\frac{C}{z-1/2}+\frac{D}{z-1/8}$$

$$A = \lim_{z\to 0} zH(z) = \frac{z^2(z-1)^3}{z^2(z-1/2)(z-1/8)}\bigg|_{z\to 0} = \frac{(-1)^3}{(-\frac{1}{2})(-\frac{1}{8})} = -16$$

$$C = \lim_{z\to 1/2}(z-1/2)\frac{H(z)}{z} = \lim_{z\to 1/2}\frac{(z-1)^3}{z^2(z-1/8)} = -\frac{4}{3}$$

$$D = \lim_{z\to 1/8}(z-1/8)\frac{H(z)}{z} = \lim_{z\to 1/8}\frac{(z-1)^3}{z^2(z-1/2)} = \frac{343}{3}$$

z = anything other than 0, ½, 1/8, 1➔ B = -112

For example, z=2

$$\lim_{z\to 0} H(z) = \lim_{z\to 0}(\frac{A}{z} + B)$$

$$\lim_{z\to 0} zH(z) = \lim_{z\to 0}(A + Bz)$$

$$\lim_{z\to 0}\frac{d(H(z)z)}{dz} = B$$

$$\Rightarrow B = \lim_{z\to 0}\frac{d}{dz}\left[\frac{(z-1)^3}{(z-\frac{1}{2})(z-\frac{1}{8})}\right]$$

$$= -112$$

$$\frac{H(z)}{z} = \frac{1^3}{4\times\frac{3}{2}\times\frac{15}{8}} = \frac{4}{45}$$

$$\frac{A}{z^2} = \frac{-16}{4} = -4$$

$$\frac{C}{z-\frac{1}{2}} = \frac{-4/3}{3/2} = -8/9$$

$$\frac{D}{z-\frac{1}{8}} = \frac{-343/3}{15/8} = \frac{343}{3}\frac{8}{15}$$

$$B = 2[\frac{4}{45} + 4 + \frac{8}{9} - \frac{343}{3}\times\frac{8}{15}] = -112$$

Example 9-2: System having a complex conjugate pole pair at $z = ae^{\pm j\theta}$

➔Transfer function

$$H(z) = \frac{z^2}{(z-ae^{+j\theta})(z-ae^{-j\theta})} = \frac{1}{(1-ae^{j\theta}z^{-1})(1-ae^{-j\theta}z^{-1})}$$

$$= \frac{z^2}{z^2 - 2a(\cos\theta)z + a^2} = \frac{1}{1-2a(\cos\theta)z^{-1} + a^2z^{-2}}$$

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

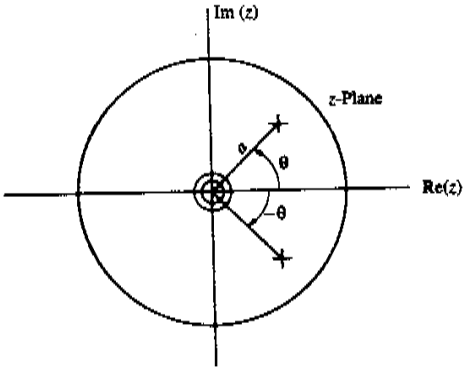
Semester: 5th

$$H(e^{j2\pi r}) = \frac{1}{1 - 2a(\cos\theta)e^{-j2\pi r} + a^2 e^{-j4\pi r}}$$

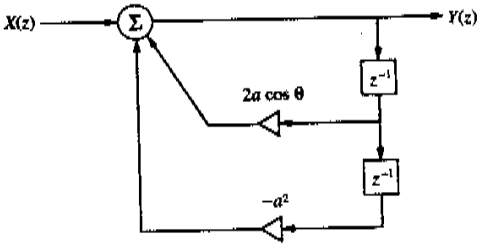
How do we calculate the amplitude response

$|H(e^{j2\pi r})|$ and $\angle H(e^{j2\pi r})$?

How the distance between the pole and the unit circle influence |H| and ∠H ?



(a) Pole-zero plot for transfer function



(b) Direct form implementation

FIGURE 9-6. Pole-zero plot and implementation for Example 9-2.

How the distance between the pole and the unit circle influence $|H|$?

$1 - a$

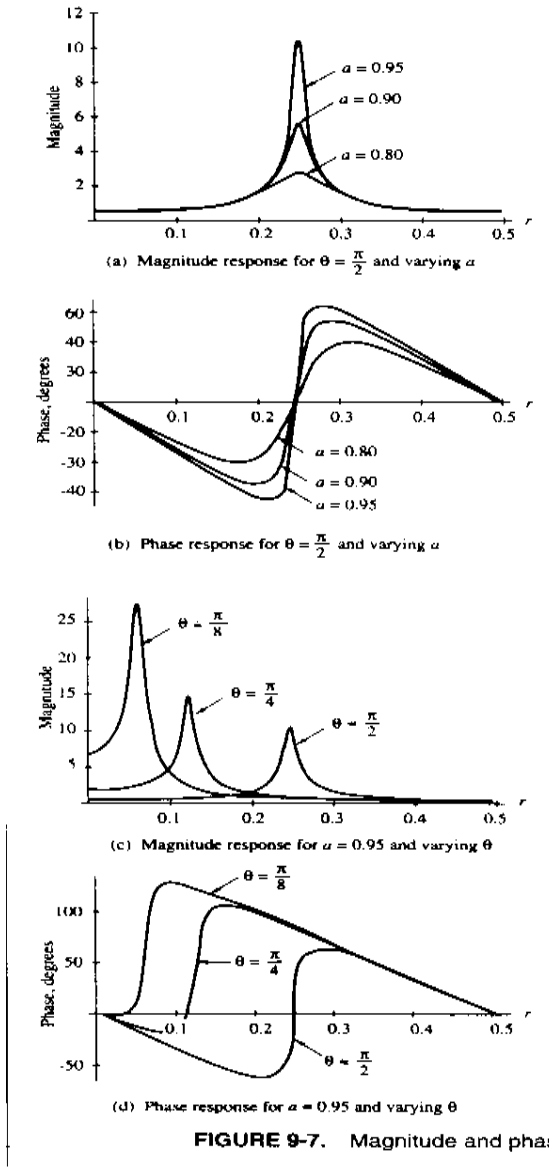
How the pole angle θ influence $|H|$ and $\angle H$?

See Fig. 9-7

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th



(a) Magnitude response for $\theta = \frac{\pi}{2}$ and varying a

(b) Phase response for $\theta = \frac{\pi}{2}$ and varying a

(c) Magnitude response for $a = 0.95$ and varying θ

(d) Phase response for $a = 0.95$ and varying θ

FIGURE 9-7. Magnitude and phase response for second-order filter.

HW: 9-2(a), 9-11, 9-13, 9-14

Discrete-Time Integration

A method of Discrete-time system Design: Approximate continuous-time system

Integrator $y(t) = \int_0^t x(\tau)d\tau$ ← a simple system

↑↑ system input Output

Discrete-time approximation of this system: discrete-time Integrator

Rectangular Integration

$$y(t) = \int_0^t x(\tau)d\tau = \int_0^{t_0^-} x(\tau)d\tau + \int_{t_0}^t x(\tau)d\tau = y(t_0) + \int_{t_0}^t x(\tau)d\tau$$

Subject Name: Digital Signal Processing

Subject Code: EC 503 **Contacts: 3L** **Credits: 3** **Lectures: 35 hrs.**

Semester: 5th

change of y from t0 to t : $\Delta y(t_0, t)$

t= nT, t0 = nT- T

$$\Rightarrow y(nT) = y(nT-T) + \int_{nT-T}^{nT} x(\tau)d\tau$$
$$= y(nT-T) + \Delta y(nT-T, nT)$$

Constant $(\tau \in [nT-T, nT])$

T small enough => $x(\tau) \approx x(nT-T)$

$$\Rightarrow \int_{nT-T}^{nT} x(\tau)d\tau \approx \int_{nT-T}^{nT} x(nT-T)d\tau$$

Constant

y(nT) : System output $\int_{nT-T}^{nT} d\tau$

x(nT) : System input , to be $T)$

$$y(nT) = y(nT-T) + Tx(nT-T)$$

A discrete-time integrator: rectangular integrator

$$\Rightarrow Y(z) = z^{-1}Y(z) + z^{-1}TX(z)$$

$$\Rightarrow H(z) = \frac{Y(z)}{X(z)} = \frac{z^{-1}T}{1-z^{-1}}$$

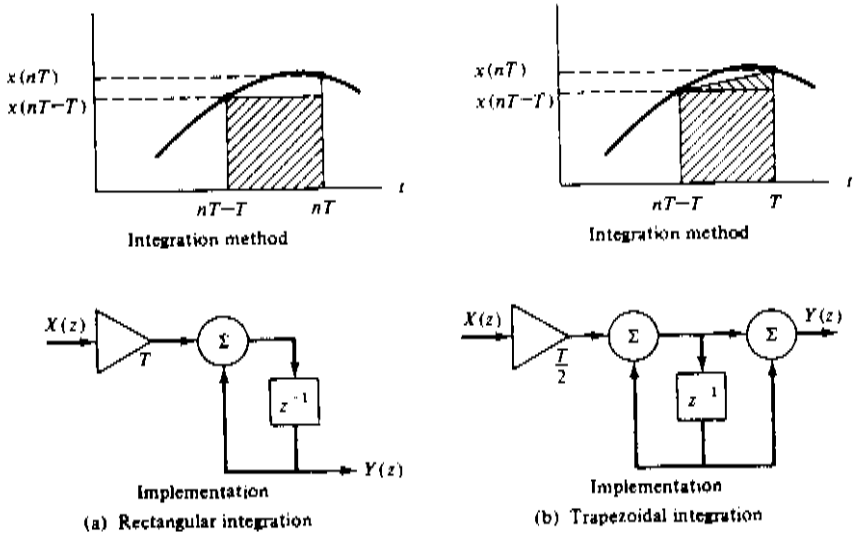


FIGURE 9-8. Rectangular and trapezoidal integration.

Trapezoidal Integration

Constants

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

$$\int_{nT-T}^{nT} x(\tau)d\tau = \int_{nT-T}^{nT} \frac{x(nT-T)+x(nT)}{2} d\tau$$

$$= \frac{T}{2}(x(nT-T)+x(nT))$$

$$y(nT) = y(nT-T) + \frac{T}{2}x(nT) + \frac{T}{2}x(nT-T)$$

$$y(nT) = y(nT-T) + Tx(nT-T) + \underbrace{\frac{T}{2}[x(nT)-x(nT-T)]}_{\substack{\text{difference} \\ \text{between} \\ \text{two} \\ \text{dicrete-time} \\ \text{integrators}}}$$

or

$$\Rightarrow Y(z) = z^{-1}Y(z) + \frac{T}{2}X(z) + \frac{T}{2}z^{-1}X(z)$$

$$H(z) = \frac{T}{2}\frac{1+z^{-1}}{1-z^{-1}}$$

Frequency Characteristics

Rectangular Integrator

$$H_r(z) = \frac{z^{-1}T}{1-z^{-1}}$$

Frequency Response

$$H_r(e^{j\omega T}) = \frac{Te^{-j\omega T}}{1-e^{-j\omega T}} = \frac{Te^{-j\omega T/2}}{e^{j\omega T/2}-e^{-j\omega T/2}} = \frac{Te^{-j\omega T/2}}{2j\sin\omega T/2}$$

Or

$$H_r(e^{j2\pi r}) = \frac{Te^{-j\pi r}}{2j\sin\pi r} \qquad \begin{pmatrix} \omega_s = 2\pi\frac{1}{T} \\ \omega/\omega_s = r \\ \Rightarrow \omega T = 2\pi r \end{pmatrix}$$

Amplitude Response

$$A_r(r) = \frac{T}{2\sin\pi r} \qquad 0 \le r \le \frac{1}{2} \quad (\sin\pi r \ge 0)$$

Phase Response

$$\Phi_r(r) = \angle e^{-j\pi r} - \angle j = -\pi r - \frac{\pi}{2} \qquad 0 \le r \le \frac{1}{2}$$

Subject Name: Digital Signal Processing

Subject Code: EC 503 **Contacts: 3L** **Credits: 3** **Lectures: 35 hrs.**

Semester: 5th

Trapezoidal Integrator

$$H_t(z) = \frac{T}{2}\frac{1+z^{-1}}{1-z^{-1}}$$

Frequency Response

$$H_t(e^{j2\pi r}) = \frac{T}{2}\frac{1+e^{-j2\pi r}}{1-e^{-j2\pi r}} = \frac{T}{2}\frac{e^{j\pi r}+e^{-j\pi r}}{e^{j\pi r}-e^{-j\pi r}}$$

$$= \frac{T}{2}\frac{2\cos\pi r}{2j\sin\pi r} = \frac{T}{2}\frac{\cos\pi r}{j\sin\pi r}$$

Amplitude: $A_t(r) = \dfrac{T\cos\pi r}{2\sin\pi r}$ $\qquad 0 \le r \le \dfrac{1}{2}$

Phase:

$$\Phi_t(r) = -\frac{\pi}{2} \qquad 0 \le r \le \frac{1}{2} \quad \left(=> \begin{matrix} \sin\pi r > 0 \\ \cos\pi r \ge 0 \end{matrix} \right)$$

Versus Ideal Integrator

Ideal (continuous-time) Integrator

$$H(j\omega) = \frac{1}{j\omega} \qquad \omega = 2\pi f = 2\pi r f_s$$

$$H(r) = \frac{1}{j2\pi r f_s}$$

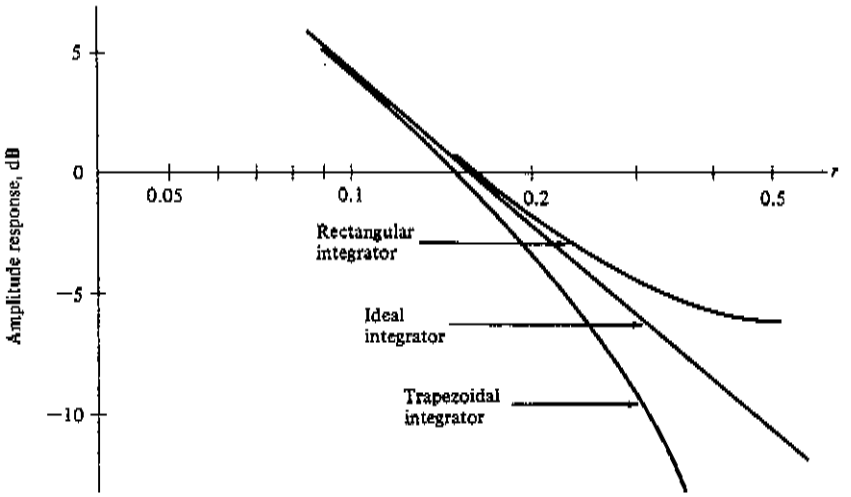
$$A(r) = \frac{1}{2\pi r f_s} \qquad \Phi(r) = -\frac{\pi}{2}$$

when T=1 second (Different plots and relationships will result if T is different.)

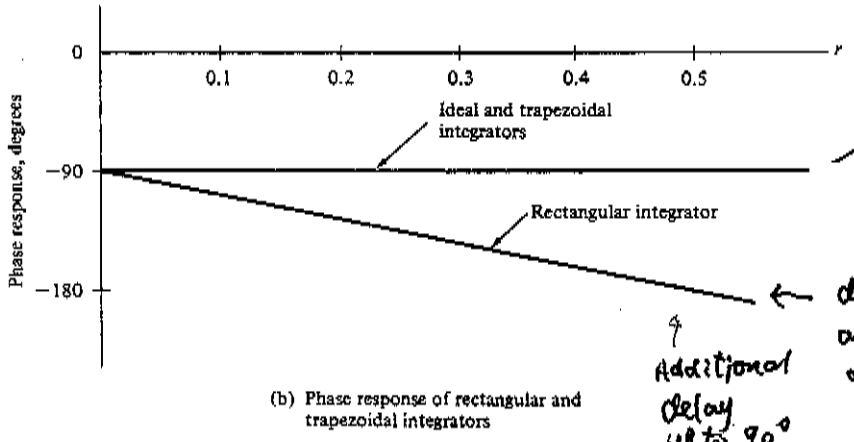
Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th



(a) Amplitude response of rectangular and trapezoidal integrators



(b) Phase response of rectangular and trapezoidal integrators

delay the input by $\frac{2\pi}{4}$ ($\frac{T}{4}$) (Should be)

← delay increases as the frequency of input increases

Additional delay up to 90°

FIGURE 9-9. Amplitude and phase responses of rectangular and trapezoidal integrators.

Low Frequency Range $\omega T << 1 \qquad (2\pi r << 1)$

(Frequency of the input is much lower than the sampling frequency:

It should be!)

$$\frac{\cos \pi r}{\sin \pi r} \approx \frac{1}{\pi r} \Rightarrow A_t(r) \approx \frac{T}{2}\frac{1}{\pi r} = \frac{1}{2\pi r f_s}$$

$$\frac{1}{\sin \pi r} \approx \frac{1}{\pi r} \Rightarrow A_r(r) \approx \frac{T}{2\pi r} = \frac{1}{2\pi r f_s}$$

High Frequency: Large error (should be)

Example 9-4 Differential equation (system)

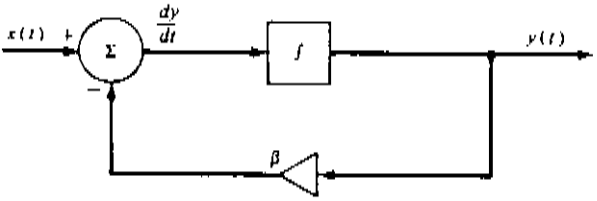
Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

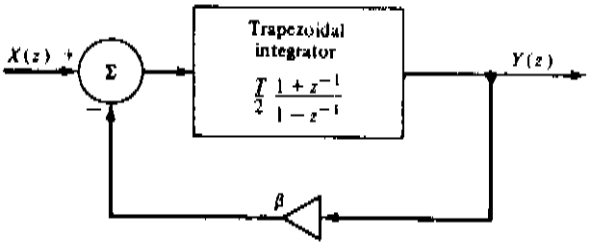
Semester: 5th

$$\frac{dy}{dt} = x(t) - \beta y(t)$$

Determine a digital equivalent.



(a) Analog system



(b) Digital equivalent using trapezoidal integration

Solution

(1) Block Diagram of the original system

(2) An equivalent

(3) Transfer Function Derivation

$$Y(z) = \frac{T}{2}\frac{1+z^{-1}}{1-z^{-1}}(X(z) - \beta Y(z))$$

$$\left(1 + \frac{T}{2}\beta\frac{1+z^{-1}}{1-z^{-1}}\right)Y(z) = \frac{T}{2}\frac{1+z^{-1}}{1-z^{-1}}X(z)$$

$$H(z) = \frac{Y(z)}{Z(z)} = \frac{T(1+z^{-1})}{(\beta T + 2) + (\beta T - 2)z^{-1}}$$

Design: 9-4 Find Equivalence of a given analog filter (IIR):

Including methods in Time Domain and Frequency Domain.

9-5 No analog prototype, from the desired frequency response: FIR

9-6 Computer-Aided Design

Infinite Impulse Response (IIR) Filter Design

(Given H(s) → Hd(z))

9-4A Synthesis in the Time-Domain: Invariant Design

Impulse – Invariant Design

Design Principle

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

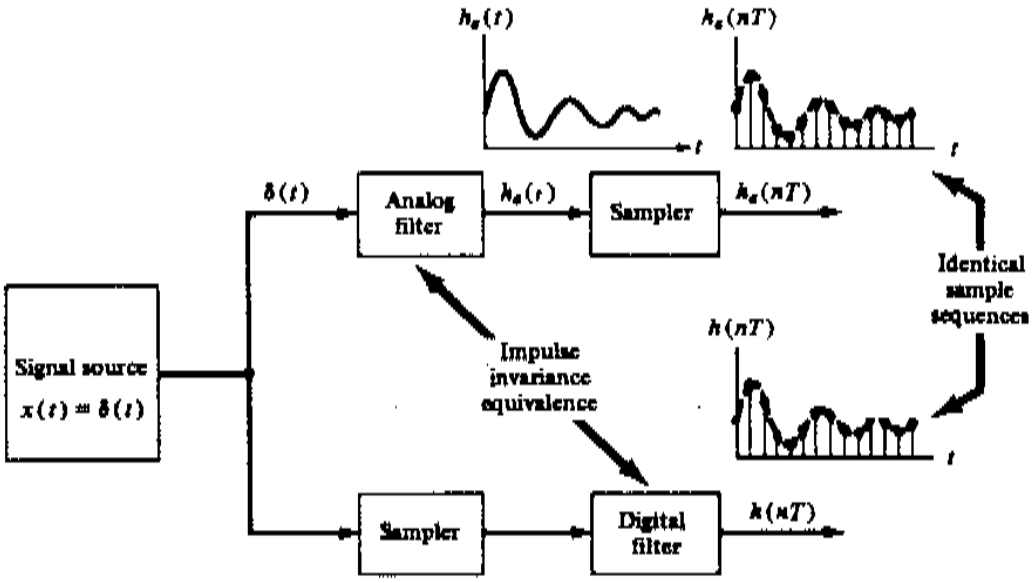


FIGURE 9-11. The concept of impulse invariance.

Illustration of Design Mechanism (Not General Case)

Assume:

Given analog filter (Transfer Function)

$$H_a(s) = \sum_{i=1}^{m} \frac{K_i}{s + s_i}$$ (a special case)

(2) Sampling Period T (sample ha(t) to generate ha(nT))

Derivation:

Impulse Response of analog filter

$$h_a(t) = L^{-1}(H_a(s)) = \sum_{i=1}^{m} k_i e^{-s_i t}$$

ha(nT): sampled impulse response of analog filter

$$h_a(nT) = \sum_{i=1}^{m} k_i e^{-s_i nT} = \sum_{i=1}^{m} k_i (e^{-s_i T})^n$$

z-transform of ha(nT)

Sampled impulse response of analog filter

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

$$Z(h_a(nT)) = \sum_{n=0}^{\infty} h_a(nT) z^{-n} = \sum_{n=0}^{\infty} \sum_{i=1}^{m} k_i (e^{-s_i T})^n z^{-n}$$

$$= \sum_{i=1}^{m} k_i \sum_{n=0}^{\infty} (e^{-s_i T} z)^{-n} = \sum_{i=1}^{m} k_i \frac{1}{1 - e^{-s_i T} z^{-1}} = \sum_{i=1}^{m} \frac{k_i}{1 - e^{-s_i T} z^{-1}}$$

Impulse-Invariant Design Principle

$$h(nT) \Leftrightarrow h_a(nT) \Rightarrow Z(h(nT)) \Leftrightarrow Z(h_a(nT))$$

↑

Digital filter is so designed that its impulse response h(nT)

equals the sampled impulse response of the analog filter ha(nT)

Hence, digital filter must be designed such that

z-transfer function $H(z)$ of the digital filter. Of course, the z-transfer function of its

$$Z(h(nT)) \Leftrightarrow Z(h_a(nT)) = \sum_{i=1}^{m} \frac{k_i}{1 - e^{-s_i T} z^{-1}}$$

$$\Rightarrow H(z) \Leftrightarrow \sum_{i=1}^{m} \frac{k_i}{1 - e^{-s_i T} z^{-1}}$$

(5) $$TZ(h(nT))\Big|_{z=e^{j\omega T}} \xrightarrow[T \to 0]{} L(h_a(t))$$ (scaling)

=> $$H(z) = T \sum_{i=1}^{m} \frac{k_i}{1 - e^{-s_i T} z^{-1}}$$

(3) Characteristics

(1) $H(z)\big|_{z=e^{j\omega T}} \to H_a(j\omega)$ when T→0

↑

frequency response of digital filter

$T \neq 0$ $\quad H(z)\big|_{z=e^{j\omega T}} \neq H_a(j\omega)$

Design: Optimized for T = 0

Not for T ≠ 0 (practical case) (due to the design principle)

(4) Realization: Parallel

Subject Name: Digital Signal Processing

Subject Code: EC 503 **Contacts: 3L** **Credits: 3** **Lectures: 35 hrs.**

Semester: 5th

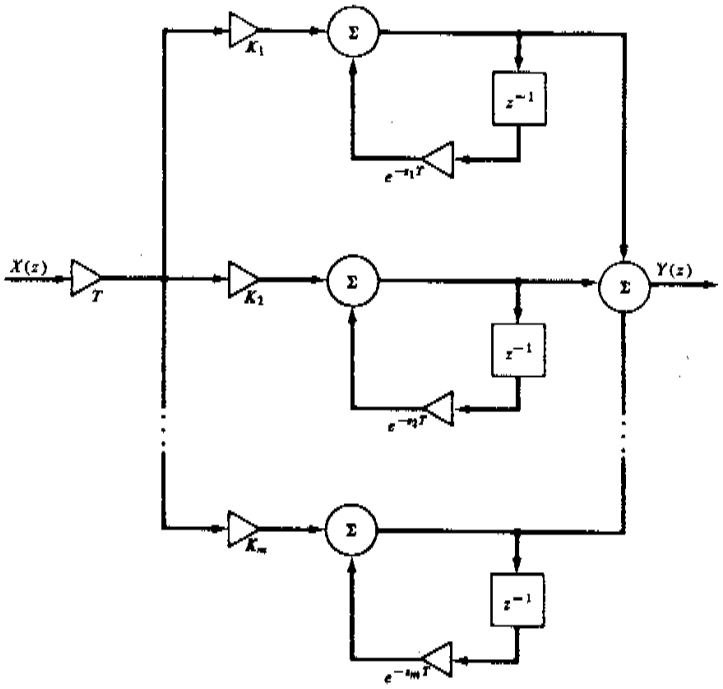


FIGURE 9-12. Impulse-invariant digital filter equivalent to (9-43).

(5) Design Example $H_a(s) = \dfrac{1}{s+1} \qquad T = 2s$

Solution: $m = 1, K_1 = 1, s_1 = -1$

$$H(z) = T\sum_{i=1}^{m} \frac{K_i}{1 - e^{-s_i T} z^{-1}} = 2\frac{1}{1 - e^{2} z^{-1}} = \frac{2}{1 - e^{2} z^{-1}}$$

General Time – Invariant Synthesis

Design Principle

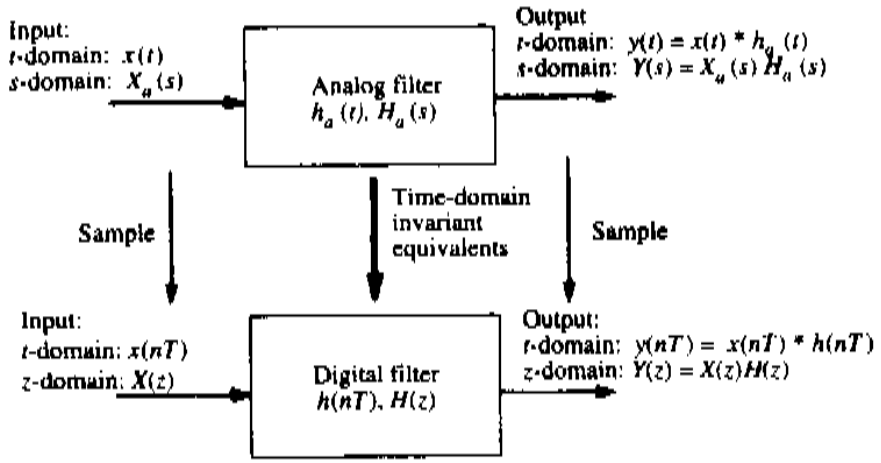


FIGURE 9-13. Time-domain invariance.

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

Derivation

Given: Ha(s) transfer function of analog filter

Xa(s) Lapalce transform of input signal of analog Filter

T sampling period

Find H(z) z-transfer function of digital filter

Response of analog filter xa(t)

$$y_a(t) = L^{-1}[H_a(s)X_a(s)]$$

(2) ya(nT) sampled signal of analog filter output

$$y_a(nT) = [L^{-1}[H_a(s)X_a(s)]\big|_{t=nT}$$

(3) z-transform of ya(nT)

$$Z[y_a(nT)] = Z\{[L^{-1}[H_a(s)X_a(s)]\big|_{t=nT}\}$$

(4) Time – invariant Design Principle

$$y(nT) = y_a(nT) \Rightarrow Z(y(nT)) \Leftrightarrow Z[y_a(nT)]$$

↑

Digital filter is so designed

that its output equals the sampled

output of the analog filter

Incorporate the scaling :

$$\underbrace{Z[y(nT)]}_{H(z)X(z)} \Leftrightarrow \underbrace{G}_{G=T} Z[y_a(nT)]$$

↑

z-transfer function of digital filter ← T

(5) Design Equation

$$H(z) = \frac{G}{X(z)} Z\{[L^{-1}(H_a(s)X_a(s))]\big|_{t=nT}\}$$

special case X(z)=1, Xa(s) = 1 (impulse)

$$\Rightarrow H(z) = GZ\{[L^{-1}(H_a(s))]\big|_{t=nT}\}$$

(6) Design procedure

A: Find $L^{-1}[H_a(s)X_a(s)] = y_a(t)$ (output of analog filter)

B: Find $y_a(nT) = y_a(t)\big|_{t=nT}$

C: Find $Z(y_a(nT))$

D: $H(z) = GZ(y_a(nT))$

Example 9-5 $$H_a(s) = \frac{0.5(s+4)}{(s+1)(s+2)}$$

Find digital filter H(z) by impulse - invariance.

Solution of design:

Find $L^{-1}[H_a(s)X_a(s)] = y_a(t)$

$$X_a(s) = 1$$

$$H_a(s) = \frac{0.5(s+4)}{(s+1)(s+2)} = \frac{1.5}{s+1} - \frac{1}{s+2}$$

$$y_a(t) = L^{-1}[H_a(s)X_a(s)] = 1.5e^{-t} - e^{-2t}$$

Find $y_a(nT) = y_a(t)\big|_{t=nT}$

$$y_a(nT) = 1.5(e^{-T})^n - (e^{-2T})^n$$

Find $Z(y_a(nT))$

$$Z(y_a(nT)) = \frac{1.5}{1-e^{-T}z^{-1}} - \frac{1}{1-e^{-2T}z^{-1}}$$

Find z-transfer function of the digital filter

$$H(z) = GZ(y_a(nT))$$

$$= G\left[\frac{1.5}{1-e^{-T}z^{-1}} - \frac{1}{1-e^{-2T}z^{-1}}\right]$$

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

use G = T

$$H(z) = \frac{1.5T}{1-e^{-T}z^{-1}} - \frac{T}{1-e^{-2T}z^{-1}}$$

Implementation

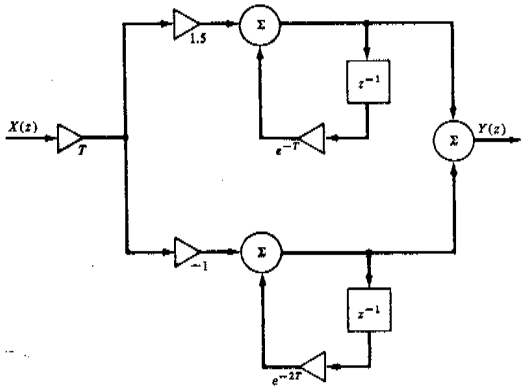


FIGURE 9-14. Impulse-invariant filter for Example 9-5.

1. In general, an FIR system is described by the difference equation $y(n)=\sum_{k=0}^{M-1} b_k x(n-k)$.

a) True

b) False

Answer: a

Explanation: The difference equation a href="https://www.sanfoundry.com/wp-content/uploads/2015/10/digital-signal-processing-questions-Answers-structures-fir-systems-1-q1.png"> $y(n)=\sum_{k=0}^{M-1} b_k x(n-k)$. describes the FIR system.

2. What is the general system function of an FIR system?

a) $\sum_{k=0}^{M-1} b_k z^{k}$
b) $\sum_{k=0}^{M} b_k z^{-k}$
c) $\sum_{k=0}^{M-1} b_k z^{-k}$
d) None of the mentioned

Answer: c

Explanation: We know that the difference equation of an FIR system is given by

$=>h(n)=b_k=>H(z)=\sum_{k=0}^{M-1} b_k z^{-k}$.

3. Which of the following is an method for implementing an FIR system?

a) Direct form

b) Cascade form

c) Lattice structure

d) All of the mentioned

Answer: d

Explanation: There are several structures for implementing an FIR system, beginning with the simplest structure, called the direct form. There are several other methods like cascade form realization, frequency sampling realization and lattice realization which are used for implementing and FIR system.

4. In cascade form of realization, how many bits should be used to represent the FIR filter coefficients in order to avoid the quantization effect on filter coefficients?

a. 5 to 10

b. 12 to 14

c. 20 to 24

d. 28 to 40

Answer: b

Explanation:

No explanation is available for this question!

5. The direct form realization is often called a transversal or tapped-delay-line filter.

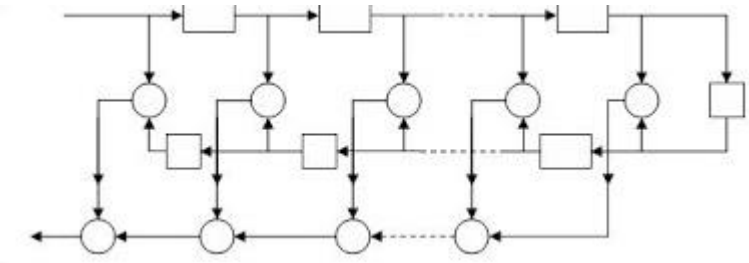
a) True

b) False

Answer: a

Explanation: The structure of the direct form realization, resembles a tapped delay line or a transversal system.

6. What is the condition of M, if the structure according to the direct form is as follows?



a) M even

b) M odd

c) All values of M

d) Doesn’t depend on value of M

Answer: b

Explanation: When the FIR system has linear phase, the unit sample response of the system satisfies either the symmetry or asymmetry condition, h(n)= ±h(M-1-n)

For such a system the number of multiplications is reduced from M to M/2 for M even and to (M-1)/2 for M odd. Thus for the structure given in the question, M is odd.

7. By combining two pairs of poles to form a fourth order filter section, by what factor we have reduced the number of multiplications?

a) 25%

b) 30%

c) 40%

d) 50%

Answer: d

Explanation: We have to do 3 multiplications for every second order equation. So, we have to do 6 multiplications if we combine two second order equations and we have to perform 3 multiplications by directly calculating the fourth order equation. Thus the number of multiplications are reduced by a factor of 50%.

8. The desired frequency response is specified at a set of equally spaced frequencies defined by the equation:

a) $\pi/2M(k+\alpha)$

b) $\pi/M(k+\alpha)$

c) $2\pi/M(k+\alpha)$

d) None of the mentioned

Answer: c

Explanation: To derive the frequency sampling structure, we specify the desired frequency response at a set of equally spaced frequencies, namely $\omega k=2\pi/M(k+\alpha)$,k=0,1…(M-1)/2 for M odd

k=0,1….(M/2)-1 for M even

α=0 or 0.5.

9. The realization of FIR filter by frequency sampling realization can be viewed as cascade of how many filters?

a) Two

b) Three

c) Four

d) None of the mentioned

Answer: a

Explanation: In frequency sampling realization, the system function H(z) is characterized by the set of frequency samples {H(k+ α)} instead of {h(n)}. We view this FIR filter realization as a cascade of two filters. One is an all-zero or a comb filter and the other consists of parallel bank of single pole filters with resonant frequencies.

10. What is the system function of all-zero filter or comb filter?

a) $\frac{1}{M}(1+z^{-M}e^{j2\pi\alpha})$

b) $\frac{1}{M}(1+z^{M}e^{j2\pi\alpha})$

c) $\frac{1}{M}(1-z^{M}e^{j2\pi\alpha})$

d) $\frac{1}{M}(1-z^{-M}e^{j2\pi\alpha})$

Answer: d

Explanation: The system function H(z) which is characterized by the set of frequency samples is obtained as

$$H(z)=\frac{1}{M}(1-z^{-M}e^{j2\pi\alpha})\sum_{k=0}^{M-1}\frac{H(k+\alpha)}{1-e^{j2\pi(k+\alpha)/M}z^{-1}}$$

We view this FIR realization as a cascade of two filters, $H(z)=H_1(z).H_2(z)$

Here $H_1(z)$ represents the all-zero filter or comb filter whose system function is given by the equation

$$H_1(z)=\frac{1}{M}(1-z^{-M}e^{j2\pi\alpha}).$$

11. The zeros of the system function of comb filter are located:

a) Inside unit circle

b) On unit circle

c) Outside unit circle

d) None of the mentioned

2) In the frequency response characteristics of FIR filter, the number of bits per coefficient should be _____in order to maintain the same error.

a. Increased

b. Constant

c. Decreased

d. None of the above

Answer: b

Explanation: The system function of the comb filter is given by the equation

$$H_1(z)=\frac{1}{M}(1-z^{-M}e^{j2\pi\alpha})$$

Its zeros are located at equally spaced points on the unit circle at

$$z_k=e^{j2\pi(k+\alpha)/M} \quad k=0,1,2....M-1$$

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

12. What is the system function of the second filter other than comb filter in the realization of FIR filter?

a) $\sum_{k=0}^{M} \frac{H(k+\alpha)}{1-e^{\frac{j2\pi(k+\alpha)}{M}}z^{-1}}$

b) $\sum_{k=0}^{M-1} \frac{H(k+\alpha)}{1+e^{\frac{j2\pi(k+\alpha)}{M}}z^{-1}}$

c) $\sum_{k=0}^{M-1} \frac{H(k+\alpha)}{1-e^{\frac{j2\pi(k+\alpha)}{M}}z^{-1}}$

d) None of the mentioned

Answer: c

Explanation: The system function H(z) which is characterized by the set of frequency samples is obtained as

$$H(z)=\frac{1}{M}(1-z^{-M}e^{j2\pi\alpha})\sum_{k=0}^{M-1}\frac{H(k+\alpha)}{1-e^{j2\pi(k+\alpha)/M}z^{-1}}$$

We view this FIR realization as a cascade of two filters, $H(z)=H_1(z).H_2(z)$
Here $H_1(z)$ represents the all-zero filter or comb filter, and the system function of the other filter is given by the equation

$$H_2(z)=\sum_{k=0}^{M-1}\frac{H(k+\alpha)}{1-e^{\frac{j2\pi(k+\alpha)}{M}}z^{-1}}.$$

13. Where does the poles of the system function of the second filter locate?

a) $e^{j2\pi(k+\alpha)M}$
b) $e^{j2\pi(k+\alpha)/M}$
c) $e^{j2\pi(k-\alpha)/M}$
d) $e^{j\pi(k+\alpha)/M}$

Answer: b

Explanation: The system function of the second filter in the cascade of an FIR realization by frequency sampling method is given by

$$H_2(z)=\sum_{k=0}^{M-1}\frac{H(k+\alpha)}{1-e^{\frac{j2\pi(k+\alpha)}{M}}z^{-1}}$$

We obtain the poles of the above system function by equating the denominator of the above equation to zero.

$$=>1-e^{\frac{j2\pi(k+\alpha)}{M}}z^{-1}=0$$
$$=>z=p_k=e^{\frac{j2\pi(k+\alpha)}{M}},\ k=0,1....M-1$$

14. When the desired frequency response characteristic of the FIR filter is narrowband, most of the gain parameters {H(k+α)} are zero.

a) True

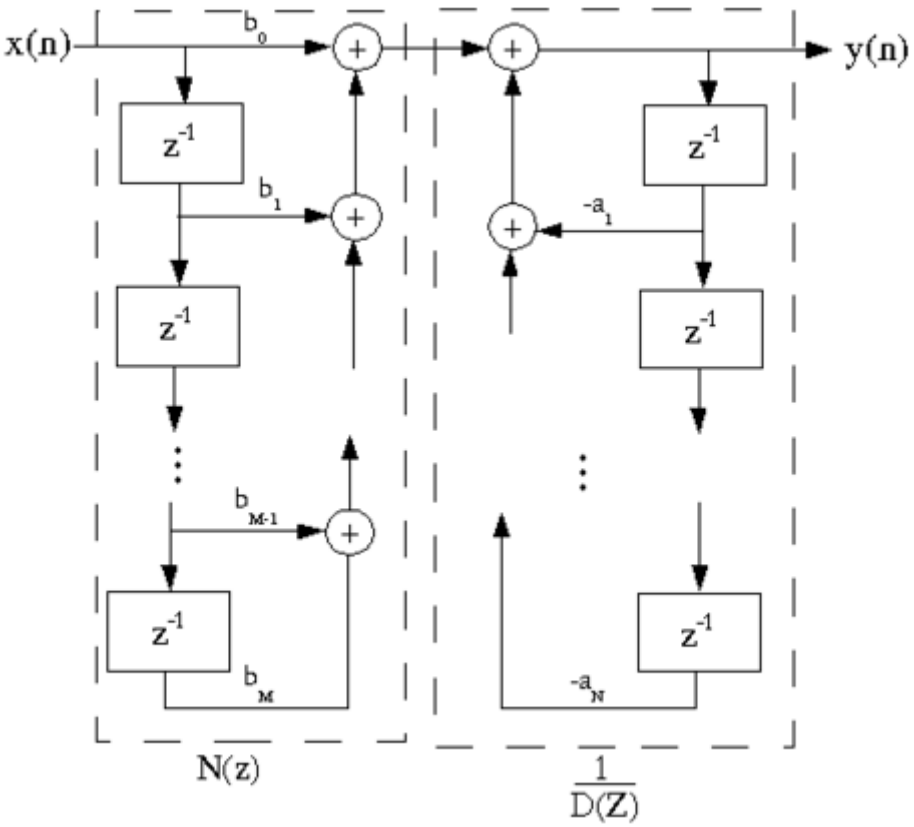
b) False

Answer: a

Explanation: When the desired frequency response characteristic of the FIR filter is narrowband, most of the

gain parameters {H(k+α)} are zero. Consequently, the corresponding resonant filters can be eliminated and only the filters with nonzero gains need be retained.

15. Which of the following filters have a cascade realization as shown below?



a) IIR filter

b) Comb filter

c) High pass filter

d) FIR filter

Answer: d

Explanation: The system function of the FIR filter according to the frequency sampling realization is given by the equation

$$H(z)=\frac{1}{M}(1-z^{-M}e^{j2\pi\alpha})\sum_{k=0}^{M-1}\frac{H(k+\alpha)}{1-e^{j2\pi(k+\alpha)/M}z^{-1}}$$

The above system function can be represented in the cascade form as shown in the above block diagram.

16. In FIR filters, which among the following parameters remains unaffected by the quantization effect?

a. Magnitude Response

b. Phase Characteristics

c. Both a and b

d. None of the above

Answer: Phase Characteristics

Explanation:

No explanation is available for this question!

17. Which filters exhibit their dependency upon the system design for the stability purpose?

a. FIR

b. IIR

c. Both a and b

d. None of the above

Answer: IIR

Explanation:

No explanation is available for this question!

18. In linear phase realization, equal valued coefficients are taken common for reducing the requisite number of _____.

a. adders

b. subtractors

c. multipliers

d. dividers

Answer: multipliers

Explanation:

No explanation is available for this question!

19. How is/are the roundoff errors reduced in the digital FIR filter?

a. By representation of all products with double-length registers

b. By rounding the results after acquiring the final sum

c. Both a and b

d. None of the above

Answer: Both a and b

Explanation:

No explanation is available for this question!

20. Which is/are the correct way/s for the result quantization of an arithmetic operation?

a. Result Truncation

b. Result Rounding

c. Both a and b

d. None of the above

Answer: Both a and b

Explanation:

No explanation is available for this question!

21. Which among the following has/have a provision to support an adaptive filtering mechanism?

a. IIR

b. FIR

c. Both a and b

d. None of the above

Answer: Both a and b

Explanation:

No explanation is available for this question!

22. How is the sensitivity of filter coefficient quantization for FIR filters?

a. Low

b. Moderate

c. High

d. Unpredictable

Answer: Low

Explanation:

No explanation is available for this question!

23. In tapped delay line filter, the tapped line is also known as _____

a. Pick-on node

b. Pick-off node

c. Pick-up node

d. Pick-down node

Answer: Pick-off node

Explanation:

No explanation is available for this question

Referrences:

1) Digital Signal Processing : Principles, Algorithms, and Applications 4 Edition (English, Paperback, Dimitris G Manolakis, John G. Proakis)

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

2) https://www.mathworks.com › Examples Home › Digital and Analog Filters

3) Digital Signal Processing, Alan V. Oppenheim

4) Digital Signal Processing by P.RAMESH BABU

Module IV

Finite word length effect Introduction

Introduction:

Practical digital filters must be implemented with finite precision numbers and arithmetic. As a result, both the filter coefficients and the filter input and output signals are in discrete form. This leads to four types of finite word length effects. Discretization (quantization) of the filter coefficients has the effect of perturbing the location of the filter poles and zeroes. As a result, the actual filter response differs slightly from the ideal response. This deterministic frequency response error is referred to as coefficient quantization error. The use of finite precision arithmetic makes it necessary to quantize filter calculations by rounding or truncation. Round off noise is that error in the filter output that results from rounding or truncating calculations within the filter. As the name implies, this error looks like low-level noise at the filter output. Quantization of the filter calculations also renders the filter slightly nonlinear. For large signals this nonlinearity is negligible and round off noise is the major concern. However, for recursive filters with a zero or constant input, this nonlinearity can cause spurious oscillations called limit cycles. With fixed-point arithmetic it is possible for filter calculations to overflow. The term overflow oscillation, sometimes also called adder overflow limit cycle, refers to a high-level oscillation that can exist in an otherwise stable filter due to the nonlinearity associated with the overflow of internal filter calculations.

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

Coefficient Quantization

The coefficient quantization results in FIR filter changing its transform function. The position of FIR filter zeros is also changed, whereas the position of its poles remains unchanged as they are located in z=0. Quantization has no effect on them. The conclusion is that quantization of FIR filter coefficients cannot cause a filter to become instable as is the case with IIR filters. Even though there is no danger of FIR filter destabilization, it may happen that transfer function is deviated to such an extent that it no longer meets the specifications, which further means that the resulting filter is not suitable for intended implementation. The FIR filter quantization errors cause the stopband attenuation to become lower. If it drops below the limit defined by the specifications, the resulting filter is useless. Transfer function changes occurring due to FIR filter coefficient quantization are more effective for high-order filters. The reason for this is the fact that spacing between zeros of the transfer function get smaller as the filter order increases and such slight changes of zero positions affect the FIR filter frequency response.[2]

We have seen that a digital filter, or a general digital signal processing system, operates on an input sampled-data signal to produce an output sampled-data signal by means of a computational algorithm. Since the sampled-data signals are represented by number sequences, these are quantized and encoded using binary codes, and the processor algorithm can be implemented either in software using a general purpose computer or in dedicated hardware. The latter approach is becoming increasingly popular due to the advances in very large scale intergration (VLSI) and the resulting availability of integrated circuit modules and special purpose hardward with sufficient memory size, complexity and speed so as to render the hardware implementation of digital filters, operating in real time, an attractive technique.

Now, regardless of the type of implementation, the numbers processed by the digital system are ultimately stored in (memory) registers with finite capacity. Therefore, all digital networks operate with only a finite number of binary digits (bits); thus resulting in an inherent limitation on the accuracy of processing. In this chapter, we discuss the effect of using finite word-lengths to represent the numbers and the arithmetic operations, on the accuracy of digital signal processors in general and digital filters in particular .

The errors due to the use of finite word-lengths to represent the pertinent numbers can be of the following types:

Input signal quantization effects

These are errors due to the representation of the input signal by a set of values with discrete amplitudes and subsequently by binary numbers with finite word-lengths. Typical word-lengths are 32 or 64 bits.

Coefficient quantization effects

These result from the representation of each filter coefficient: ar, br in the expression

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

$$H(z) = \frac{\sum_{r=0}^{M} a_r z^{-r}}{1 + \sum_{r=0}^{N} b_r z^{-r}} \qquad 4.1$$

by a quantized discrete-amplitude value and subsequently in binary form by a finite number of bits. We recall that the solution to the approximation problem leads to a transfer function of the general form (4.1) in which the coefficients (ar, br) are assumed capable of being represented to an arbitrary degree of accuracy. However, prior to implementation, these coefficients are quantized and each represented by a finite number of bits; again typically 32 or 64 bits. Consequently, the filter response may deviate considerably from the desired characteristic. Moreover, the error due to coefficient quantization may cause the poles of the transfer function in (4.1) to change their positions in the z-plane, perhaps moving to points on the unit circle or exterior to it and causing the filter to become unstable.

i. Product quantization and accumulation errors

In the practical implementation of digital filters using dedicated hardware or a microprocessor, a fixed length l of shift registers is chosen for use throughout the filter. In the arithmetic operation of multiplication a signal represented by an l-bit number is, however, multiplied by a coefficient represented by another, say, l-bit number also. This results in a product having 2l digits. But this number must be rounded or truncated to l bits, which is the uniform length of each register. This results in errors known as product quantization errors, which accumulate as the operations of multiplication are performed to implement the filter.

ii. Auto-oscillations due to overflow and limit cycles

In performing arithmetic operations, overflow may result, giving rise to self-sustained oscillations, which represent a type of instability of the filter. These may occur if the capacity of the memory is exceeded when logic saturation devices are not used. Furthermore, auto-oscillations may occur even when the filter input is nominally zero, because even in the absence of data an error signal exists which is caused by the quantization of the internal data before storage in the memory. This effect is called limit cycle oscillations.

Numbers are represented in either fixed-point or floating-point format. In either case, number quantization is performed by rounding or truncation. In fixed-point arithmetic with l-bit word-length, number quantization is assumed to be a random process with the resulting error having uniform probability density functions in the appropriate range as shown in Figure 4.1. This is a reasonable assumption since it is clear that there is no 'preferred' range of the error. For floating-point arithmetic with l-bit mantissa, the corresponding error probability density functions are shown in Figure4 .2. In our discussion of finite word- length effects, we concentrate on the case of fixed-point arithmetic.

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

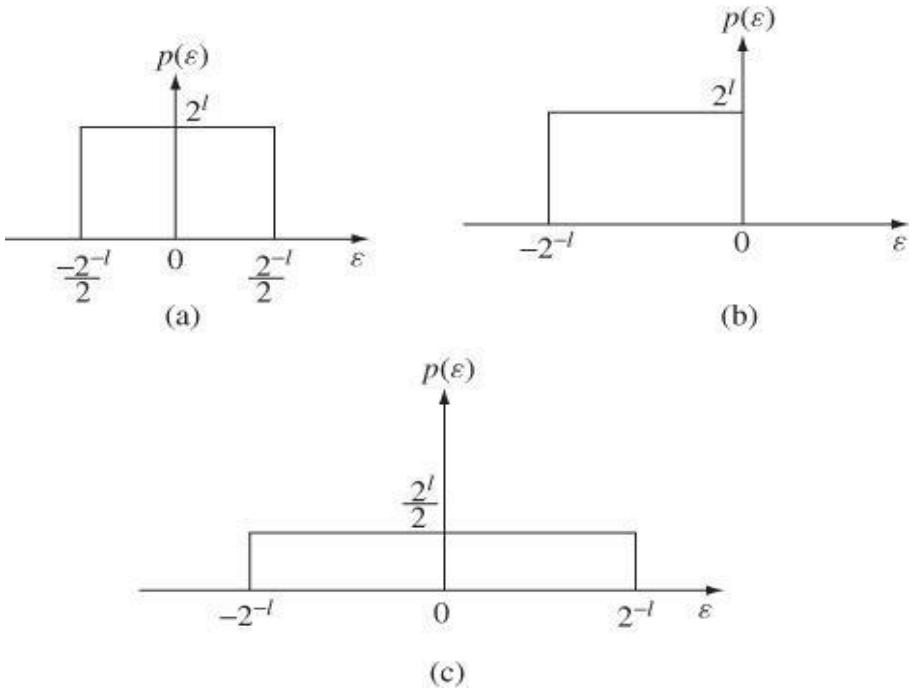


Figure 4.1 Error probability density functions in fixed-point arithmetic: (a). rounding, (b) truncation with two's complement and (c) truncation with one's complement or signed magnitude[1]

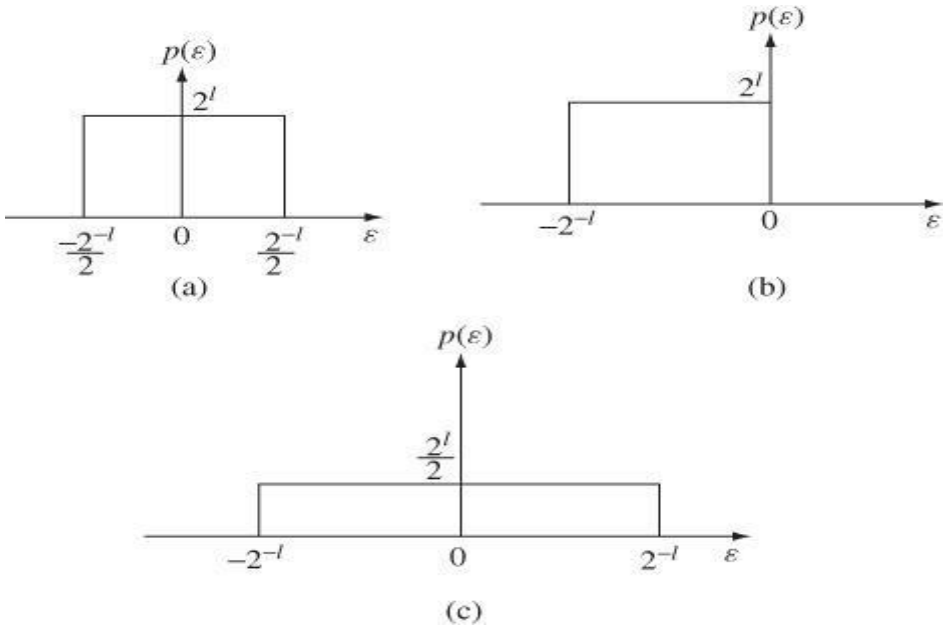


Figure4 .2 Error probability density functions in floating-point arithmetic: **(a)** rounding, **(b)** truncation with two's complement and **(c)** truncation with one's complement or signed-magnitude[1]

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

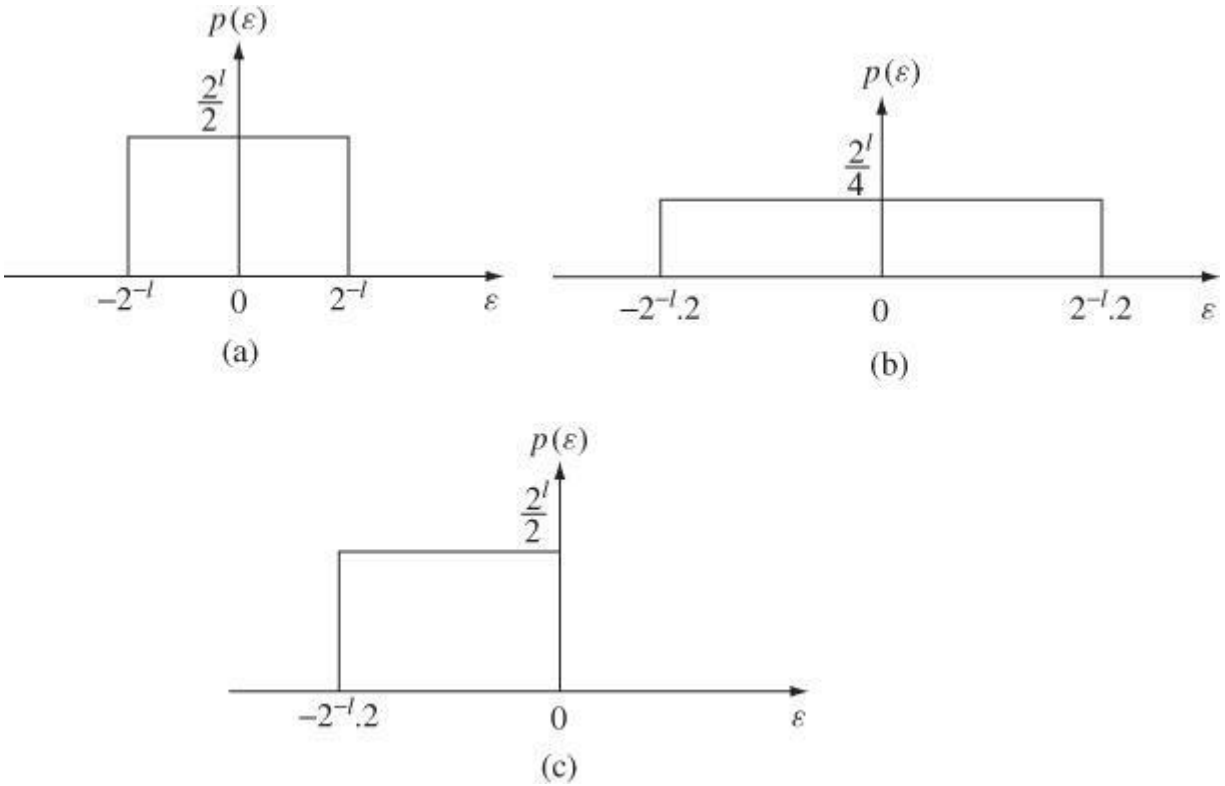


Figure 4.3 (a) The output $fq(t)$ of a quantizer with input $f(t)$. (b) The quantization error

4.1 Input Signal Quantization Errors

The process of analog to digital conversion involves sampling the continuous-time signal $f(t)$, to give the sampled version $f(nT)$, which is quantized in amplitude to give fq(nT); this is finally encoded to produce the input to the digital system. The process of sampling, if the pertinent rules are followed, does not produce errors. In contrast, the difference between the actual *analog* input sample and the corresponding binary-coded quantized value, is called ***quantization error*** **or** ***quantization noise*****.** This is the first source of degradation in the processing, which is now studied in some detail, concentrating on the use of fixed-point arithmetic. [1]Typically, an A/D converter used in conjunction with an implementation of a digital filter, has an (l + 1)-bit fixed-point output, including a sign bit. This corresponds to a resolution of about one part in 2^l. The quantizing step q is, therefore

$$q = 1/2^l. \qquad \textbf{4.2}$$

If rounding is used in the quantization of the samples $f(nT)$, then the maximum possible quantization error is $\pm q/2$. Figure 4.3(a) shows the output of a quantizer (prior to encoding) with the input as the original continuous signal $f(t)$. Figure .3(b) shows the corresponding quantization error, and the amplitude of the error signal lies between $-q/2$ and $q/2$. The amplitudes of the quantized signal are called the *decision amplitudes*. It is generally reasonable to assume that the round-off error ε is white noise with uniform probability density $p(\varepsilon)$ as shown in Figure 4.4(a). The variance (power) of the error signal can be taken as a measure of the degradation suffered by the signal due to quantization. When the quantizing step is very

$$\begin{aligned}\sigma_0^2 &= \frac{1}{\tau}\int_{-\tau/2}^{\tau/2}\varepsilon^2(t)\,\mathrm{d}t \\ &= \frac{q^2}{\tau^3}\int_{-\tau/2}^{\tau/2}t^2\mathrm{d}t\end{aligned}$$

small compared with the signal variations, the error signal can be considered equivalent to the sum of basic error signals, each approximated by a straight line segment as shown in Figure 4.4(b). The average power (variance) of a basic error signal of width τ is
or

$$\sigma_0^2 = \frac{q^2}{12} = \frac{2^{-2l}}{12}. \qquad \text{.....4.3}$$

Now, after quantization, the input to the filter is

$$f_{\mathrm{q}}(nT) = f(nT) + \varepsilon(nT). \qquad \text{.....4.4}$$

Since the filter is a linear system, its output is the sum of two components: one due to the signal $f(nT)$ and the other due to the quantization error $\varepsilon(nT)$. Thus, the error is also filtered by the transfer function $H(z)$ of the filter. Since the error $\varepsilon(nT)$ is assumed white noise with zero mean and variance $q^2/12$, then the steady state output component due to $\varepsilon(nT)$ is a zero-mean wide-sense stationary process with power spectral density given, in the z-domain, by

$$P_d(z) = \frac{1}{12} q^2 H(z) H(z^{-1}) \qquad \text{.....4.5}$$

so that

$$p_d(w) = \frac{1}{12} q^2 H(\exp(-\mathrm{j}\omega T)) H(\exp(\mathrm{j}\omega T)). \qquad \text{.....4.6}$$

Here, we have neglected the effect of coefficient quantization and round-off accumulation since their effect on the response due to $\varepsilon(nT)$ is much smaller than that due to $f(nT)$. Thus, the output mean-square error (output noise power) is given by

$$\sigma_{\mathrm{out}}^2 = \frac{q^2}{12} \sum_{n=0}^{\infty} [h(nT)]^2 \qquad \text{.....4.7}$$

where $\{h(nT)\}$ is the impulse response sequence of the filter, which is assumed causal. Alternatively, the mean square value of the output error due to input quantization may be obtained from (4.5) and Parseval's relation in the z-domain

$$\sigma_{\mathrm{out}}^2 = \frac{q^2}{12} \frac{1}{\mathrm{j}2\pi} \oint_c H(z) H(z^{-1}) \frac{\mathrm{d}z}{z} \qquad \text{.....4.8}$$

where the integration is carried out over a closed contour c enclosing all the singularities of $H(z)$. The integral can be evaluated using the method of the residues, numerically, algebraically or by a computer programme.

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

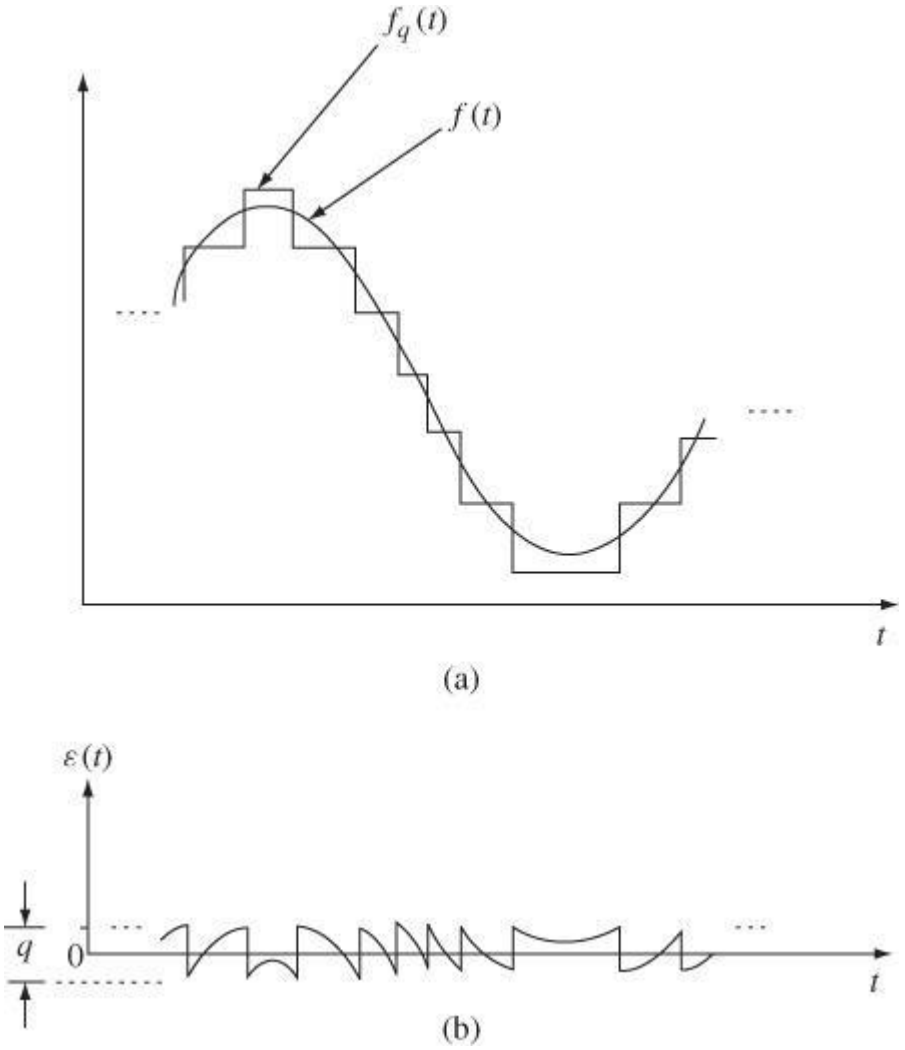
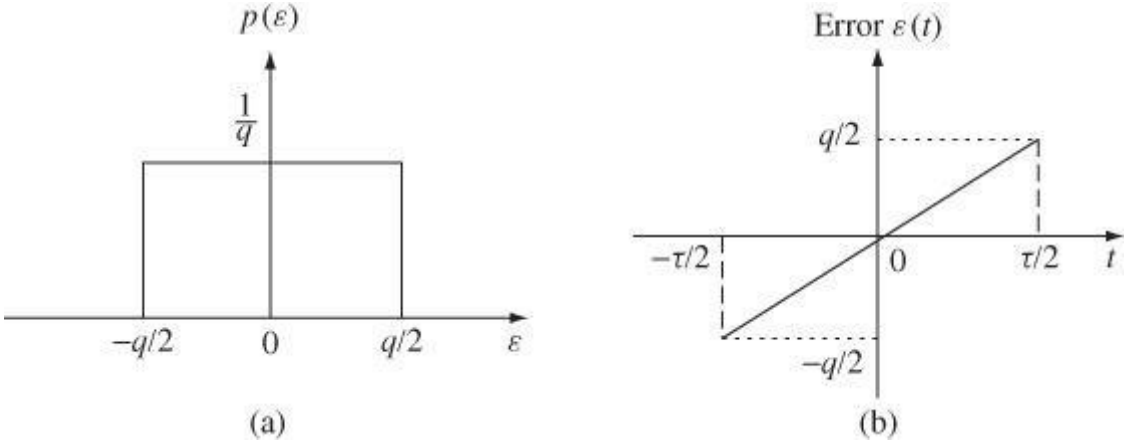


Figure 4.4 **(a)** Probability density of round-off error. **(b)** A basic error signal approximation



Now, let each member of the quantized sequence $f_q(nT)$ be represented by an l-bit binary number. Then, the maximum number of quantized amplitudes which can be encoded into binary form is 2^l. It follows that the range of amplitudes A which can be encoded lies in the range

$$q \leq A \leq q2^l. \qquad \text{.....4.9}$$

Therefore, any amplitude value exceeding $q2^l$ cannot be represented, and the signal is *clipped* which results in degradation.

Let the range of signal amplitudes be $[-A_m, A_m]$ so that

$$A_m = \frac{q2^l}{2}. \quad \text{.....4.10}$$

If rounding is used in the quantization operation, the error signal is

$$|\varepsilon(nT)| \leq A_m 2^{-l}. \quad \text{.....4.11}$$

Now, define the *peak power* of a coder as the power of the sinusoidal signal with the maximum possible amplitude A_m which the coder can pass without clipping (see Figure 4.5). Thus, the peak power is given by

$$\begin{aligned} P_c &= \frac{1}{2}\left(\frac{q2^l}{2}\right)^2 \\ &= q^2 2^{2l-3}. \end{aligned} \quad \text{.....4.12}$$

The coding dynamic range R_c is defined as the ratio of the peak power to the quantization noise power. Hence, using (4.3) and (4.12) we have

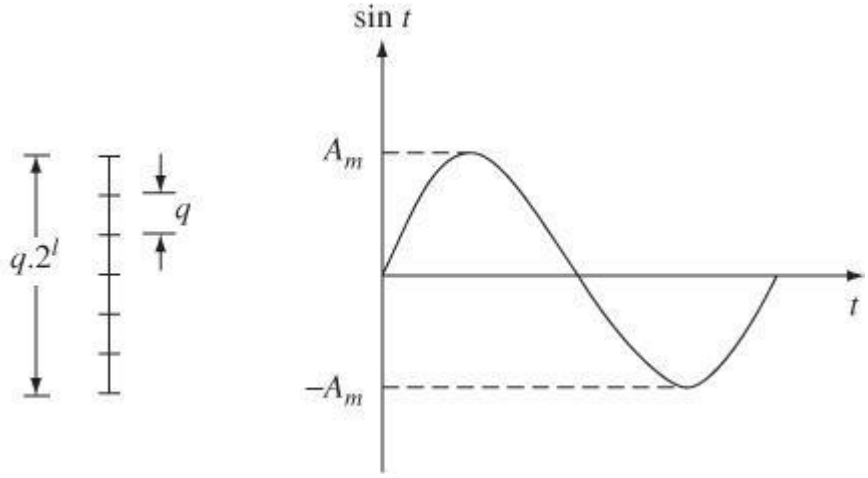
$$\begin{aligned} R_c &= \frac{P_c}{\sigma_0^2} \\ &= 3(2^{2l-1}) \end{aligned} \quad \text{.....4.13}$$

or

$$10 \log \left(\frac{P_c}{\sigma_0^2}\right) = 6.02l + 1.76 \text{ dB}. \quad \text{..... 4.14}$$

For example, an 6-bit coder has a dynamic range of about 50 dB while a 16-bit coder has a dynamic range of about 96 dB.

Figure 4.5 Pertinent to the definition of the coding dynamic range of the A/D converter



If the range of amplitudes of the input signal exceeds the coder dynamic range, then scaling of the signal prior to quantization can be applied to reduce the amplitude range; thus eliminating clipping. Consequently, in the model of quantization expressed by (4.4), a scaling factor K is incorporated such that

$$f_q(nT) = Kf(nT) + \varepsilon(nT) \quad \text{.....4.15}$$

where

$$0 < K < 1. \quad \text{.....4.16}$$

The signal to noise ratio in this case is given by

$$\mathrm{SNR} = 10\log\left(\frac{k^2\sigma_f^2}{\sigma_0^2}\right)\mathrm{dB} \quad \text{.....4.17}$$

where $K^2\sigma_f^2$ is the power of the scaled signal, and σ_0^2 is the quantization noise power. It is found that negligible clipping results with the choice

$$K = 1/5\sigma_f \quad \text{.....4.14}$$

so that

$$\begin{aligned}\mathrm{SNR} &= 10\log\left(\frac{1}{25\sigma_0^2}\right)\\ &= 10\log\left(\frac{12}{25q^2}\right)\mathrm{dB}\end{aligned} \quad \text{.....4.19}$$

and using (4.2) the above expression becomes

$$\mathrm{SNR} = 6.02l - 3.1876\,\mathrm{dB}. \quad \text{.....4.20}$$

Thus, for an 6-bit A/D converter, the signal to noise ratio is about 45 dB, whereas for a 16-bit converter it is about 100 dB.

4.2 Coefficient Quantization Effects

When binary words with finite length are used to represent the filter coefficients a_r and b_r, each coefficient is replaced by its *l*-bit representation in either fixed-point or floating-point form. Thus, a coefficient a_r is replaced by $(a_r)_q$ where for the fixed-point representation[1]

$$(a_r)_q = a_r + \alpha_r \quad \text{.....4.21}$$

while in the floating-point form

$$(a_r)_q = \alpha_r(1 + \alpha_r) \quad \text{.....4.22}$$

with

$$|\alpha_r| \le 2^l. \quad \text{.....4.23}$$

Similarly b_r becomes

$$(b_r)_q = b_r + \beta_r \quad \text{.....4.24}$$

in the fixed-point form, or

$$(b_r)_q = b_r(1 + \beta_r) \quad \text{.....4.25}$$

in the floating-point representation. Again

$$|\beta_r| \le 2^{-l}. \quad \text{.....4.26}$$

It now follows that the filter characteristics deviate from the desired ones. To study this effect we can compute the frequency response of the actual filter with *l*-bit rounded coefficients, that is using the actual transfer function

$$H_{\mathrm{q}}(z) = \frac{\sum_{r=0}^{M} (\alpha_r)_{\mathrm{q}} z^{-r}}{1 + \sum_{r=1}^{N} (b_r)_{\mathrm{q}} z^{-r}} \qquad \text{.....4.27}$$

to evaluate

$$H_{\mathrm{q}}(\exp(\mathrm{j}\omega T) = |H_q(\exp(\mathrm{j}\omega T))| \exp(\mathrm{j}\psi_{\mathrm{q}}(\omega T)). \qquad \text{.....4.28}$$

The result is then compared with the desired theoretical response $H(\exp(\mathrm{j}\omega T))$ of the original design obtained from the solution to the approximation problem. Naturally, the longer the word used to represent the numbers, the closer the actual response to the desired one.

An alternative approach to the study of the above effects is to calculate the movements of the poles and zeros of the transfer function due to coefficient rounding, then apply sensitivity theory to examine the changes in the filter response. Let the poles $H(z)$ be $pi(i = 1, 2, \ldots N)$ and the poles of $H\mathrm{q}(z)$ be at $(pi + \Delta pi)$. Then it has been shown [20] that the variation in the typical pole position is given by

$$\Delta P_i = \sum_{r=1}^{N} \frac{p_i^{r+1}}{\prod_{\substack{m=1 \\ m \neq i}} \left(1 - \frac{p_i}{p_m}\right)} \Delta a_r \qquad \text{.....4.29}$$

where Δar is the change in the coefficient due to rounding, which is either αr or $ar \quad \alpha r$. Similar results can be obtained for the movement of the zeros. From these perturbations, the deviations in the overall filter response can be examined.

In addition to altering the frequency response, coefficient quantization can also affect the stability of an IIR filter. If the poles happen to be close to the unit circle in the *z*-plane, coefficient quantization can cause their positions to be sufficiently perturbed so as to move to points on the unit circle or exterior to it, thus producing instability. It has been shown

that for an *N*th order low-pass IIR filter operating at a sampling frequency of $1/T$ with distinct poles at ($\cos \omega_r T \pm \mathrm{j} \sin \omega_r T$), stability is guaranteed if the number of bits satisfies the, rather pessimistic, inequality

$$l > -\log_2 \left(\frac{5\sqrt{N}}{2^{N+2}} \prod_{r=1}^{N} \omega_r T \right). \quad \text{.....4.30}$$

Example 4.1

Consider the transfer function

$$H(z) = \frac{z}{z^2 - 1.7z + 0.745}$$

and determine the coefficient word-length required to avoid

instability. Solution. The poles occur at z = (0.65 ± j0.15), that is

$$\begin{aligned} \omega_1 T &= \tan^{-1}\left(\frac{0.15}{0.85}\right) \\ &= 0.1747 \text{ rad} \\ &= \omega_2 T. \end{aligned}$$

Thus, using the estimate in (4.30) we obtain

$$l > -\log_2 \left(\frac{5\sqrt{2}}{2^4} (0.1747)^2 \right)$$

That is

$$l > -\log_2(0.0135)$$

so that l > 6.21 and we take l = 7.0. As noted before, the estimate is somewhat pessimistic, particularly since it is based on quantization by truncation rather than rounding.

Generally, the effect of coefficient quantization is more significant for a high degree filter realized in direct form, than the corresponding cascade or parallel realizations. Therefore the parallel or cascade form should be used for high degree filters since the saving in the required word-length is quite significant.

Finally, we note that due to coefficient quantization, the output sequence will differ from the desired one. The analysis of the resulting errors will be undertaken in the next section, in conjunction with round-off accumulation.

4.3 Effect of Round-off Accumulation

To begin with, we shall examine the effect of round-off accumulation due to product quantization without taking into account the effect of coefficient quantization. Then the combined errors due to both effects are studied. In our discussion, we concentrate on the case of fixed-point arithmetic.

4.3.1 Round-off Accumulation without Coefficient Quantization

These errors are dependent on the particular form of realization. Hence each form is treated separately.

4.3.1.1 Direct Form

An IIR filter transfer function

$$H(z) = \frac{\sum_{r=0}^{M} a_r z^{-r}}{1 + \sum_{r=1}^{N} b_r z^{-r}} = \frac{N(z^{-1})}{D(z^{-1})} \quad \text{.....4.}$$

31 corresponds to the difference equation

$$g(n) = \sum_{r=0}^{M} a_r f(n-r) - \sum_{r=1}^{N} b_r g(n-r) \quad \text{.....4.32}$$

where the sampling period T has been dropped for convenience. In fixed-point arithmetic the result of multiplying two l-bit numbers is a 2l-bit product; this is rounded to an l-bit word. In dealing with this type of product quantization error we assume that the signal levels throughout the filter are much larger than the quantizing step q. This allows us to treat product quantization errors at the output of the multipliers as uncorrelated (statistically independent) random variables, each being white noise with power spectral density q2/12. Therefore, the total error ε(n) is the sum of all those due to the multiplications. If μ is the number of coefficients ar which are neither 0 or 1, and v is the number of coefficients br which are neither 0 nor 1, then the total number of multiplications is (μ + v).
Therefore the total error is

$$\varepsilon(n) = \frac{(\mu + v)q^2}{12}. \quad \text{.....4.33}$$

For the most general case expressed in (4.31) we may write

$$\mu = M + 1$$
$$v = N$$

.....4.34 so that

$$\varepsilon(n) = \frac{(M + N + 1)q^2}{12}. \quad \text{.....4.35}$$

Therefore the quantized (rounded) filter output sequence is given by

$$g_q(n) = \sum_{r=0}^{M} a_r f(n-r) - \sum_{r=1}^{N} b_r g_q(n-r) + \varepsilon(n). \quad \text{.....4.36}$$

Now, the error in the filter output can be defined as

$$e(n) = g_q(n) - g(n). \qquad \text{.....4.37}$$

Substituting from (4.32) and (4.36) in (4.37) we have

$$e(n) = \varepsilon(n) - \sum_{r=1}^{N} b_r e(n-r) \qquad \text{.....4.38}$$

which describes a linear discrete system with input ε(n) and output e(n). Hence, taking the z-transform of (4.36) we can form the transfer function

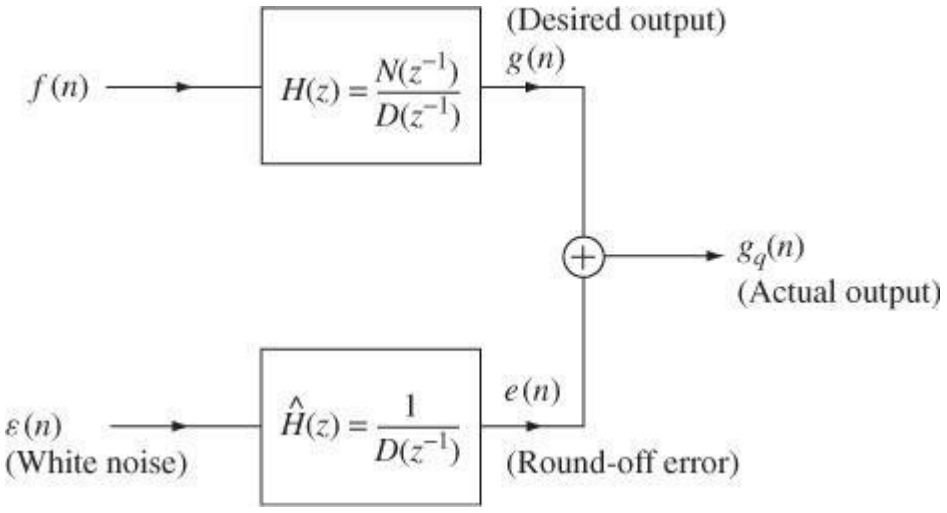
$$\hat{H}(z) = \frac{E(z)}{E(z)} = \frac{Z\{e(n)\}}{Z\{\varepsilon(n)\}} = \frac{1}{1 + \sum_{r=1}^{N} b_r z^{-r}}$$

.....4.39a or

$$\hat{H}(z) = \frac{1}{D(z^{-1})} \qquad \text{.....4.39b}$$

where D(z−1) is the denominator of the filter transfer function H(z). It follows that D(z−1) is the part of the transfer function which contributes to the error noise in the filter output. We can, therefore, construct a model of the filter which takes account of product round-off error accumulation as shown in Figure 4.6.

Figure 4.6 A model of an IIR filter, in direct form, taking round-off error accumulation into account



The average output noise power is

$$\begin{aligned}\sigma^2 &= \frac{(N+M+1)q^2}{12}\sum_{n=0}^{\infty}(\hat{h}(n))^2 \\ &= \frac{(N+M+1)q^2}{12}\frac{1}{\mathrm{j}2\pi}\oint_c \frac{1}{D(z)D(z^{-1})}\,\frac{\mathrm{d}z}{z}.\end{aligned} \qquad \ldots\ldots\ldots\ldots\ldots\ldots 4.40$$

Example 4.2

Calculate the output noise due to round-off accumulation, for the filter described by the transfer function

$$H(z) = \frac{1}{(1-0.4z^{-1})(1-0.5z^{-1})}.$$

Solution

$$\begin{aligned}D(z^{-1}) &= (1-0.4z^{-1})(1-0.5z^{-1}) \\ D(z) &= (1-0.4z)(1-0.5z).\end{aligned}$$

Here, M = 0 and N = 2. Therefore, (4.40) gives

$$\sigma^2 = \frac{q^2}{8\mathrm{j}\pi}\oint_c \frac{z\mathrm{d}z}{(0.4-z)(0.5-z)(1-0.4)(1-0.5z)}.$$

The integral is evaluated by summing the residues due to the poles inside the contour of integration, which in this case is the unit circle. This gives

$$\sigma^2 = \frac{q^2}{8\mathrm{j}\pi} \times \mathrm{j}2\pi \sum \text{residues}.$$

The poles of the integrand inside the unit circle are those at z = 0.4 and z = 0.5. The corresponding residues are

$$\begin{aligned}k_1 &= \frac{0.4}{(0.4-0.5)(1-0.4^2)(1-0.5\times 0.4)} = -5.9524 \\ k_2 &= \frac{0.5}{(0.5-0.4)(1-0.4\times 0.5)(1-0.5^2)} = 8.333.\end{aligned}$$

Therefore, the round-off accumulation noise is

$$\begin{aligned}\sigma^2 &= \frac{q^2}{4}(8.333-5.924) \\ &= 0.595q^2.\end{aligned}$$

Now, consider the direct non-recursive realization of an FIR filter described by

$$H(z) = \sum_{r=0}^{M} a_r z^{-r}.$$

…….4.41

In this case, the round-off accumulation noise can be calculated using 4.40 with D(z−1) = 1 and N = 0. Thus, the
average output noise power is

$$\sigma^2 = \frac{(M+1)q^2}{12} \frac{1}{\mathrm{j}2\pi} \oint_c \frac{\mathrm{d}z}{z}$$

That is

$$\sigma^2 = \frac{(M+1)}{12} q^2.$$

…….4.42

4.3.1.2 Cascade Realization

As explained in Chapter 4, this method proceeds by expressing the transfer function in the form

$$H(z) = \prod_{k=1}^{m} H_k(z)$$

…….4.43 where Hk(z) is a second-order factor of the form

$$\begin{aligned} H_k(z) &= \frac{\alpha_{0k} + \alpha_{1k} z^{-1} + \alpha_{2k} z^{-2}}{1 + \beta_{1k} z^{-1} + \beta_{2k} z^{-2}} \\ &= \frac{A(z^{-1})}{B(z^{-1})} \end{aligned}$$

…….4.44

with the possibility of one first-order factor for an odd-degree function; this is of the same form in (4.44) with α2k = β2k = 0.

The realization is then the cascade connection of the sections described by the second-order terms (and a possible first-order one) which realize each typical transfer function given by (4.44). Without taking round-off errors into account, the realization takes the form shown in Figure 4.5. To include round-off errors in the analysis, the model of Figure 4.4 is used for each transfer function described by (4.44) and the results are connected in cascade. This gives the model of Figure 4.7. The quantization error inputs εk(n) produce noise power components σ_k^2 at the outputs resulting in a total noise power given by

$$\sigma_{\mathrm{T}}^2 = \sum_{k=1}^{m} \sigma_k^2.$$

…… 4.45

Examination of the model in Figure 4.7 shows that each noise input εk(n) produces an output ek(n) such that an
error transfer function may be formed as

$$\hat{H}_k(z) = \frac{Z[e_k(n)]}{Z[\varepsilon_k(n)]} = \frac{1}{B_k(z^{-1})} \prod_{r=k+1}^{m} \hat{H}_r \quad k = 1, 2, \ldots (m-1)$$

……4.46a and

$$\hat{H}_m(z) = \frac{1}{B_m(z^{-1})}.$$

……4.46b

Hence, using Parseval's theorem and a relation similar to (4.40) the kth quantization error component produces a noise power given by

$$\sigma_k^2 = \frac{\mu_k q^2}{12} \frac{1}{\mathrm{j}2\pi} \oint \frac{1}{B_k B_{k^*}} \prod_{r=k+1}^{m} H_r H_{r^*} \frac{\mathrm{d}z}{z} \text{ for } \quad k = 1, 2, \ldots (m-1)$$

….. 4.47a

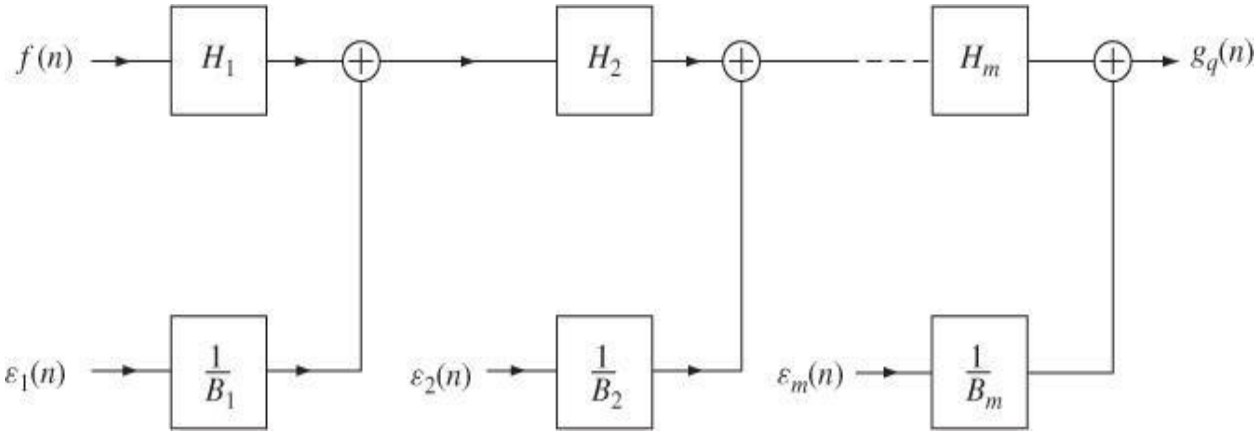
and

$$\sigma_m^2 = \frac{\mu_m q^2}{12} \frac{1}{\mathrm{j}2\pi} \oint \frac{1}{B_m B_{m^*}} \frac{\mathrm{d}z}{z}$$

…….4.47b

where the lower asterisk denotes replacement of z by z−1 and μk is the number of multiplications in the kth section. μk = 5 for a general second-order section and μk = 3 for a general first-order one. Finally the total output noise power is obtained from (4.46) and (4.47).

Figure 4.7 A model for an IIR filter, in cascade form, taking round-off error accumulation into account



Example 4.3

Find the output noise power due to round-off accumulation in the cascade realization of the transfer function of Example 4.2. [1]

Solution. Let us realize the transfer function as a cascade of two first-order sections. This is possible because both poles are real. Thus

$$H(z) = H_1(z)H_2(z)$$

where

$$H_1(z)\frac{z^{-1}}{(1-0.4z^{-1})} = \frac{1}{z-0.4}$$

$$H_2(z) = \frac{z^{-1}}{(1-0.5z^{-1})} = \frac{1}{z-0.5}.$$

Using (4.47) we obtain with μ1 = 1

$$\sigma_1^2 = \frac{q^2}{12}\frac{1}{\mathrm{j}2\pi}\oint \frac{5z\,\mathrm{d}z}{(z-0.4)(z-0.5)(2.5-z)(2-z)}.$$

The two poles inside the unit circle are at z = 0.4 and z = 0.5. The corresponding residues are −5.9524 and 4.333. Thus

$$\begin{aligned}\sigma_1^2 &= \frac{q^2}{12}\frac{1}{\mathrm{j}2\pi}\mathrm{j}2\pi(-5.9524+8.333)\\ &= 0.1984q^2.\end{aligned}$$

Similarly for $\hat{H}_2$ with μ2 = 1

$$= \sigma_2^2 = 0.111q^2.$$

Therefore the total output noise power is

$$\begin{aligned}\sigma_{\mathrm{T}}^2 &= \sigma_1^2+\sigma_2^2\\ &= 0.3095q^2.\end{aligned}$$

Comparison with the direct realization in Example 4.2, it is seen that for this transfer function, the cascade form produces lower round-off noise, for the same quantizing step.

4.3.1.3 Parallel Realization

In this method we express the transfer function as

$$H(z) = \sum_{k=1}^{m} H_k(z)$$

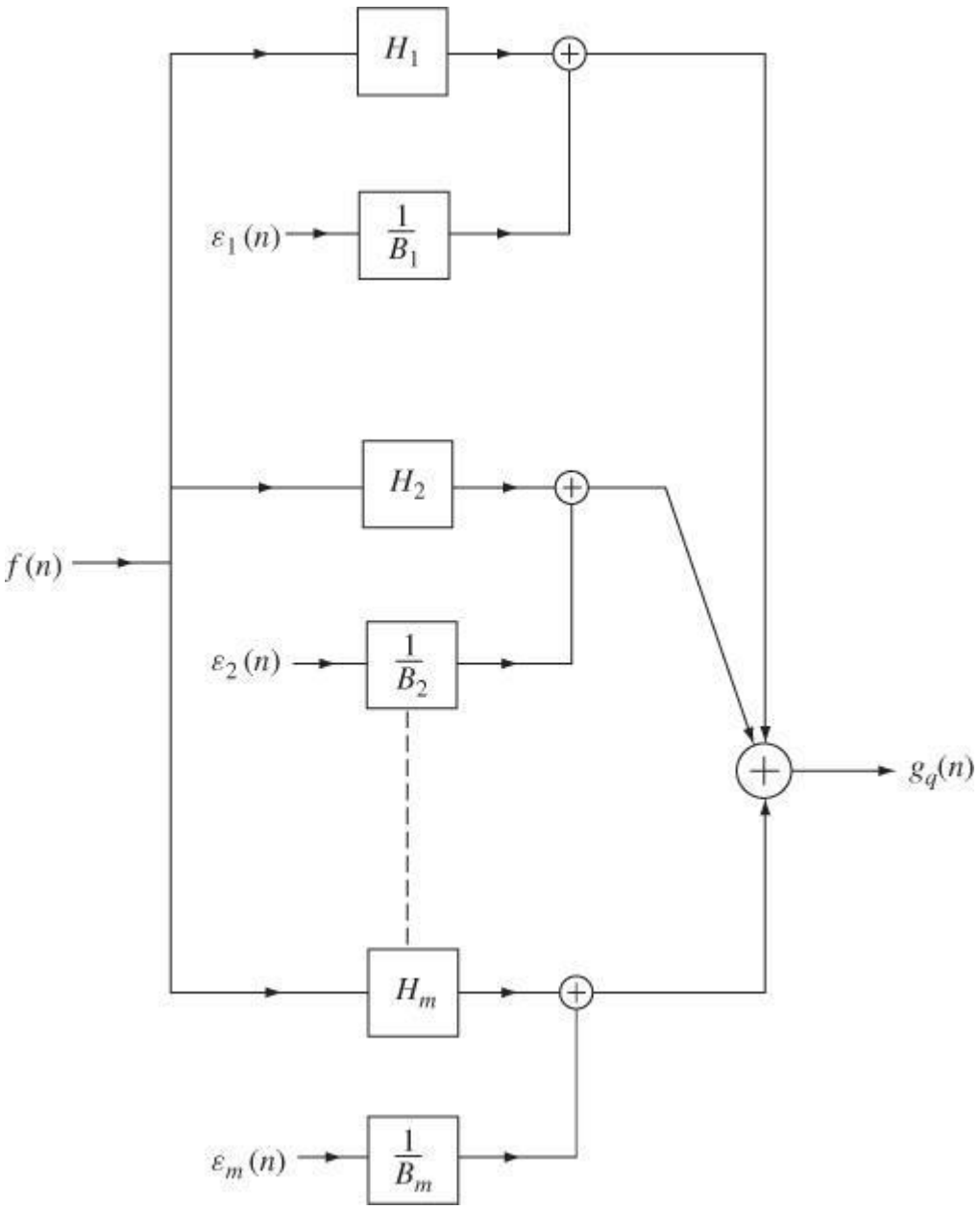
…….4.48 with

$$H_k = \frac{\alpha_{0k} + \alpha_{1k} z^{-1}}{1 + \beta_{1k} z^{-1} + \beta_{2k} z^{-2}} = \frac{A_k(z)}{B_k(z)}.$$

…….4.49

As before, we can develop the model of Figure 4.6 which gives the parallel form of realization taking the round-off noise into account.

Figure 4.6 A model for an IIR filter, in parallel form, taking round-off error accumulation into account[1]



Each noise error input εk(n) produces a quantization noise power component σ_k^2 at the output. The total output round-off noise power is

$$\sigma_{\mathrm{T}}^2 = \sum_{k=1}^{m} \sigma_k^2$$

…….4.50 where

$$\sigma_k^2 = \frac{\mu_k q^2}{12} \frac{1}{\mathrm{j}2\pi} \oint_c \frac{1}{B_k B_{k^*}} \frac{\mathrm{d}z}{z} \quad k = 1, 2, \ldots m$$

…….4.51 and μk is the number of multiplications in the kth section.

Example 4.4

Find the output round-off noise power in the parallel realization of the same transfer function of Examples 4.2 and 4.3.

Solution

$$H(z) = \frac{1}{(z - 0.4)(z - 0.5)}$$

which can be realized as two first-order sections in the parallel form, because both poles are real. Thus

$$H(z) = H_1(z) + H_2(z)$$

where

$$\begin{aligned} H_1(z) &= \frac{10}{z - 0.5} \\ H_2(z) &= \frac{-10}{z - 0.4}. \end{aligned}$$

Thus

$$\begin{aligned} \sigma_1^2 &= \frac{q^2}{12} \frac{1}{\mathrm{j}2\pi} \oint_c \frac{1}{(z - 0.4)(z^{-1} - 0.4)} \frac{\mathrm{d}z}{z} \\ &= 0.0992 q^2 \end{aligned}$$

and

$$\begin{aligned} \sigma_2^2 &= \frac{q^2}{12} \frac{1}{\mathrm{j}2\pi} \oint_c \frac{1}{(z - 0.5)(z^{-1} - 0.5)} \frac{\mathrm{d}z}{z} \\ &= 0.111 q^2 \end{aligned}$$

so that

$$\sigma_T^2 = \sigma_1^2 + \sigma_2^2 = 0.2103q^2$$

which is even smaller than in the cascade form, for this transfer function.

4.4.2 Round-off Accumulation with Coefficient Quantization

We now repeat the same analysis of the previous section with the additional assumption that the filter coefficients are also rounded . Here also, it is assumed that fixed-point arithmetic is used.

4.4.2.1 Direct Realization

With the notation of (4.21)–(4.26) for the rounded coefficients, define[1]

$$\Delta(n) = -\sum_{k=1}^{M} b_k \Delta(n-k) + u(n)$$

……4.52 where

$$u(n) = \sum_{k=0}^{N} \alpha_k f(n-k) - \sum_{k=1}^{M} \beta_k g(n-k) + \varepsilon(n)$$ ……4.53

$$A(z) = \sum_{k=0}^{M} \alpha_k z^{-k}$$ …….4.54

$$B(z) = \sum_{k=1}^{N} \beta_k z^{-k}$$ …….. 4.55

$$C(z) = A(z) - H(z)B(z)$$ …….4.56

$$D(z) = 1 + \sum_{k=1}^{N} b_k z^{-k}.$$ …….…….4.57

Now, with the assumption that the input sequence {f(n)} is zero-mean and wide-sense stationary, it has an autocorrelation sequence {Rff(n)} and power spectral density ϕff(z). It follows that the output sequence {g(n)} is also zero-mean and wide-sense stationary with power spectral density ϕgg(z) given by (4.147) as

$$\phi_{gg}(z) = H(z)H_*(z)\phi_{ff}(z).$$ …….4.58

It can be shown that {u(n)}, as defined by (4.53), is also zero-mean and wide-sense stationary with an autocorrelation function given by

$$\phi_{uu}(z) = C(z)C_*(z)\phi_{ff}(z) + \frac{(M+N+1)q^2}{12DD_*}.$$ …….4.59

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

The output quantization error Δ(n) which includes both coefficient quantization and product round-off accumulation, is also zero-mean and wide-sense stationary with autocorrelation

$$\phi_{\Delta\Delta}(z) = \frac{1}{D(z)D_*(z)}\phi_{uu}(z)$$

…….4.60 which, upon use of (4.59), becomes

$$\phi_{\Delta\Delta}(z) = \frac{C(z)C_*(z)}{D(z)D_*(z)}\phi_{ff}(z) + \frac{(M+N+1)}{12}q^2. \qquad \text{…….4.61}$$

Therefore, the mean-square error, or output noise power, is given by

$$\sigma_{\mathrm{T}}^2 = \frac{1}{\mathrm{j}2\pi}\oint \phi_{\Delta\Delta}(z)\frac{\mathrm{d}z}{z}. \qquad \text{…….4.62}$$

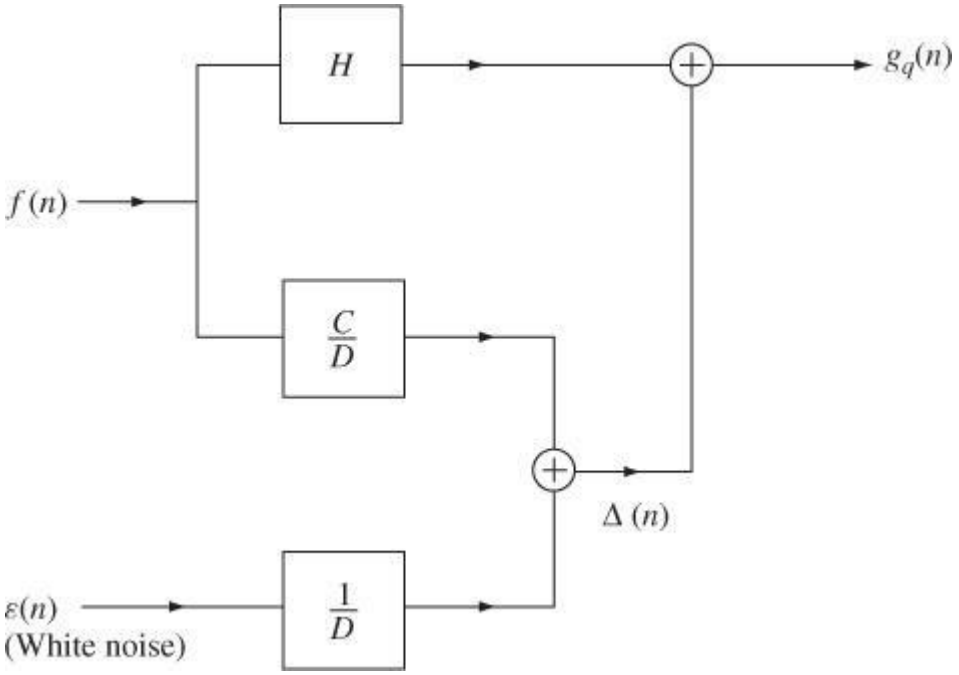
Clearly, if there is no coefficient rounding error, the first term in

(4.61) is absent and expressions (4.61) to (4.62) reduce to (4.40) obtained before. However, if there is no round-off error, then the second term in (4.61) is absent. It follows that the total noise at the output, due to internal quantization, is the sum of two components, one is due to round-off accumulation and the other due to coefficient rounding to l bits. The component due to round-off accumulation is uncorrelated with either the input sequence
{f(n)} or the theoretical output sequence {g(n)}. A model for the direct realization of the transfer function incorporating these errors can be easily obtained from (4.61) and (4.62). This is shown in Figure 4.9, of which Figure
4.6 is a special case.
Figure 4.9 A model for an IIR filter in direct form, taking into account both round-off accumulation and coefficient quantization[1]

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th



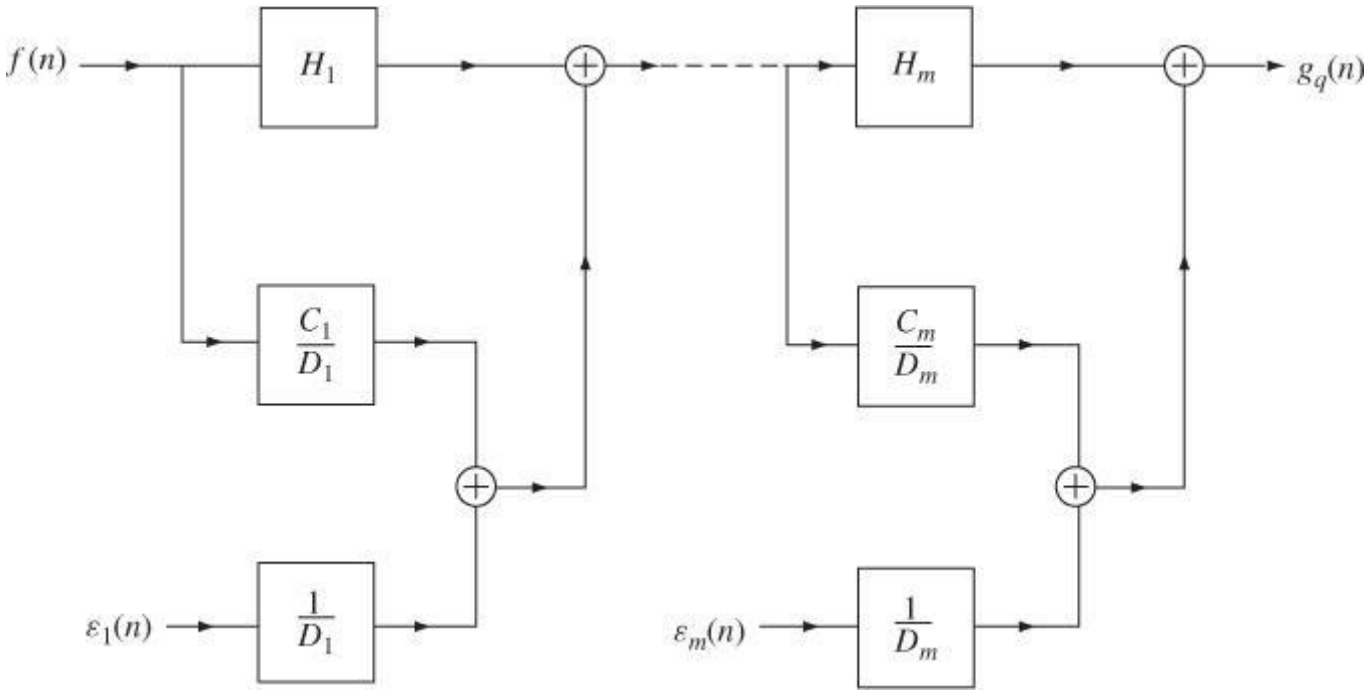
4.4.2.2 Cascade Realization

With H(z) factored as in (4.43), the model in Figure 4.9 can be applied to each factor Hk, with M = N = 2 (or 1). The notation in (4.21) to (4.36) is used for the quantized coefficients for each second- or first-order section. Thus, with

$$\begin{aligned} \hat{A}_k(z) &= \alpha_{0k} + \alpha_{1k}z^{-1} + \alpha_{2k}z^{-2} \\ \hat{B}_k(z) &= \beta_{1k}z^{-1} + \beta_2 k z^{-2} \\ C_k(z) &= \hat{A}_k(z) - H_k(z)\hat{B}_k(z) \end{aligned} \qquad \text{.....4.63}$$

we may construct the model in Figure 4.10 of which Figure 4.7 is a special case when coefficient quantization is neglected.

Figure 4.10 A model for an IIR filter in cascade form taking into account both round-off accumulation and coefficient quantization



Thus, for the cascade realization, the power spectral density of the output error is given by

$$\phi_{\Delta\Delta}(z) = \phi_{ff}\sum_{k=1}^{m}\frac{C_kC_{k^*}}{D_kD_{k^*}}\prod_{\substack{r=1\\r\neq k}}^{m}H_rH_{r^*} + \frac{q^2}{12}\left(\frac{\mu_m}{D_mD_{m^*}}+\sum_{k=1}^{m-1}\frac{\mu_k}{D_kD_{k^*}}\prod_{\substack{r=1\\r\neq k}}^{m}H_rH_{r^*}\right) \quad \text{.....4.64}$$

where μk is the number of multiplications in the kth section.

Finally, the total output noise power is obtained from (4.62) and (4.64). We also note that for zero coefficient rounding, (4.64) when used in (4.62) reduces to (4.47).

4.4.2.3 Parallel Realization

With H(z) expressed as in (4.46), the model in Figure 4.9 can be applied to each second-order (or first-order) term Hk(z). The notation in (4.21) to (4.26) is used for the quantized coefficients of each term. Thus, with

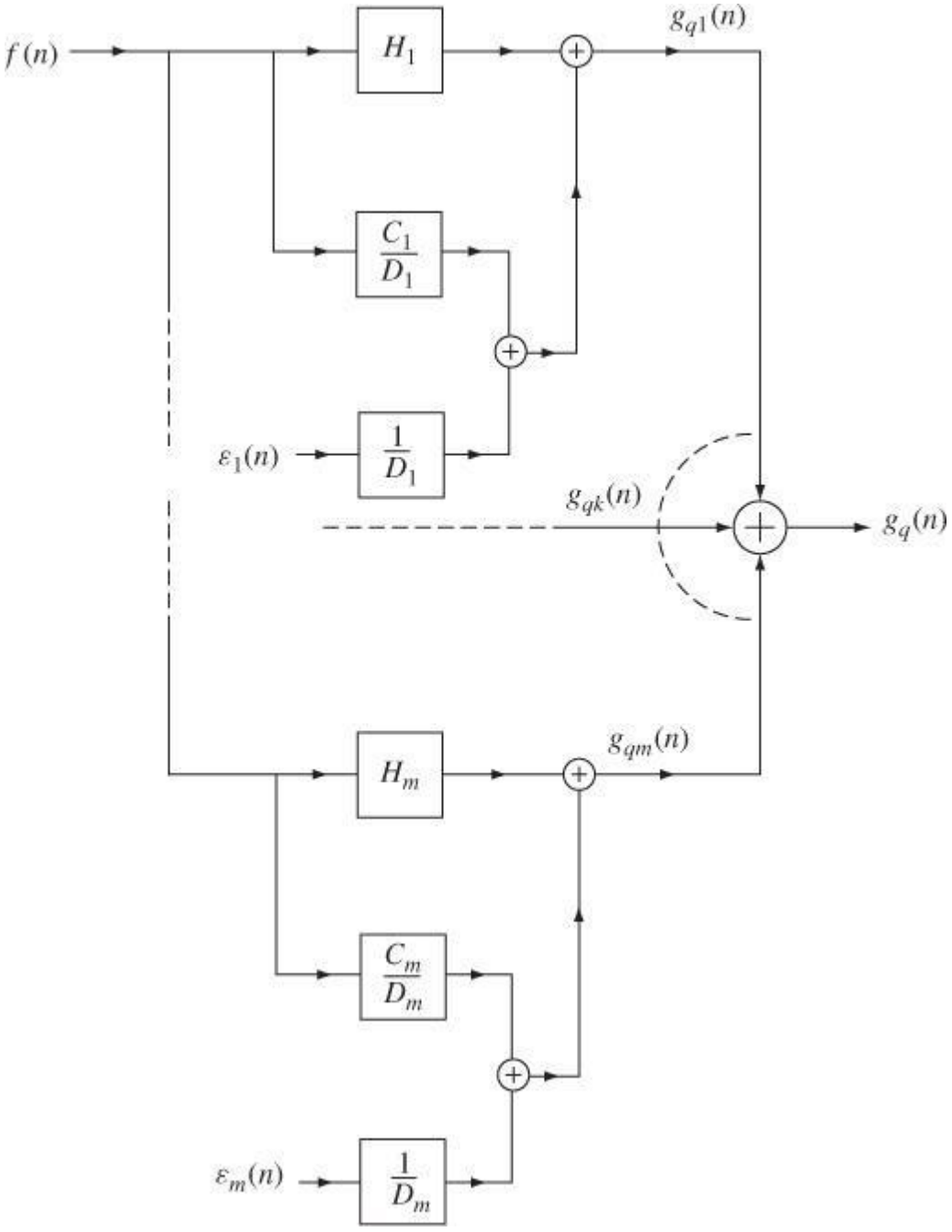
$$\begin{aligned}\hat{A}_k(z) &= \alpha_{0k}+\alpha_{1k}z^{-1}\\ \hat{B}_k(z) &= \beta_{0k}z^{-1}+\beta_{1k}z^{-2}.\\ C_k(z) &= \hat{A}_k(z)-H_k(z)\hat{B}_k(z)\end{aligned} \quad \text{.....4.65}$$

we may construct the model in Figure 4.11 of which Figure 4.4 is a special case when coefficient quantization is absent. Thus, for the parallel form of realization, the power spectral density of the output error is given by

$$\phi_{\Delta\Delta}(z) = \phi_{ff}(z)\left(\sum_{k=1}^{m}\frac{C_k}{D_k}\right)\left(\sum_{k=1}^{m}\frac{C_{k^*}}{D_{k^*}}\right) + \frac{q^2}{12}\sum_{k=1}^{m}\frac{\mu_k}{D_k D_{k^*}} \qquad \text{…..4.66}$$

where μk is the number of multiplications in the kth section. The total output noise power is obtained from (4.66) and (4.62). For no coefficient quantization, the result reduces to (4.51).

Figure 4.11 A model for an IIR filter in parallel form taking into account both round-off accumulation and coefficient quantization[1]



4.5 Auto-oscillations: Overflow and Limit Cycles

4.5.1 Overflow Oscillations

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

Consider the second-order section as the basic building block in the realization of a digital transfer function [1]. For the section shown in Figure 4.12 described by the difference equation

$$g(n) = f(n) - \beta_1 g(n-1) - \beta_2 g(n-2)$$

……4.67 the corresponding transfer function is an all-pole one

$$\begin{aligned} H(z) &= \frac{Z\{g(n)\}}{Z\{f(n)\}} \\ &= \frac{1}{1+\beta_1 z^{-1} + \beta_2 z^{-2}}. \end{aligned}$$

……4.68

The condition of stability is obtained by finding the poles as

$$P_{1,2} = \frac{-\beta_1 \pm (\beta_1^2 - 4\beta_2)^{1/2}}{2}$$

……4.69

so that for stability p1, 2 must be inside the unit circle in the z-plane. This condition can be expressed in a (β1, β2)- plane as shown in Figure 4.13. For complex poles, the domain of stability is bounded by the parabola

$$\beta_2 = \frac{\beta_1^2}{4}$$

……4.70 and for stability

$$0 \le \beta_2 < 1.$$

……4.71

For real poles (4.69) gives

$$-\frac{\beta_1}{2} + \frac{1}{2}(\beta_1^2 - 4\beta_2)^{1/2} < 1$$
$$-1 < \frac{\beta_1}{2} - \frac{1}{2}(\beta_1^2 - 4\beta_2)^{1/2}$$

……4.72

That is

$$\beta_2 > -\beta_1 - 1, \quad \beta_2 > -1 + \beta_1$$

……4.73 or

$$|\beta_1| < 1 + \beta_2.$$

……..4.74

In this case the domain of stability is the triangle defined by the three straight lines

$$\beta_2 = 1, \quad \beta_2 = -\beta_1 - 1, \quad \beta_2 = \beta_1 - 1$$

……4.75 as shown in Figure 4.13.

Figure 4.12 A second-order filter section described by (4.67) and (4.66)

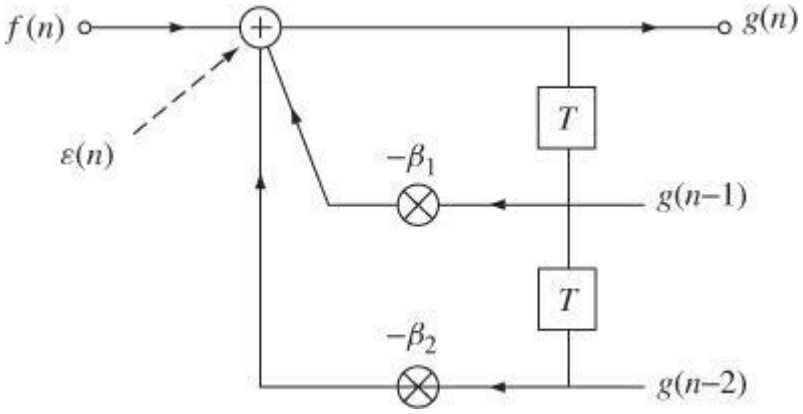
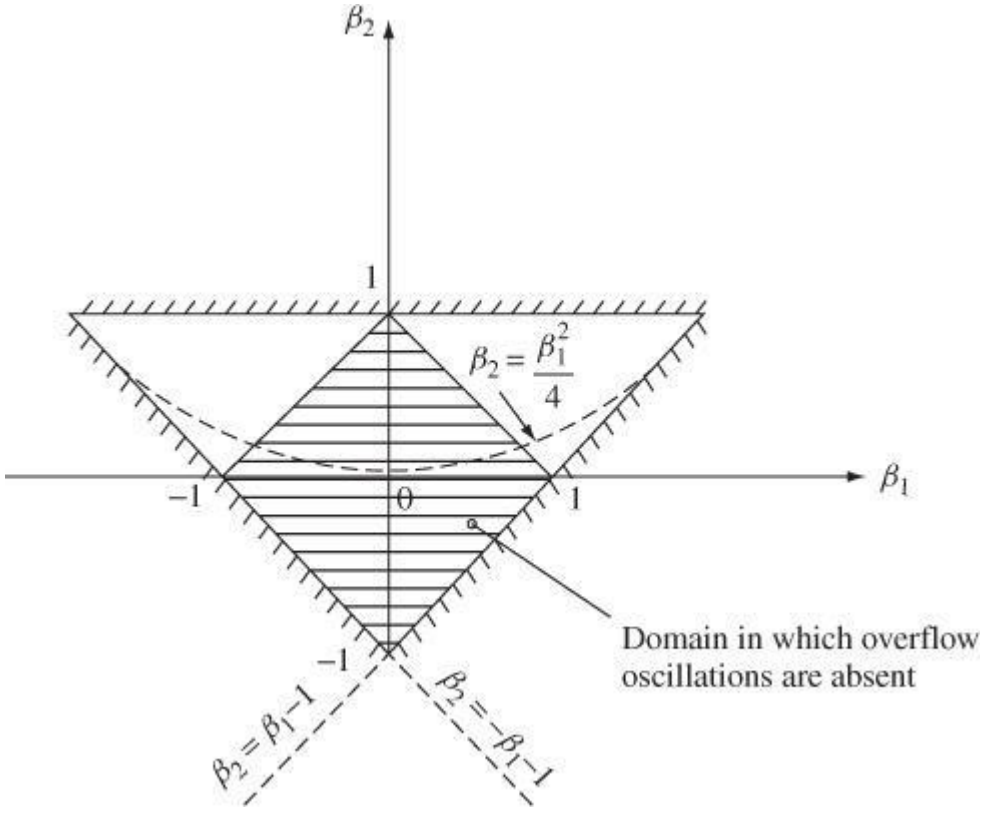


Figure 4.13 Pertinent to the discussion of auto-oscillations



Now, even for zero input, and even if the stability conditions are satisfied, self-sustained oscillations may occur. The zero-input difference equation is obtained by setting f(n) = 0 in (4.67); thus

$$g(n) + \beta_1 g(n-1) + \beta_2 g(n-2) = 0$$

or

$$g(n) = -[\beta_1 g(n-1) + \beta_2 g(n-2)]. \qquad \text{……4.76}$$

When using fixed-point arithmetic in the adder of Figure 4.12 overflow may occur. For example, the transfer characteristic of the two's complement adder is shown in Figure 4.14. Clearly overflow will occur if the adder receives at its input numbers whose sum is outside the range (−1, 1). The condition for no overflow is obtained from (4.76) as

$$|g(n)| < 1$$

……4.77 so that

$$|\beta_1 g(n-1) + \beta_2 g(n-2)| < 1 \quad \text{……4.78}$$

and since g(n − 1) and g(n − 2) are constrained to be less than unity, then a necessary and sufficient condition for the
absence of overflow is

$$|\beta_1| + |\beta_2| < 1. \quad \text{……4.79}$$

This defines a square in the (β1, β2)-plane within the triangle of stability as shown in Figure 4.13. It has been shown that if (4.79) is not satisfied the adder will operate in a non-linear fashion, producing a filter output even when the input is zero. This output can be either a constant or a periodic signal which is generally called an overflow oscillation. The solution to this problem is to use saturating adders. The transfer characteristic of such an adder is shown in Figure 4.15 and does not allow the result to exceed the specified dynamic range.

Figure 4.14 Transfer characteristic of a two's complement adder

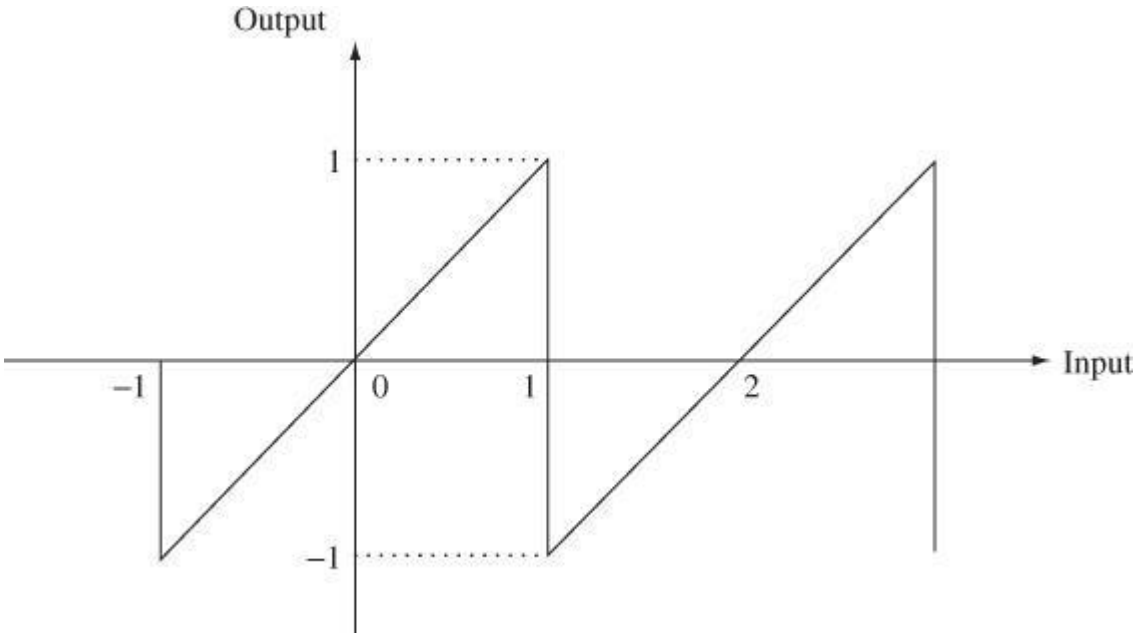
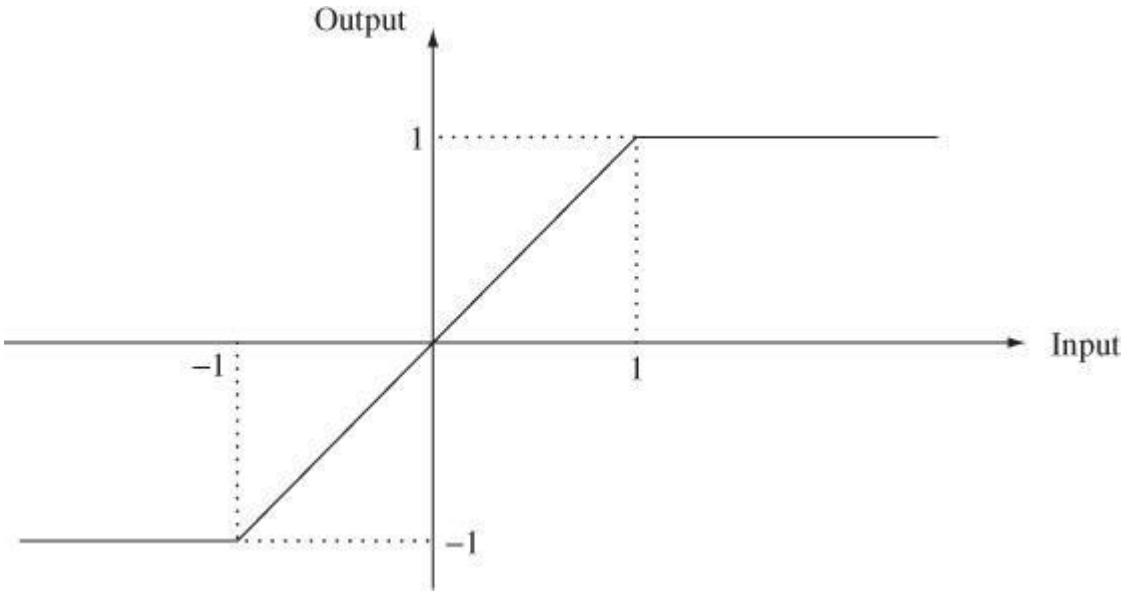


Figure 4.15 Transfer characteristic of a two's complement saturating adder



4.5.2 Limit Cycles and the Dead-band Effect

If the input to a digital filter is zero or a constant, the arithmetic round-off errors cannot be treated as uncorrelated random processes in the manner given in Section 4.4. Instead, the round-off noise is dependent on the input signal, and even when the input is switched-off there will be an output determined by this round-off noise. This results in self-sustained oscillations known as limit cycles. These are now illustrated by a first-order filter described by the first- order difference equation

$$g(n) = \alpha g(n-1) + f(n)$$

…… 4.80 so that its transfer function is

$$H(z) = \frac{1}{1 - \alpha z^{-1}}$$ ……4.81

and stability requires

$$|\alpha| < 1.$$ ……4.82

Taking α = 0.94, the initial condition g(−1) = 11 and assuming that the input is switched-off, that is f(n) = 0, we have

$$g(n) = 0.94g(n-1).$$ ……4.83

The output is calculated assuming infinite precision arithmetic, then rounded to the nearest integer, giving the values in Table 4.1. These show that although the exact value of g(n) decays exponentially, the rounded value reaches a steady state value of 6. This is what is meant by the limit cycle response to zero input. If we repeat the calculations in Table 4.1 with α in 4.60 being −0.94 we obtain Table 4.2 revealing that the rounded value of g(n) oscillates between 6 and −6. The range [−6, 6] is called the dead-band and the phenomenon is called the dead-band effects.

Table 4.1 Output of (4.83) calculated assuming infinite precision arithmetic and rounded to the nearest integer

n	Exact *g*(*n*)	Rounded *g*(*n*)
0	10.34	10
1	9.719 6	9
2	9.136 424	6
3	6.566 236 6	6
4	6.072 944 2	6
5	7.566 567 6	6

Table 4.2 Calculations in Table 4.1 with α in (6.60) being −0.94

n	Exact *g*(*n*)	Rounded *g*(*n*)
0	−10.34	−10
1	9.719 6	9
2	−9.136 424	−6
3	6.566 236 6	6
4	−6.072 944 2	−6
5	7.566 567 6	6

In general, for a first-order filter, the dead-band [−*D*, *D*] is obtained from

$$D = \text{integral part of} \left\{ \frac{0.5}{1 - |\alpha|} \right\} \quad \text{.....4.84}$$

where:

i. for α > 0 the limit cycles are constant.

ii. for α < 0 the signs alternate with a frequency of oscillation ω*N*/2.

The analysis of limit-cycles in a second-order section shows that it has two modes of auto- oscillations. The first is similar to the first-order case resulting in either a constant output or oscillations with ω*N*/2. In the second mode, the filter behaves as though it possessed a pair of conjugate poles on the unit circle.

It must be observed that limit-cycle oscillations occur due to quantization before storage in the registers. They generally have small amplitudes in well-designed systems with a sufficiently large number of bits and a sufficiently small quantizing step. Therefore they are generally smaller in amplitude than possible overflow oscillations in the absence of logic saturation devices. An upper bound on limit cycle amplitudes can be easily obtained by noting that the quantization error signal is bounded by

$$\varepsilon(n) \leq \frac{q}{2}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 4.85$$

Therefore, if the filter impulse response sequence is {*h*(*n*)}, then

$$|g(n)| \leq \frac{q}{2} \sum_n |h(n)| \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 4.86$$

which is, in fact, too pessimistic. A more realistic estimate of the limit cycle amplitudes is given by

$$A_{1c} = \frac{q}{2} \max |H(\exp(\mathrm{j}\omega T))| \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 4.87$$

where *H*(exp(jω*T*)) is the transfer function of the filter section. For example, a second- order section described by

$$H(z) = \frac{1}{1 + \beta_1 z^{-1} + \beta_2 z^{-2}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 4.88$$

has poles as

$$P_{1,2} = |r| \exp(\mathrm{j} \pm \theta) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 4.89$$

so that (4.86) gives

$$|g(n)| \leq \frac{q}{2} \frac{1}{(1-r)\sin\theta}$$

while (4.87) gives

$$A_{1c} = \frac{q}{2} \frac{1}{(1-r^2)\sin\theta}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 4.90$$

Multiple choice questions:

1) dimensional signal is a function of

a. Multiple independent variables

b. Single independent variable

c. Multiple dependent variables

d. Single dependent variable

ANSWER: (b) Single independent variable

2) Superposition of signals in a linear system refers to the

a. Output that is product of all the signals

b. Output that is sum of all the signals

c. Output that is of highest amplitude of all the signals

d. Output that is of largest spectrum of all the signals

3) The effects caused due to finite word lengths are

1) Coefficient quantization error

2) Adder overflow limit cycle

3) Round off noise

4) Limit cycles

a. 1, 2 and 3 are correct

b. 1 and 3 are correct

c. 1 and 4 are correct

d. All the four are correct

ANSWER: (d) All the four are correct

4) The error in the filter output that results from rounding or truncating calculations within the filter is called

a. Coefficient quantization error

b. Adder overflow limit cycle

c. Round off noise

d. Limit cycles

ANSWER: (c) Round off noise

5) How is/are the roundoff errors reduced in the digital FIR filter?

a. By representation of all

b. products with double-length registers

c. By rounding the results after acquiring the final sum

d. Both a and b

e. None of the above

ANSWER: (c) Both a and b

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

Module V
Application of DSP

Subband coding

The Subband coding is basically the decomposing the input signal into different frequency bands and after the input is decomposed to its constituents, we can use the coding technique best suited to each constituent to improve the compression performance Each component may have different perceptual characteristics.

Given signal x[n] to compress…

Idea: Split signal into Msignals x1[n], x2[n], …, xM[n] such that each signal can be more easily/effectively compressed.

Goal: signals x1[n], x2[n], …, xM[n] should be made such that –Each xi[n] is uncorrelated… •then using SQ on each is a viable (though still suboptimal) approach –Some xi[n] have smaller dynamic range •Then can use fewer bits for a given desired distortion –Should be a clear way to exploit psychological effects (for audio and video) or other effects that imply some xi[n] are

– Quantization errors that are objectionable in one component may be acceptable in a different component[1]

Idea: decompose a signal into components by applying frequency-selective filtering. Then select the best coding technique that best suits each component (subjectively and objectively).

- Example: slow- and fast-varying components.[1]

$$y[n] = (x[n] + x[n-1])/2 \quad z[n] = (x[n] - x[n-1])/2$$

$$\text{The signal can be recovered: } x[n] = y[n] + z[n]$$

$$\text{The filters are: } h[n] = (\delta[n] + \delta[n-1])/2 \quad g[n] = (\delta[n] - \delta[n-1])/2$$

If we use the same number of bits for each of y[n] and z[n], we are transmitting twice as many samples, doubling the bit rate.

- We can avoid this by sending every other value of y[n] and z[n] (e.g., even numbered elements)

Analysis

- Source output is passed through a bank of filters (analysis filters)
- Analysis filters cover the range of frequencies that make up source output
- Passband of the filters can be non-overlapping or overlapping
- Output of filters are then subsampled (also called decimation or

downsampling)

- Justification for subsampling: Nyquist rule (range of frequencies of output of the filter is less than input to the filter)

Quantization, coding and bit allocation

- Selection of compression scheme and allocation of bits between subbands is important and can have significant impact on the quality of the final reconstruction

Synthesis

- Encoded samples from each subband are decoded
- Decoded values are then upsampled by inserting an appropriate number of 0s between samples
- Upsampled signals are passed through a bank of reconstruction filters
- Output of reconstruction filters are added to give final output
- Three major components of subband system are: 1.

Analysis and synthesis filters

– Simple to implement, good separation between frequency bands 2.

Bit allocation (quantization)

– Can have s significant affect on the quality of the reconstruction 3.

Encoding scheme

– Based on the characteristics of each of the subbands, we can use a separate compression
scheme

– Human perception is frequency dependent. We can use this fact to design our compression
scheme so that the frequency bands that are most important to perception are reconstructed most accurately

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

Telecommunications: telephone line modems, FAX, cellular telephones, wireless networks, speaker phones, answering machines • Voice/Speech: speech digitization and compression, voice mail, speaker verification, and speech synthesis • Automotive: engine control, antilock brakes, active suspension, airbag control, and system diagnosis • Control Systems: head positioning servo systems in disk drives, laser printer control, robot control, engine and motor control, and numerical control of automatic machine tools Military: radar and sonar signal processing, navigation systems, missile guidance, HF radio frequency modems, secure spread spectrum radios, and secure voice • Medical: hearing aids, MRI imaging, ultrasound imaging, and patient monitoring • Instrumentation: spectrum analysis, transient analysis, signal generators • Image Processing: HDTV, image enhancement, image compression and transmission, 3-D rotation, and animation[1,3]

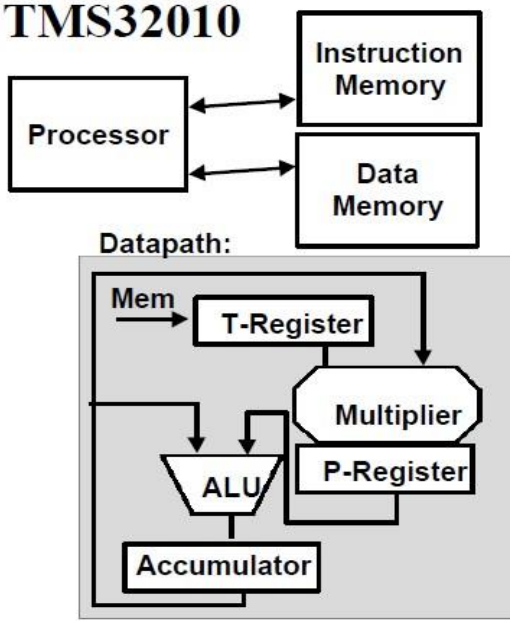
Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

First Commercial DSP (1982): Texas Instruments TMS32010

- 16-bit fixed-point arithmetic
- Introduced at 5Mhz (200ns) instruction cycle.
- "Harvard architecture"
  - separate instruction, data memories
- Accumulator
- Specialized instruction set
  - Load and Accumulate
- Two-cycle (400 ns) Multiply-Accumulate (MAC) time.



First Commercial DSP[1]

A conventional DSP: TI TMSC54xx

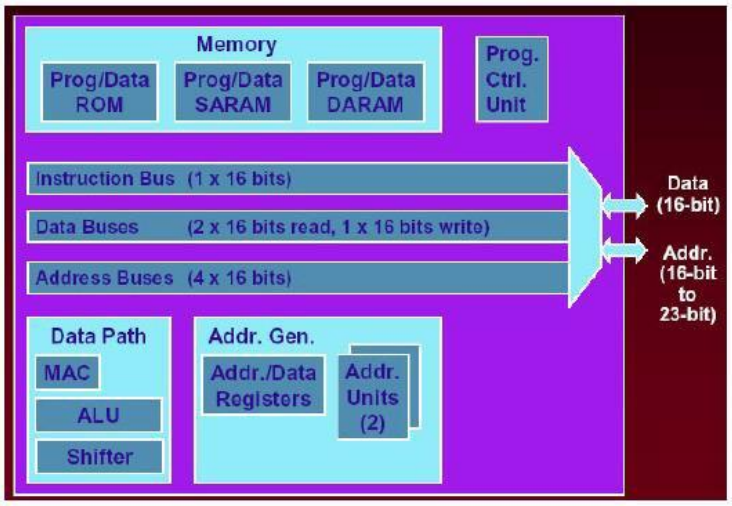
- 16-bit fixed point DSP
- Issues one 16-bit instruction/cycle
- Modified Harvard memory architecture
- Peripherals typical of conventional DSPs:
  - ---- 2-3 synch. Serial ports,parallel port
  - ---- Bit I/O,timer,DMA
- Low cost
- Low power

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

A Current Conventional DSP: TI TMSC54xx



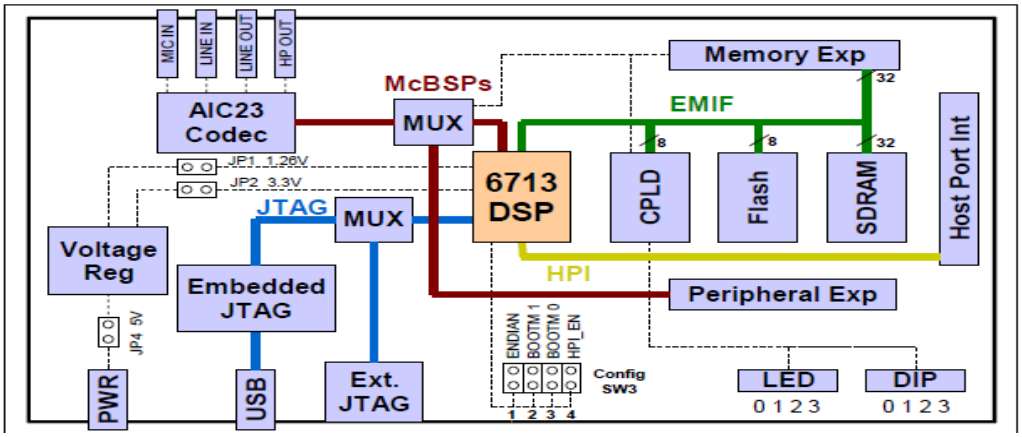


Fig Block Diagram C6713 DSK [3]

TMS320C6713 DSK Features

- A TMS320C6713 DSP operating at 225 MHz.
- An AIC23 stereo codec with Line In, Line Out, MIC, and headphone stereo jacks
- 16 Mbytes of synchronous DRAM
- 512 Kbytes of non-volatile Flash memory (256 Kbytes usable in default con_guration)
- 4 user accessible LEDs and DIP switches
- Software board con_guration through registers implemented in CPLD
- Con_gurable boot options
- Expansion connectors for daughter cards
- JTAG emulation through on-board JTAG emulator with USB host interface or external emulator

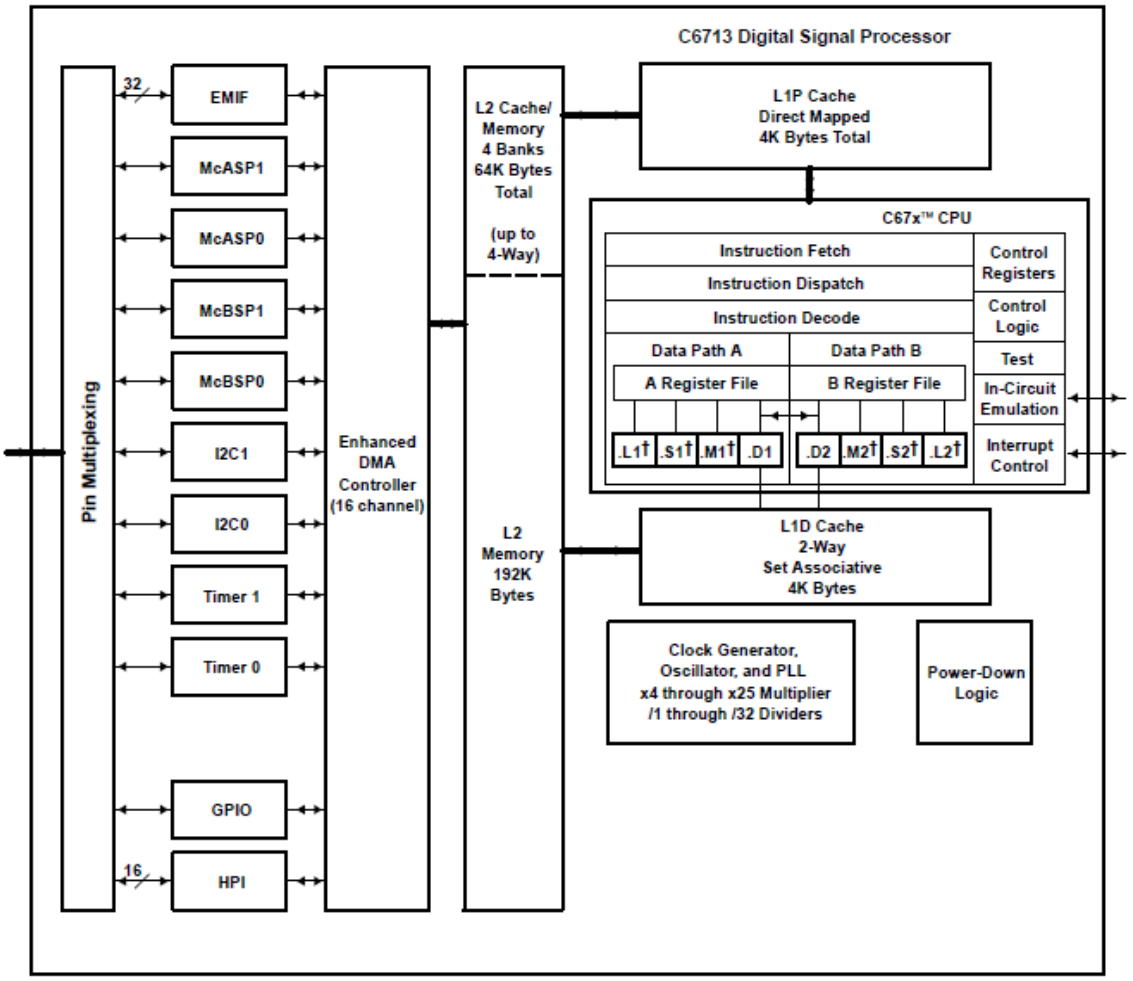
Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

TMS320C6713 Architecture

functional block and CPU (DSP core) diagram



† In addition to fixed-point instructions, these functional units execute floating-point instructions.

EMIF interfaces to:
–SDRAM
–SBSRAM
–SRAM,
–ROM/Flash, and
–I/O devices

McBSPs interface to:
–SPI Control Port
–High-Speed TDM Codecs
–AC97 Codecs
–Serial EEPROM

McASPs interface to:
–I2S Multichannel ADC, DAC, Codec, DIR
–DIT: Multiple Outputs

Fig: TMS320C6713 Architecture[3]

References:

[1] meseec.ce.rit.edu-Digital Signal Processor (DSP) Architecture

[2] www.ti.com

[3] ece.umd.edu

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

Multiple choice questions:

1) In DSP Processor, what kind of queuing is undertaken/executed through instruction register and instruction cache?

a. Implicate
b. Explicate
c. Both a and b
d. None of the above
ANSWER: (a) Implicate

2) In TMS 320 C6x processor architecture, which functional unit is adopted for transferring the data from register to and from control register?

a. L_2
b. M_2
c. S_2
d. D_2
ANSWER: (c) S_2

3) In TMS 320 C6x processor architecture, which operation/s is/are performed by 'M' functional unit?

a. Bit expansion
b. Bit interleaving & deinterleaving
c. Rotation & Variable shifting
d. All of the above
ANSWER: (d) All of the above

4) In C6X processor, which external device/s get/s acquire/s an interface support by EMIF peripheral?

a. Synchronous burst
b. Asynchronous devices
c. Externally shared memory devices
d. All of the above
ANSWER: (d) All of the above

5) Which peripheral on C 6 X processor allows buffering of serial samples in memory by port automatically & especially with an assistance of EDMA controller?

a. Boot Loader
b. HPI
c. EMIF
d. McBSP
ANSWER:(d) McBSP

Subject Name: Digital Signal Processing

Subject Code: EC 503 Contacts: 3L Credits: 3 Lectures: 35 hrs.

Semester: 5th

CO-PO Mapping:

COs	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
CO1	3	3	1	1	1	2	1	2	2	2	2	3
CO2	3	2	2	1	2	3	1	2	2	1	2	3
CO3	3	3	1	3	2	3	1	2	2	1	2	3
CO4	3	2	1	3	1	3	1	3	1	1	1	3
CO5	3	2	1	1	1	1	1	1	1	2	2	1
CO6	3	3	3	1	3	3	2	3	3	2	3	2