IMPLEMENTATION STRUCTURES FOR DISCRETE-TIME SYSTEMS

FINITE PRECISION NUMERICAL EFFECTS-NUMBER REPRESENTATIONS

QUANTIZATION IN IMPLEMENTING SYSTEMS

REALIZABLE POLE LOCATIONS

DIRECT FORM-I, DIRECT FORM-II

CASCADE FORM

PARALLEL FORMS

TRANSPOSED FORMS

FIR STRUCTURES

GENERALIZED LINEAR PHASE FIR STRUCTURES

DETERMINATION OF THE SYSTEM FUNCTION FROM A FLOW GRAPH

“Although two structures may be equivalent with regard to their input-output characteristics for infinite precision representations of coefficients and variables, they may have vastly different behavior when the numerical precision is limited.”

Oppenheim, Schafer, 3rd ed., p. 403

FINITE PRECISION NUMERICAL EFFECTS

NUMBER REPRESENTATIONS

A real number in two's complement form (infinite precision)

$$x = X_m\left(-b_0 + \sum_{i=1}^{\infty} b_i 2^{-i}\right)$$

X_m: arbitrary scale factor

$$b_0 = 0 \qquad \Rightarrow \qquad 0 \leq x \leq X_m$$

$$b_0 = 1 \qquad \Rightarrow \qquad -X_m \leq x \leq 0$$

Quantized form (+1 bits, finite precision)

$$\hat{x} = X_m\left(-b_0 + \sum_{i=1}^{B} b_i 2^{-i}\right)$$

$$= X_m \hat{x}_B$$

$$= X_m (b_0\ b_1\ b_2\ b_3 \dots b_B)$$

Quantization step size,

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

THE ROLE OF X_m

In <u>A/D conversion</u>

$$[-X_m, X_m] \quad \text{volts} \quad \leftrightarrow \quad -1 \le \hat{x}_B \le 1 \quad \text{binary numbers}$$

Ex: A 14 bit A/D converter is specified to have a dynamic range of ± 5 volts. Assuming uniform quantization what are the values of 14 binary bits when its input is 3.111 Volt?

Solution:

$$\begin{aligned} X_m &= 5 \\ B &= 13 \\ \Delta &= X_m 2^{-B} \\ &= 5 \times 2^{-13} \end{aligned}$$

$$\frac{3.111}{\Delta} = 5097.1 \ldots$$

$$5097 = 2^{12} + 2^9 + 2^8 + 2^7 + 2^6 + 2^5 + 2^3 + 2^0$$

$$\Rightarrow \; b_0 = 0$$

$$b_1 = b_4 = b_5 = b_6 = b_7 = b_8 = b_{10} = b_{13} = 1$$

$$b_2 = b_3 = b_9 = b_{11} = b_{12} = 0$$

```
MATLAB code to check

x = 5*(2^12+2^9+2^8+2^7+2^6+2^5+2^3+2^0)*2^-13
d = 5*2^-13
(3.111-x)/d
Result
x = 3.110961914062500
d = 6.103515625000000e-04
ans = 0.062400000000343
```

In <u>fixed-point arithmetic</u>, it is common to assume that each binary number has a scale factor of

$$X_m = 2^c$$

For example

$$c = 2 \qquad \Rightarrow \qquad \hat{x}_B = b_0\, b_1\, b_2 . b_3 \ldots b_B$$

$$binary\ point$$

In floating-point arithmetic,

$$\hat{x} = \underbrace{X_m}_{characteristic} \quad \underbrace{\hat{x}_B}_{mantissa}$$

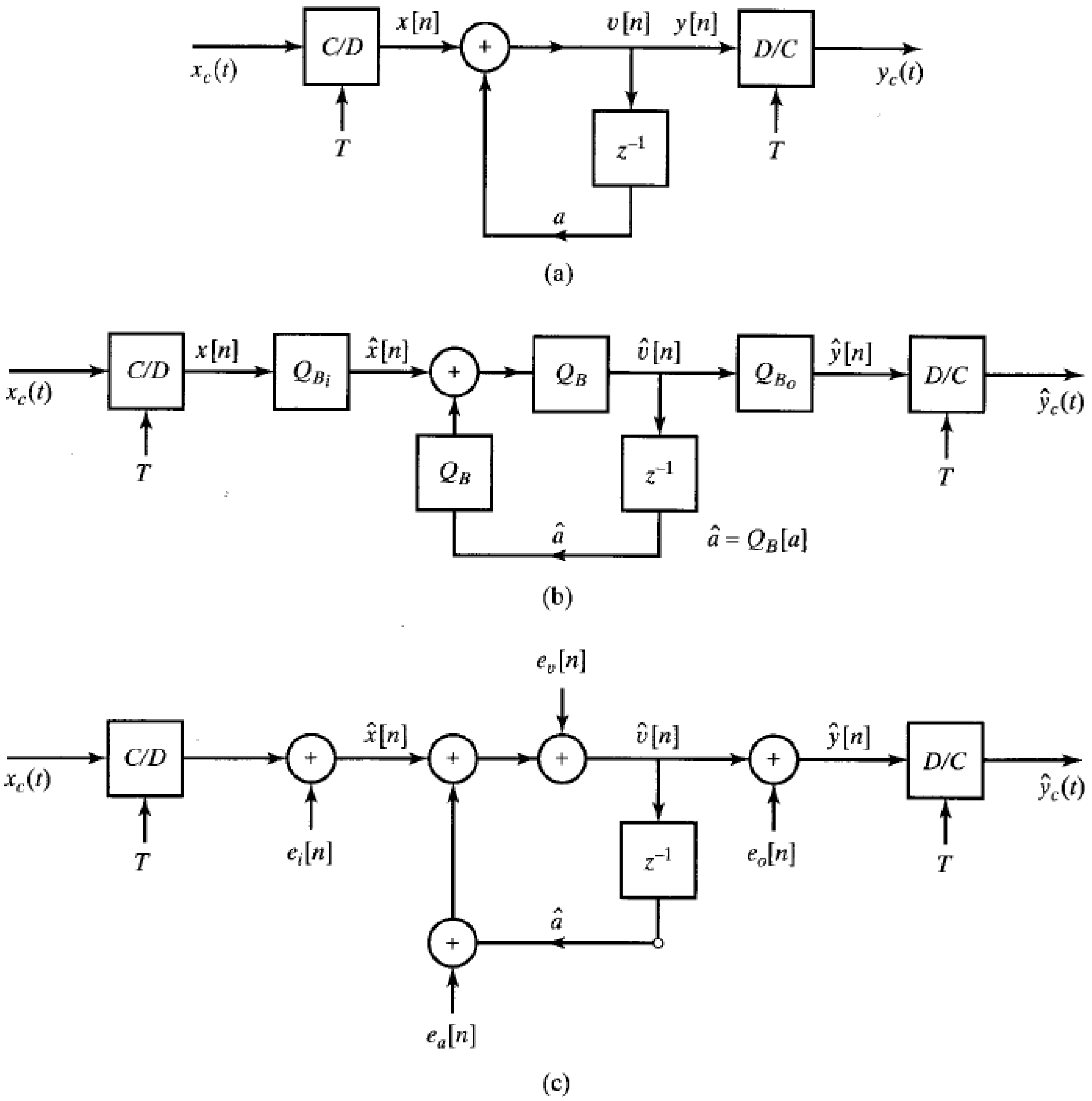


Figure 6.46 Implementation of discrete-time filtering of an analog signal. (a) Ideal system. (b) Nonlinear model. (c) Linearized model.

REALIZABLE POLE LOCATIONS

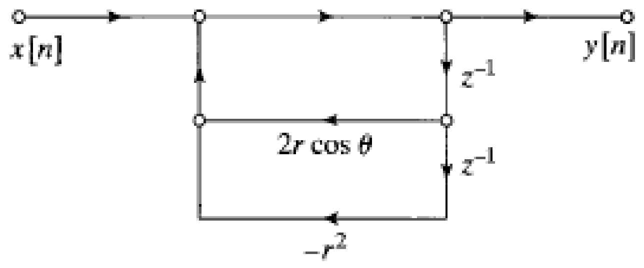
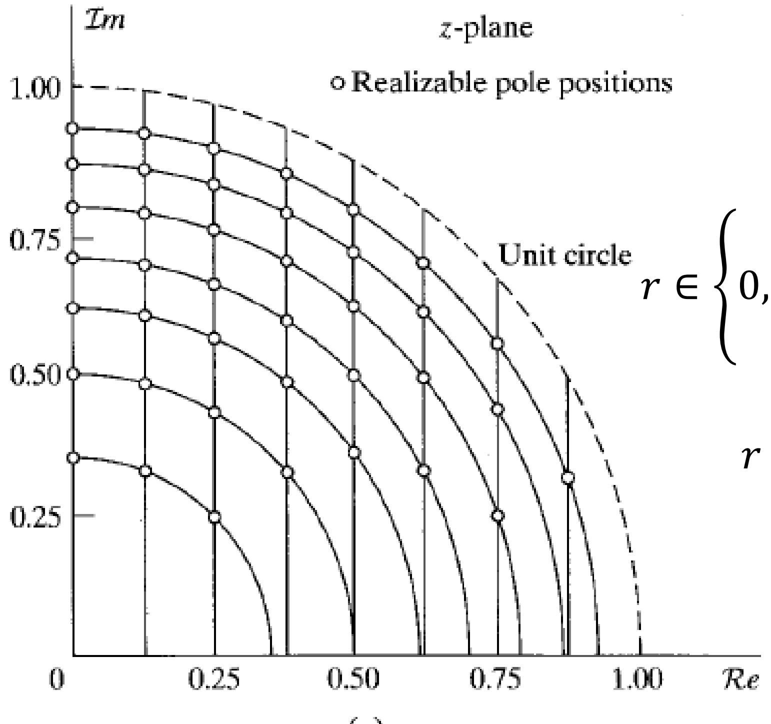


Figure 6.49 Direct form implementation of a complex-conj pole pair.

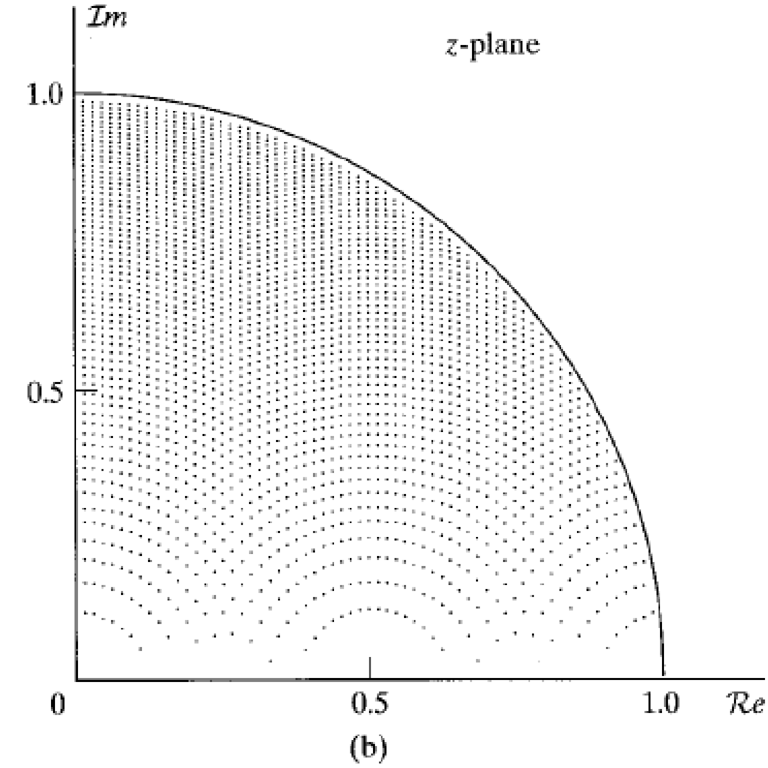


(a)

$$r^2 \in \left\{0, \frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \frac{4}{8}, \frac{5}{8}, \frac{6}{8}, \frac{7}{8}\right\}$$

$$r \in \left\{0, \frac{1}{\sqrt{8}}, \frac{1}{2}, \sqrt{\frac{3}{8}}, \frac{1}{\sqrt{2}}, \sqrt{\frac{5}{8}}, \frac{\sqrt{3}}{2}, \sqrt{\frac{7}{8}}\right\}$$

$$r\cos\theta \in \left\{0, \frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \frac{4}{8}, \frac{5}{8}, \frac{6}{8}, \frac{7}{8}\right\}$$



(b)

Figure 6.50 Pole-locations for the 2nd-order IIR direct form system of Figure 6.49. (a) Four-bit quantizatio coefficients. (b) Seven-bit quantiza

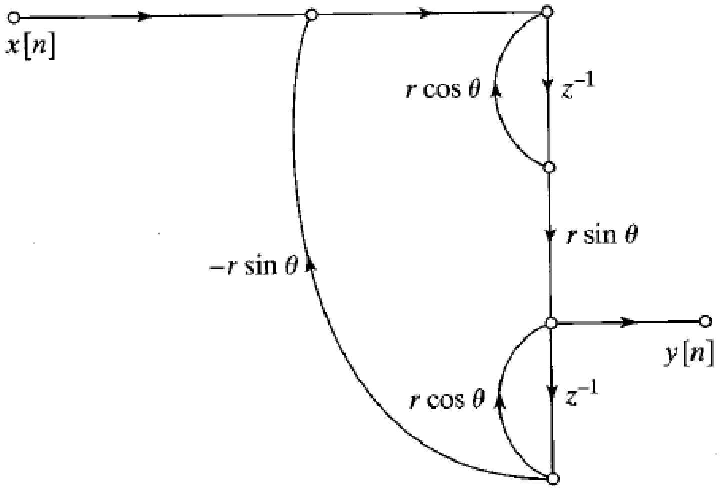


Figure 6.51 Coupled form implementation of a complex-conjugate pole pair.

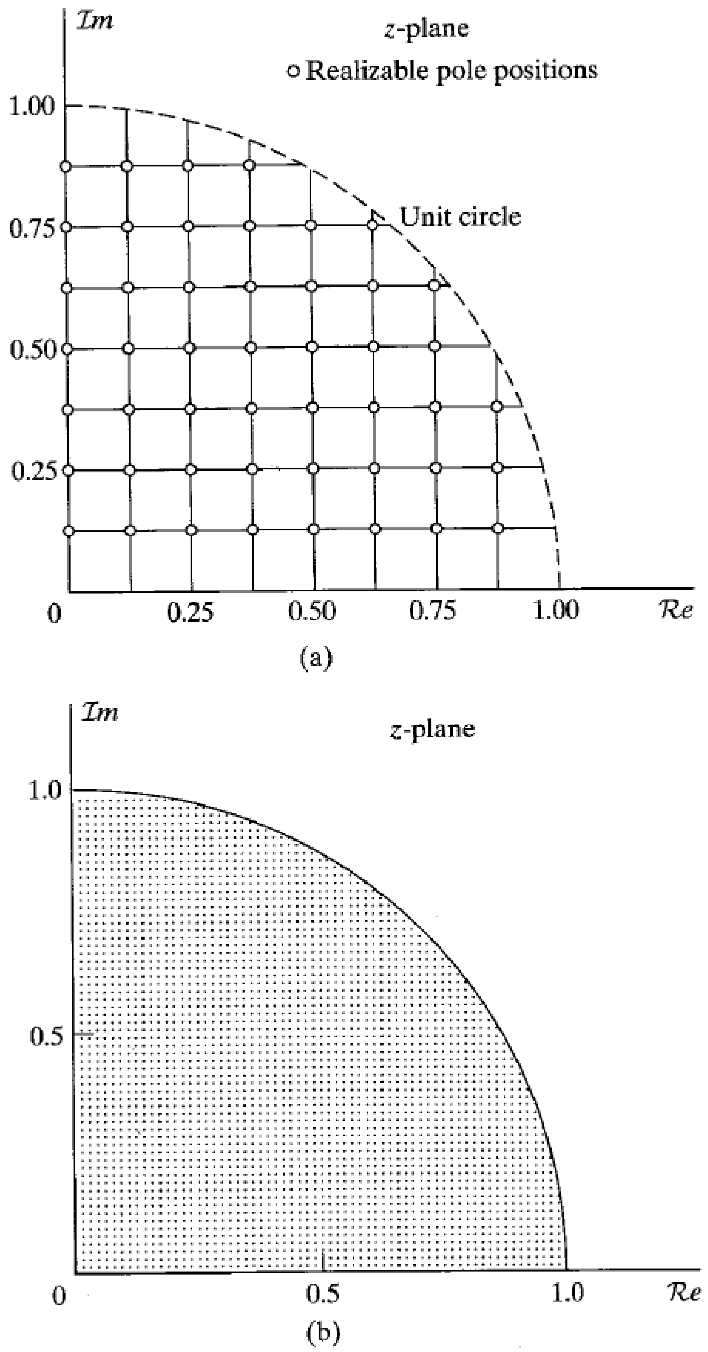


Figure 6.52 Pole locations for coupled form 2nd-order IIR system of Figure 6.51. (a) Four-bit quantization of coefficients. (b) Seven-bit quantization

TABLE 6.1 UNQUANTIZED DIRECT-FORM COEFFICIENTS FOR A 12TH-ORDER ELLIPTIC FILTER

k	b_k	a_k
0	0.01075998066934	1.00000000000000
1	-0.05308642937079	-5.22581881365349
2	0.16220359377307	16.78472670299535
3	-0.34568964826145	-36.88325765883139
4	0.57751602647909	62.39704677556246
5	-0.77113336470234	-82.65403268814103
6	0.85093484466974	88.67462886449437
7	-0.77113336470234	-76.47294840588104
8	0.57751602647909	53.41004513122380
9	-0.34568964826145	-29.20227549870331
10	0.16220359377307	12.29074563512827
11	-0.05308642937079	-3.53766014466313
12	0.01075998066934	0.62628586102551

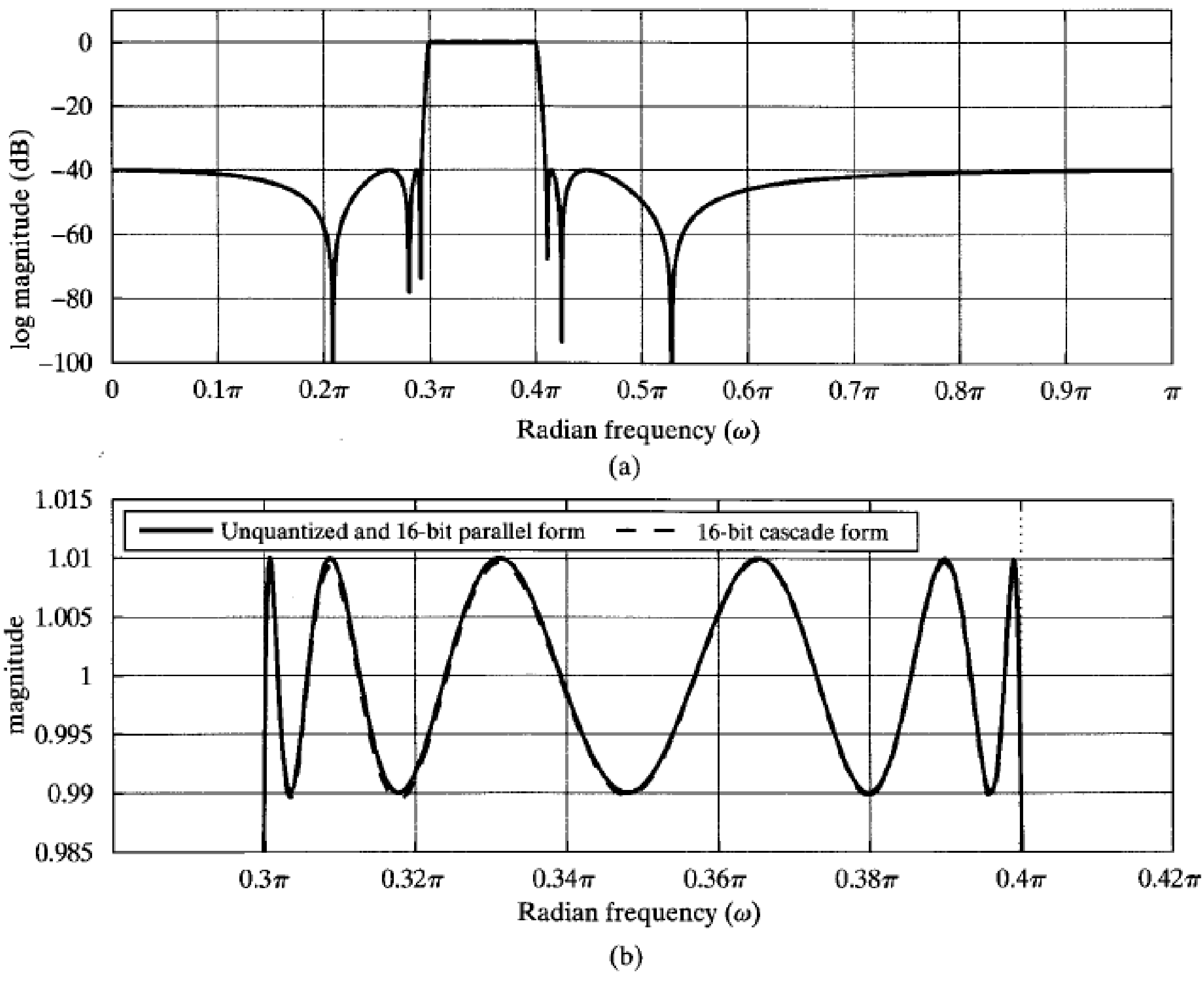


Figure 6.47 IIR coefficient quantization example. (a) Log magnitude for unquantized elliptic bandpass filter. (b) Magnitude in passband for unquantized (solid line) and 16-bit quantized cascade form (dashed line).

TABLE 6.2 ZEROS AND POLES OF UNQUANTIZED 12TH-ORDER ELLIPTIC FILTER.

k	$\lvert c_k \rvert$	$\angle c_k$	$\lvert d_k \rvert$	$\angle d_{1k}$
1	1.0	$\pm$ 1.65799617112574	0.92299356261936	$\pm$1.15956955465354
2	1.0	$\pm$ 0.65411612347125	0.92795010695052	$\pm$1.02603244134180
3	1.0	$\pm$ 1.33272553462313	0.96600955362927	$\pm$1.23886921536789
4	1.0	$\pm$ 0.87998582176421	0.97053510266510	$\pm$0.95722682653782
5	1.0	$\pm$ 1.28973944928129	0.99214245914242	$\pm$1.26048962626170
6	1.0	$\pm$ 0.91475122405407	0.99333628602629	$\pm$0.93918174153968

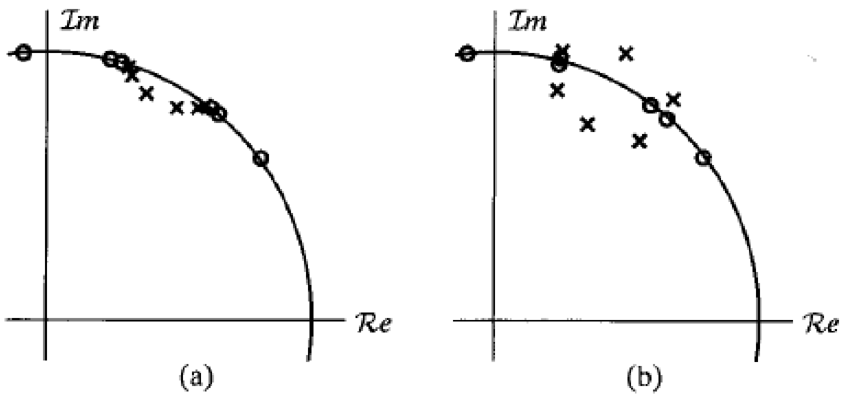


Figure 6.48 IIR coefficient quantization example. (a) Poles and zeros of $H(z)$ for unquantized coefficients. (b) Poles and zeros for 16-bit quantization of the direct form coefficients.

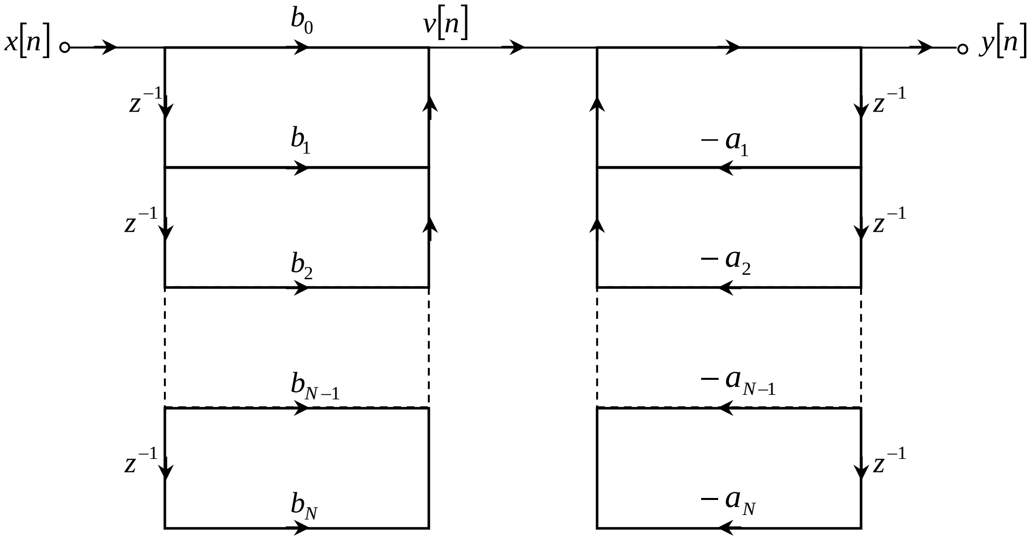
DIRECT FORM-I, DIRECT FORM-II

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_M z^{-M}}{1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_N z^{-N}}$$

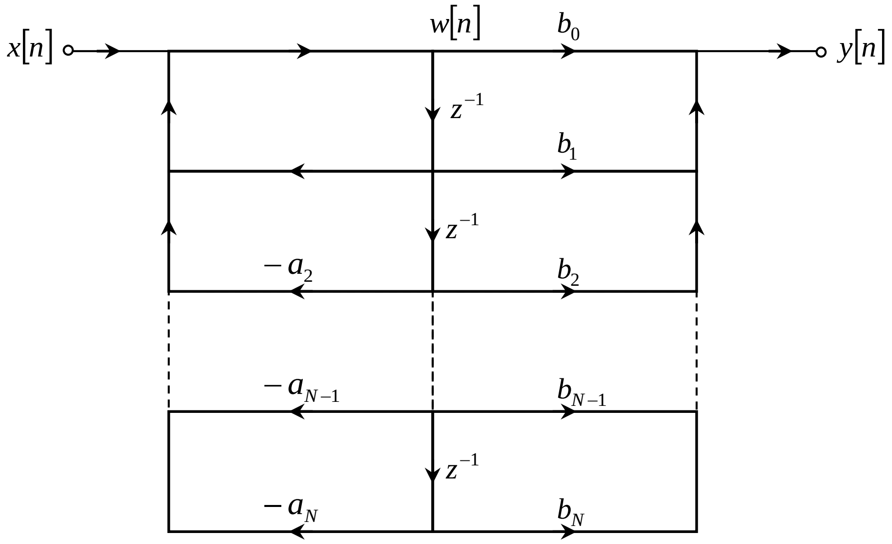
$$= \frac{\sum_{k=0}^{M} b_k z^{-k}}{\sum_{k=0}^{N} a_k z^{-k}}$$

with $a_0 = 1$,

Direct Form - I



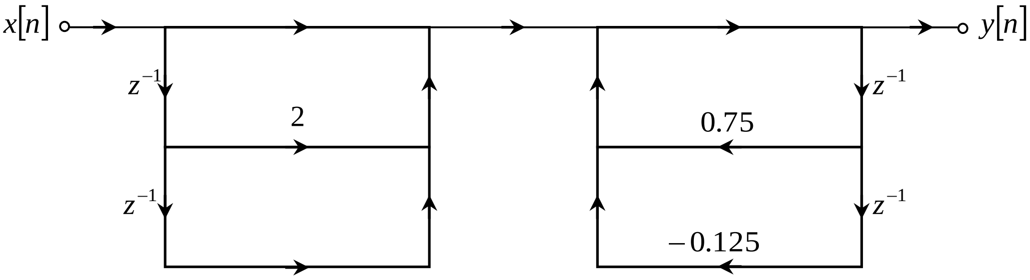
Direct Form - II



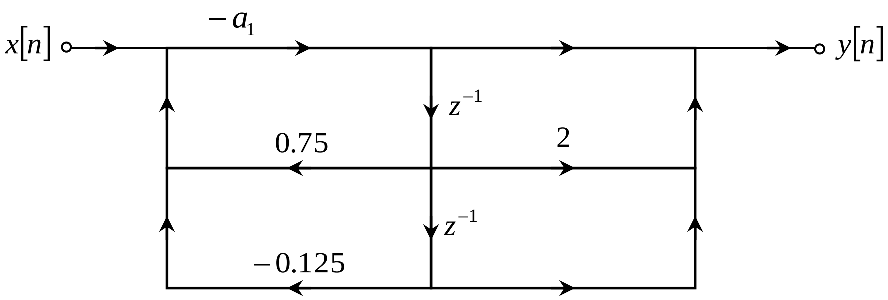
Ex:

$$H(z)=\frac{1+2z^{-1}+z^{-2}}{1-0.75z^{-1}+0.125z^{-2}}$$

Direct Form - I



Direct Form - II



CASCADE FORM

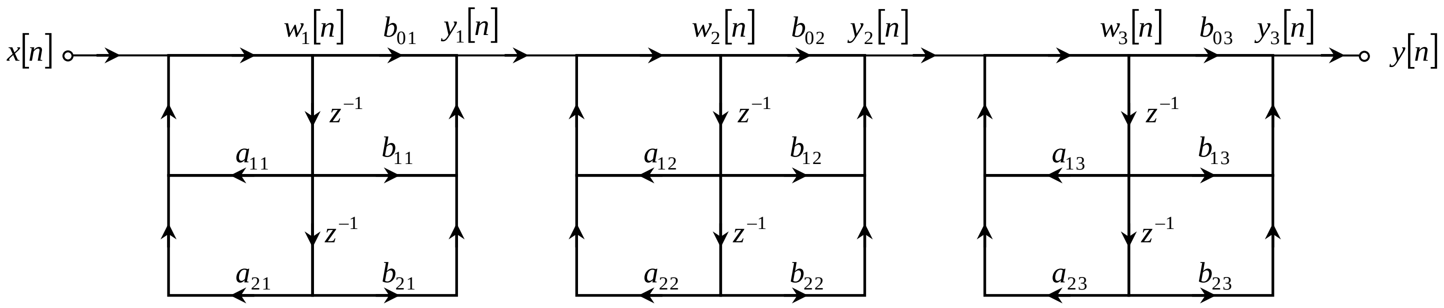
$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_M z^{-M}}{1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_N z^{-N}}$$

$$= A \frac{\prod_{k=1}^{M_1} \left(1 - f_k z^{-1}\right) \prod_{k=1}^{M_2} \left(1 - g_k z^{-1}\right)\left(1 - g_k^* z^{-1}\right)}{\prod_{k=1}^{N_1} \left(1 - c_k z^{-1}\right) \prod_{k=1}^{N_2} \left(1 - d_k z^{-1}\right)\left(1 - d_k^* z^{-1}\right)}$$

$$= \prod_{k=1}^{N_S} \frac{b_{0k} + b_{1k} z^{-1} + b_{2k} z^{-2}}{1 - a_{1k} z^{-1} - a_{2k} z^{-2}}$$

Ex: Cascade form of a 6th order system.

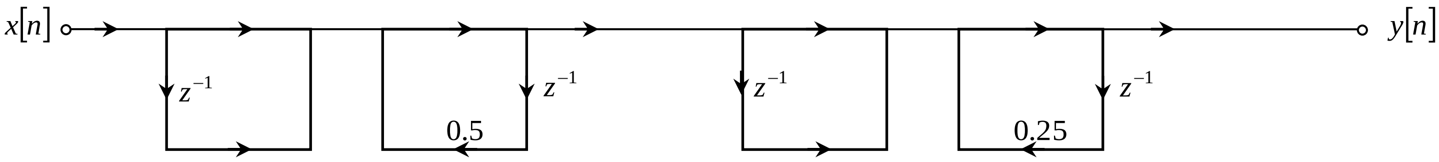
2nd order subsystems have Direct Form-II realizations.



Ex: Cascade form of a 2nd order system.

$$H(z)=\frac{1+2z^{-1}+z^{-2}}{1-0.75z^{-1}+0.125z^{-2}}$$

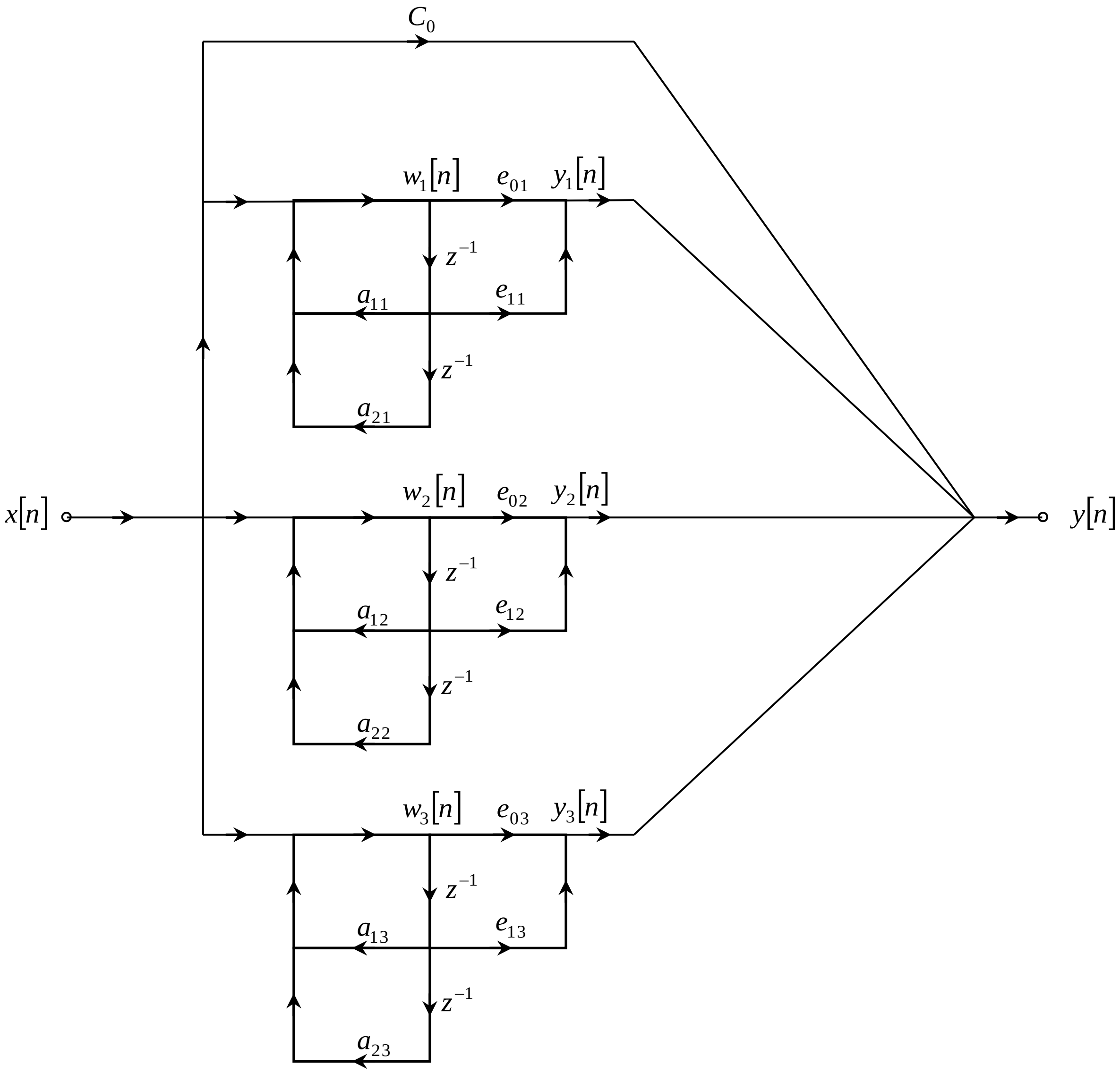
$$=\frac{\left(1+z^{-1}\right)\left(1+z^{-1}\right)}{\left(1-0.5z^{-1}\right)\ \left(1-0.25z^{-1}\right)}$$



PARALLEL FORMS

$$H(z)=\sum_{k=0}^{N_P} C_k z^{-1}+\sum_{k=1}^{N_1}\frac{A_k}{1-c_k z^{-1}}+\sum_{k=1}^{N_2}\frac{B_k\left(1-e_k z^{-1}\right)}{\left(1-d_k z^{-1}\right)\left(1-d_k^* z^{-1}\right)}$$

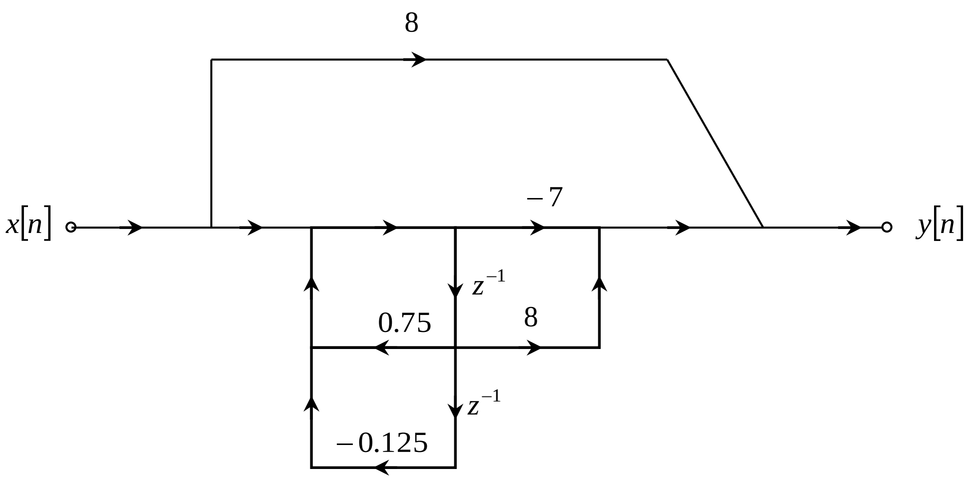
$$N_P=M-N \qquad M=M_1+2M_2 \qquad N=N_1+2N_2$$


C_0
$w_1[n]$
e_{01}
$y_1[n]$
z^{-1}
a_{11}
e_{11}
z^{-1}
a_{21}
$x[n]$
$w_2[n]$
e_{02}
$y_2[n]$
$y[n]$
z^{-1}
a_{12}
e_{12}
z^{-1}
a_{22}
$w_3[n]$
e_{03}
$y_3[n]$
z^{-1}
a_{13}
e_{13}
z^{-1}
a_{23}

Ex:

$$H(z) = \frac{1 + 2z^{-1} + z^{-2}}{1 - 0.75z^{-1} + 0.125z^{-2}}$$

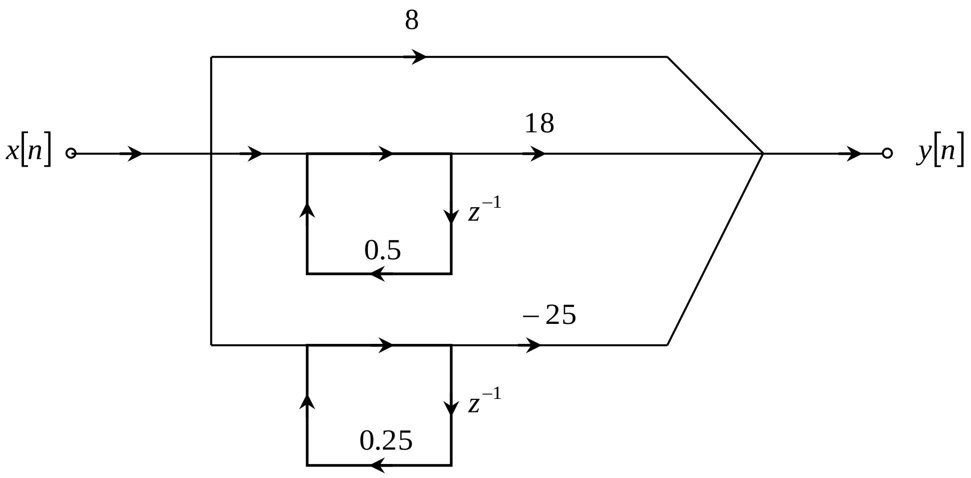
$$= 8 + \frac{-7 + 8z^{-1}}{1 - 0.75z^{-1} + 0.125z^{-2}}$$



Ex:

$$H(z) = \frac{1 + 2z^{-1} + z^{-2}}{1 - 0.75z^{-1} + 0.125z^{-2}}$$

$$= 8 + \frac{18}{1 - 0.5z^{-1}} - \frac{25}{1 - 0.25z^{-1}}$$



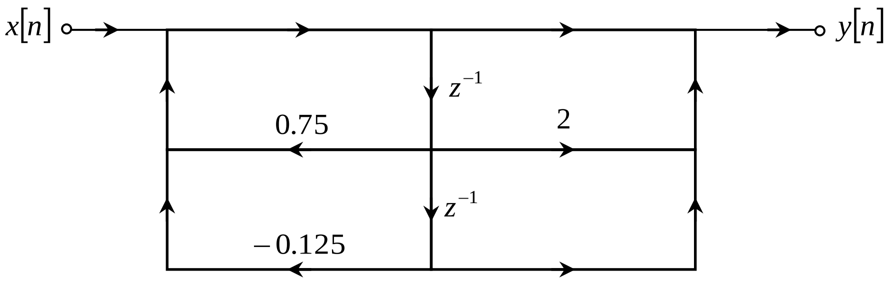
TRANSPOSED FORMS

For a single input, single output (SISO) linear flow graph: "Reverse all branch directions, interchange the input and output node assignments, keep transmittences the same, then the system function remains unchanged"

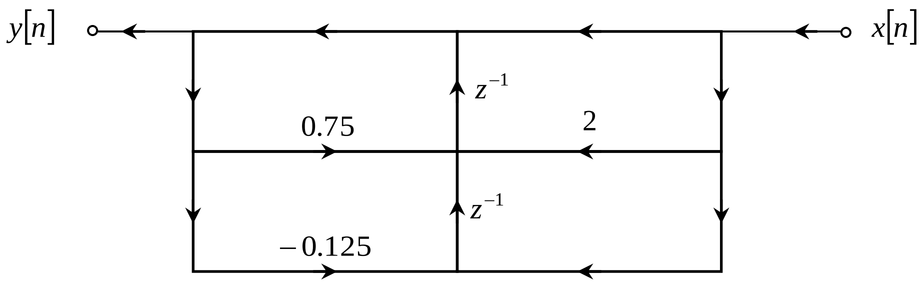
Ex:

$$H(z)=\frac{1+2z^{-1}+z^{-2}}{1-0.75z^{-1}+0.125z^{-2}}$$

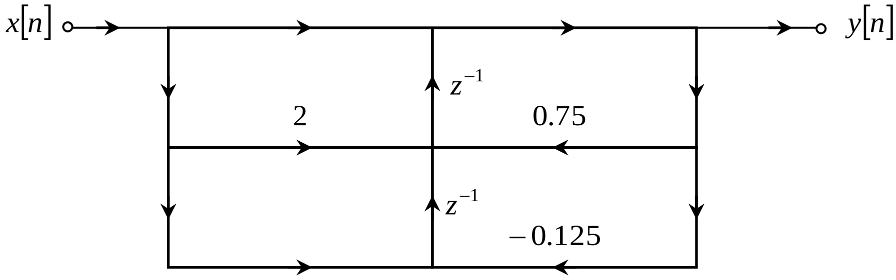
Direct Form II



Transposed Direct Form II

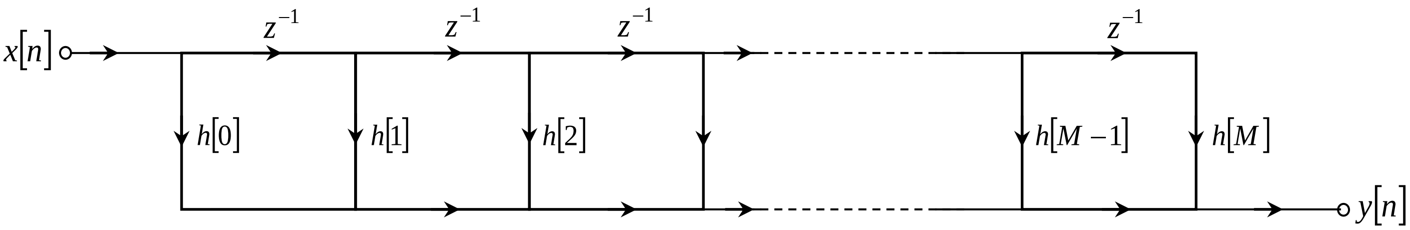


Transposed Direct Form II, redrawn.

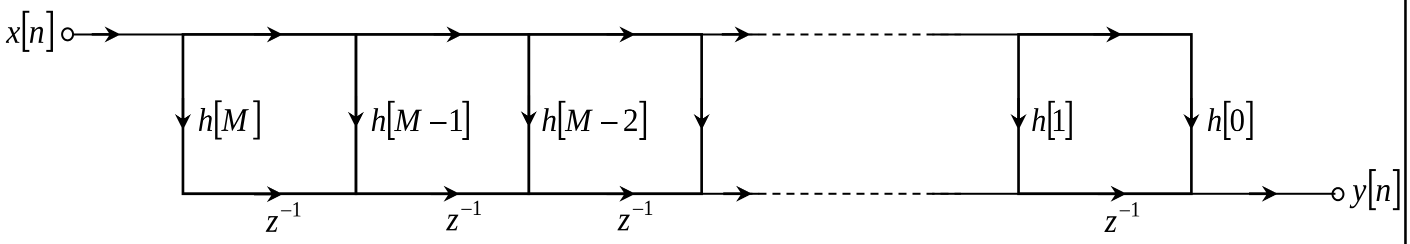


$$y[n] = \sum_{k=0}^{M} h[k]\; x[n-k]$$

Direct Form

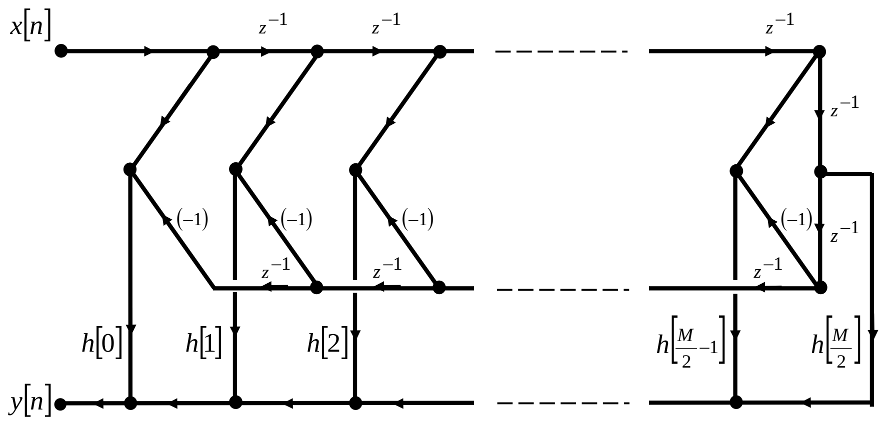


Transposed Direct Form



Odd Length Filters (Type-I and Type-III)

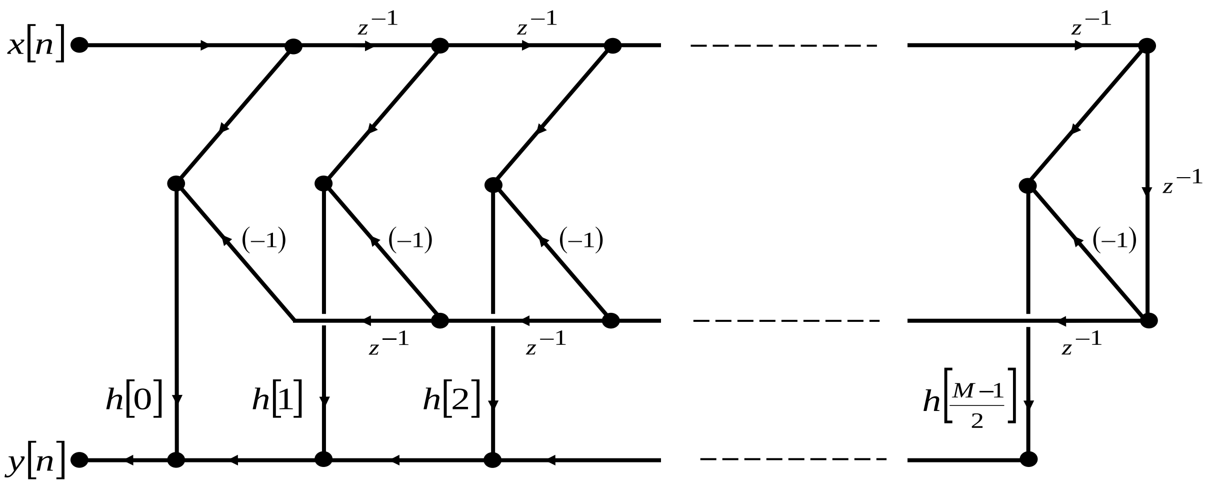
M: even (filter order)



Note that $h\left[\frac{M}{2}\right]=0$ for Type-III filters!

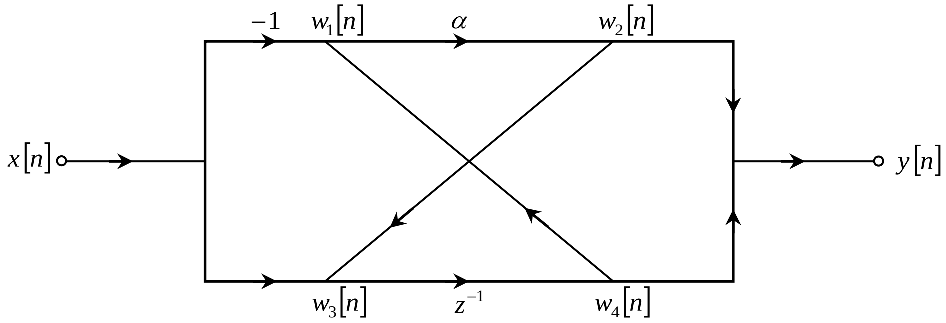
-1 multiplications in parentheses are for Type-III (odd symmetry) filters!

M: odd (filter order)



-1 multiplications in parentheses are for Type-IV (odd symmetry) filters!

DETERMINATION OF THE SYSTEM FUNCTION FROM A FLOW GRAPH



$$w_1[n]=w_4[n]-x[n]$$

$$W_1(z)=W_4(z)-X(z) \qquad \text{(a)}$$

$$W_2(z)=\alpha W_1(z) \qquad \text{(b)}$$

$$W_3(z)=W_2(z)+X(z) \qquad \text{(c)}$$

$$W_4(z)=z^{-1}W_3(z) \qquad \text{(d)}$$

$$Y(z)=W_2(z)+W_4(z) \qquad \text{(e)}$$

a → b $$W_2(z)=\alpha\left(W_4(z)-X(z)\right) \qquad \text{(f)}$$

c → d $$W_4(z)=z^{-1}\left(W_2(z)+X(z)\right) \qquad \text{(g)}$$

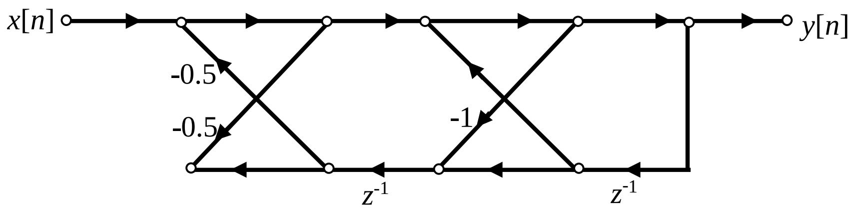
f,g $$W_2(z)=\frac{\alpha\left(z^{-1}-1\right)}{1-\alpha z^{-1}}X(z) \qquad \text{(h)}$$

f,g $$W_4(z)=\frac{z^{-1}\left(1-\alpha\right)}{1-\alpha z^{-1}}X(z) \qquad \text{(i)}$$

h,i → e
$$Y(z)=\frac{\alpha\left(z^{-1}-1\right)+z^{-1}\left(1-\alpha\right)}{1-\alpha z^{-1}}X(z)$$

$$=\frac{z^{-1}-\alpha}{1-\alpha z^{-1}}X(z)$$

Ex: **a)** Given the following flow graph of an LTI filter, determine its transfer function $H(z)$.



b) Plot the Direct Form II structure for the filter $H_1(z)=(1-2z^{-1})H(z)$, where $H(z)$ is the filter in part-a.

Ex: Consider the following system function with real valued coefficients

$$H(z) = \frac{b_2 + b_1 z^{-1} + z^{-2}}{1 + b_1 z^{-1} + b_2 z^{-2}}$$

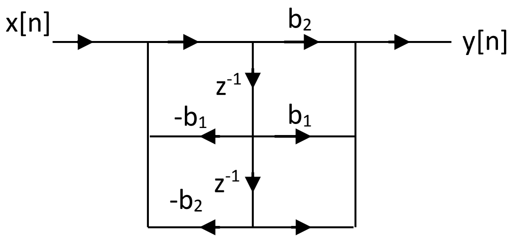
a) Find and plot the direct form II structure for $H(z)$. Determine the number of multiplications, additions and delay terms.

b) Find and plot the signal flow graph of a new filter structure such that there are two multiplications only. You can have more delay terms than those in part a. (multiplication by 1 or -1 does not count).

a) num. of multiplications=4

num. of additions=4

num. of delay terms = 2



b)