Structures For Discrete-Time Systems

- Realization (implementation) of digital filters
- Structures of IIR and FIR filters, their advantages and disadvantages – efficiency and error

Example: Given $H(z) = \dfrac{b_0 + b_1 z^{-1}}{1 - az^{-1}}$ $\quad |z| > |a|$

$\Rightarrow\ h[n] = b_0 a^n u[n] + b_1 a^{n-1} u[n-1]$ IIR

It is not possible to implement the system by discrete convolution!

$\Rightarrow\ y[n] = ay[n-1] + b_0 x[n] + b_1 x[n-1]$

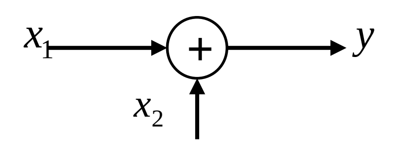
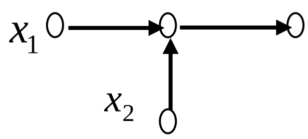
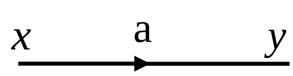
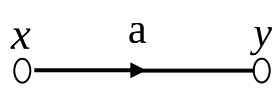
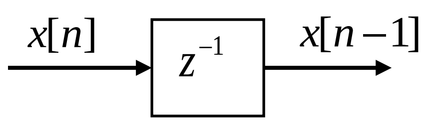
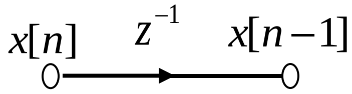
Actually, an unlimited variety of computational structures result in the same relation between y[n] and x[n]!

When the numerical precision is limited
$\Rightarrow$ different structures may have vastly different behavior.

(i) Finite-precision representation of the system coefficients
(ii) Truncation or rounding of intermediate computations.

✧ Block Diagram and Signal Flow Graph

- Three elements in LTI discrete-time systems:

	Block diagram	Signal flow graph
Adder		
Scalar (Multiplication by a constant)		
Unit delay		

- *Nodes* and *branches* are keys in a signal flow graph

Source node: No entering branches

Sink node: Only entering branches

✧ Basic Structures for IIR Systems

- **Direct Forms**

(1) Direct Form I

$$H(z)=\frac{\sum_{k=0}^{M} b_k z^{-k}}{1-\sum_{k=1}^{N} a_k z^{-k}} \Leftrightarrow y[n]-\sum_{k=1}^{N} a_k y[n-k]=\sum_{k=0}^{M} b_k x[n-k]$$

$$H(z)=H_2(z)H_1(z)=\frac{Y(z)}{X(z)}=H_2(z)V(z)\frac{H_1(z)}{V(z)}$$

$$Y(z)=H_2(z)V(z)\leftrightarrow y[n]-\sum a_k y(n-k)=v[n]$$

$$X(z)=V(z)/H_1(z)$$

$$V(z)=H_1(z)X(z)\leftrightarrow v[n]=\sum b_k x[n-k]$$

$$V(z)=\sum b_k X(z)z^{-k}$$

$$Y(z)-\sum a_k Y(z)z^{-k}=V(z)$$

$$Y(z)=V(z)+\sum a_k Y(z)z^{-k}$$

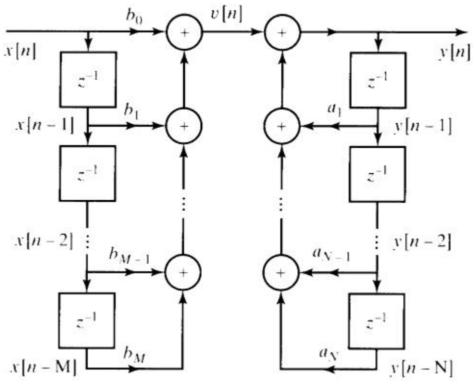


Figure 6.3 Block diagram representation for a general Nth-order difference equation.

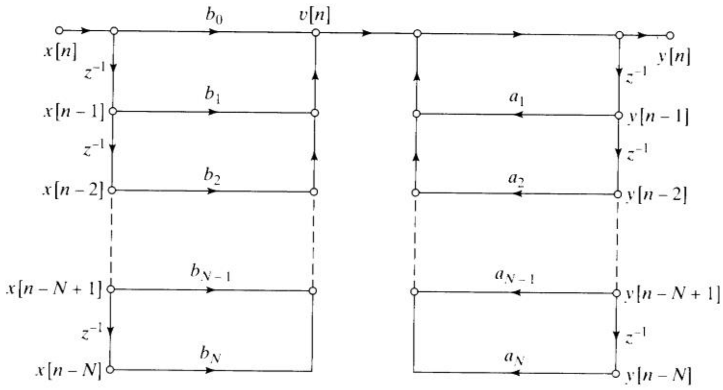


Figure 6.14 Signal flow graph of direct form I structure for an Nth-order system.

(2) Direct Form II (Canonic form)

-- Interchange 1st and 2nd "segments" and merge the delay lines (z^{-1})

-- Number of delay = max(N,M) ← "Canonic"

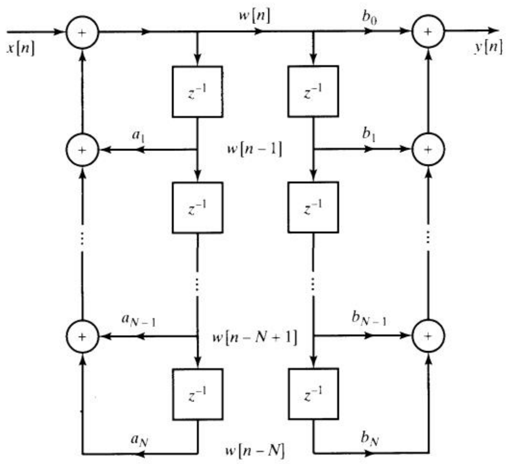


Figure 6.4 Rearrangement of block diagram of Figure 6.3. We assume for convenience that $N = M$. If $N \neq M$, some of the coefficients will be zero.

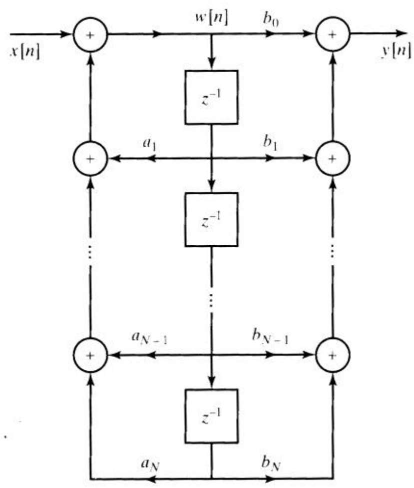


Figure 6.5 Combination of delays in Figure 6.4.

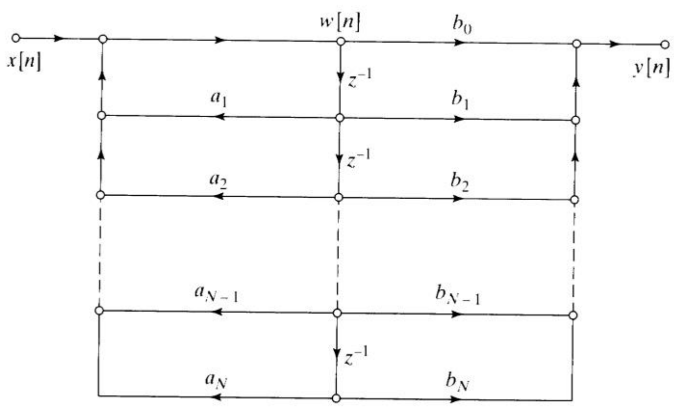


Figure 6.15 Signal flow graph of direct form II structure for an Nth-order system.

- **Cascade Form**

  -- Serial connection of 1st order and 2nd order factors

$$H(z)=\prod_{k=1}^{N_x}\frac{b_{ok}+b_{1k}z^{-1}+b_{2k}z^{-2}}{1-a_{1k}z^{-1}-a_{2k}z^{-2}}$$

Remark: Each factor is a Direct Form II.

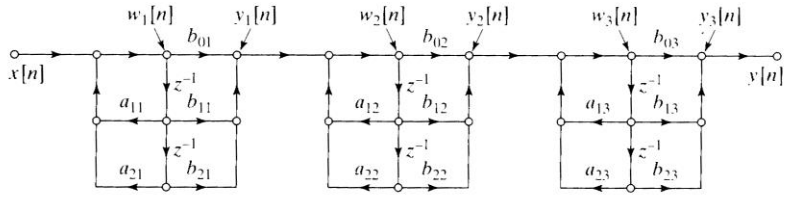


Figure 6.18 Cascade structure for a sixth-order system with a direct form II realization of each second-order subsystem.

If there are N_S second-order sections

$\Rightarrow (N_S!)^2$ different pairings and orderings!

- $\prod_{k=1}^{N_x}\frac{b_{ok}+b_{1k}z^{-1}+b_{2k}z^{-2}}{1-a_{1k}z^{-1}-a_{2k}z^{-2}}$ **needs 5 constant multipliers for each section.**
- $b_0\prod_{k=1}^{N_x}\frac{1+\tilde{b}_{1k}z^{-1}+\tilde{b}_{2k}z^{-2}}{1-a_{1k}z^{-1}-a_{2k}z^{-2}}$ **needs 4 constant multipliers for each section.**

The 5-multiplier sections are commonly used when implemented with fix-point arithmetic.

- **Parallel Form**

  -- Parallel connection of 1st order and 2nd order factors

$$H(z)=\sum_{k=0}^{N_p}C_k z^{-k}+\sum_{k=1}^{N_s}\frac{e_{ok}+e_{1k}z^{-1}}{1-a_{1k}z^{-1}-a_{2k}z^{-2}}$$

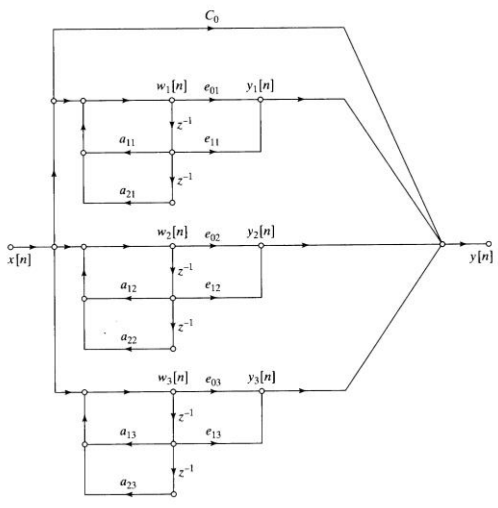


Figure 6.20 Parallel-form structure for sixth-order system ($M = N = 6$) with the real and complex poles grouped in pairs.

- Feedback in IIR Systems

  -- Basic formula of a feedback system (negative feedback)

$$H(z) = \frac{F(z)}{1 + F(z)B(z)}$$

  -- If a system has poles, a corresponding block diagram or signal flow graph will have feedback loops.

  (BUT neither poles in the system function nor loops in the network are sufficient for the impulse response to be infinitely long.)

  -- A delay element is necessary in the feedback loop; otherwise, it is *noncomputable*.

  (The structure should be modified to eliminate the noncompuable loops.)

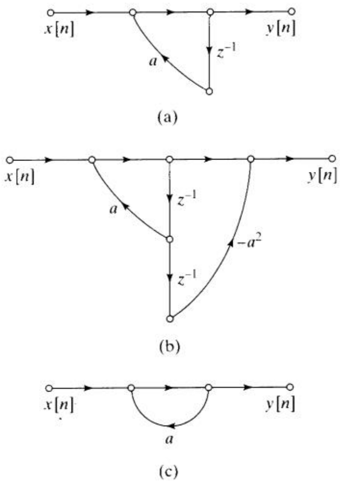


Figure 6.23 (a) System with feedback loop. (b) FIR system with feedback loop. (c) Noncomputable system.

- Transpose Forms

  -- *Transposition* of a flow graph is reversing the *directions* of *all* branches in the network while keeping the branch transmittances (as they were) and reversing the roles of the input and output (so that source nodes become sink nodes and vice versa).

  -- ***Flow Graph Reversal Theorem***

  For single-input, single-output systems, the transposed flow graph has the same system function as the original graph if the input nodes and output nodes are interchanged.

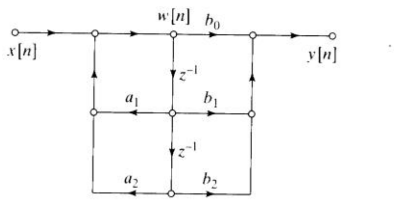


Figure 6.27 Direct form II structure for Example 6.8.

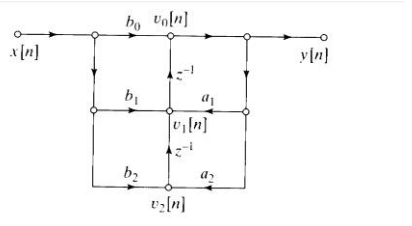


Figure 6.28 Transposed direct form II structure for Example 6.8.

✧ Basic Structures for FIR Systems

- **Direct Form**

  -- *Transversal filter* or *tapped delay line*

  $$y[n] = \sum_{k=0}^{M} b_k x[n-k]$$ -- convolution

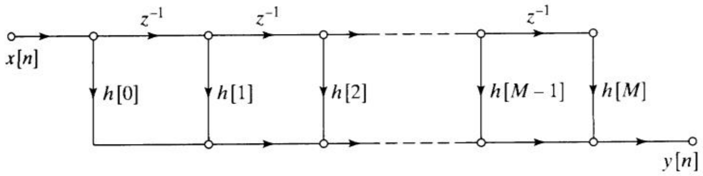


Figure 6.31 Direct-form realization of an FIR system.

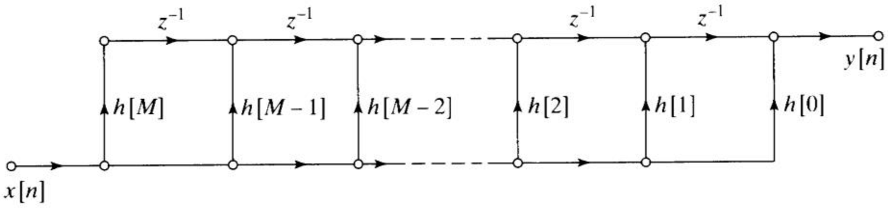


Figure 6.32 Transposition of the network of Figure 6.31.

- **Cascade Form**

  -- Serial connection of 1st order and 2nd order factors

  $$H(z) = \prod_{k=1}^{M_s} \left(b_{ok} + b_{1k} z^{-1} + b_{2k} z^{-2} \right)$$

  Remark: Each factor is a Direct Form.

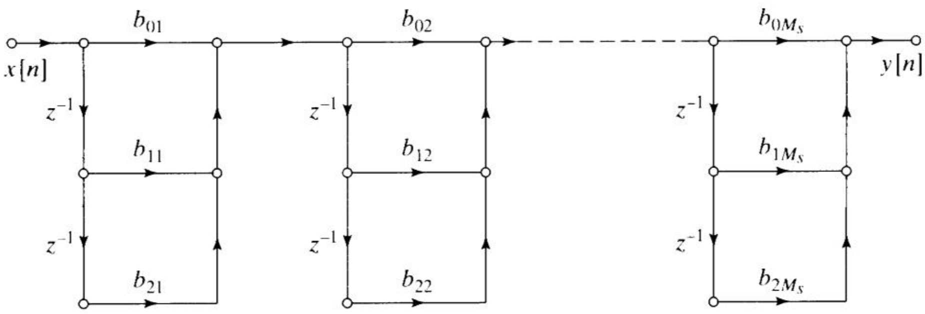


Figure 6.33 Cascade-form realization of an FIR system.

- **Linear Phase FIR Filters**

■ Take the advantage of the symmetry property of the impulse response

$$h[M-n]=h[n]$$

$$h[M-n]=-h[n]$$

$$y[n]=\sum_{k=0}^{M} h[k]x[n-k]$$

■ **Type I or III:** *M* even (order odd)

$$y[n]=\sum_{k=0}^{M/2-1} h[k]x[n-k]+h\left[\frac{M}{2}\right]x\left[n-\frac{M}{2}\right]+\sum_{k=\frac{M}{2}+1}^{M} h[k]x[n-k]$$

$$=\sum_{k=0}^{\frac{M}{2}-1} h[k](x[n-k]\pm x[n-M=k])+k\left[\frac{M}{2}\right]x\left[n-\frac{M}{2}\right]$$

■ **Type II or IV:** *M* odd (order even)

$$y[n]=\sum_{k=0}^{\frac{M-1}{2}} h[k](x[n-k]\pm x[n-M+k])$$

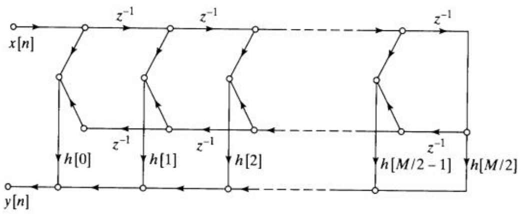


Figure 6.34 Direct-form structure for an FIR linear-phase system when *M* is an even integer.

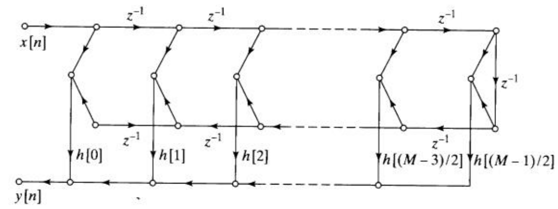


Figure 6.35 Direct-form structure for an FIR linear-phase system when *M* is an odd integer.

■ Linear-phase FIR filters can also be implemented as a cascade of 1st-order, 2nd-order, and 4th-order real-coefficient systems. (The 4th-order system is formed by grouping the conjugate and the conjugate reciprocal zeros together.)

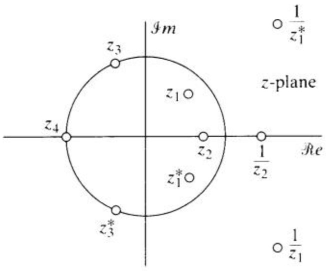
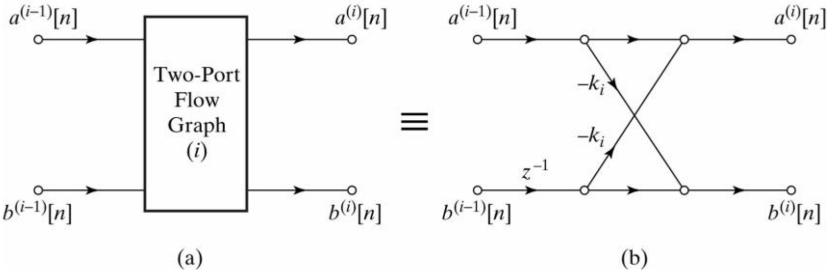


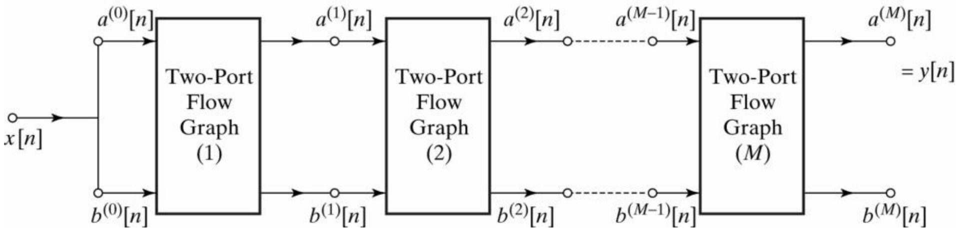
Figure 6.36 Symmetry of zeros for a linear-phase FIR filter.

- **Lattice Filters**

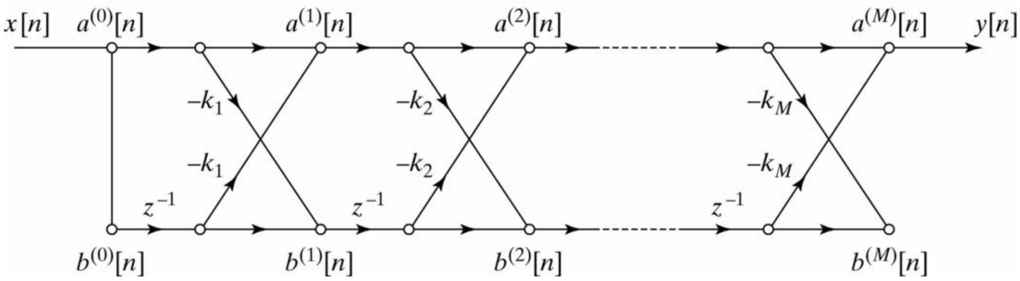
Two-port building block



Cascade connection of M basic building block sections



FIR Lattice Filters



k-parameters: $k_1, k_2, \cdots, k_M$

$$a^{(0)}[n] = b^{(0)}[n] = x[n]$$
$$a^{(i)}[n] = a^{(i-1)}[n] - k_i b^{(i-1)}[n-1] \qquad i = 1,2,\ldots,M$$
$$b^{(i)}[n] = b^{(i-1)}[n] - k_i a^{(i-1)}[n-1] \qquad i = 1,2,\ldots,M$$
$$y[n] = a^{(M)}[n]$$

Let x[n] = δ[n].

The transfer function between the input and the upper ith node is

$$A^{(i)}(z) = \sum_{n=0}^{i} a^{(i)}[n] z^{-n} = 1 - \sum_{m=1}^{i} \alpha_m^{(j)} z^{-m}$$

$$a^{(i)}[n] = \begin{cases} 1 & n = 0 \\ -\alpha_n^{(i)} & 1 \le n \le i \\ 0 & \text{otherwise} \end{cases}$$

$$A_0(z) = B_0(z) = 1$$
$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-1} B^{(i-1)}(z)$$
$$B^{(i)}(z) = -k_i A^{(i-1)}(z) + z^{-1} B^{(i-1)}(z)$$

$$A^{(1)}(z) = A^{(0)}(z) - k_1 z^{-1} B^{(0)}(z) = 1 - k_1 z^{-1}$$
$$B^{(1)}(z) = -k_1 A^{(0)}(z) + z^{-1} B^{(0)}(z) = -k_1 + z^{-1} = z^{-1} A^{(1)}(1/z)$$

$$A^{(2)}(z) = A^{(1)}(z) - k_2 z^{-1} B^{(1)}(z) = 1 - k_1 z^{-1} - k_2 z^{-2}(1 - k_1 z)$$
$$= 1 - k_1(1-k_2) z^{-1} - k_2 z^{-2}$$
$$B^{(2)}(z) = -k_2 A^{(1)}(z) + z^{-1} B^{(1)}(z) = -k_2(1 - k_1 z^{-1}) + z^{-2}(1 - k_1 z)$$
$$= z^{-2}(1 - k_1(1-k_2) z - k_2 z^2) = z^{-2} A^{(2)}(1/z)$$

By induction, we can prove that the relationship between $B^{(i)}(z)$ and $A^{(i)}(z)$ is

$$B^{(i)}(z) = z^{-i} A^{(i)}(1/z)$$

or $$A^{(i)}(z) = z^{-i} B^{(i)}(1/z)$$

Since $A^{(i)}(z) = 1 - \sum_{m=1}^{i} \alpha_m^{(j)} z^{-m}$, we have

$$A^{(i-1)}(z) = 1 - \sum_{m=1}^{i-1} \alpha_m^{(j-1)} z^{-m} \quad \text{and}$$
$$B^{(i-1)}(z) = z^{-(i-1)} A^{(i-1)}(1/z)$$
$$= z^{-(i-1)}[1 - \sum_{m=1}^{i-1} \alpha_m^{(j-1)} z^{m}]$$

$$\Rightarrow \quad A^{(i)}(z) = (1 - \sum_{m=1}^{i-1} \alpha_m^{(i-1)} z^{-m}) - k_i z^{-1}(z^{-(i-1)}[1 - \sum_{m=1}^{i-1} \alpha_m^{(j-1)} z^{m}])$$

$$\Rightarrow \quad A^{(i)}(z) = 1 - \sum_{m=1}^{i-1} [\alpha_m^{(i-1)} - k_i \alpha_{i-m}^{(i-1)}] z^{-m} - k_i z^{-i}$$

$$\Rightarrow \quad \begin{aligned} &\alpha_m^{(i)} = [\alpha_m^{(i-1)} - k_i \alpha_{i-m}^{(i-1)}] \quad m = 1, \ldots, i-1 \\ &\alpha_i^{(i)} = k_i \end{aligned}$$

In matrix form, if we define

$$\boldsymbol{\alpha}_{i-1} = \begin{bmatrix} \alpha_1^{(i-1)} & \alpha_2^{(i-1)} & \cdots & \alpha_{i-1}^{(i-1)} \end{bmatrix}^T \text{ and } \breve{\boldsymbol{\alpha}}_{i-1} = \begin{bmatrix} \alpha_{i-1}^{(i-1)} & \alpha_{i-2}^{(i-1)} & \cdots & \alpha_1^{(i-1)} \end{bmatrix}^T$$

$$\boldsymbol{\alpha}_i = \begin{bmatrix} \boldsymbol{\alpha}_{i-1} \\ \cdots \\ 0 \end{bmatrix} - k_i \begin{bmatrix} \breve{\boldsymbol{\alpha}}_{i-1} \\ \cdots \\ -1 \end{bmatrix}$$

We can use the above equations recursively to compute the transfer function of successively higher-order FIR filters until we get

$$A(z) = 1 - \sum_{m=1}^{M} \alpha_m z^{-m} = \frac{Y(z)}{X(z)} \quad \text{where} \quad \alpha_m = \alpha_m^{(M)} \qquad m = 1,2,\ldots,M$$

k-Parameters-to-Coefficients Algorithm

Given $k_1, k_2, \ldots, k_M$
for $i = 1, 2, \ldots, M$
  $\alpha_i^{(i)} = k_i$ Eq. (6.66b)
  if $i > 1$ then for $j = 1, 2, \ldots, i-1$
    $\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}$ Eq. (6.66a)
  end
end
$\alpha_j = \alpha_j^{(M)} \quad j = 1, 2, \ldots, M$ Eq. (6.68b)

On the contrary, if we are given A(z), we follow the follow algorithm to get the *k*-parameters.

Coefficients-to-k-Parameters Algorithm

Given $\alpha_j^{(M)} = \alpha_j \quad j = 1, 2, \ldots, M$
$k_M = \alpha_M^{(M)}$ Eq. (6.69)
for $i = M, M-1, \ldots, 2$
  for $j = 1, 2, \ldots, i-1$
    $\alpha_j^{(i-1)} = \dfrac{\alpha_j^{(i)} + k_i \alpha_{i-j}^{(i)}}{1 - k_i^2}$ Eq. (6.71a)
  end
  $k_{i-1} = \alpha_{i-1}^{(i-1)}$ Eq. (6.71b)
end

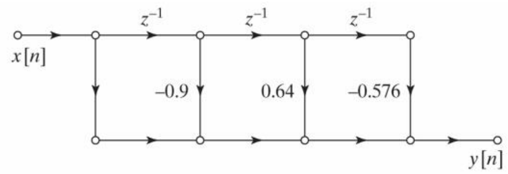
Example: $A(z) = 1 - 0.9z^{-1} + 0.64z^{-2} - 0.576z^{-3}$

M = 3 $\quad \alpha_1^{(3)} = 0.9 \quad \alpha_2^{(3)} = 0.64 \quad \alpha_3^{(3)} = 0.576 \Rightarrow k_3 = \alpha_3^{(3)} = 0.576$

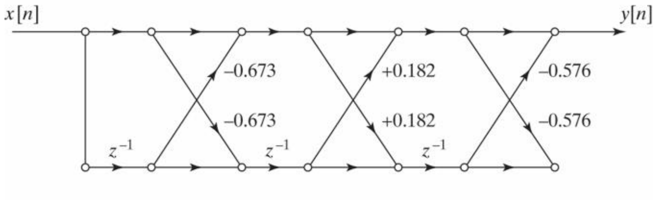
$$\alpha_1^{(2)} = \frac{\alpha_1^{(3)} + k_3 \alpha_2^{(3)}}{1 - k_3^2} = 0.795$$

$$\alpha_2^{(2)} = \frac{\alpha_2^{(3)} + k_3 \alpha_1^{(3)}}{1 - k_3^2} = -0.182 \Rightarrow k_2 = \alpha_2^{(2)} = -0.182$$

$$\alpha_1^{(1)} = \frac{\alpha_1^{(2)} + k_2 \alpha_1^{(2)}}{1 - k_2^2} = 0.673 \Rightarrow k_1 = \alpha_1^{(1)} = 0.673$$



(a)



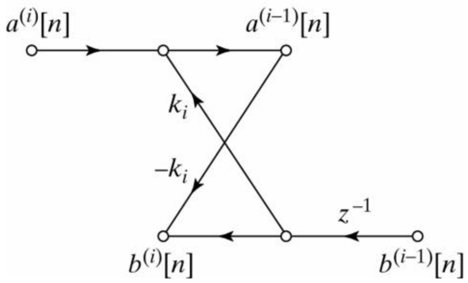
(b)

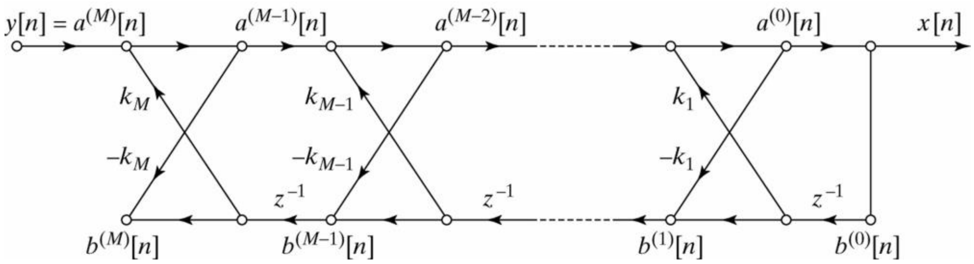
All-Pole Lattice Filters

$H(z) = 1/A(z)$

Assume we are given $y[n] = a^{(M)}[n]$ and we wish to compute the input $a^{(0)}[n]=x[\mathrm{n}]$.

$$A^{(i-1)}(z) = A^{(i)}(z) + k_i z^{-1} B^{(i-1)}(z)$$
$$B^{(i)}(z) = -k_i A^{(i-1)}(z) + z^{-1} B^{(i-1)}(z)$$

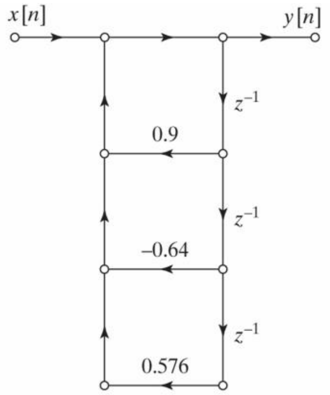




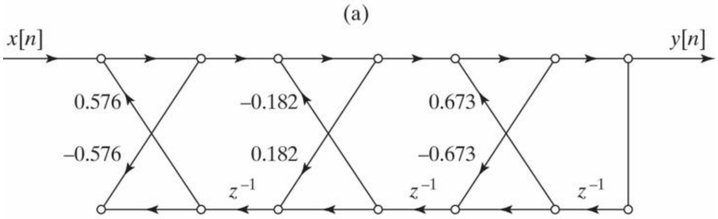
$$a^{(M)}[n] = y[n]$$
$$a^{(i-1)}[n] = a^{(i)}[n] + k_i b^{(i-1)}[n-1] \qquad i = M, (M-1), \ldots, 1$$
$$b^{(i)}[n] = b^{(i-1)}[n-1] - k_i a^{(i-1)}[n] \qquad i = M, (M-1), \ldots, 1$$
$$x[n] = a^{(0)}[n] = b^{(0)}[n]$$

Example:

$$H(z)=\frac{1}{1-0.9z^{-1}+0.64z^{-2}-0.576z^{-3}}=\frac{1}{(1-0.8jz^{-1})(1+0.8jz^{-1})(1-0.9z^{-1})}$$



(a)



(b)

✧ Finite-precision Numerical Effects

- Due to finite-precision (finite-word length) of computational and/or storage devices.
  -- Parameter quanitization
  -- Round-off error
  -- Limit cycle (IIR) ← zero input!
- Number Representation **-- Two's complement** number representation

$$\hat{x} = X_m \underbrace{\left(-b_0 + \sum_{i=1}^{B} b_i 2^{-i} \right)}_{\hat{x}_B = b_o b_1 b_2 \ldots b_B}$$

$$(-1 \le \hat{x}_B < 1)$$

$$\therefore \quad -X_m \le \hat{x} < X_m$$

$$\Delta = X_m 2^{-B}$$

$\hat{x} = Q_B[x]$: quantized value of x

$\hat{x}_B$: normalized quantized value of x; normalized value of $\hat{x}$

Δ: quantization stepsize

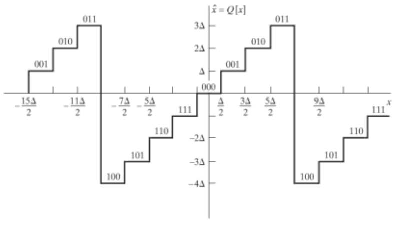
b_0: sign bit; b_0 =0, if x is non-negative; b_0 =1, if x is negative.

- Quantization error $e = Q_B[x] - x$

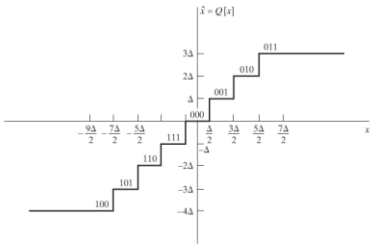
(1) Overflow: if $x > X_m$. This can be a serious problem if, for example, 0111→1000, and we don't check it first. (This is *natural overflow*.) We first clip the input. It becomes *saturation*.

(2) -- Rounding: nearest integer $-\Delta/2 < e \le \Delta/2$

-- Truncation: smaller integer $-\Delta < e \le 0$



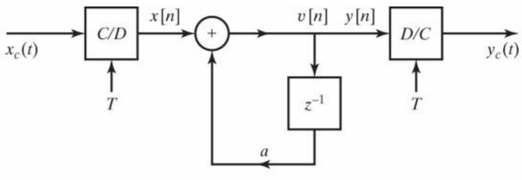
(a)



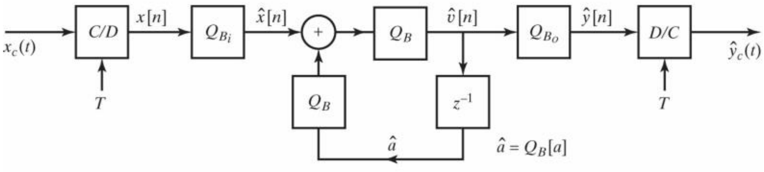
(b)

- **Quantization in implementing systems**

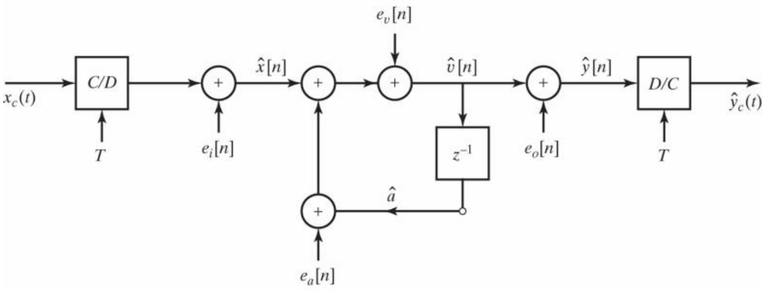
$$H(z)=\frac{1}{1-az^{-1}}$$



(a)



(b)



(c)

✧ Effects of Coefficient Quantization

- **Coefficient Quantization in IIR Systems**

  -- depends on the filter structure

  - **Direct form**

$$H(z)=\frac{\sum_{k=0}^{M} b_k z^{-k}}{\underbrace{1-\sum_{k=1}^{N} a_k z^{-k}}_{A(z)}} \rightarrow \widehat{H}(z)=\frac{\sum_{k=0}^{M} \widehat{b}_k z^{-k}}{1-\sum_{k=1}^{N} \widehat{a}_k z^{-k}}$$

Note: $\hat{a}_k = a_k + \Delta a_k$; $\hat{b}_k = b_k + \Delta b_k$

<u>Effect on pole locations</u>

(→ affect frequency response and stability)

Compare
$$\begin{cases} A(z)=1-\sum_{k=1}^{N} a_k z^{-k} = \prod_{j=1}^{N}\left(1-z_j z^{-1}\right) \\ \widehat{A}(z)=1-\sum \widehat{a}_k z^{-k} = \prod\left(1-\widehat{z}_j z^{-1}\right) \end{cases}$$

The change of pole location: $\hat{z}_j = z_j + \Delta z_j, \quad \mathrm{j}=1,\ldots,\mathrm{N}$

Δz_k is affected by all $\{\Delta a_k\}$.

$$\Delta z_i \approx \sum_{k=1}^{N} \left(\frac{\partial z_i}{\partial a_k} \right) \Delta a_k, \qquad \mathrm{i}=1,2,\ldots,\mathrm{N}$$

Remark: This formula is approximately true when Δa_k and Δz_k are small.

Note that if $\frac{\partial z_i}{\partial a_k}$ is large, then a small Δa_k leads to a large Δz_k. If so, this is a *sensitive* system. (Undesirable)

One step further, $\dfrac{\partial z_i}{\partial a_k} = \dfrac{z_i^{\,N-k}}{\prod_{j=1, j\neq i}^{N} (z_i - z_j)}$ $\left(\begin{array}{cc} \left(\dfrac{\partial A(z)}{\partial z_i} \right)_{z=z_i} \dfrac{\partial z_i}{\partial a_k} = & \left. \dfrac{\partial A(z)}{\partial a_k} \right|_{z=z_i} \\ \Downarrow & \Downarrow \\ \prod_{i\neq j} \left(1 - z_i^{-1} \cdot z_j\right) & z_i^{-k} \end{array} \right)$

That is, if $(z_i - z_j)$ is small, then $\frac{\partial z_i}{\partial a_k}$ is large; for example, narrow bandwidth lowpass and bandpass filters which have clustered poles.

Remark: The preceding analysis can be applied to zeros.

- **Parallel and cascade forms**

-- consists of 1st-order and 2nd-order sections.

Errors in a particular pole pairs (section) are independent of the other poles (sections).

This is also true for zeros in cascade form. ➔ In general, both the *cascade form* and the *parallel form* are less sensitive to coefficient quantization (because zeros are often widely distributed the unit circle).

Example: Bandpass IIR elliptic filter

$0.99 \le |H(e^{jw})| \le 1.01 \qquad 0.3\pi \le \omega \le 0.4\pi$

$|H(e^{jw})| \le 0.01 \quad (\text{-}40\text{dB}) \qquad \omega \le 0.29\pi$

$|H(e^{jw})| \le 0.01 \quad (\text{-}40\text{dB}) \qquad 0.41\pi \le \omega \le \pi$

TABLE 6.3 UNQUANTIZED CASCADE-FORM COEFFICIENTS FOR A 12TH-ORDER ELLIPTIC FILTER

k	a_{1k}	a_{2k}	b_{0k}	b_{1k}	b_{2k}
1	0.737904	-0.851917	0.137493	0.023948	0.137493
2	0.961757	-0.861091	0.281558	-0.446881	0.281558
3	0.629578	-0.933174	0.545323	-0.257205	0.545323
4	1.117648	-0.941938	0.706400	-0.900183	0.706400
5	0.605903	-0.984347	0.769509	-0.426879	0.769509
6	1.173028	-0.986717	0.937657	-1.143918	0.937657

TABLE 6.2 SIXTEEN-BIT QUANTIZED CASCADE-FORM COEFFICIENTS FOR A 12TH-ORDER ELLIPTIC FILTER

k	a_{1k}	a_{2k}	b_{0k}	b_{1k}	b_{2k}
1	24196×2^{-15}	-27880×2^{-15}	17805×2^{-17}	3443×2^{-17}	17805×2^{-17}
2	31470×2^{-15}	-28180×2^{-15}	18278×2^{-16}	-29131×2^{-16}	18278×2^{-16}
3	20626×2^{-15}	-30522×2^{-15}	17556×2^{-15}	-8167×2^{-15}	17556×2^{-15}
4	18292×2^{-14}	-30816×2^{-15}	22854×2^{-15}	-29214×2^{-15}	22854×2^{-15}
5	19831×2^{-15}	-32234×2^{-15}	25333×2^{-15}	-13957×2^{-15}	25333×2^{-15}
6	19220×2^{-14}	-32315×2^{-15}	15039×2^{-14}	-18387×2^{-14}	15039×2^{-14}

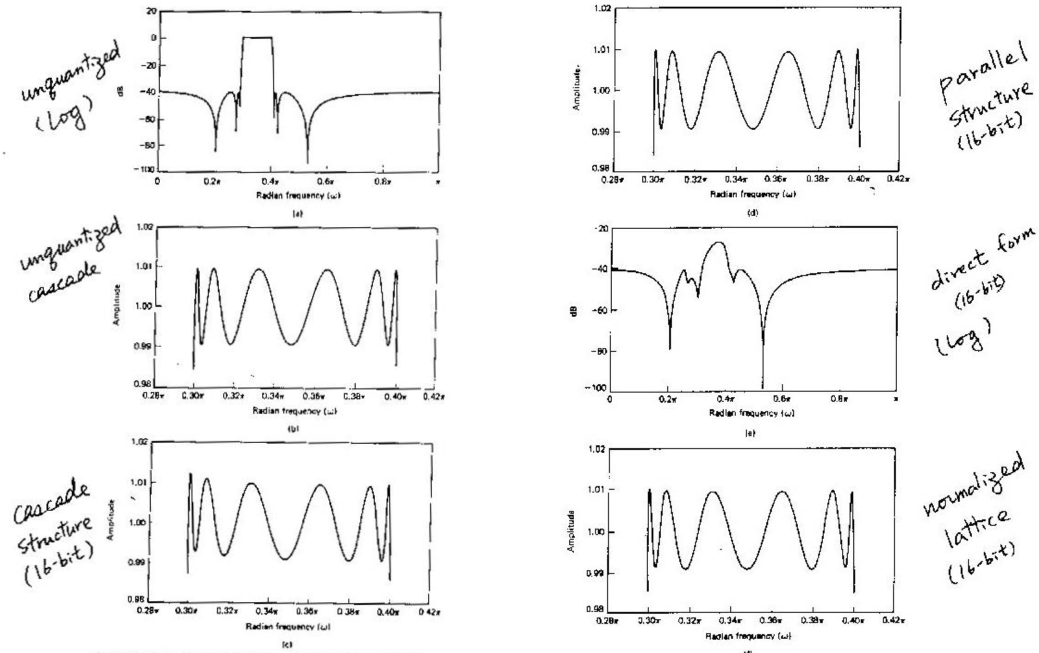


Figure 6.47 IIR coefficient quantization example. (a) Log magnitude for unquantized elliptic bandpass filter. (b) Passband for unquantized cascade case. (c) Passband for cascade structure with 16-bit coefficients

Figure 6.47 (*continued*) (d) Passband for parallel structure with 16-bit coefficients. (e) Log magnitude for direct form with 16-bit coefficients. (f) Passband for normalized lattice with 16-bit coefficients.

- **Second-order Section**

  (1) Direct form

  Poles: $(re^{j\theta}, re^{-j\theta})$

$$\frac{1}{\left(1-re^{j\theta}z^{-1}\right)\left(1-re^{-j\theta}z^{-1}\right)} = \frac{1}{1-2r\cos\theta z^{-1}+r^2z^{-2}}$$

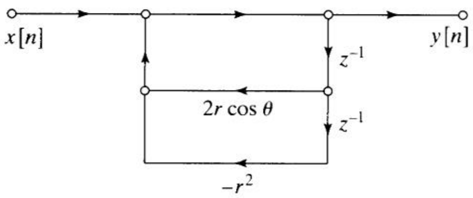
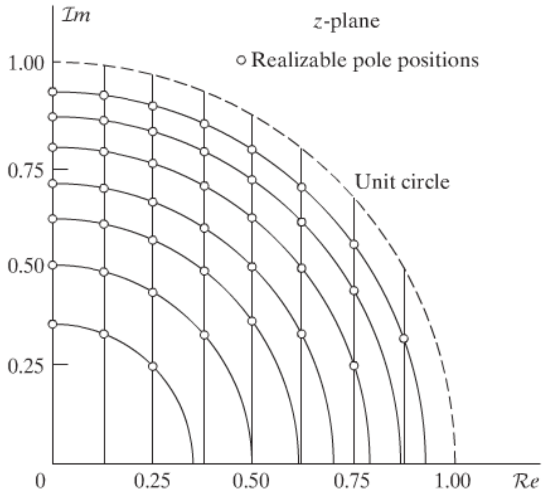
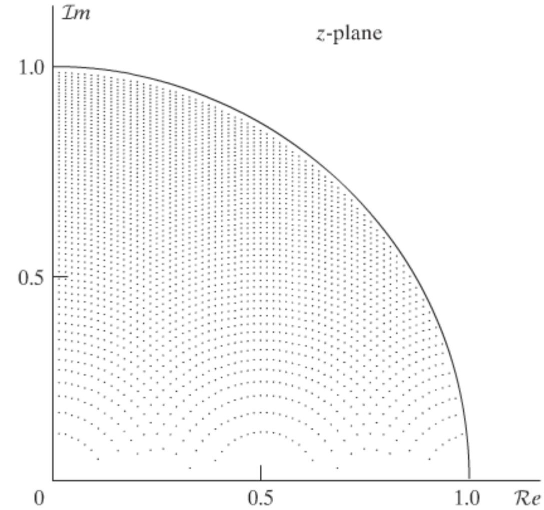


Figure 6.41 Direct-form implementation of a complex-conjugate pole pair.



(a)



(b)

(2) Coupled form

Poles: $(re^{j\theta}, re^{-j\theta})$

$$\begin{cases} Y_1 = X + r\cos\theta\, z^{-1}Y_1 - r\sin\theta\, z^{-1}Y \\ Y = r\sin\theta\, z^{-1}Y_1 + r\cos\theta\, z^{-1}Y \end{cases}$$

$$\Rightarrow Y_1 = \left(z - r\sin\theta\, z^{-1}Y\right) / 1 - r\cos\theta\, z^{-1}$$

$$\frac{Y}{X} = \frac{(r\sin\theta)z^{-1}}{1 - 2r\cos\theta\, z^{-1} + r^2 z^{-2}}$$

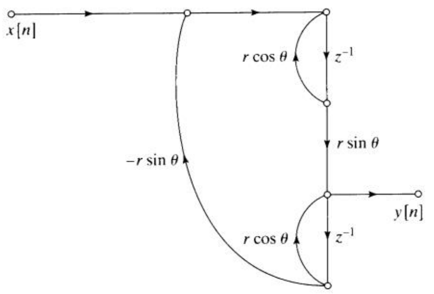
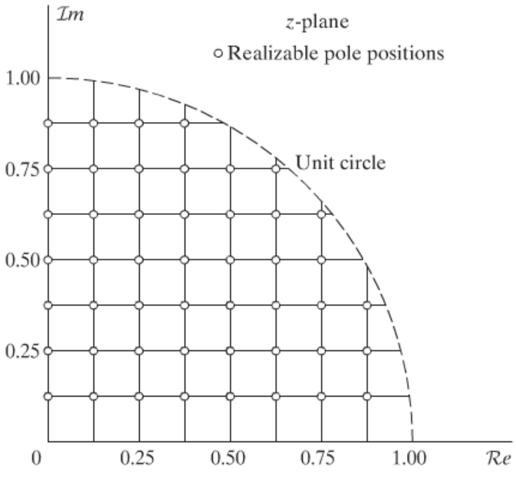
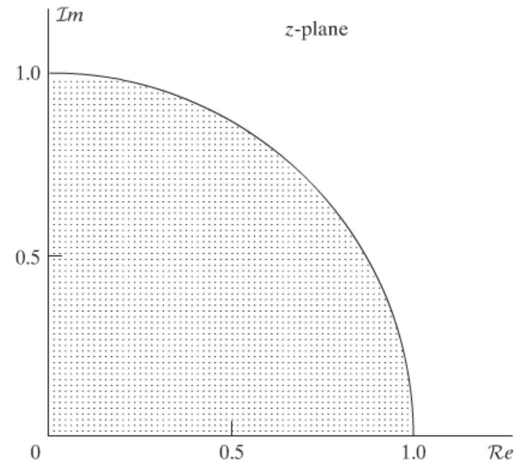


Figure 6.43 Coupled-form implementation of a complex-conjugate pole pair.



(a)



(b)

➔ The pole location distribution is even.

🡐 The price: The number of multiplications is doubled!

- **Coefficient Quantization in FIR Systems**
  - **Direct form**

$$H(z)=\sum_{n=0}^{M}h[n]z^{-n}\rightarrow \hat{H}(z)=\sum \hat{h}[n]z^{-n}$$

$$=H(z)+\underbrace{\Delta H(z)}_{\sum \Delta h[n]z^{-n}}$$

$$\Delta H(z)=\sum_{n=0}^{M}\Delta h[n]z^{-n}$$

<u>Effect on the zero locations</u>

The sensitivity function of this form is similar to that of the direct form I IIR filter. That is, if the zeros are tightly clustered, their locations will then be highly sensitive to quantization errors. However, for most linear phase FIR systems, the zeros are more or less uniformly spread in the z-plane.

<u>Effect on $H(e^{j\omega})$</u>

After scaling, each $h[n]$ is represented by (B+1) bits 2's complement number; i.e.,

$$-2^{-(B+1)} < \Delta h[n] \le 2^{-(B+1)}.$$

$$\Delta H\left(e^{j\omega}\right) = \sum_{n=0}^{M} \Delta h[n] e^{-j\omega n}$$

$$\left|\Delta H\left(e^{j\omega}\right)\right| = \left|\sum_{n=0}^{M} \Delta h[n] e^{-j\omega n}\right| \le \sum_{n=0}^{M} \left|\Delta h[n]\right| \left|e^{-j\omega n}\right|$$

$$\underbrace{\le (M+1) 2^{-(B+1)}}_{worst\ case!}$$

<u>Effect on linear phase</u>

Not affect the linear phase property as long as $\hat{h}[n] = \hat{h}[M-n]$.

Example: Linear Phase Lowpass Filter

$$0.99 \le \left|H(e^{jw})\right| \le 1.01 \qquad 0 \le \omega \le 0.4\pi$$

$$\left|H(e^{jw})\right| \le 0.001 \quad (\text{-}60\text{dB}) \qquad 0.6\pi \le \omega \le \pi$$

TABLE 6.3 UNQUANTIZED AND QUANTIZED COEFFICIENTS FOR AN OPTIMUM FIR LOWPASS FILTER ($M = 27$)

Coefficient	Unquantized	16 bits	14 bits	13 bits	8 bits
$h[0] = h[27]$	1.359657×10^{-3}	45×2^{-15}	11×2^{-13}	6×2^{-12}	0×2^{-7}
$h[1] = h[26]$	-1.616993×10^{-3}	-53×2^{-15}	-13×2^{-13}	-7×2^{-12}	0×2^{-7}
$h[2] = h[25]$	-7.738032×10^{-3}	-254×2^{-15}	-63×2^{-13}	-32×2^{-12}	-1×2^{-7}
$h[3] = h[24]$	-2.686841×10^{-3}	-88×2^{-15}	-22×2^{-13}	-11×2^{-12}	0×2^{-7}
$h[4] = h[23]$	1.255246×10^{-2}	411×2^{-15}	103×2^{-13}	51×2^{-12}	2×2^{-7}
$h[5] = h[22]$	6.591530×10^{-3}	216×2^{-15}	54×2^{-13}	27×2^{-12}	1×2^{-7}
$h[6] = h[21]$	-2.217952×10^{-2}	-727×2^{-15}	-182×2^{-13}	-91×2^{-12}	-3×2^{-7}
$h[7] = h[20]$	-1.524663×10^{-2}	-500×2^{-15}	-125×2^{-13}	-62×2^{-12}	-2×2^{-7}
$h[8] = h[19]$	3.720668×10^{-2}	1219×2^{-15}	305×2^{-13}	152×2^{-12}	5×2^{-7}
$h[9] = h[18]$	3.233332×10^{-2}	1059×2^{-15}	265×2^{-13}	132×2^{-12}	4×2^{-7}
$h[10] = h[17]$	-6.537057×10^{-2}	-2142×2^{-15}	-536×2^{-13}	-268×2^{-12}	-8×2^{-7}
$h[11] = h[16]$	-7.528754×10^{-2}	-2467×2^{-15}	-617×2^{-13}	-308×2^{-12}	-10×2^{-7}
$h[12] = h[15]$	1.560970×10^{-1}	5115×2^{-15}	1279×2^{-13}	639×2^{-12}	20×2^{-7}
$h[13] = h[14]$	4.394094×10^{-1}	14399×2^{-15}	3600×2^{-13}	1800×2^{-12}	56×2^{-7}

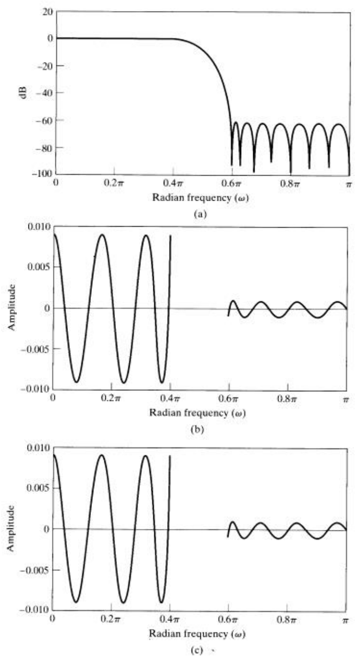


Figure 6.46 FIR quantization example. (a) Log magnitude for unquantized case. (b) Approximation error for unquantized case. (Error not defined in transition band.) (c) Approximation error for 16-bit quantization.

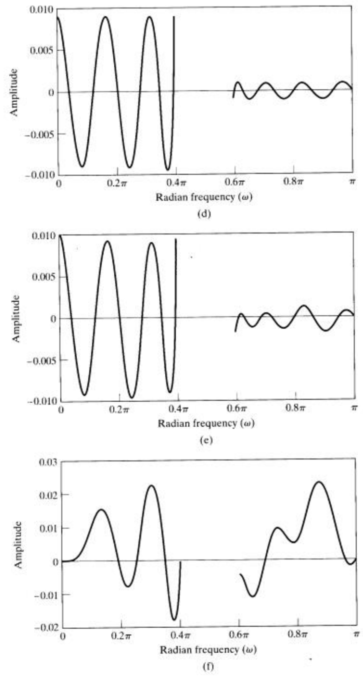


Figure 6.46 (*continued*) (d) Approximation error for 14-bit quantization. (e) Approximation error for 13-bit quantization. (f) Approximation error for 8-bit quantization.

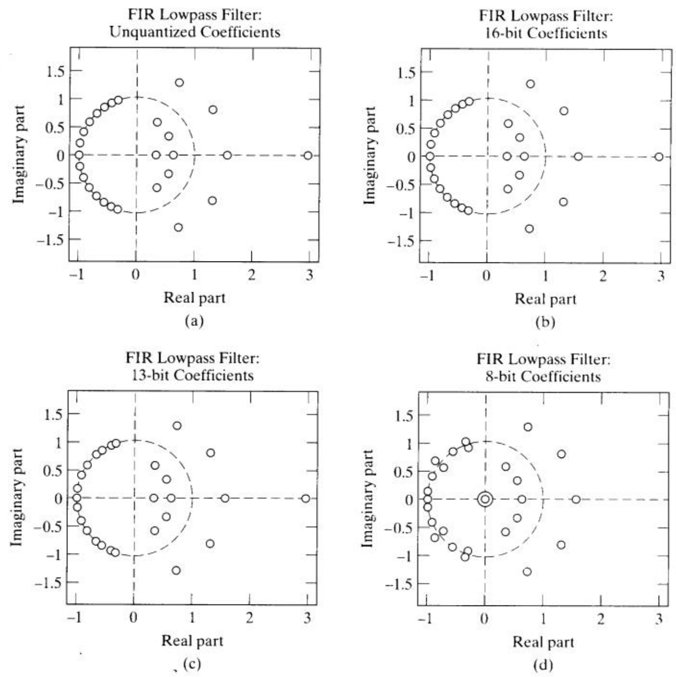


Figure 6.47 Effect of impulse response quantization on zeros of $H(z)$. (a) Unquantized. (b) Sixteen-bit quantization. (c) Thirteen-bit quantization. (d) Eight-bit quantization.

- **Cascade form**

  -- less sensitive because it isolate the quantization errors from the other sections.

  -- To preserve linear phase each section is linear phase.

  (a) Conjugate 2nd-order sections for conjugate zero pairs on the unit circle. $(1+az^{-1}+z^{-2})$

  (b) Real zero 2nd-order sections for a real zero inside the unit circle and its reciprocal (outside the unit circle).

  (c) Zeros at ± 1.

  (d) 4th-order sections for conjugate zero pairs inside the unit circle and their associated reciprocals (outside the unit circle).

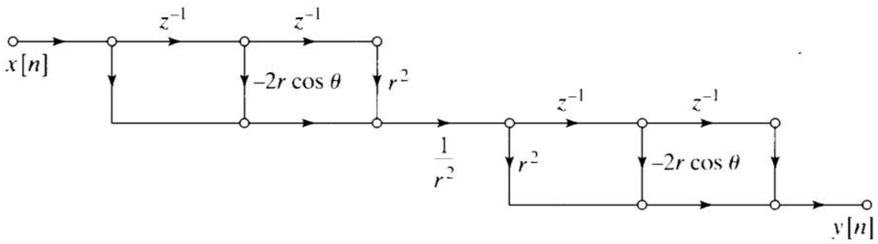


Figure 6.48 Subnetwork to implement fourth-order factors in a linear-phase FIR system such that linearity of the phase is maintained independently of parameter quantization.

$$(1-2r\cos\theta\cdot z^{-1}+r^2\cdot z^{-2})\frac{1}{r^2}(r^2-2r\cos\theta\cdot z^{-1}+z^{-2})$$