Open access • Proceedings Article • DOI:10.1109/ISCAS.1990.112711

Multiplier-free decimator algorithms for superresolution oversampled converters

Tapio Saramaki, T. Karema, T. Ritoniemi, Hannu Tenhunen

Institutions: Tampere University of Technology

Published on: 01 May 1990 - International Symposium on Circuits and Systems

Topics: Decimation, Oversampling, Digital filter and Finite impulse response

MULTIPLIER-FREE HALF-BAND FILTERS

- This pile of lecture notes shows how to design half-band FIR filters without general multipliers.
- These filters can be used as building blocks for constructing multiplier-free superresolution decimators and interpolators.
- They have been used in the article (a copy of this article as well as the conference talk are included):
- T. Saramäki, T. Karema, T. Ritoniemi, and H. Tenhunen, "Multiplier-free decimator algorithms for superresolution oversampled converters," in *Proc.* 1990 *IEEE International Symposium on Circuits and Systems* (New Orleans, Louisiana), pp. 3275–3278, May 1990.

What Are Half-Band FIR Filters?

- For a half-band FIR filter, the transfer function is of the form

$$H(z) = \sum_{n=0}^{2M} h[n]z^{-n}, \quad h[2M-n] = h[n],$$

where M is odd.

- For these filters,

$$h[M] = 1/2$$

$$h[M+2r] = 0 \quad \text{for} \quad r = \pm 1, \pm 2, \ldots, \pm(M-1)/2.$$

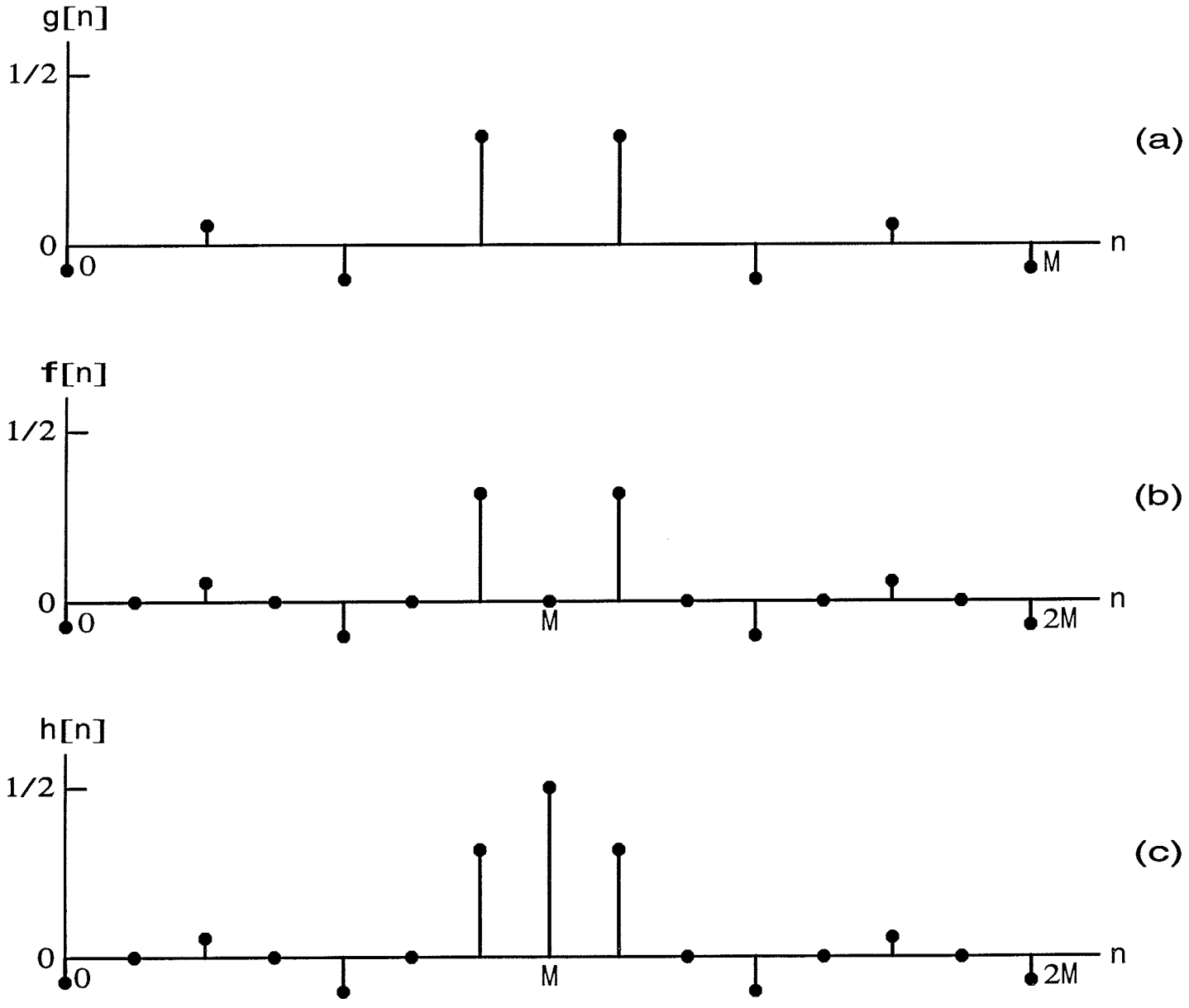
- A filter satisfying these conditions can be generated in two steps by starting with a Type II (M is odd) transfer function

$$G(z) = \sum_{n=0}^{M} g[n]z^{-n}, \quad g[n] = g[M-n].$$

- In the first step, zero-valued impulse-response values are inserted between the $g[n]$'s [see Figures (a) and (b) in the following transparency], giving the following Type I transfer function of order $2M$:

$$F(z) = \sum_{n=0}^{2M} f[n]z^{-n} = G(z^2) = \sum_{n=0}^{M} g[n]z^{-2n}.$$

Generation of the Impulse Response of a Half-Band Filter



- The second step is then to replace the zero-valued impulse-response value at $n = M$ by $1/2$ [see Figure (c) in the previous transprency], resulting in the desired transfer function

$$H(z) = \sum_{n=0}^{2M} h[n]z^{-n} = \frac{1}{2}z^{-M} + F(z) = \frac{1}{2}z^{-M} + \sum_{n=0}^{M} g[n]z^{-2n}.$$

- This gives $h[M] = 1/2$, $h[n] = g[n/2]$ for n even, and $h[n] = 0$ for n odd and $n \neq M$, as is desired.

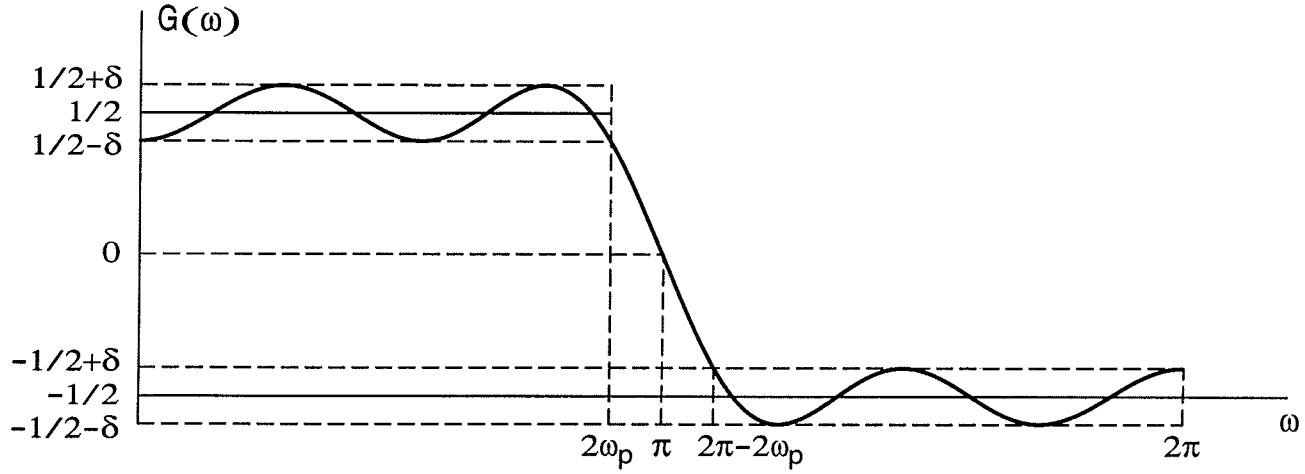
Filter Design

- The zero-phase frequency responses of $H(z)$, $F(z)$, and $G(z)$ are related through

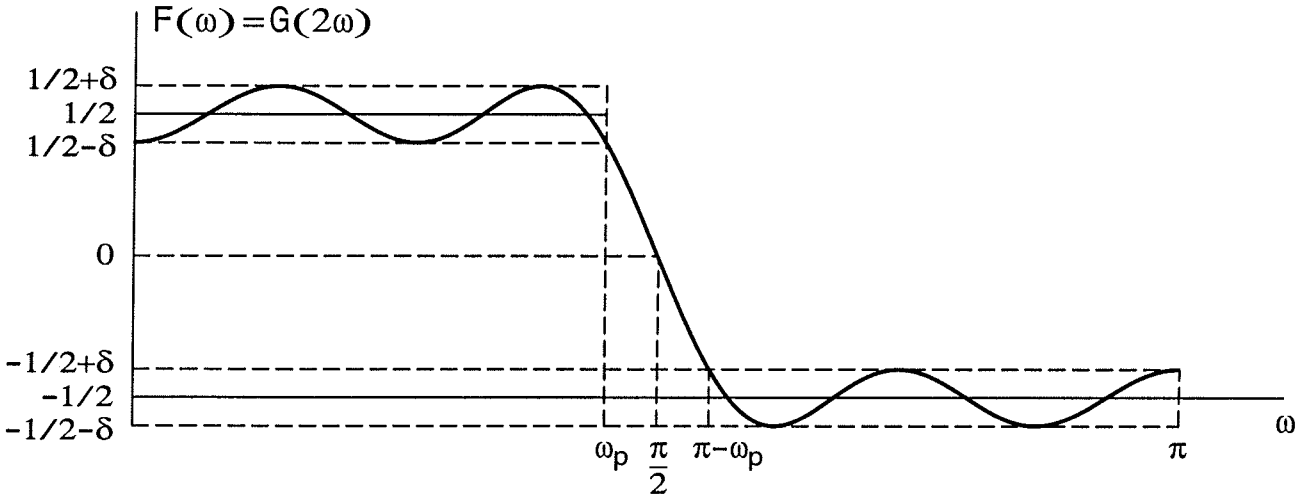
$$H(\omega) = 1/2 + F(\omega) = 1/2 + G(2\omega).$$

- Based on these relations, the design of a low-pass half-band filter with passband edge at ω_p and passband ripple of δ can be accomplished by determining $G(z)$ such that $G(\omega)$ oscillates within $1/2 \pm \delta$ on $[0, 2\omega_p]$ [see Figure (a) in the following transprency].
- Since $G(z)$ is a Type II transfer function, it has one fixed zero at $z = -1$ $(\omega = \pi)$.
- $G(z)$ can be designed directly with the aid of the Remez algorithm using only one band $[0, 2\omega_p]$, $D(\omega) = 1/2$, and $W(\omega) = 1$.
- Since $G(z)$ has a single zero at $z = -1$, $G(\omega)$ is odd about $\omega = \pi$.
- Hence, $G(2\pi - \omega) = -G(\omega)$ and $G(\omega)$ oscillates within $-1/2 \pm \delta$ on $[2\pi - 2\omega_p, 2\pi]$.

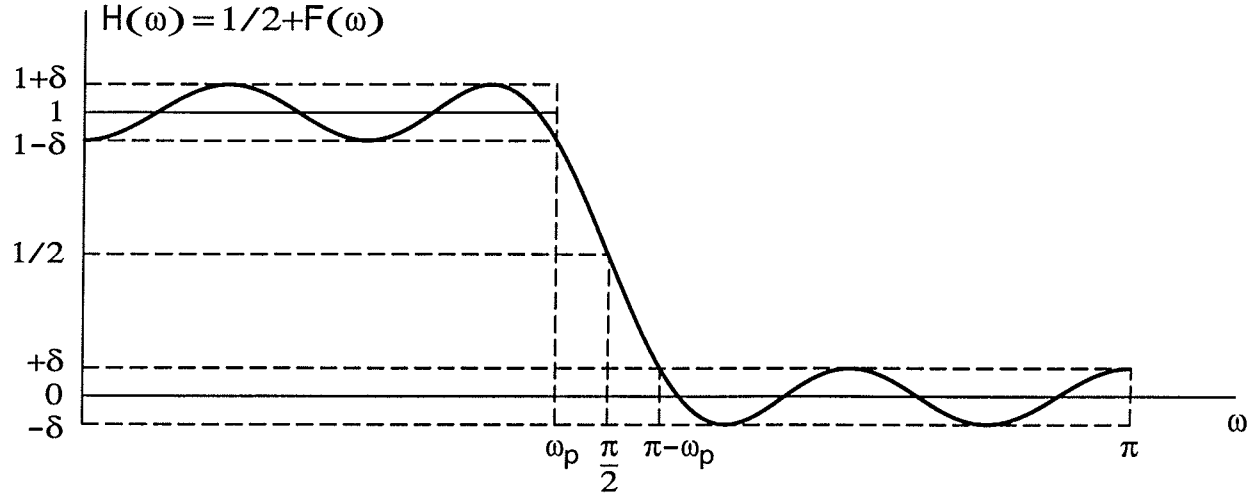
Design of A Lowpass Half-Band Filter



(a)



(b)



(c)

- The corresponding $F(\omega) = G(2\omega)$ stays within $1/2 \pm \delta$ on $[0, \omega_p]$ and within $-1/2 \pm \delta$ on $[\pi - \omega_p, \pi]$ [see Figure (b) in the previous transparency].
- Finally, $H(\omega)$ approximates unity on $[0, \omega_p]$ with tolerance δ and zero on $[\pi - \omega_p, \pi]$ with the same tolerance δ [see Figure (c) in the previous transparency].
- For the resulting $H(\omega)$, the passband and stopband ripples are thus the same and the passband and stopband edges are related through $\omega_s = \pi - \omega_p$.
- In general, $H(\omega)$ satisfies

$$H(\omega) + H(\pi - \omega) = 1.$$

- This makes $H(\omega)$ symmetric about the point $\omega = \pi/2$ such that the sum of the values $H(\omega)$ at $\omega = \omega_0 < \pi/2$ and at $\omega = \pi - \omega_0 > \pi/2$ is equal to unity [see Figure (c) in the previous transparency].

Subfilter Approach for Designing Half-Band Filters

- According to the previous discussion, a half-band filter transfer function of order $2M$ with M odd is expressible as

$$H(z) = \frac{1}{2}z^{-M} + G(z^2),$$

where $G(z)$ is a Type II transfer function of odd order M (having allways a zero at $z = -1$).

- Furthermore, the design of $H(z)$ in such a way that $H(\omega)$ approximates on $[0, \omega_p]$ unity with deviation δ and on $[\pi - \omega_p, \pi]$ zero with the same deviation δ can be converted into the design of $G(z)$ such that $G(\omega)$ approximates $1/2$ on $[0, 2\omega_p]$ with deviation δ.

- In order to construct such a transfer function without general multipliers also for a small value of δ, we generate $G(z)$ as follows:

$$G(z) = \sum_{l=0}^{L} a_l z^{(L-l)K} [F(z)]^{2l+1}, \qquad (A)$$

where

$$F(z) = \sum_{n=0}^{K} f[n] z^{-n}, \quad f[K-n] = f[n].$$

- Here, $F(z)$ is a Type II transfer function of odd order K.
- The order of $G(z)$ is thus $(2L+1)K$ and the delay of each term in the summation of Eq. (A) is $(2L+1)K/2$, as is desired to guarantee the linear-phase performance.
- Efficient implementations of the proposed overall filter for decimation and interpolation purposes are depicted in the following transparency.
- The zero-phase frequency response of the above $G(z)$ is expressible as

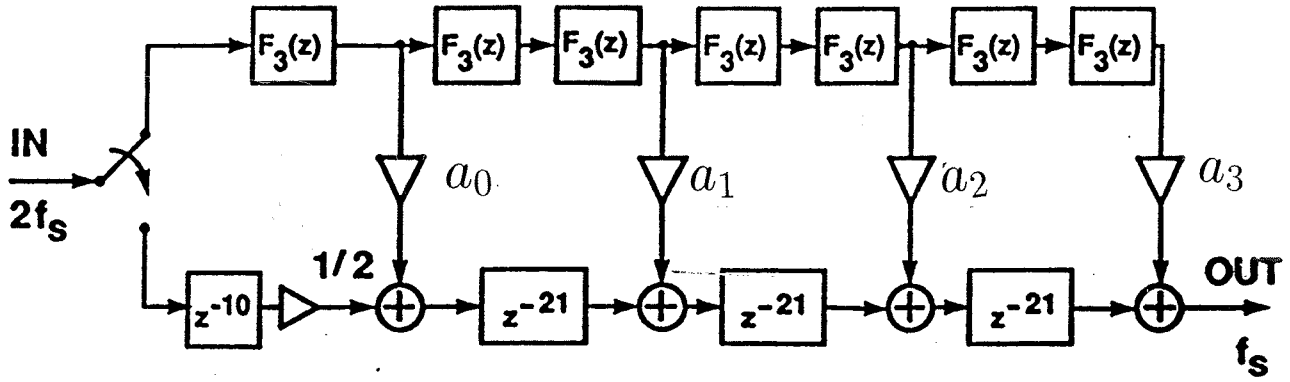
$$G(\omega) = \sum_{l=0}^{L} a_l [F(\omega)]^{2l+1},$$

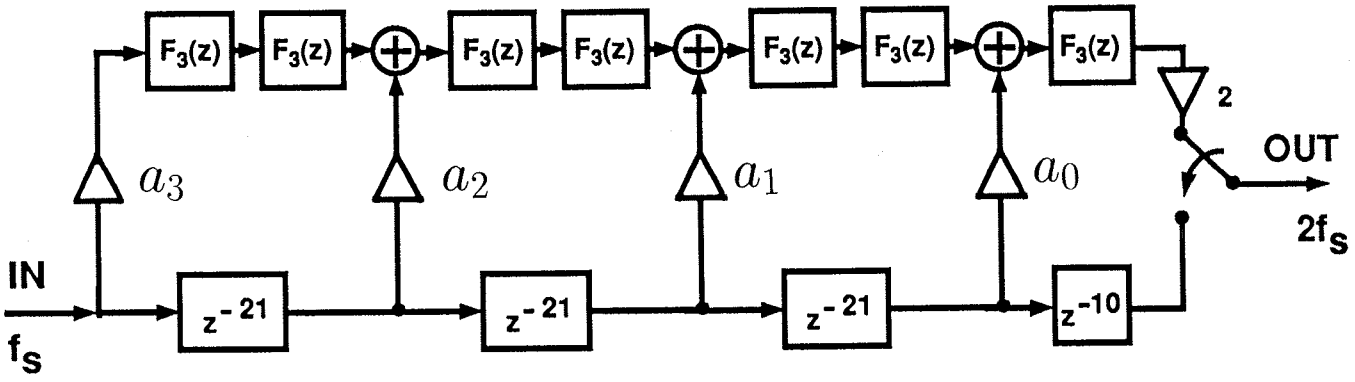
where

$$F(\omega) = 2 \sum_{n=0}^{(K-1)/2} f[(K-1)/2 - n] \cos[(n+1/2)\omega].$$

Implentations of the Proposed Half-Band Filters for Sampling Rate Alteration by a Factor of Two

- The first and second figures show the decimator and interpolator structures for $L = 3$ and $K = 21$.
- These are commutative structures where delay terms have been shared $(F(z) \equiv F_3(z))$.





Simple Design Technique

- Here, we concentrate on the cases with $L = 1$, $L = 2$, and $L = 3$.
- For $L = 3$,

$$G(z) = a_0 F(z) + a_1[F(z)]^3 + a_2[F(z)]^5 + a_3[F(z)]^7$$

and

$$G(\omega) = a_0 F(\omega) + a_1[F(\omega)]^3 + a_2[F(\omega)]^5 + a_3[F(\omega)]^7.$$

- For $L = 2$, $a_3 \equiv 0$ and for $L = 1$, $a_3 = a_2 \equiv 0$.
- Assume that $F(\omega)$ oscillates within $1 - \epsilon_1$ and $1 + \epsilon_2$ with $\epsilon_1 > 0$ and $\epsilon_2 > 0$ on $[0,\ 2\omega_p]$.
- Then, we state the following problem: Given L and δ, find the adjustable parameters a_l as well as ϵ_1 and ϵ_2 such that $G(\omega)$ oscillates within $1/2 \pm \delta$ on $[0,\ 2\omega_p]$.
- If the value of $F(\omega)$ is $1 + \epsilon$, where ϵ is either positive or negative, then the corresponding value of $G(\omega)$ is

$$G(\omega) = a_0[1 + \epsilon] + a_1[1 + \epsilon]^3 + a_2[1 + \epsilon]^5 + a_3[1 + \epsilon]^7,$$

where

$$[1+\epsilon]^3 = 1 + 3\epsilon + \epsilon^2,$$

$$[1+\epsilon]^5 = 1 + 5\epsilon + 10\epsilon^2 + 10\epsilon^3 + \epsilon^4,$$

and

$$[1+\epsilon]^7 = 1 + 7\epsilon + 21\epsilon^2 + 35\epsilon^3 + 35\epsilon^4 + 21\epsilon^5 + 7\epsilon^6 + \epsilon^7$$

- For $L = 1$, the selection

$$a_0 = 3/4, \quad a_1 = -1/4$$

gives

$$G(\omega) = 1/2 + \Delta,$$

where

$$\Delta = -(3/4)\epsilon^2 + \epsilon^3,$$

that is, the constant coefficient is equal to $1/2$ and the coefficient of ϵ is zero.

- For $L = 2$, the selection

$$a_0 = 15/16, \quad a_1 = -10/16, \quad a_2 = 3/16$$

gives

$$G(\omega) = 1/2 + \Delta,$$

where

$$\Delta = -(5/4)\epsilon^3 + (15/16)\epsilon^4 + (3/16)\epsilon^5,$$

that is, the constant coefficient is equal to $1/2$ and the coefficients of ϵ and ϵ^2 are zero.

- For $L = 3$, the selection

$$a_0 = 35/32, \;\; a_1 = -35/32, \;\; a_2 = 21/32, \;\; a_3 = -5/32$$

gives

$$G(\omega) = 1/2 + \Delta,$$

where

$$\Delta = -(35/16)\epsilon^4 - (21/8)\epsilon^5 - (35/32)\epsilon^6 - (5/32)\epsilon^7,$$

that is, the constant coefficient is equal to $1/2$ and the coefficients of ϵ, ϵ^2, and ϵ^3 are zero.

- In all the above cases, the variation of $G(\omega)$ around $1/2$, denoted by Δ, is significantly smaller than the variation of $F(\omega)$ around unity, ϵ.
- In the following, we consider two cases. In the first case, referred to as Case A, it is required that $-0.001 \leq \Delta \leq 0.001$. In the second

case, referred to as Case B, it is required that $-0.000001 \leq \Delta \leq 0.000001$.

- In Cases A and B, the stopband attenuations of the overall half-band filters are 60 dB and 120 dB, respectively.

$L = 1,\ a_0 = 2^{-1} + 2^{-2},\ a_1 = -2^{-2}$

- In this case, the deviation of $G(\omega)$ from 1/2, denoted by Δ, is related to the deviation of $F(\omega)$ from unity, denoted by ϵ, through the equation

$$\Delta = -(3/4)\epsilon^2 + \epsilon^3.$$

- The following two transparencies give plots of the above equation for both Case A ($|\Delta| \leq 0.001$) and Case B ($|\Delta| \leq 0.000001$).
- It is seen that in Case A $F(\omega)$ is allowed to vary within the limits $1 - \epsilon_1$ and $1 + \epsilon_2$, where

$$\epsilon_1 = 0.0367405, \quad \epsilon_2 = 0.0362959,$$

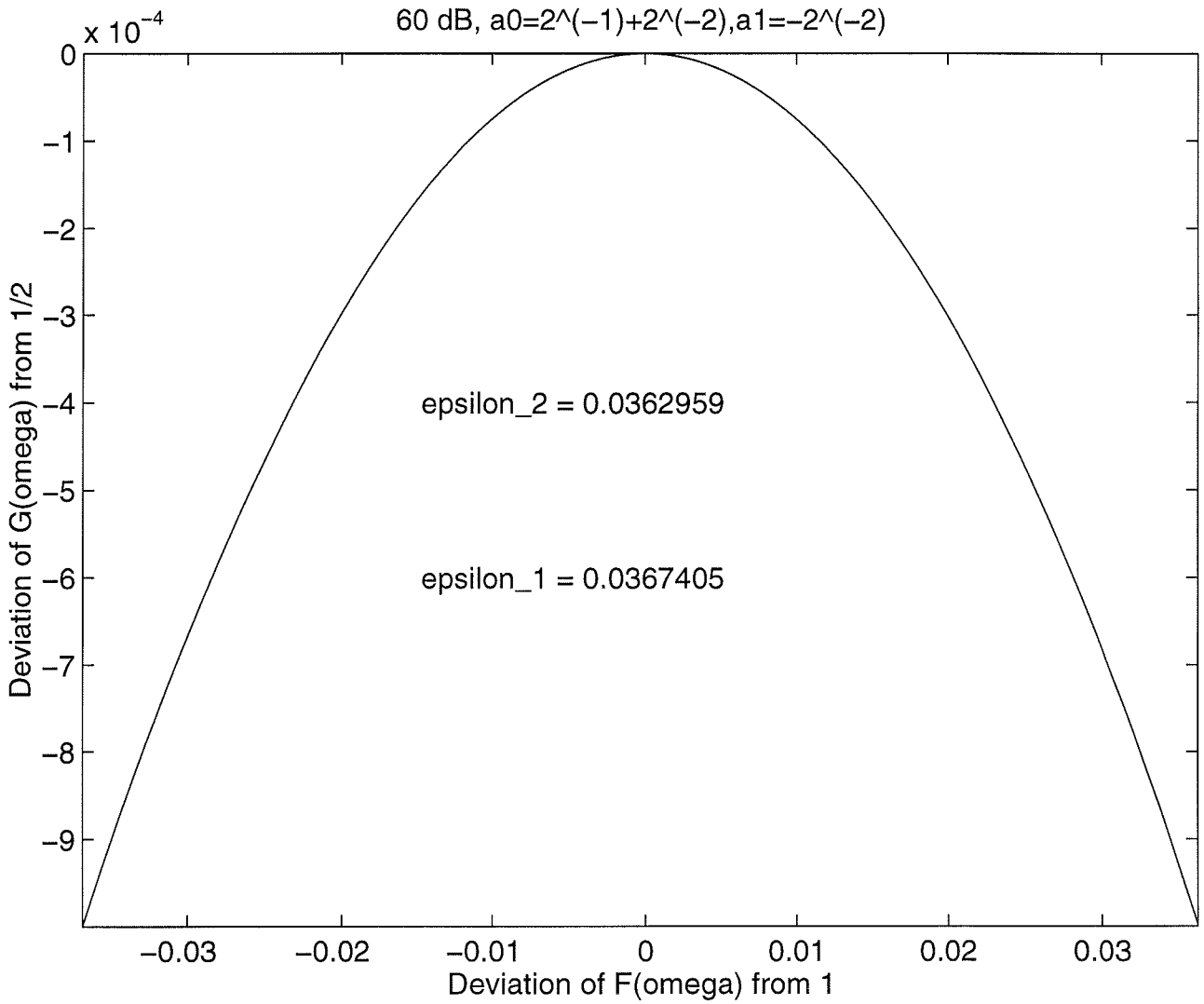
to satisfy $-0.001 \leq \Delta \leq 0.001$.

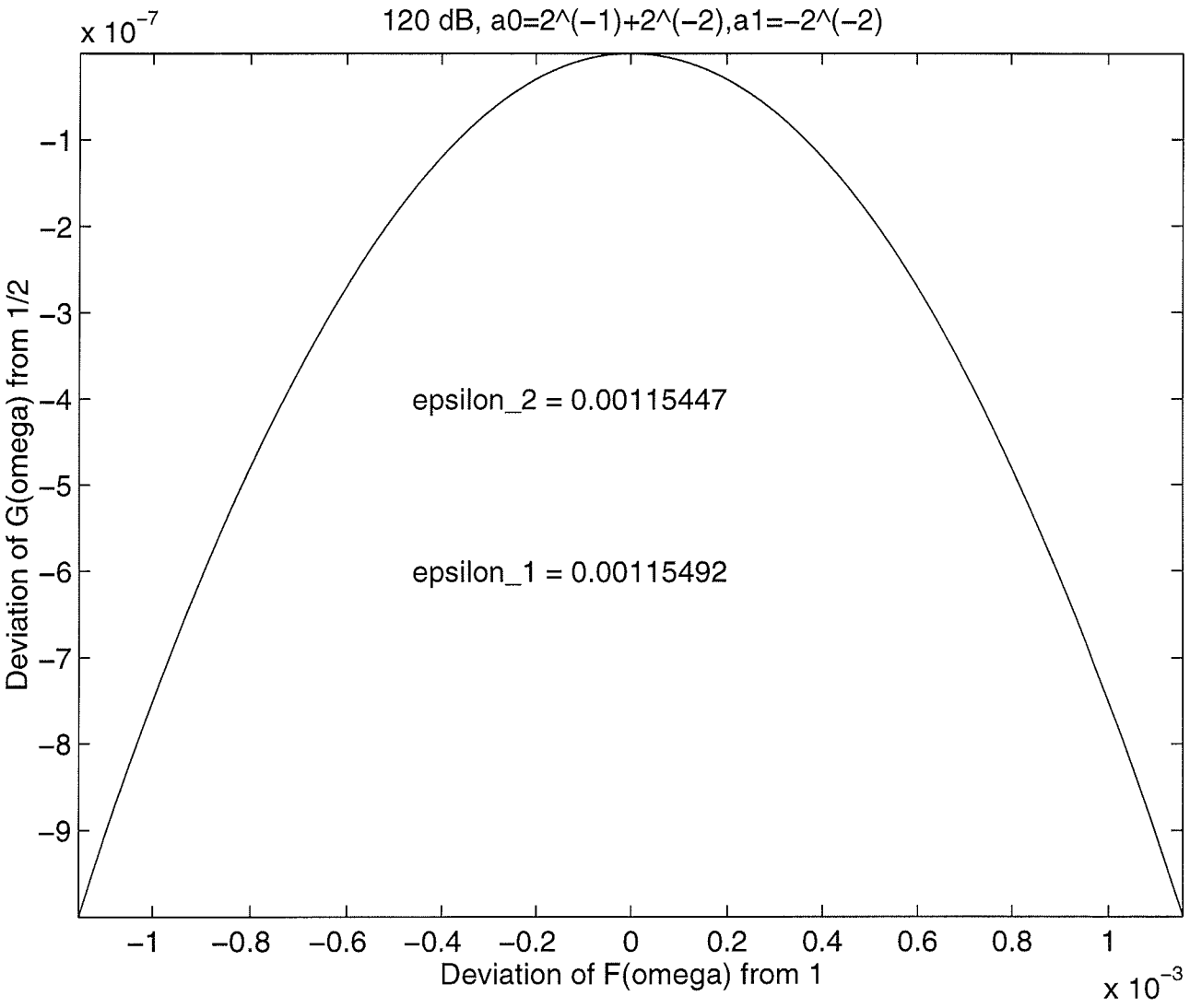
- In Case B, $F(\omega)$ is allowed to vary within the limits $1 - \epsilon_1$ and $1 + \epsilon_2$, where

$$\epsilon_1 = 0.00115492, \quad \epsilon_2 = 0.00115447,$$

to satisfy $-0.000001 \leq \Delta \leq 0.000001$.

- The disadvantage of the above selections of a_0 and a_1 is that the maximum value of Δ is zero.


60 dB, a0=2^(−1)+2^(−2),a1=−2^(−2)
x 10^−4
0
−1
−2
−3
−4
−5
−6
−7
−8
−9
Deviation of G(omega) from 1/2
epsilon_2 = 0.0362959
epsilon_1 = 0.0367405
−0.03
−0.02
−0.01
0
0.01
0.02
0.03
Deviation of F(omega) from 1


120 dB, a0=2^(−1)+2^(−2),a1=−2^(−2)
x 10^−7
−1
−2
−3
−4
−5
−6
−7
−8
−9
Deviation of G(omega) from 1/2
epsilon_2 = 0.00115447
epsilon_1 = 0.00115492
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1
Deviation of F(omega) from 1
x 10^−3

- Better results, that is, both ϵ_1 and ϵ_2 become larger, is obtained by changing a_1. This is considered next.

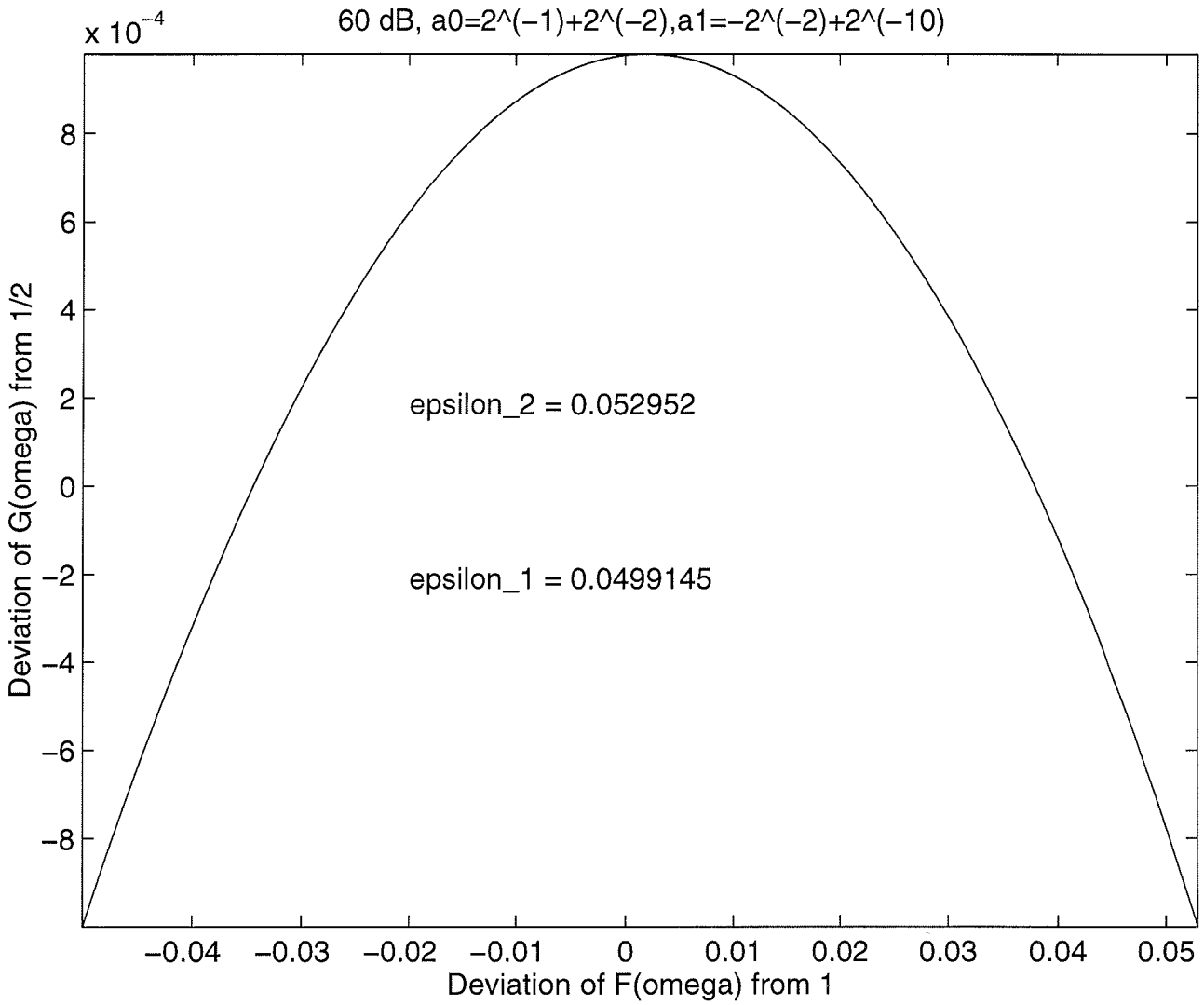
Case A: $L = 1$, $a_0 = 2^{-1} + 2^{-2}$, $a_1 = -2^{-2} + 2^{-10}$

- In this case,

$$\Delta = -(3/4)\epsilon^2 + \epsilon^3 + 2^{-10}(1+\epsilon)^3.$$

- As seen from the following transparency, in this case

$$\epsilon_1 = 0.0499145, \quad \epsilon_2 = 0.0529520.$$


60 dB, a0=2^(−1)+2^(−2),a1=−2^(−2)+2^(−10)
x 10^−4
8
6
4
2
0
−2
−4
−6
−8
Deviation of G(omega) from 1/2
epsilon_2 = 0.052952
epsilon_1 = 0.0499145
−0.04
−0.03
−0.02
−0.01
0
0.01
0.02
0.03
0.04
0.05
Deviation of F(omega) from 1

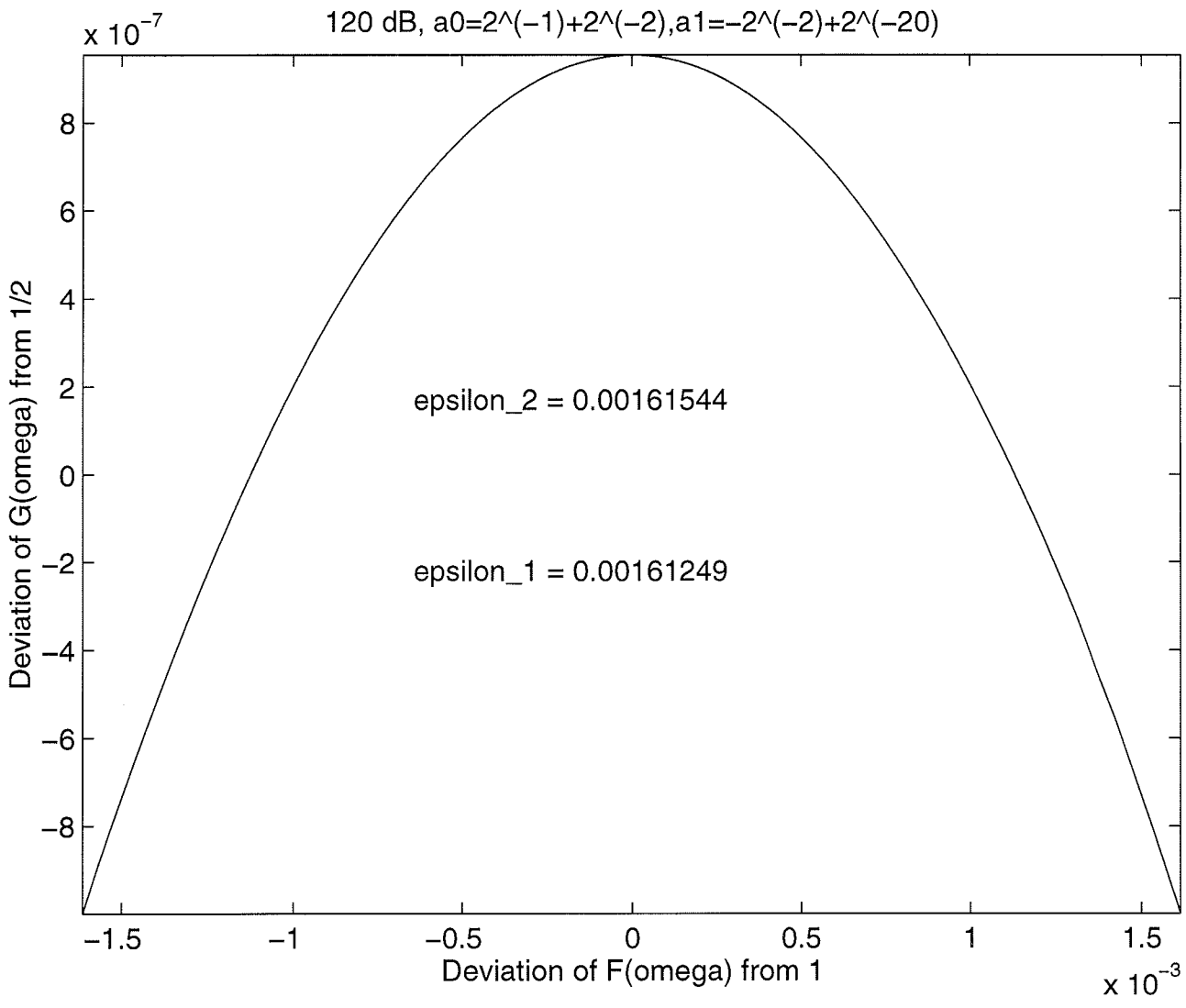
Case B: $L = 1,\ a_0 = 2^{-1} + 2^{-2},\ a_1 = -2^{-2} + 2^{-20}$

- In this case,

$$\Delta = -(3/4)\epsilon^2 + \epsilon^3 + 2^{-20}(1+\epsilon)^3.$$

- As seen from the following transparency, in this case

$$\epsilon_1 = 0.00161249, \quad \epsilon_2 = 0.00161544.$$


120 dB, a0=2^(−1)+2^(−2),a1=−2^(−2)+2^(−20)
x 10^−7
8
6
4
2
0
−2
−4
−6
−8
Deviation of G(omega) from 1/2
epsilon_2 = 0.00161544
epsilon_1 = 0.00161249
−1.5
−1
−0.5
0
0.5
1
1.5
Deviation of F(omega) from 1
x 10^−3

$L = 2,\ a_0 = 2^0 - 2^{-4},\ a_1 = -2^{-1} - 2^{-3},\ a_2 = 2^{-3} + 2^{-4}$

- In this case,

$$\Delta = -(5/4)\epsilon^3 + (15/16)\epsilon^4 + (3/16)\epsilon^5.$$

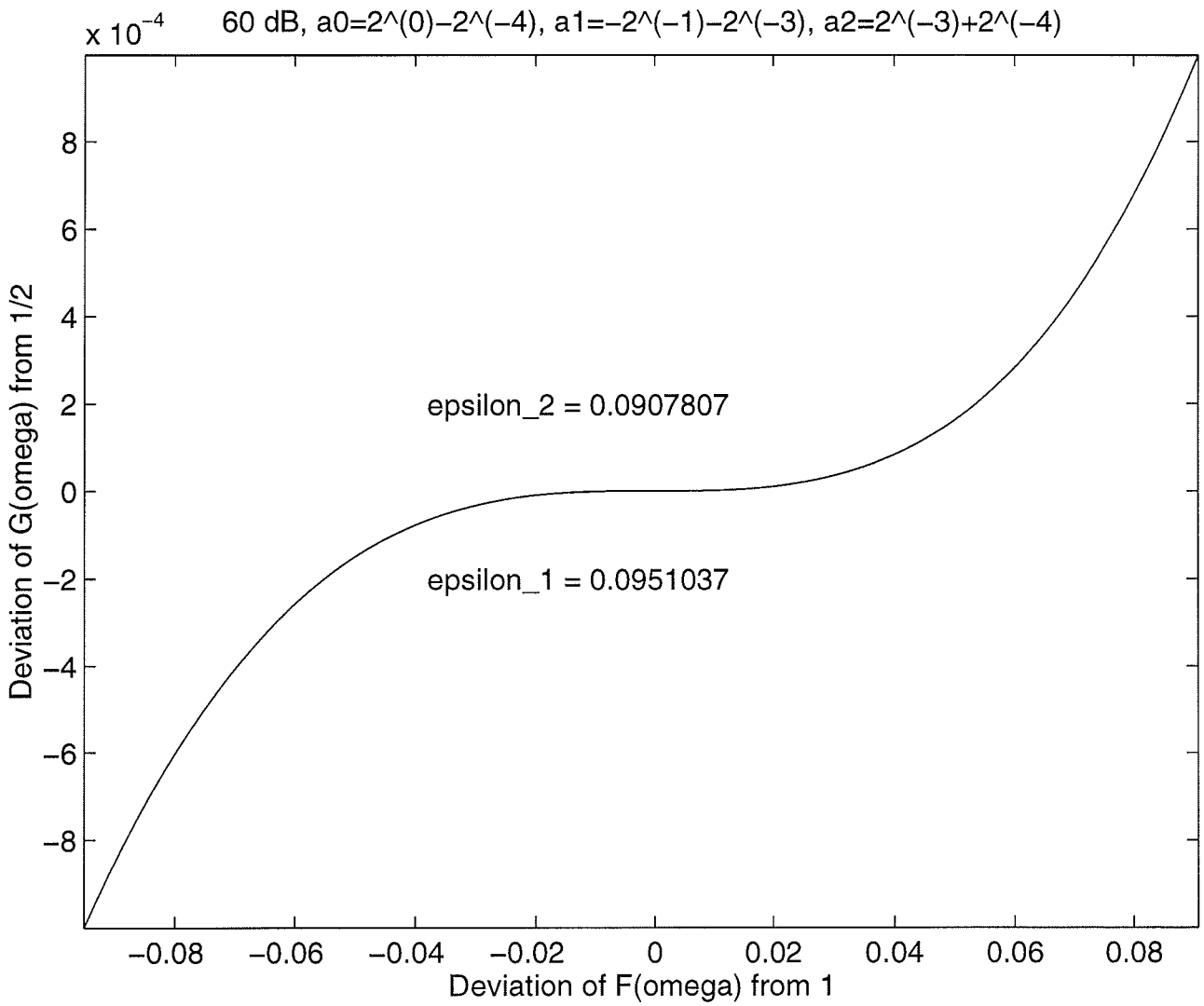
- As seen from the following two transparencies, in Case A,

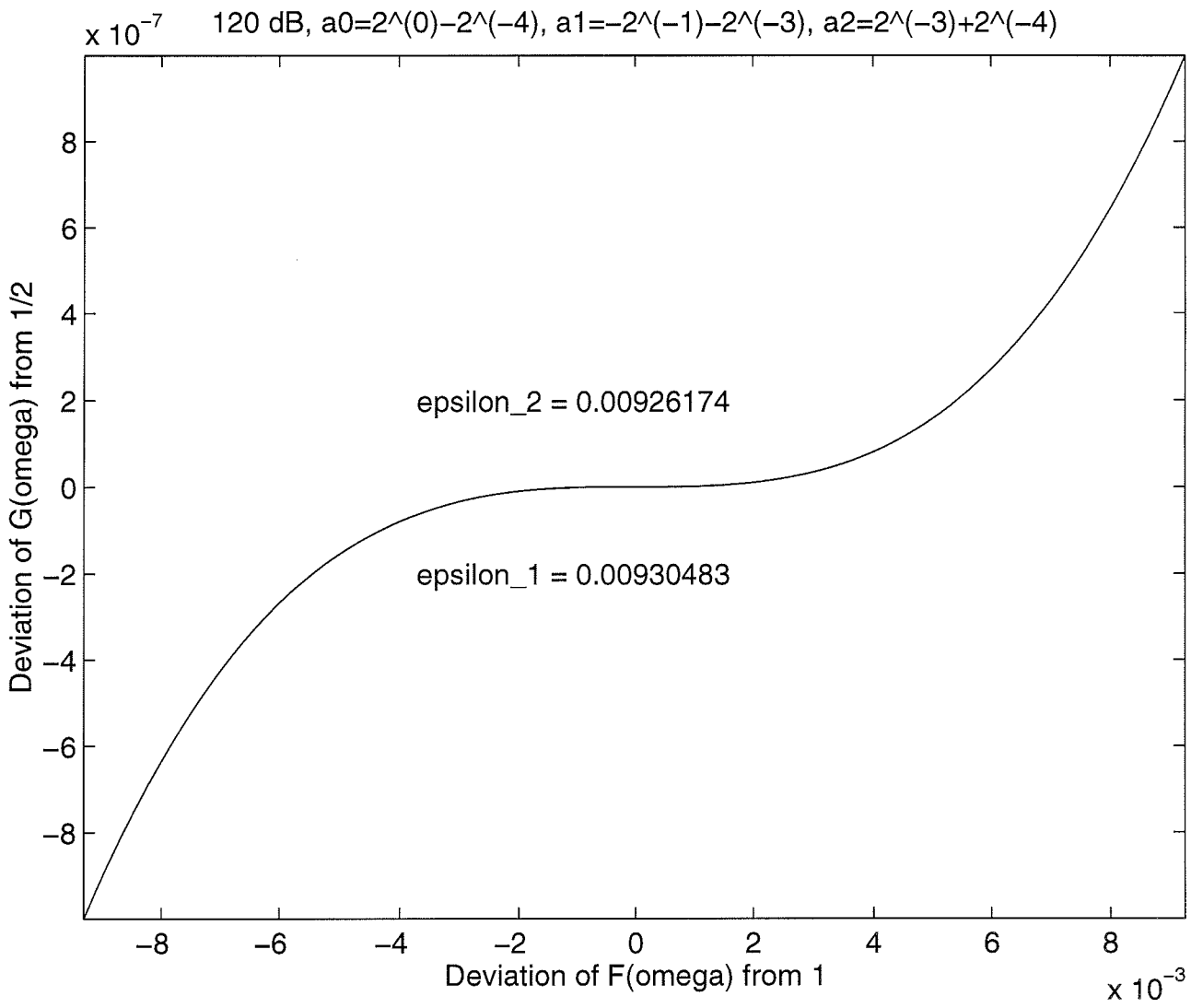
$$\epsilon_1 = 0.0951037, \quad \epsilon_2 = 0.0907807,$$

whereas in Case B,

$$\epsilon_1 = 0.00930483, \quad \epsilon_2 = 0.00926174.$$

- In these cases, Δ takes both positive and negative values.


60 dB, a0=2^(0)−2^(−4), a1=−2^(−1)−2^(−3), a2=2^(−3)+2^(−4)
x 10^−4
Deviation of G(omega) from 1/2
Deviation of F(omega) from 1
epsilon_2 = 0.0907807
epsilon_1 = 0.0951037
8
6
4
2
0
−2
−4
−6
−8
−0.08
−0.06
−0.04
−0.02
0
0.02
0.04
0.06
0.08


x 10^-7
120 dB, a0=2^(0)−2^(−4), a1=−2^(−1)−2^(−3), a2=2^(−3)+2^(−4)
8
6
4
2
0
−2
−4
−6
−8
Deviation of G(omega) from 1/2
epsilon_2 = 0.00926174
epsilon_1 = 0.00930483
−8
−6
−4
−2
0
2
4
6
8
Deviation of F(omega) from 1
x 10^-3

$L = 3,\ a_0 = 2^0 + 2^{-4} + 2^{-5},\ a_1 = -2^0 - 2^{-4} - 2^{-5},$
$a_2 = 2^{-1} + 2^{-3} + 2^{-5},\ a_3 = -2^{-3} - 2^{-5}$

- In this case,

$$\Delta = -(35/16)\epsilon^4 - (21/8)\epsilon^5 - (35/32)\epsilon^6 - (5/32)\epsilon^7.$$

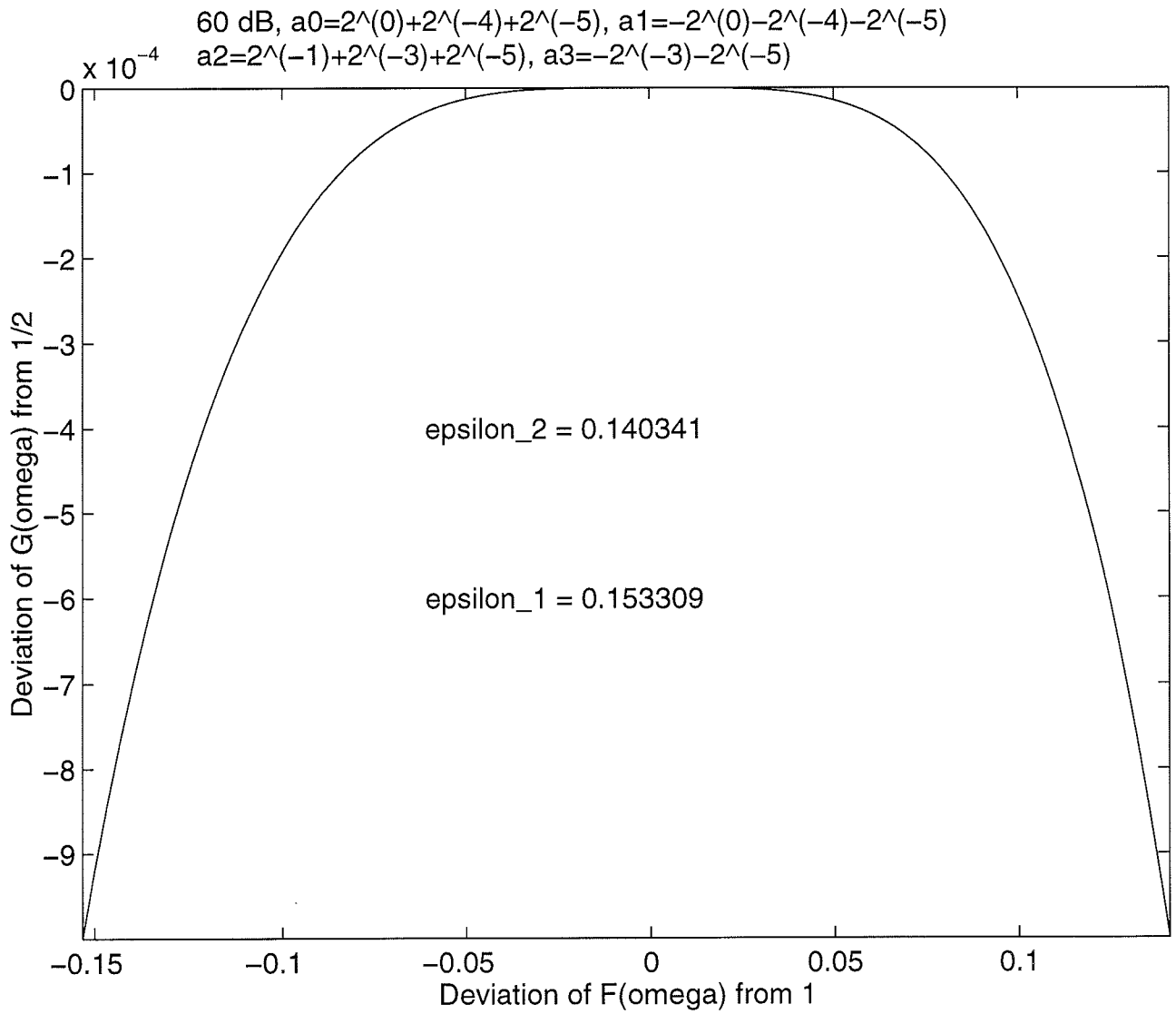
- As seen from the following two transparencies, in Case A,

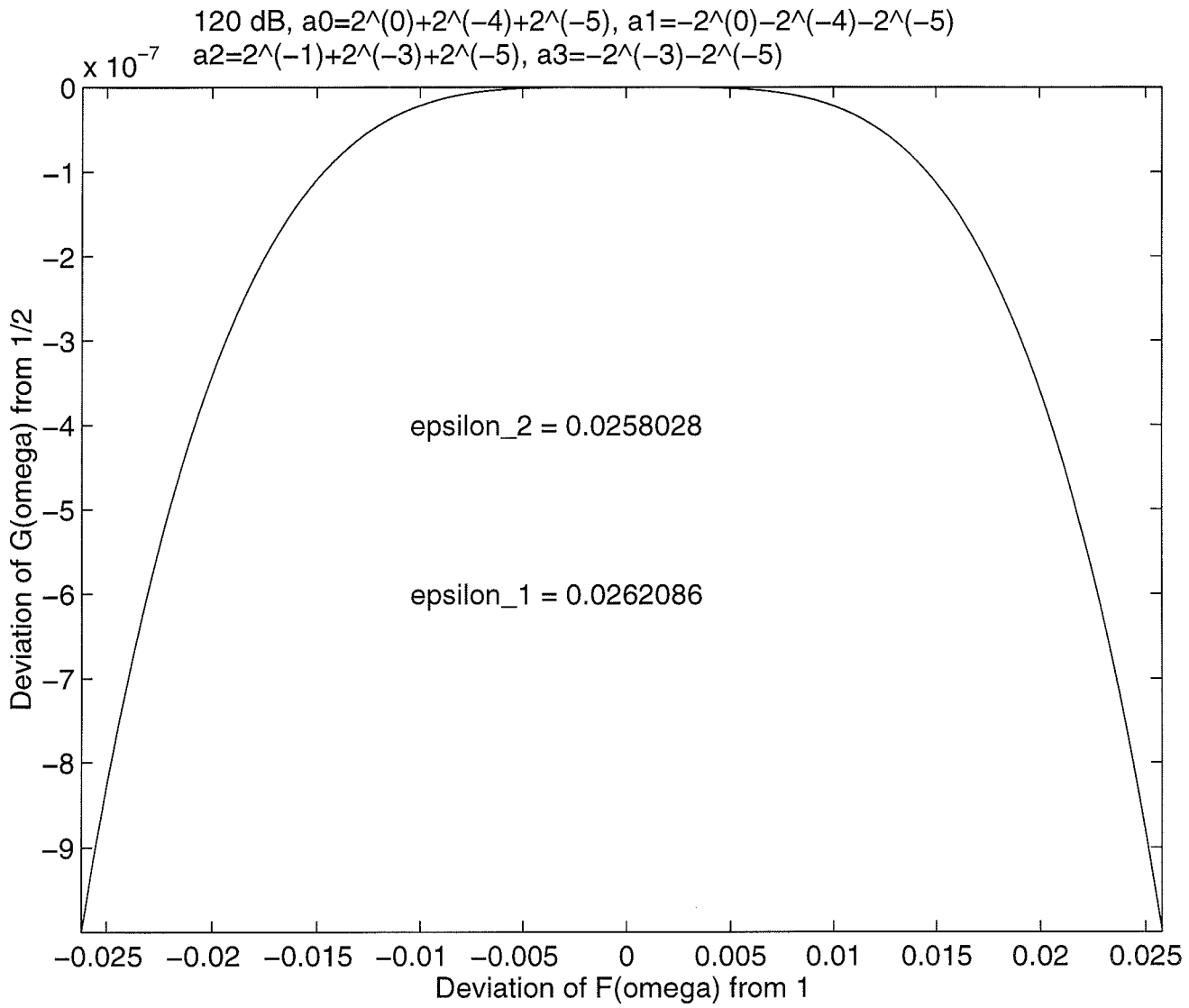
$$\epsilon_1 = 0.153309, \quad \epsilon_2 = 0.140341,$$

whereas in Case B,

$$\epsilon_1 = 0.0262086, \quad \epsilon_2 = 0.0258028$$

- The disadvantage of the above selections of a_0, a_1, a_2, and a_3 is that the maximum value of Δ is zero.
- Better results, that is, both ϵ_1 and ϵ_2 become larger, is obtained by changing a_3. This is considered next.


60 dB, a0=2^(0)+2^(−4)+2^(−5), a1=−2^(0)−2^(−4)−2^(−5)
a2=2^(−1)+2^(−3)+2^(−5), a3=−2^(−3)−2^(−5)
x 10^−4
0
−1
−2
−3
−4
−5
−6
−7
−8
−9
Deviation of G(omega) from 1/2
epsilon_2 = 0.140341
epsilon_1 = 0.153309
−0.15
−0.1
−0.05
0
0.05
0.1
Deviation of F(omega) from 1


120 dB, a0=2^(0)+2^(−4)+2^(−5), a1=−2^(0)−2^(−4)−2^(−5)
a2=2^(−1)+2^(−3)+2^(−5), a3=−2^(−3)−2^(−5)
x 10^−7
0
−1
−2
−3
−4
−5
−6
−7
−8
−9
Deviation of G(omega) from 1/2
epsilon_2 = 0.0258028
epsilon_1 = 0.0262086
−0.025
−0.02
−0.015
−0.01
−0.005
0
0.005
0.01
0.015
0.02
0.025
Deviation of F(omega) from 1

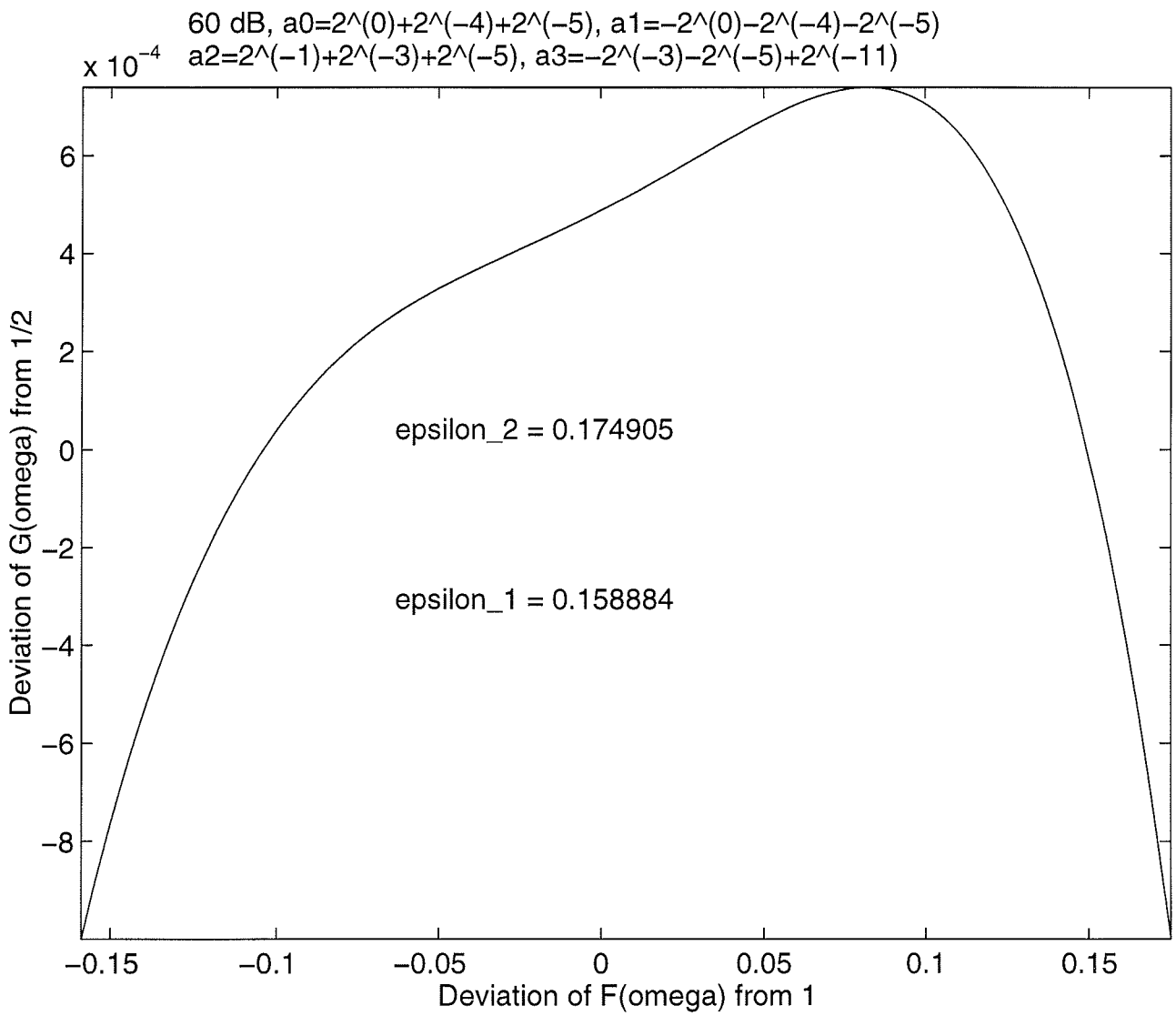
Case A: $L = 3$, $a_0 = 2^0 + 2^{-4} + 2^{-5}$, $a_1 = -2^0 - 2^{-4} - 2^{-5}$, $a_2 = 2^{-1} + 2^{-3} + 2^{-5}$, $a_3 = -2^{-3} - 2^{-5} + 2^{-11}$

- In this case,

$$\Delta = -(35/16)\epsilon^4 - (21/8)\epsilon^5 - (35/32)\epsilon^6 - (5/32)\epsilon^7 + 2^{-11}(1+\epsilon)^7.$$

- As seen from the following transparency, in this case

$$\epsilon_1 = 0.158884, \quad \epsilon_2 = 0.174905.$$


60 dB, a0=2^(0)+2^(−4)+2^(−5), a1=−2^(0)−2^(−4)−2^(−5)
a2=2^(−1)+2^(−3)+2^(−5), a3=−2^(−3)−2^(−5)+2^(−11)
x 10^−4
6
4
2
0
−2
−4
−6
−8
Deviation of G(omega) from 1/2
epsilon_2 = 0.174905
epsilon_1 = 0.158884
−0.15
−0.1
−0.05
0
0.05
0.1
0.15
Deviation of F(omega) from 1

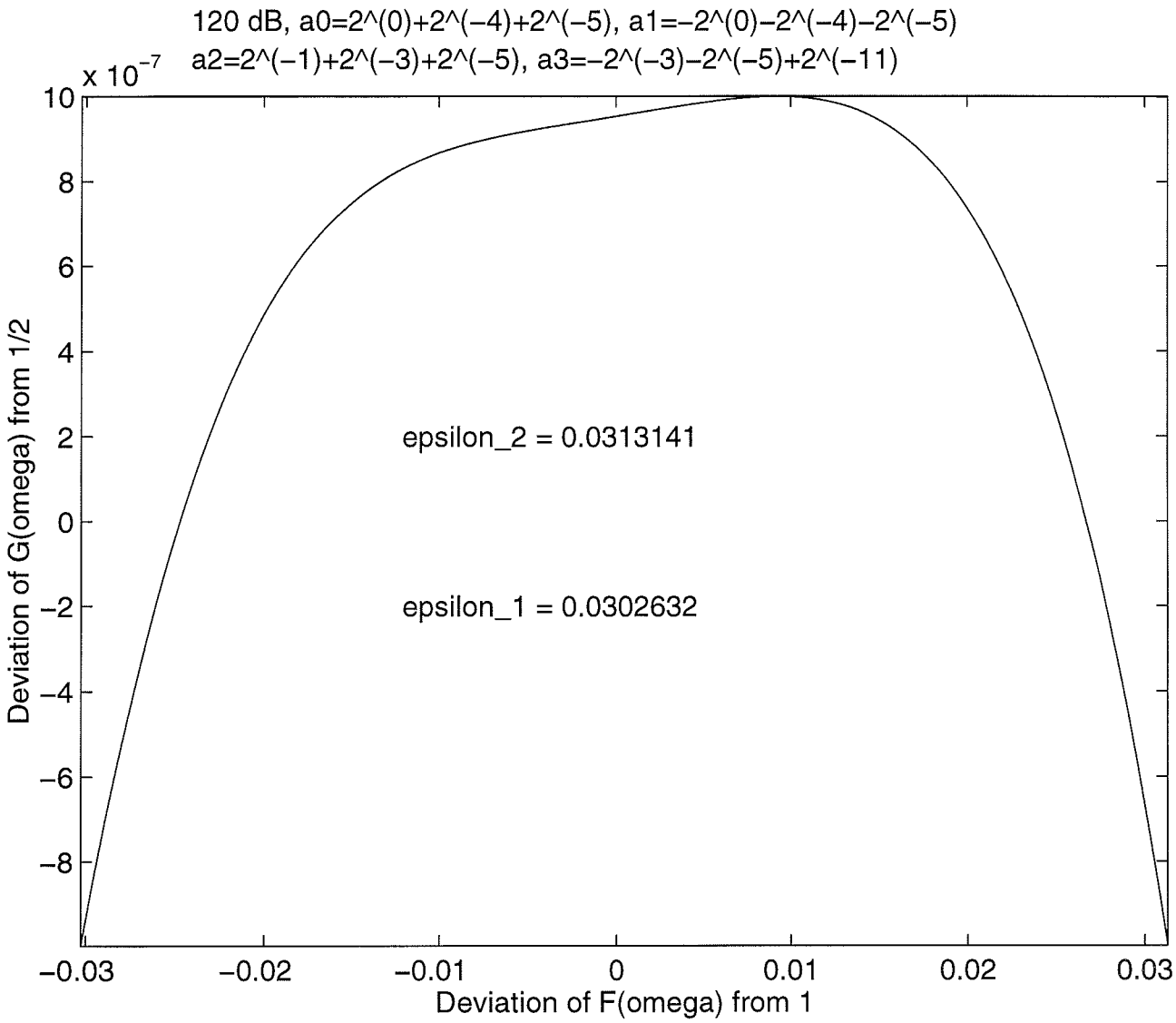
Case B: $L = 3$, $a_0 = 2^0 + 2^{-4} + 2^{-5}$, $a_1 = -2^0 - 2^{-4} - 2^{-5}$, $a_2 = 2^{-1} + 2^{-3} + 2^{-5}$, $a_3 = -2^{-3} - 2^{-5} + 2^{-20}$

- In this case,

$$\Delta = -(35/16)\epsilon^4 - (21/8)\epsilon^5 - (35/32)\epsilon^6 - (5/32)\epsilon^7 + 2^{-20}(1+\epsilon)^7.$$

- As seen from the following transparency, in this case

$$\epsilon_1 = 0.0302632, \quad \epsilon_2 = 0.0313141.$$


120 dB, a0=2^(0)+2^(−4)+2^(−5), a1=−2^(0)−2^(−4)−2^(−5)
a2=2^(−1)+2^(−3)+2^(−5), a3=−2^(−3)−2^(−5)+2^(−11)
x 10^-7
10
8
6
4
2
0
−2
−4
−6
−8
Deviation of G(omega) from 1/2
epsilon_2 = 0.0313141
epsilon_1 = 0.0302632
−0.03
−0.02
−0.01
0
0.01
0.02
0.03
Deviation of F(omega) from 1

Example

- It is desired to design a half-band decimator in such a way that
  - **1)** the sampling rate reduction ratio is 2 and the output sampling rate is 44.1 kHz.
  - **2)** Components alising into the band from 0 Hz to 20 kHz are attenuated at least 120 dB.
- In this case the problem is to design $G(z)$ such that the deviation of its zero-phase frequency response $G(\omega)$ from 1/2 is at most 0.000001 in the passband.
- The passband edge is at

$$2\omega_p = [20/(44.1/2)]\pi = 0.90703\pi.$$

- We select $L = 3$, $a_0 = 2^0 + 2^{-4} + 2^{-5}$, $a_1 = -2^0 - 2^{-4} - 2^{-5}$, $a_2 = 2^{-1} + 2^{-3} + 2^{-5}$, and $a_3 = -2^{-3} - 2^{-5} + 2^{-20}$.
- In this case, the problem is to design an odd order $F(z)$ such that its zero-phase frequency response $F(\omega)$ stays within the limits $1 - \epsilon_1$ and

$1 + \epsilon_2$ on $[0,\ 2\omega_p]$ with

$$\epsilon_1 = 0.0302632, \quad \epsilon_2 = 0.0313141.$$

- The design can be accomplished with the aid of the Remez algorithm by using a single band $[0,\ 2\omega_p]$. The desired function is $(2 - \epsilon_1 + \epsilon_2) = 1.00052545$, whereas the allowable deviation is $1 + \epsilon_2 - 1.00052545 = 0.03078865$.
- The given criteria are met by $F(z)$ of order 21 and having the impulse-response coefficient values

$$f[0] = f[21] = 2^{-6},\ f[1] = f[20] = -2^{-7} - 2^{-8},$$

$$f[2] = f[19] = 2^{-6},\ f[3] = f[18] = 2^{-6} - 2^{-7},$$

$$f[4] = f[17] = 2^{-5},\ f[5] = f[16] = -2^{-5} - 2^{-7} - 2^{-8},$$

$$f[6] = f[15] = 2^{-4} - 2^{-8},\ f[7] = f[14] = -2^{-4} - 2^{-6} - 2^{-8},$$

$$f[8] = f[13] = 2^{-3} - 2^{-8},\ f[9] = f[12] = -2^{-2} + 2^{-5} + 2^{-7},$$

and

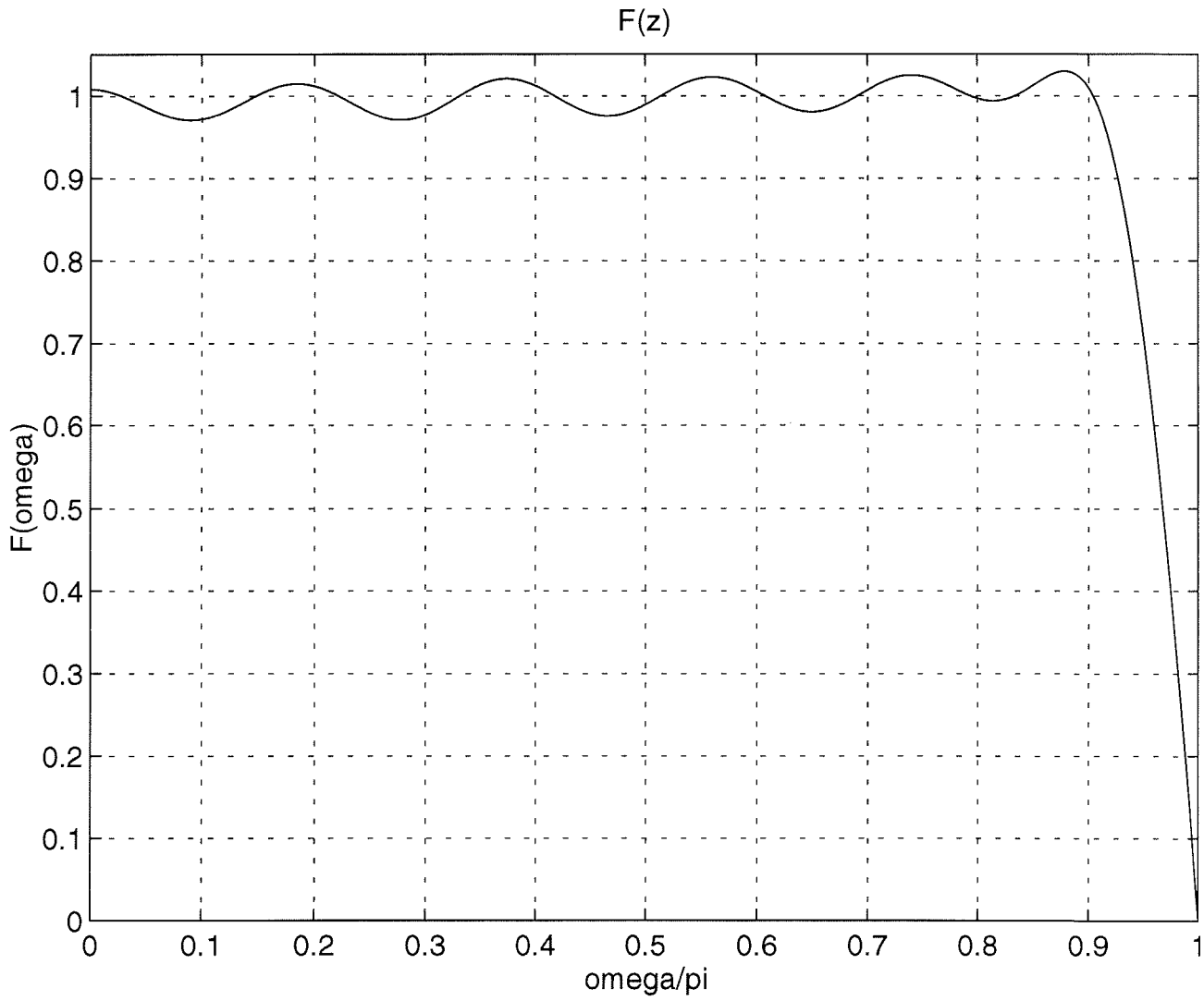
$$f[10] = f[11] = 2^{-1} + 2^{-3} + 2^{-7}.$$

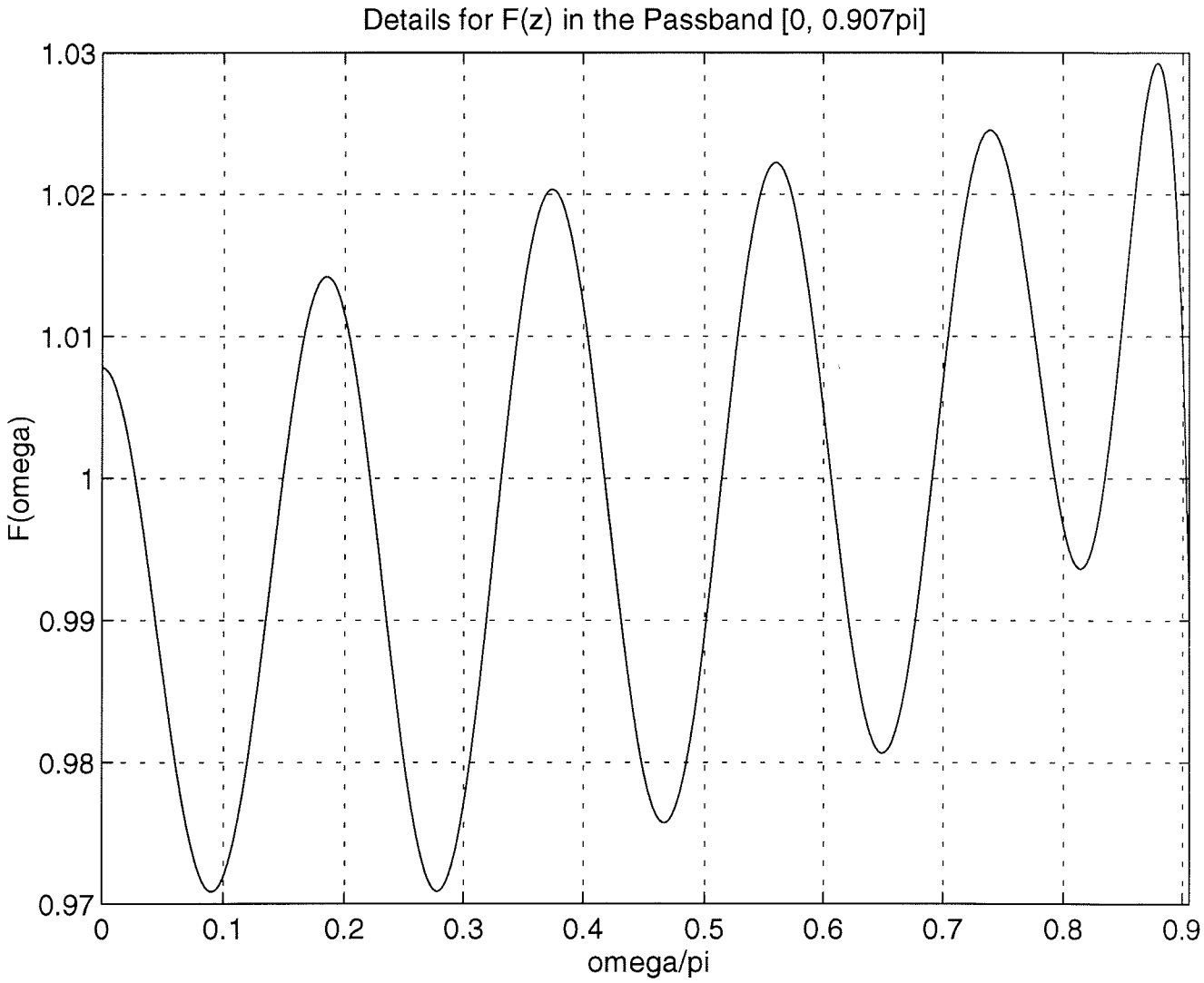
- These values have been obtained by first rounding the coefficients of $F(z)$ to 8 fractional bits.

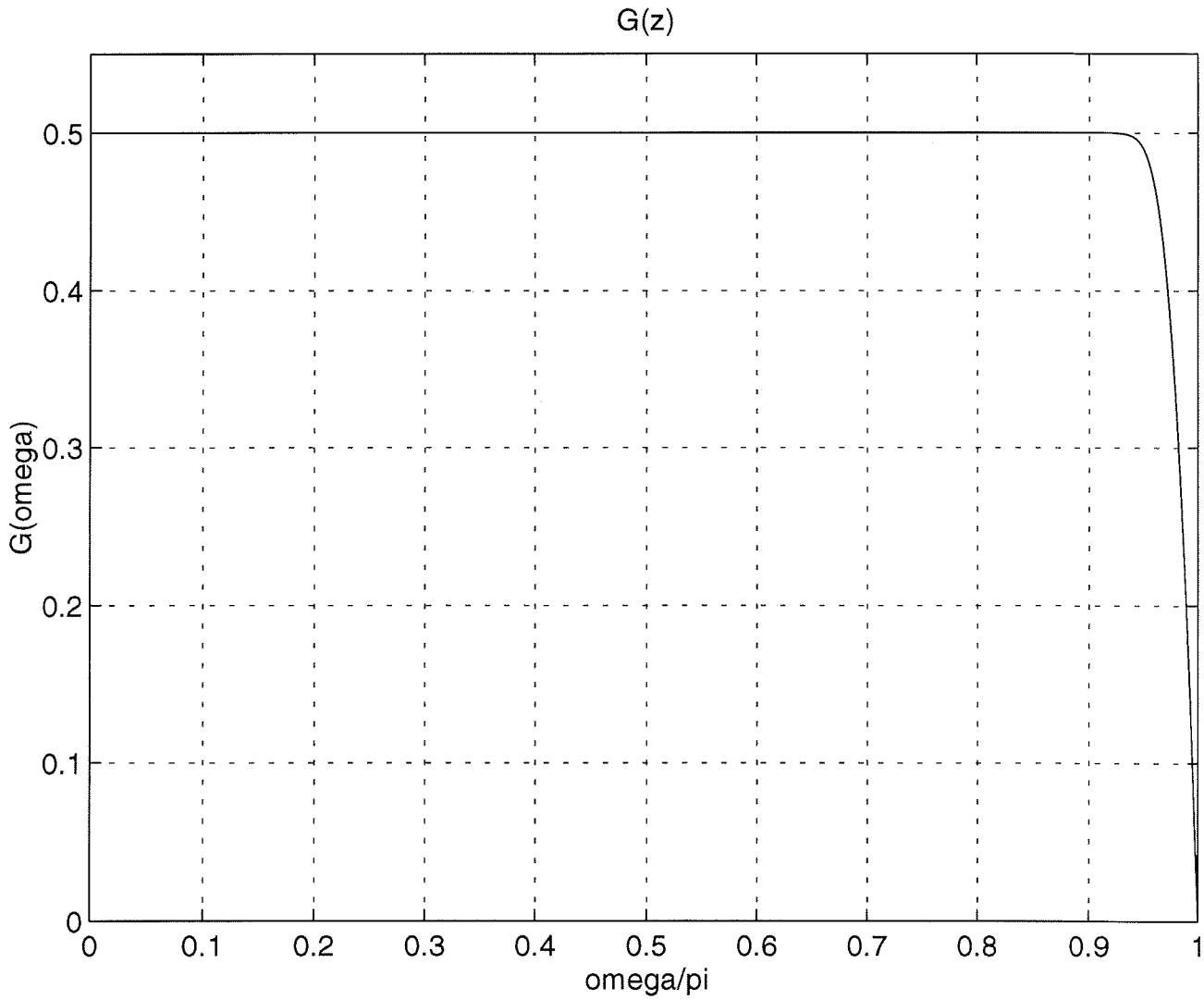
This rounding gives $f[10] = f[11] = 163 \cdot 2^{-8}$ and $f[9] = f[12] = -53 \cdot 2^{-8}$, which are not expressible as three powers of two. Therefore, they are rounded to the nearest three powers-ot-two resprentations, giving $f[10] = f[11] = 162 \cdot 2^{-8}$ and $f[9] = f[12] = -54 \cdot 2^{-8}$.

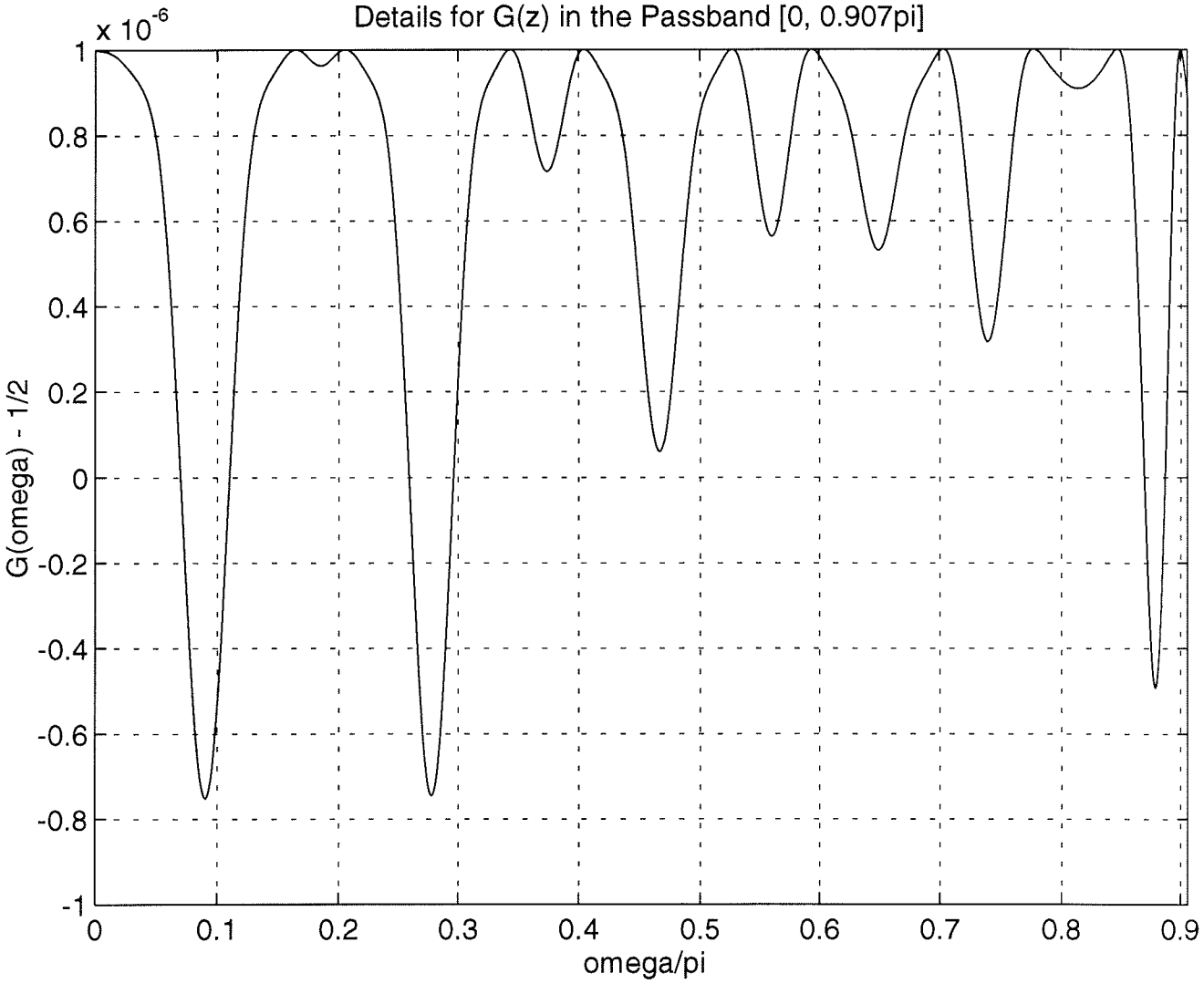
- In the following, there are six transparencies illustrating the chracteristics of $F(z)$, $G(z)$, and the resulting overall half-band FIR filter $H(z)$ with passband and stopband edges at $\omega_p = 0.453515\pi$ and $\omega_s = \pi - \omega_p = 0.546485\pi$ and a 120-dB attenuation in the stopband.
- In addition, there are two Matlab-files, halquan.m and half.m. The first file finds the above $F(z)$, whereas the second file plots the responses for $F(z)$, $G(z)$, and $H(z)$.
- In the very end of these lecture notes there are the following article as well as the corresponding conference talk:
- T. Saramäki, T. Karema, T. Ritoniemi, and H.

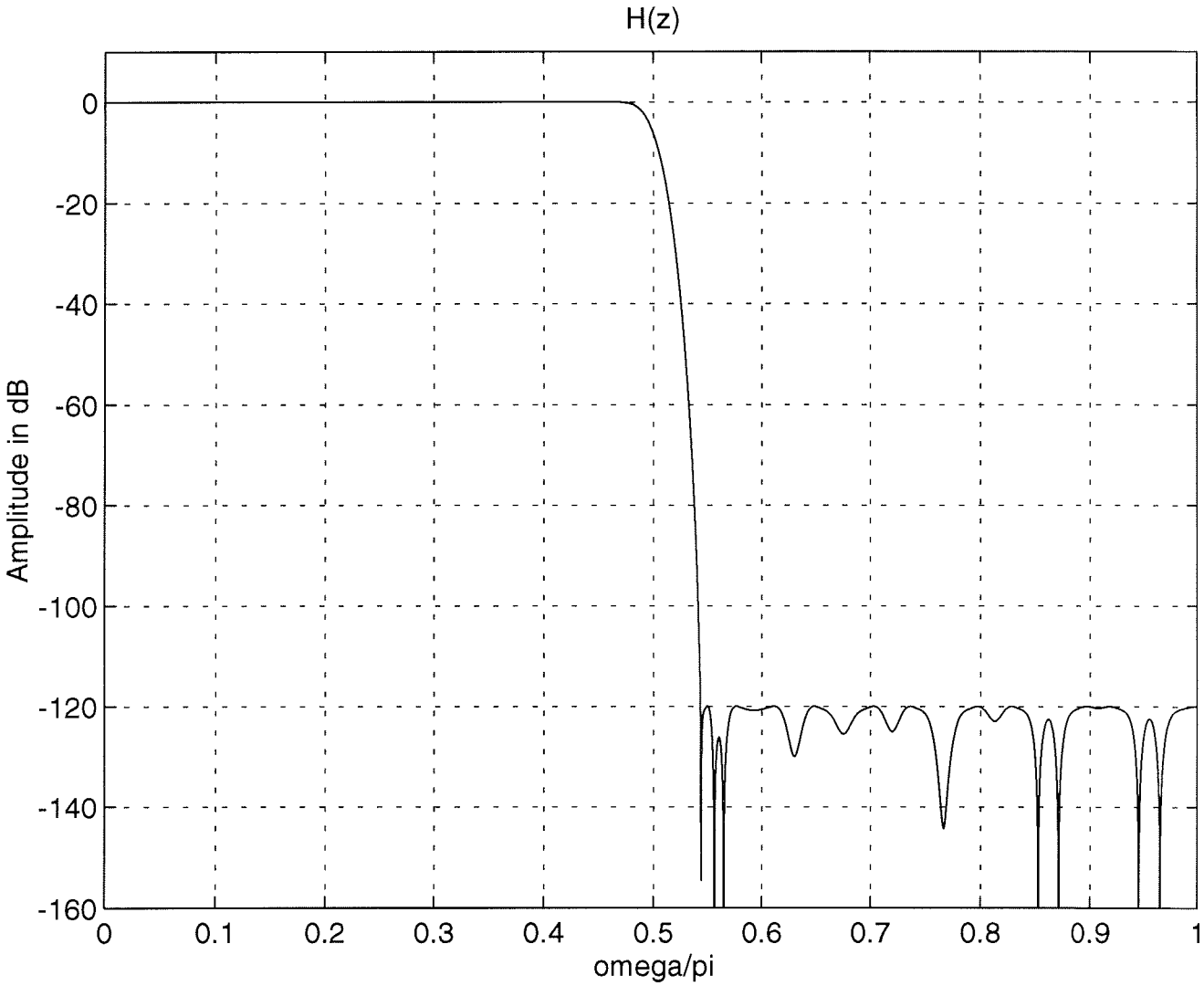
Tenhunen, "Multiplier-free decimator algorithms for superresolution oversampled converters," in *Proc.* 1990 *IEEE International Symposium on Circuits and Systems* (New Orleans, Louisiana), pp. 3275–3278, May 1990.

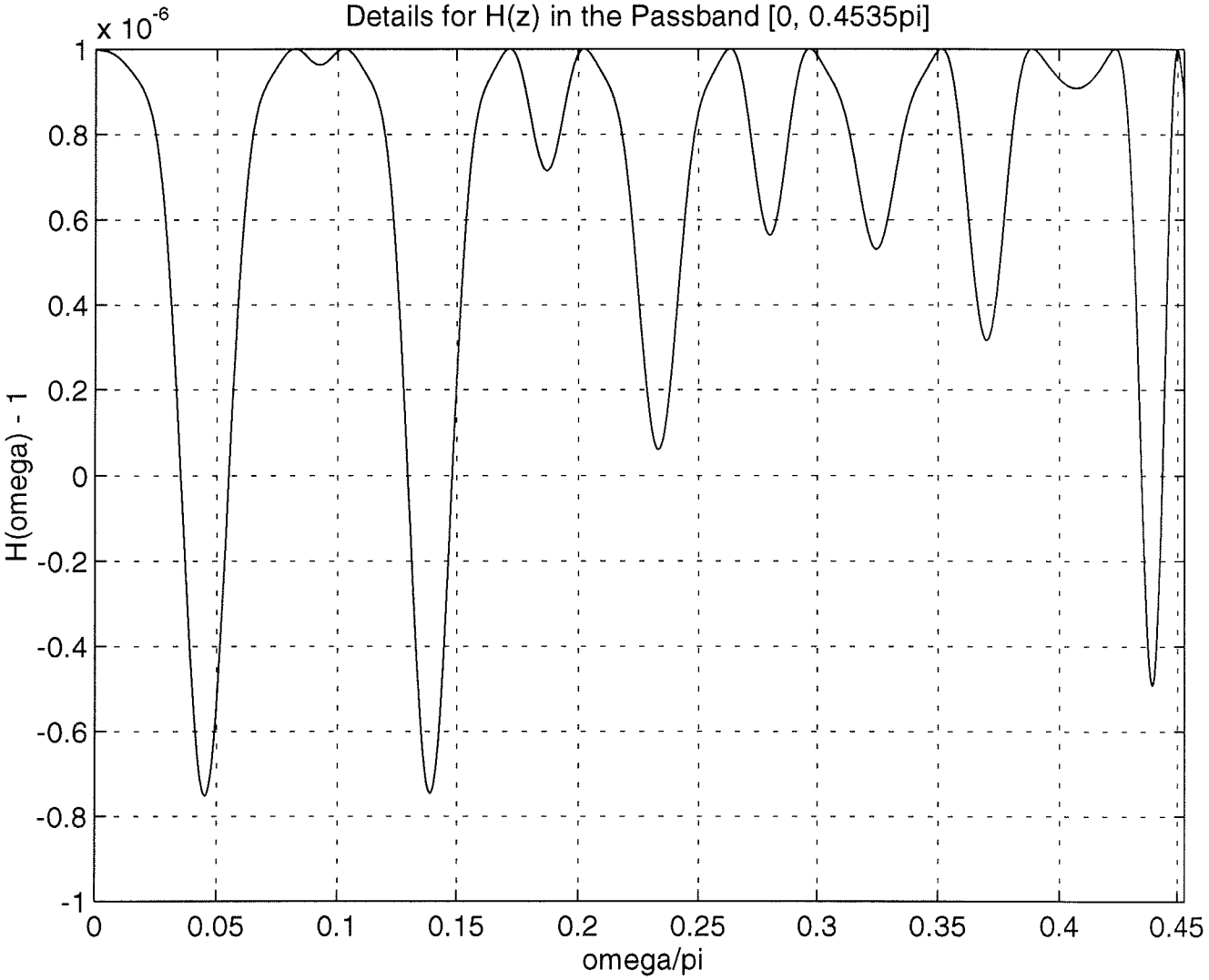

F(z)
F(omega)
omega/pi
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


Details for F(z) in the Passband [0, 0.907pi]
1.03
1.02
1.01
1
0.99
0.98
0.97
F(omega)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
omega/pi


G(z)
G(omega)
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
omega/pi


Details for G(z) in the Passband [0, 0.907pi]
x 10^-6
G(omega) - 1/2
omega/pi
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9


H(z)
Amplitude in dB
0
-20
-40
-60
-80
-100
-120
-140
-160
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
omega/pi


Details for H(z) in the Passband [0, 0.4535pi]
x 10^-6
H(omega) - 1
omega/pi
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45

```
% Matlab-file halquan.m for designing
% the example F(z)
% can be found in SUN's: ~ts/matlab/sldsp
close all;clear all
N=21;e1=0.0302632;e2=0.0313141;
fo=[0 20/(44.1/2) ];
des=(2-e1+e2)/2;mo=[des des];
w=[1];
h = remez(N,fo,mo,w);
nbit=8;
hs=round(h*2^nbit)/(2^nbit);
%
% hs(10)*2^8=hs(13)*2^8=-53
% hs(11)*2^8=hs(12)*2^8=-53
% cannot be expressed as three powers of two
% therefore, hs(10)=hs(13)=-54*2^(-8)
% hs(11)=hs(12)=162*2^(-8)
hs(10)=-54*2^(-8);hs(13)=hs(10);
hs(11)=162*2^(-8);hs(12)=hs(11);
figure(1)
[H,W]=zeroam(h,.0,1.,2000);
[H1,W1]=zeroam(hs,.0,1.,2000);
plot(W/pi,20*log10(abs(H)),'- -',W/pi,20*log10(abs(H1)));
axis([0 1 -90 10]);grid;
ylabel('Amplitude in dB'); xlabel('Angular frequency omega/pi');
title('Solid and dashed lines for quantized and ideal filters');
[H,W]=zeroam(h,fo(1),fo(2),2000);
[H1,W]=zeroam(hs,fo(1),fo(2),2000);
figure(2)
subplot(211)
plot(W/pi,H);grid
xlabel('Angular frequency omega/pi');
title('Passband: Ideal response H(omega)');
subplot(212)
plot(W/pi,H1-H);
grid;
ylabel('                    Zero-phase frequency response');
xlabel('Angular frequency omega/pi');
title('Passband: Quantization error E_b(omega)');
figure(3)
plot(W/pi,H,'- -',W/pi,H1);
```

```
title('Passband: Solid and dashed lines for quantized and ideal
filters');
ylabel('Zero-phase frequency response');
xlabel('Angular frequency omega/pi');
```

```
% Matlab-file half.m for plotting the responses
% for the example half-band FIR filter
% can be found in SUN's: ~ts/matlab/sldsp
clear all
close all
f(11)=162;f(10)=-54;f(9)=31;
f(8)=-21;f(7)=15;f(6)=-11;
f(5)=8;f(4)=-6;f(3)=4;
f(2)=-3;f(1)=4;
f=f*(2)^(-8);
for k=1:11
f(23-k)=f(k);end
figure(1)
[F,om]=zeroam(f,.0,2.,20000);
plot(om/pi,F);title('F(z)');
axis([0 1 0 1.05]);grid;
ylabel('F(omega)'),xlabel('omega/pi');
figure(2)
plot(om/pi,F);
title('Details for F(z) in the Passband [0, 0.907pi] ');
axis([0 0.90703 1-.03 1+.03]);grid;
ylabel('F(omega)'),xlabel('omega/pi');
x=F;
a1=2^(0)+2^(-4)+2^(-5);
a2=-2^(0)-2^(-4)-2^(-5);
a3=2^(-1)+2^(-3)+2^(-5);
a4=-2^(-3)-2^(-5)+2^(-20);
F1=x;
F2=F1.*F1;
F3=F1.*F2;
F5=F3.*F2;
F7=F5.*F2;
G=a1*F1+a2*F3+a3*F5+a4*F7;
figure(3)
plot(om/pi,G);title('G(z)');
axis([0 1 0 .55]);grid;
ylabel('G(omega)'),xlabel('omega/pi');
figure(4)
plot(om/pi,G-1/2);
title('Details for G(z) in the Passband [0, 0.907pi] ');
axis([0 0.90703 -.000001 .000001]);grid;
ylabel('G(omega) - 1/2'),xlabel('omega/pi');
```

```
figure(5)
plot(om/(2*pi),20*log10(abs(G+1/2)));title('H(z)');
axis([0 1 -160 10]);grid;
ylabel('Amplitude in dB'),xlabel('omega/pi');
figure(6)
plot(om/(2*pi),G-1/2);
title('Details for H(z) in the Passband [0, 0.4535pi] ');
axis([0 0.4535 -.000001 .000001]);grid;
ylabel('H(omega) - 1'),xlabel('omega/pi');
```

MULTIPLIER-FREE DECIMATOR ALGORITHMS FOR SUPERRESOLUTION OVERSAMPLED CONVERTERS

OUTLINE

* OVERSAMPLING AND DECIMATION

* FILTER REQUIREMENTS FOR 16/20 BIT RESOLUTION

* DECIMATOR REQUIREMENTS

* FILTER STRUCTURES & CHARACTERISTICS

* FILTER IMPLEMENTATIONS

* CONCLUSIONS

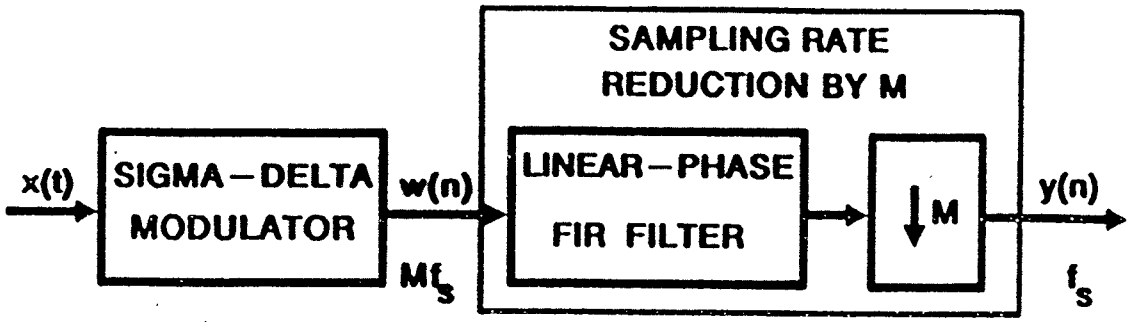
OVERSAMPLING & DECIMATION

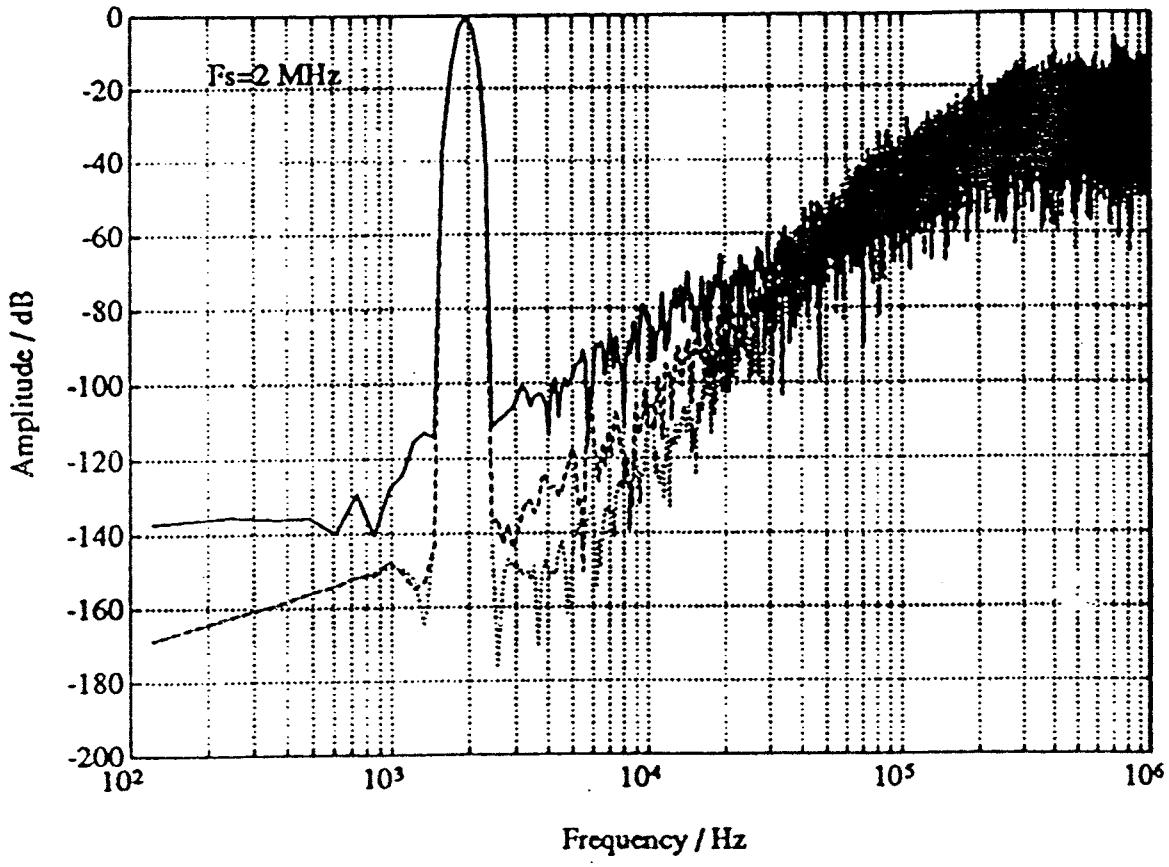
* Quantization noise attenuation

* Antialias filtering at input

* Pass band control

  -with 44.1 kHz final sampling rate band pass edge at 20 kHz



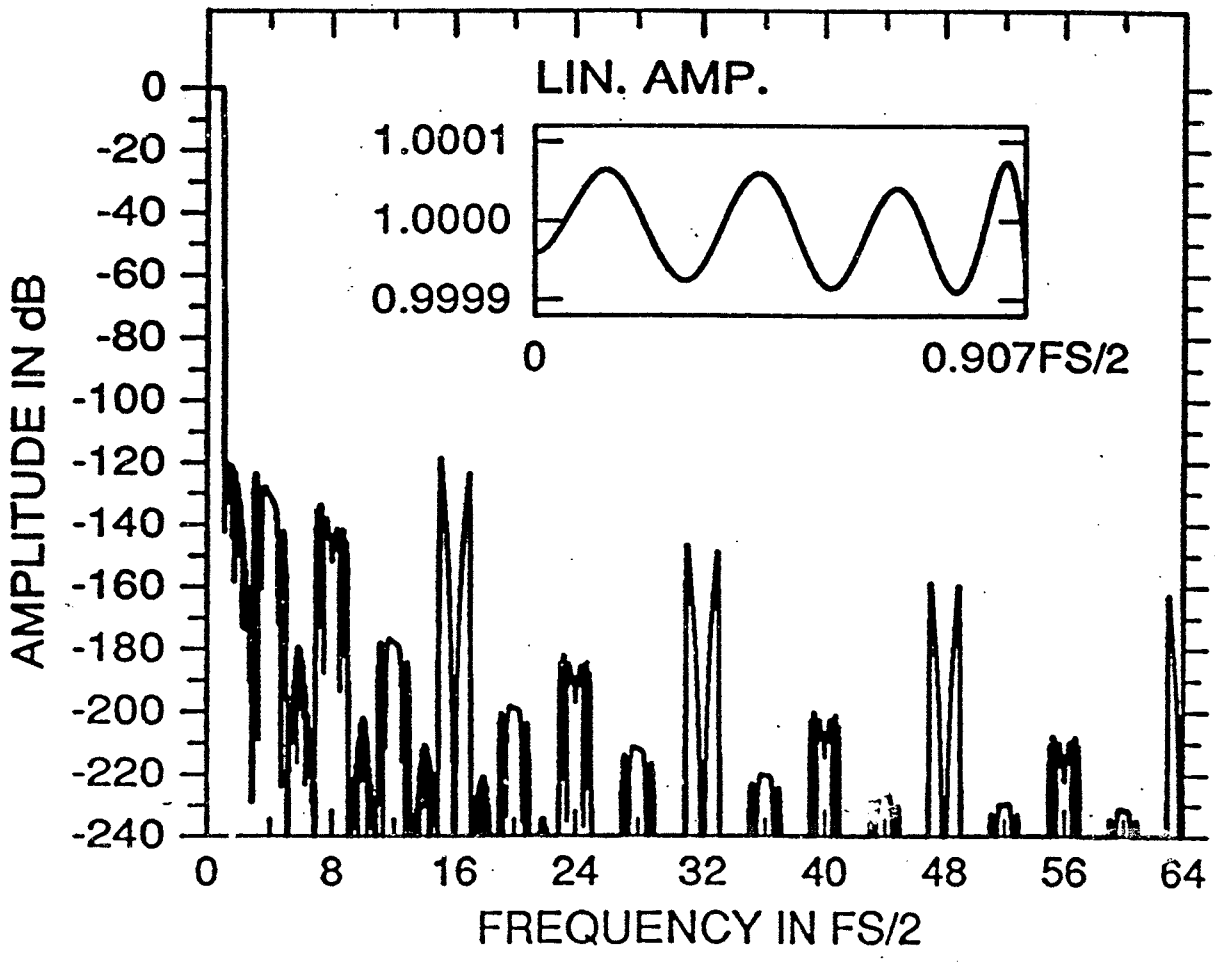


REQUIREMENTS FOR THE DECIMATOR

$$1 + \delta_p \leq |H(e^{j2\pi f/(Mf_s)})| \leq 1 - \delta_p \quad \text{for} \quad 0 \leq f \leq \alpha\frac{f_s}{2}$$

$$|H(e^{j2\pi f/(Mf_s)})| \leq \delta_s \quad \text{for} \quad (2-\alpha)\frac{f_s}{2} \leq f \leq M\frac{f_s}{2},$$

$\delta_p = 0.0001$ and $\delta_s = 0.000001$ (120-dB attenuation).



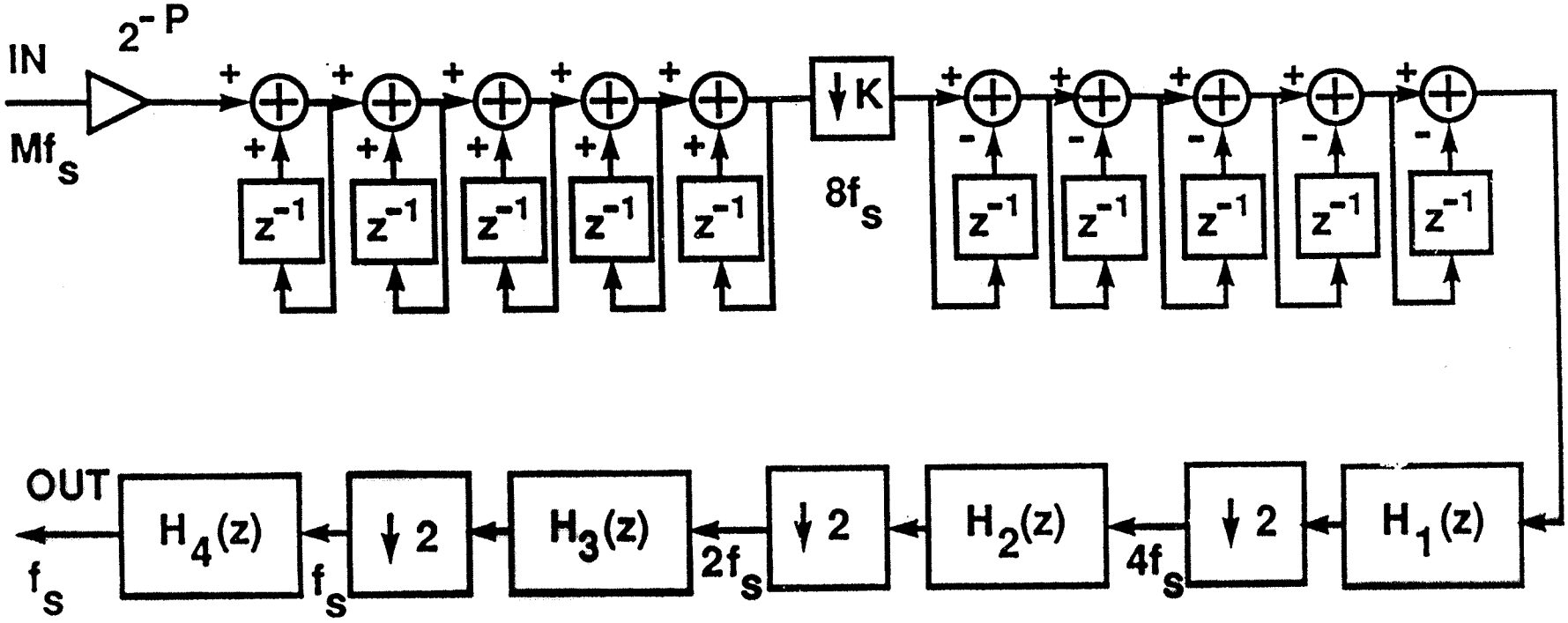
PROPOSED CLASS OF DECIMATORS

$$H(z) = H_4(z^M)H_3(z^{M/2})H_2(z^{M/4})H_1(z^{M/8})G(z),$$

$$G(z) = 2^{-P}\left[\frac{1 - z^{-4}}{1 - z^{-1}}\right]^L$$

$$K = M/8$$

$$P = L \log_2 K$$



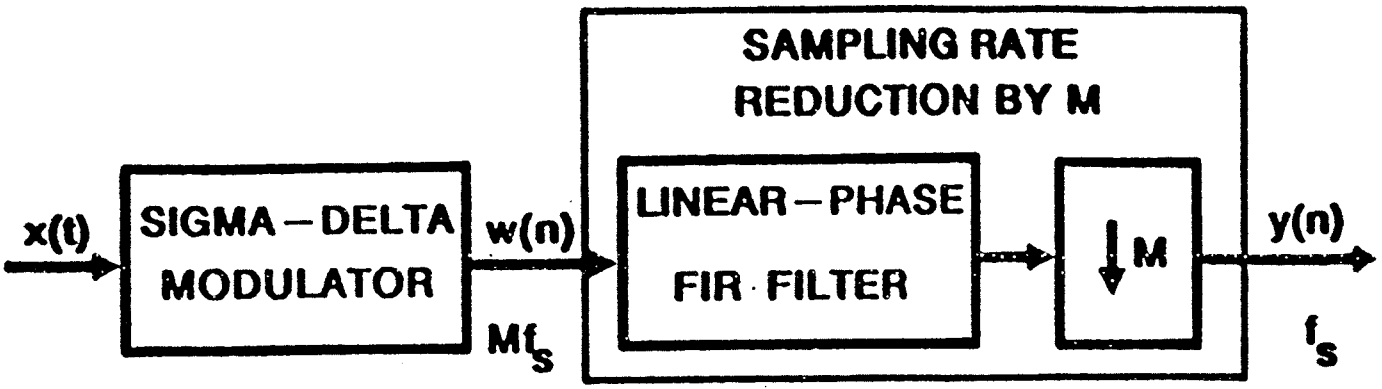
EQUIVALENT SINGLE-STAGE $H(z)$

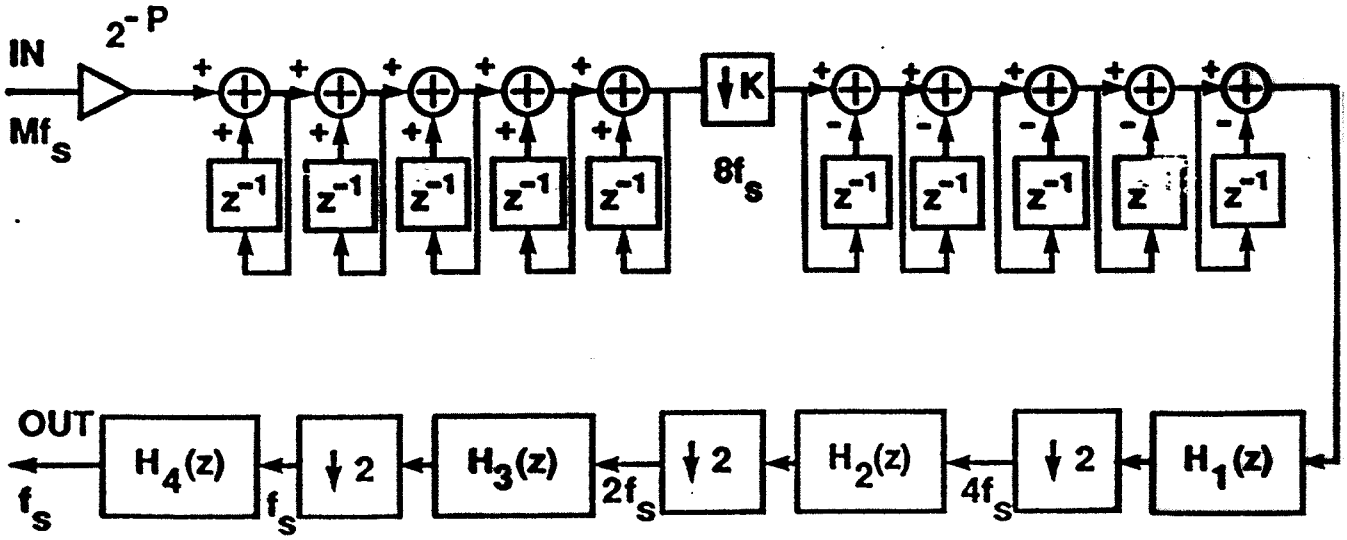
$$H(z) = H_4(z^M)H_3(z^{M/2})H_2(z^{M/4})H_1(z^{M/8})G(z),$$

$$G(z) = 2^{-P}\left[\frac{1 - z^{-4}}{1 - z^{-1}}\right]^L$$

$$K = M/8$$

$$P = L \log_2 K$$





DESIGN STRATEGIES

- Linear-phase $\Rightarrow$ FIR design
- Several oversampling ratios
- Elimination of general multipliers
- Regular structures
- Minimal silicon area

HOW TO ACHIEVE?

- Multistage design
- Recursive running sum filter for the first stage
- Special half-band designs for last decimation stages: Tapped cascaded interconnection of identical subfilters
- Special equalizer for compensating the distortion caused by the running sum filter
    - The passband ripple of each half-band filter is so small ($\delta_p = 10^{-6}$) that no equalization is required

WHY A TAPPED CASCADED INTERCONNECTION OF IDENTICAL SUBFILTERS?

General rules of thumb:

- The number of decimal bits required by FIR filter coefficients is decreased by one when the allowable quantization error is doubled.
- One bit is saved when the filter order is made one fourth

This fact is exploited:

The tap coefficients can be determined in a straightforward manner such that

- They have simple representation forms
- The allowable variation for the subfilter becomes huge.

⇒ The subfilter coefficients can be determined

- Quantize the coefficients to few decimal bits
- Select the nearest simple representation form

RUNNING SUM G(z)

$$G(z) = 2^{-P}\left[\frac{1 - z^{-4}}{1 - z^{-1}}\right]^{L}$$

$$K = M/8$$

$$P = L \log_2 K$$

* $sinc^L$ filter, efficient anti-alias filter with high attenuation on high frequencies

* Straightforward efficient implementation on Si using recursive and non-recursive sections

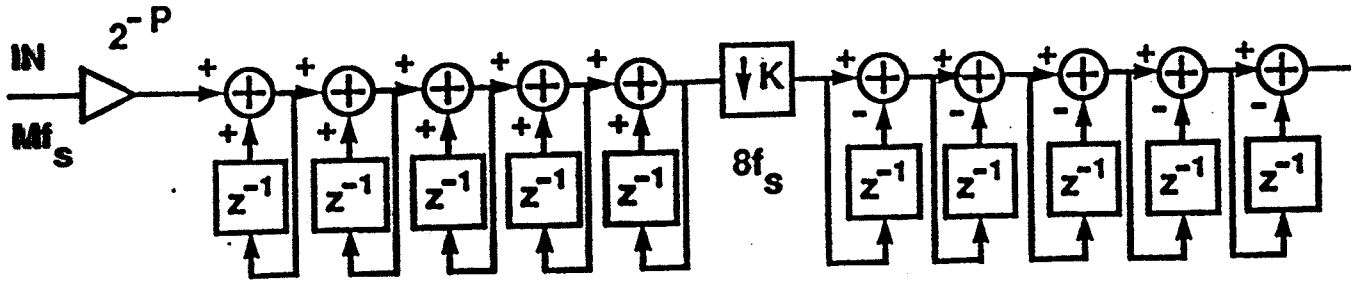
* no multipliers

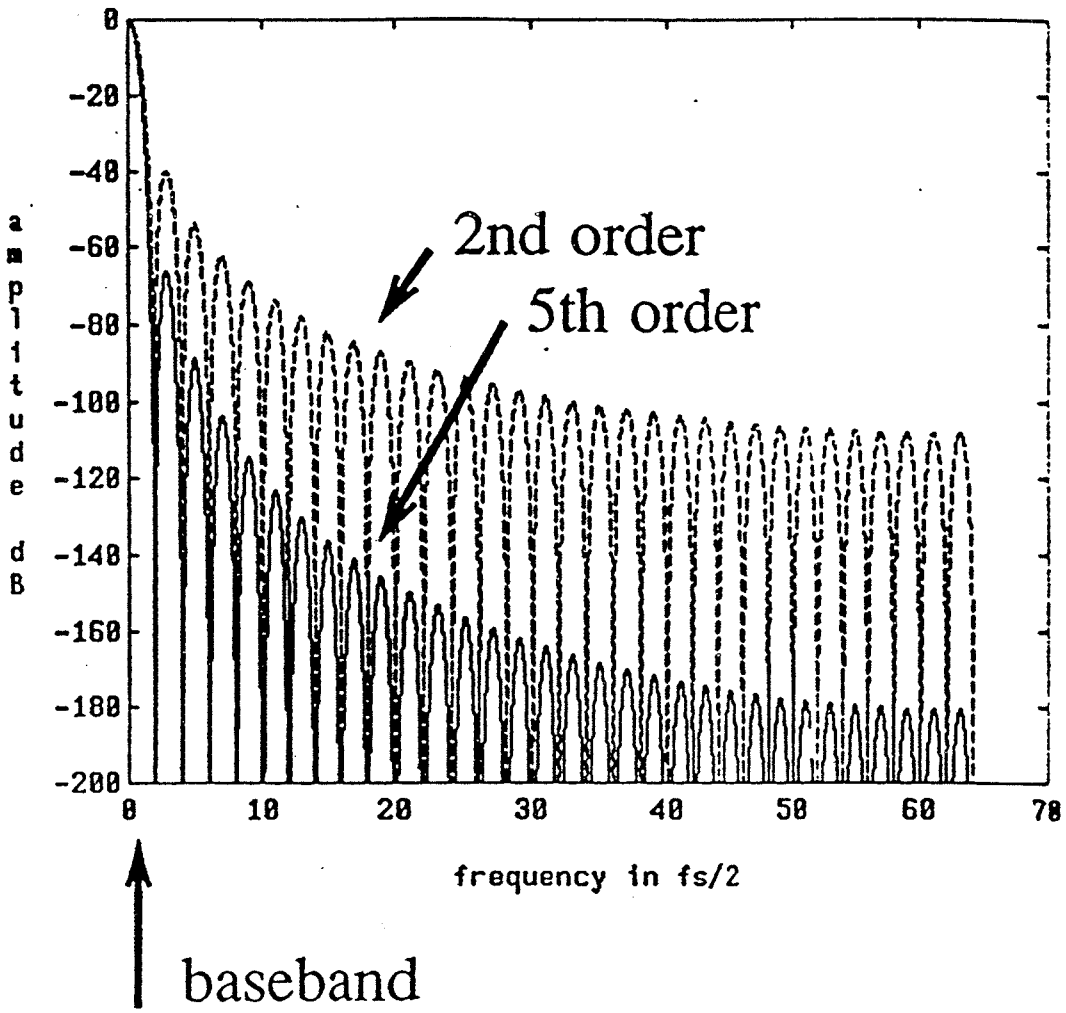
* no overfolow problems in modulo arithmetic

* Decimation by factor 8 because

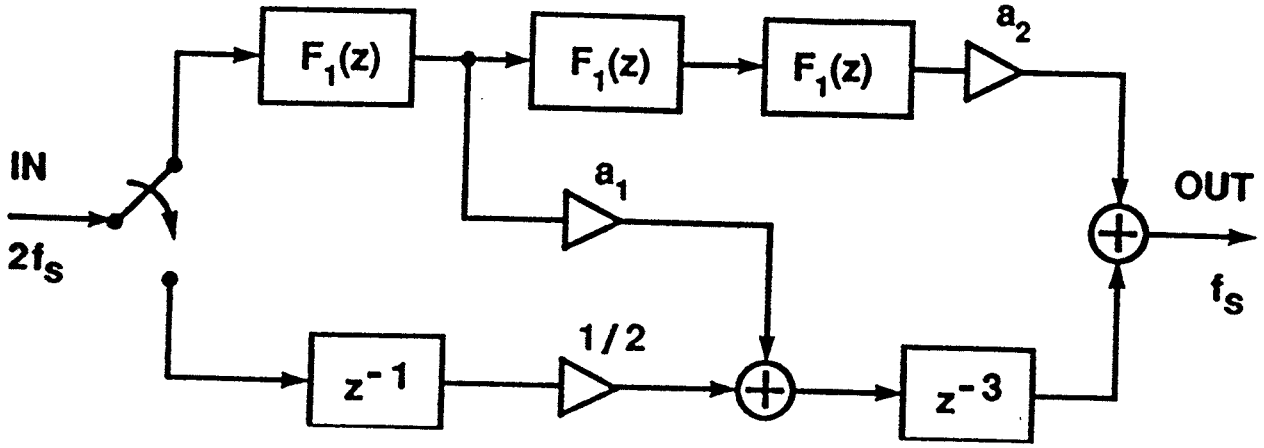
  -only weak monotonic distortion caused at pass band edges
  -with $sinc^5$ 120 dB stop band attenuation cannot achieved for higher decimation ratios

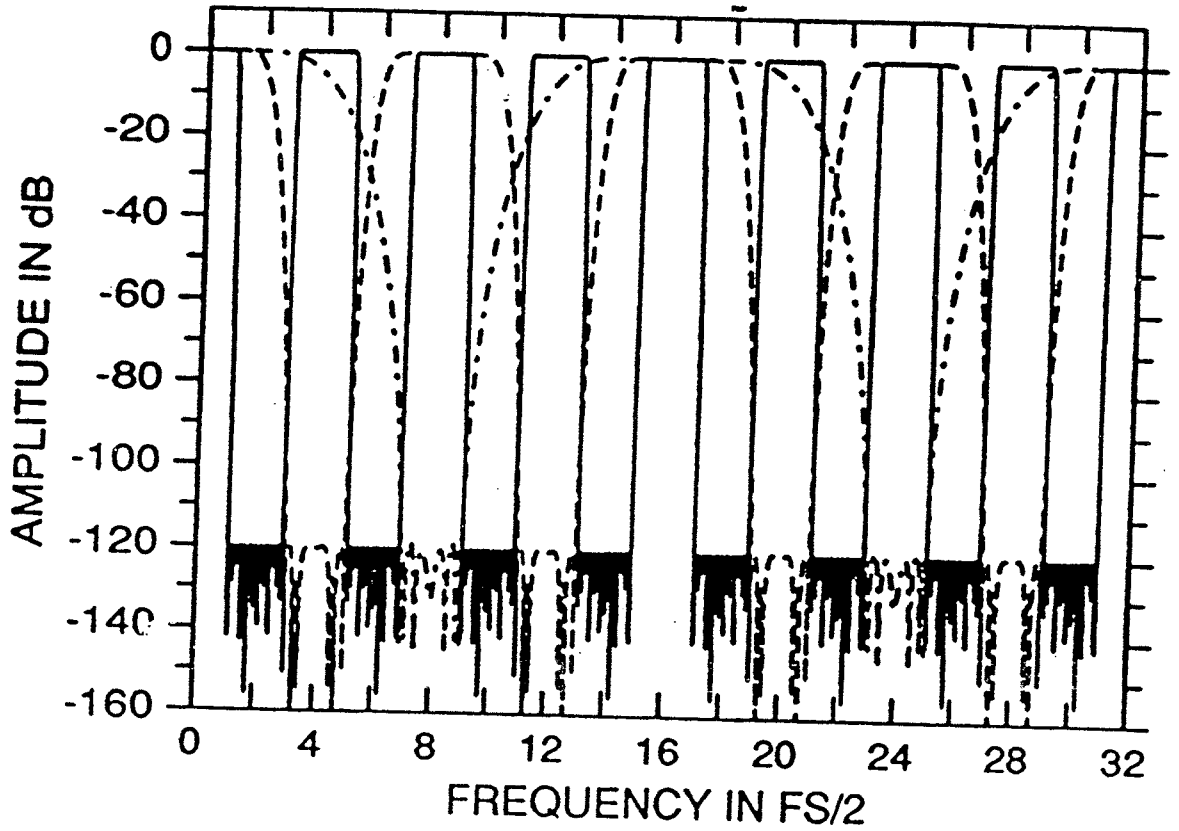
RUNNING SUM G(z)





FILTER $H_1(z)$


IN
2fs
F1(z)
F1(z)
F1(z)
a2
a1
OUT
fs
z-1
1/2
z-3


AMPLITUDE IN dB
0
-20
-40
-60
-80
-100
-120
-140
-160
0
4
8
12
16
20
24
28
32
FREQUENCY IN FS/2

FILTER $H_1(z)$

Required ripples: $\delta_p = \delta_s = 0.000001$

$$H_1(z) = 2^{-1}z^{-9} + \widehat{H}_1(z^2),$$

where

$$\widehat{H}_1(z) = F_1(z)[(2^{-1} + 2^{-2})z^{-3} + (-2^{-2} + 2^{-20})[F_1(z)]^2]$$

with

$$\begin{aligned} F_1(z) =& (-2^{-4} - 2^{-7} - 2^{-11})(1 + z^{-3}) \\ &+ (2^{-1} + 2^{-4} + 2^{-7})(z^{-1} + z^{-2}). \end{aligned}$$

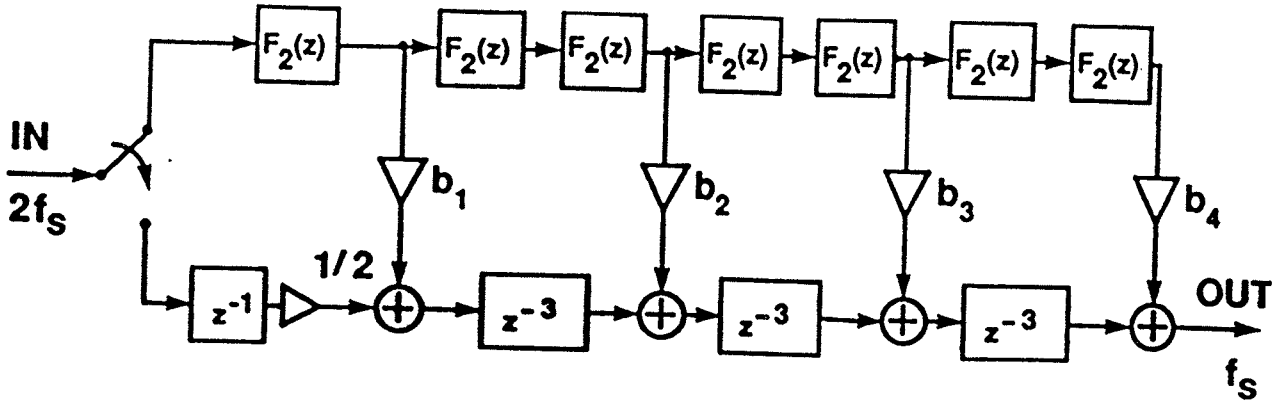
Required ripple for $F_1(z)$ is 0.0016.

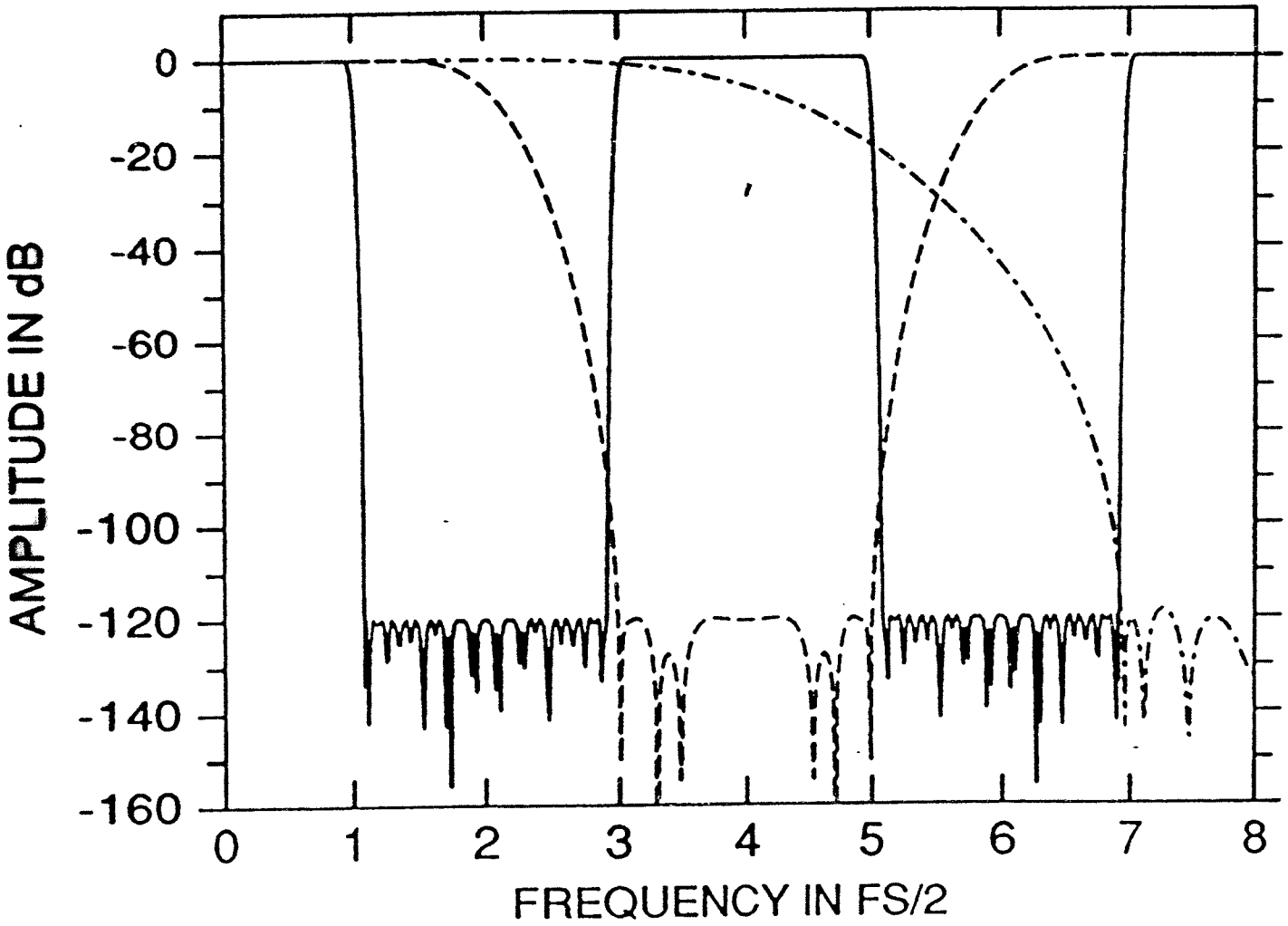
COMPARISON WITH CONVENTIONAL DIRECT-FORM HALF-BAND FILTER

$$H_1(z) = 2^{-1}z^{-7} + \widehat{H}_1(z^2)$$

- $\widehat{H}_1(z)$ is of order 7
- 22 decimal bits are required compared to 11 bits
- The overall order is 20 percent lower
- Longer filter has to be implemeted instead of several copies of a low-order subfilter

FILTER $H_2(z)$


IN
2fS
F2(z)
F2(z)
F2(z)
F2(z)
F2(z)
F2(z)
F2(z)
b1
b2
b3
b4
1/2
z-1
z-3
z-3
z-3
OUT
fS


AMPLITUDE IN dB
0
-20
-40
-60
-80
-100
-120
-140
-160
0
1
2
3
4
5
6
7
8
FREQUENCY IN FS/2

FILTER $H_2(z)$

Required ripples: $\delta_p = \delta_s = 0.000001$

$$H_2(z) = 2^{-1}z^{-21} + \widehat{H}_2(z^2),$$

where

$$\begin{aligned}\widehat{H}_2(z) =& F_2(z)[(2^0 + 2^{-4} + 2^{-5})z^{-9} \\ &+ (-2^0 - 2^{-4} - 2^{-5})z^{-6}[F_2(z)]^2 \\ &+ (2^{-1} + 2^{-3} + 2^{-5})z^{-3}[F_2(z)]^4 \\ &+ (-2^{-3} - 2^{-5} + 2^{-20})[F_2(z)]^6]\end{aligned}$$

with

$$\begin{aligned}F_2(z) =& (-2^{-4} - 2^{-5})(1 + z^{-3}) \\ &+ (2^{-1} + 2^{-4} + 2^{-5})(z^{-1} + z^{-2}).\end{aligned}$$

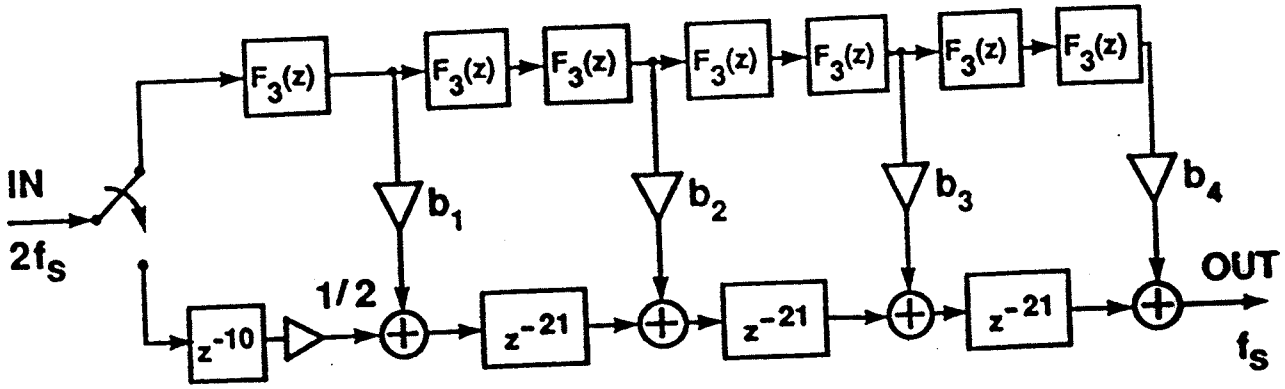
Required ripple for $F_3(z)$ is 0.031.

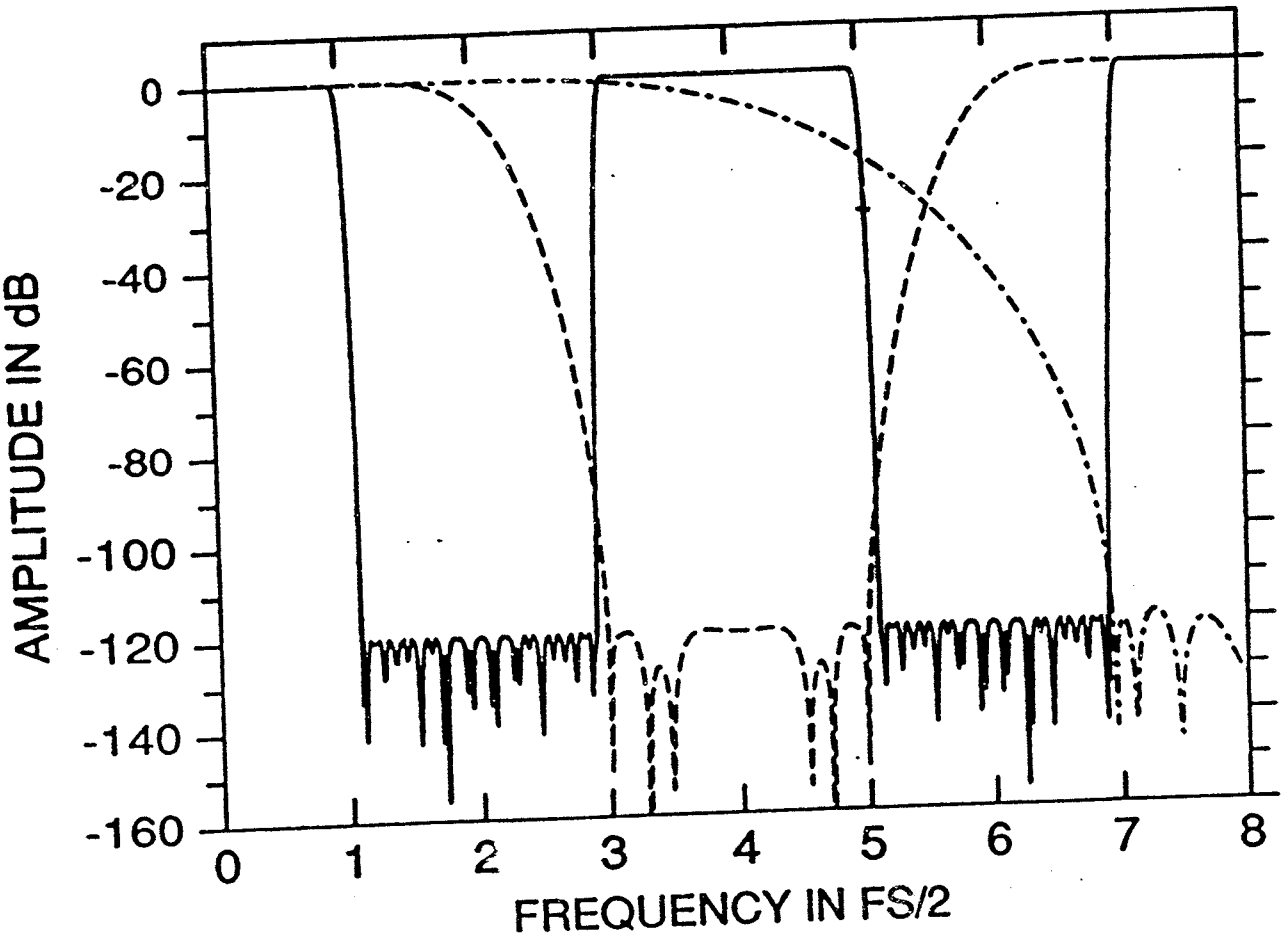
COMPARISON WITH CONVENTIONAL DIRECT-FORM HALF-BAND FILTER

$$H_2(z) = 2^{-1}z^{-15} + \widehat{H}_2(z^2)$$

- $\widehat{H}_2(z)$ is of order 15
- 21 decimal bits are required compared to 5 bits
- The overall order is 30 percent lower
- Long filter has to be implemeted instead of several copies of a low-order subfilter

FILTER $H_3(z)$


IN
2fS
F3(z)
F3(z)
F3(z)
F3(z)
F3(z)
F3(z)
F3(z)
b1
b2
b3
b4
1/2
z-10
z-21
z-21
z-21
OUT
fS


AMPLITUDE IN dB
0
-20
-40
-60
-80
-100
-120
-140
-160
0
1
2
3
4
5
6
7
8
FREQUENCY IN FS/2

FILTER $H_3(z)$

Required ripples: $\delta_p = \delta_s = 0.000001$

$$H_3(z) = 2^{-1}z^{-147} + \widehat{H}_3(z^2)$$

where

$$\begin{aligned}\widehat{H}_3(z) =& F_3(z)[(2^0 + 2^{-4} + z^{-5})z^{-63} \\ &+ (-2^0 - 2^{-4} - 2^{-5})z^{-42}[F_3(z)]^2 \\ &+ (2^{-1} + 2^{-3} + 2^{-5})z^{-21}[F_3(z)]^4 \\ &+ (-2^{-3} - 2^{-5} + 2^{-20})[F_3(z)]^6]\end{aligned}$$

with

$$\begin{aligned}F_3(z) =& 2^{-6}(1 + z^{-21}) + (-2^{-7} - 2^{-8})(z^{-1} + z^{-20}) \\ &+ 2^{-6}(z^{-2} + z^{-19}) + (-2^{-6} - 2^{-7})(z^{-3} + z^{-18}) \\ &+ 2^{-5}(z^{-4} + z^{-17}) + (-2^{-5} - 2^{-7} - 2^{-8})(z^{-5} + z^{-16}) \\ &+ (2^{-4} - 2^{-8})(z^{-6} + z^{-15}) + (-2^{-4} - 2^{-6} - 2^{-8})(z^{-7} + z^{-14}) \\ &+ (2^{-3} - 2^{-8})(z^{-8} + z^{-13}) + (-2^{-2} + 2^{-5} + 2^{-7})(z^{-9} + z^{-12}) \\ &+ (2^{-1} + 2^{-3} + z^{-7})(z^{-10} + z^{-11})\end{aligned}$$

Required ripple for $F_3(z)$ is 0.031.

COMPARISON WITH CONVENTIONAL DIRECT-FORM HALF-BAND FILTER

$$H_3(z) = 2^{-1} z^{-81} + \widehat{H}_3(z^2)$$

- $\widehat{H}_3(z)$ is of order 81
- 24 decimal bits are required compared to 8 bits
- The overall order is 40 percent lower
- Long filter has to be implemeted instead of several copies of a low-order subfilter

EQUALIZER $H_4(z)$

$$H_4(z) = z^{-7} + (2^{-6} + 2^{-10})\widehat{H}_4(z),$$

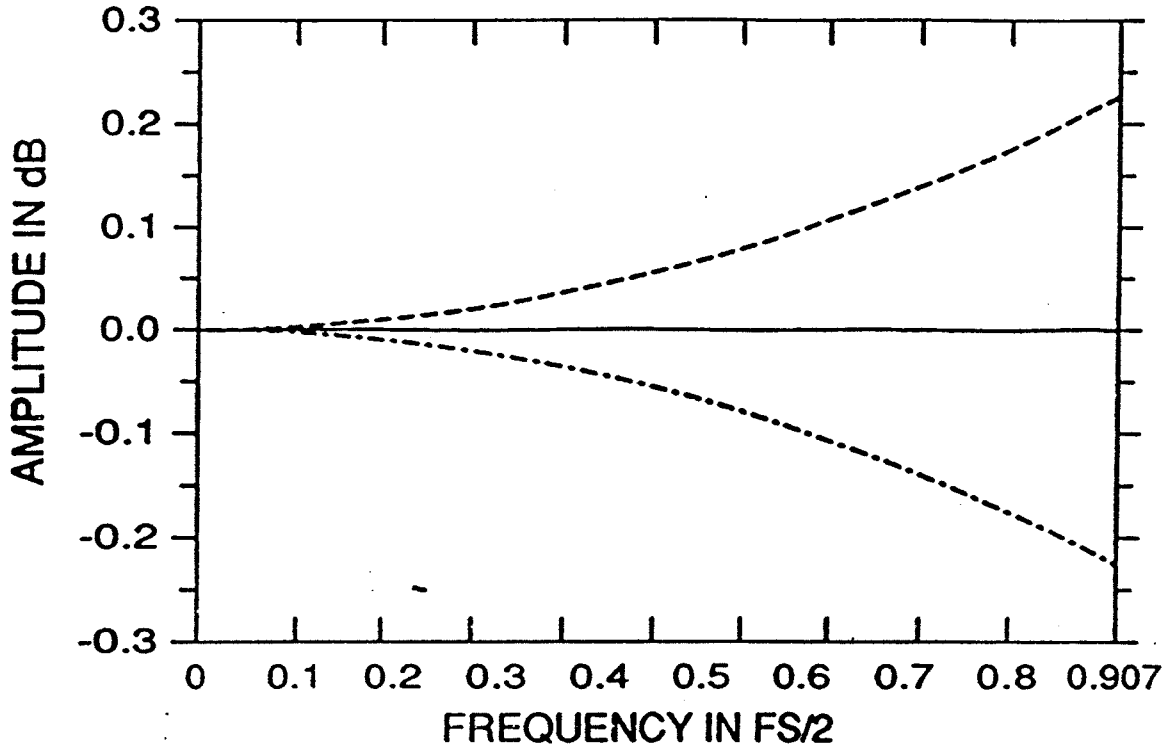
where

$$\begin{aligned}
\widehat{H}_4(z) =& (-2^{-8} - 2^{-11})(1 + z^{-14}) \\
&+ (2^{-8} + 2^{-9} + 2^{-11})(z^{-1} + z^{-13}) \\
&+ (-2^{-7} - 2^{-8} + 2^{-11})(z^{-2} + z^{-12}) \\
&+ (2^{-6} + 2^{-8} + 2^{-9})(z^{-3} + z^{-11}) \\
&+ (-2^{-5} - 2^{-7} - 2^{-8})(z^{-4} + z^{-10}) \\
&+ (2^{-3} - 2^{-6} - 2^{-8})(z^{-5} + z^{-9}) \\
&+ (-2^{-1} + 2^{-4} + 2^{-9})(z^{-6} + z^{-8}) \\
&+ (2^{-1} + 2^{-2} - 2^{-5})z^{-7}.
\end{aligned}$$

The equalizer is implemented in the above form based on the fact the central impulse response value is very close to unity and the other values are very small

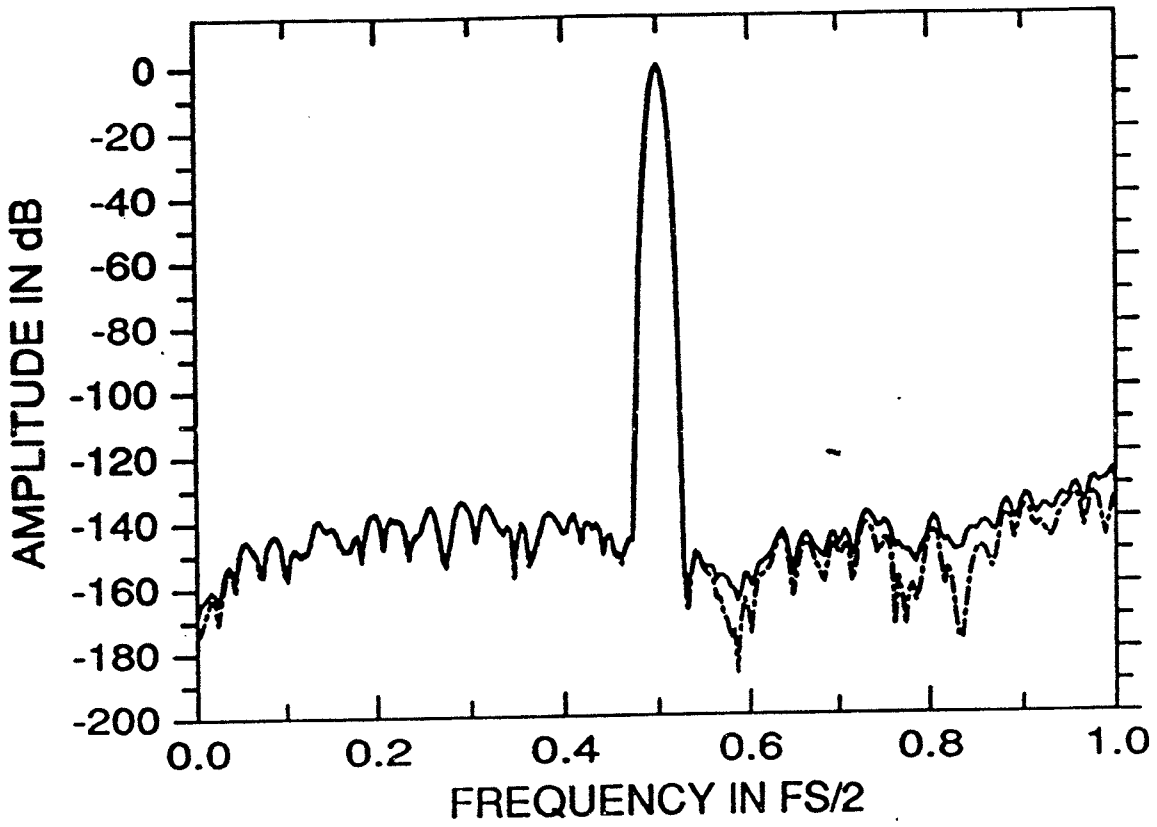
EQUALIZER $H_4(z)$

$$H_4(z) = z^{-7} + (2^{-6} + 2^{-10})\widehat{H}_4(z),$$



5TH ORDER DECIMATED NOISE SPECTRA IN BASEBAND

* Aliased quantization noise from 5th order Sigma Delta modulator after decimation by 64

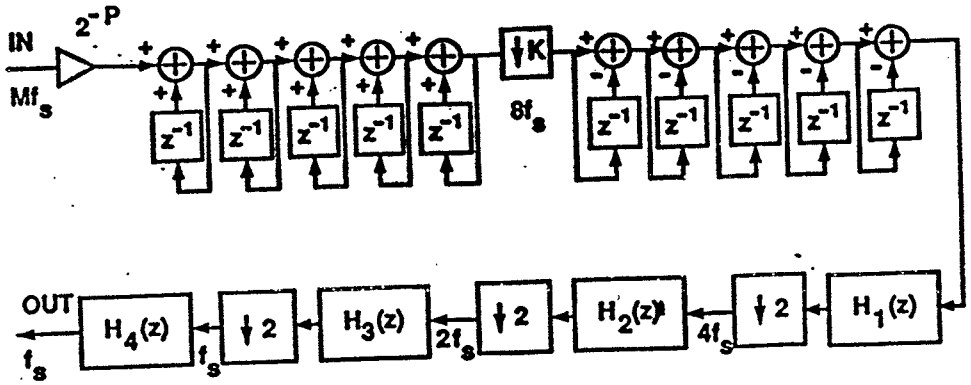


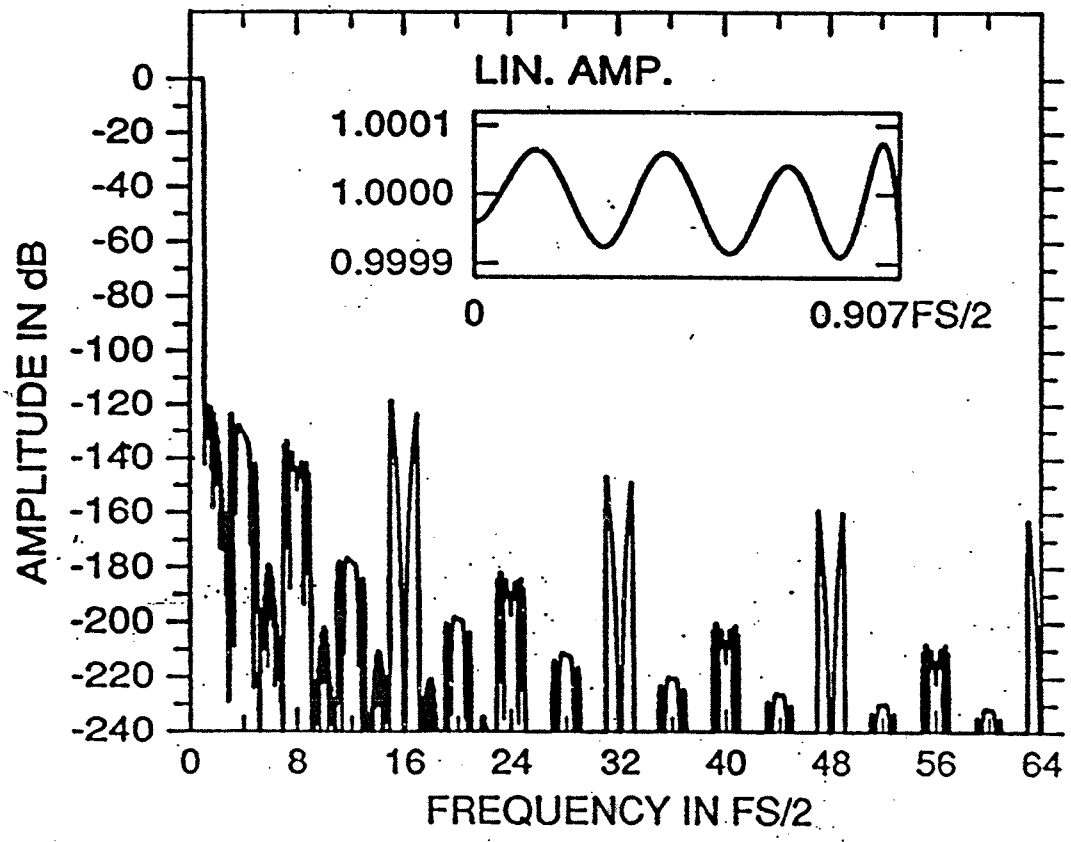
CONCLUSIONS

$$1+\delta_p \le |H(e^{j2\pi f/(Mf_s)})| \le 1-\delta_p \quad \text{for} \quad 0 \le f \le \alpha\frac{f_s}{2}$$

$$|H(e^{j2\pi f/(Mf_s)})| \le \delta_s \quad \text{for} \quad (2-\alpha)\frac{f_s}{2} \le f \le M\frac{f_s}{2},$$

$\delta_p = 0.0001$ and $\delta_s = 0.000001$ (120-dB attenuation).





Multiplier-Free Decimator Algorithms for Superresolution Oversampled Converters

Tapio Saramäki, Teppo Karema, Tapani Ritoniemi, and Hannu Tenhunen

Signal Processing Laboratory
Tampere University of Technology
P. O. Box 527, SF-33101 Tampere, Finland

Abstract – This paper introduces a class of efficient linear-phase FIR decimators for attenuating the out-of-band noise generated by a high-order sigma-delta analog-to-digital modulator. The stopband attenuation of these decimators is more than 120 dB. The decimators contain no general multipliers and a few data memory locations, thereby making them easily VLSI-realizable. This is achieved by using several decimation stages with each stage containing a small number of delays and arithmetic operations. Some of the stages have been constructed using low-order building blocks which are combined to give a selective filter using a few additional tap coefficients and adders. The output sampling rate of these decimators is the minimum possible one and the proposed decimators can be used, with very slight changes, for many oversampling ratios. Futhermore, these decimators attenuate highly the undesired out-of-band signal components of the input signal, thus significantly relaxing the anti-aliasing prefilter requirements.

I. Introduction

Efficient high resolution analog-to-digital conversion is obtained by using oversampled sigma-delta modulation with one-bit quantization. Modulation together with oversampling moves most of the quantization noise out of the baseband. The noise lying out of the baseband can then be reduced by using a decimator.

The bottleneck in superresolution sigma-delta A/D conversion has been so far the analog sigma-delta modulator. A high-order noise shaping is needed to achieve both a high resolution and a wide baseband [1], [2]. The main problem with high-order modulators is their stability, which seems to be solved in few years resulting modulator structures that will reach the technology limits [3]. It is obvious that high-quality digital decimators are needed in such systems. In order to effectively attenuate the out-of-band noise and to simulatenously serve as a selective anti-alias filter, the stopband attenuation of the decimator must be high. This results in a high filter order.

Recently, the authors have proposed a class of high performance linear-phase FIR decimator structures which can be easily integrated in a small area [4], [5]. Similar structures for efficient interpolators have been given in [6]. The proposed decimators have been designed to work at the output of a second-order sigma-delta modulator. In order to optimize both the decimator performance (noise, baseband frequency) and the VLSI realizability (circuit area, power, speed), the proposed decimators have been designed to consist of several stages with each stage requiring a small number of arithmetic operations. The optimization is performed in such a way that no general multipliers are required. To achieve this goal, some of the filter stages are constructed as a tapped cascaded interconnection of low-order subfilter. The overall filter is constructed using a fixed part and an adjustable part. With slight changes in the adjustable filter part, the overall filter can be used for many oversampling ratios. Moreover, the output sampling rate is the minimum possible one and the proposed decimators attenuate highly the undesired input signal components lying out of the baseband, thereby relaxing the anti-aliasing prefilter requirements.

In this paper, we show how ideas similar to those given in [4]–[6] can be used for designing very selective decimators for high-order sigma-delta modulators which have enhanced noise shaping characteristics. If the resolution of the overall converter is desired to be at least 20 bits, then it is advisable to have at least a 120-dB stopband attenuation, whereas the maximum passband deviation from unity for the amplitude response is desired to be less than 0.0001. In such cases, conventional FIR filters require very many bits for coefficient representations but the proposed decimator algorithm can still be determined such that no general multipliers are needed. This makes the proposed filter structures even more attractive in cases where very high selectivity is needed.

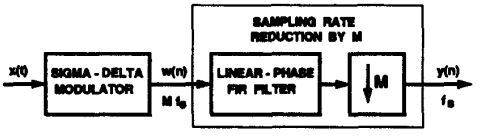


Fig. 1. *Block diagram for the A/D converter consisting of an oversampled sigma-delta modulator and a decimator filter.*

II. Statement of the Problem

The block diagram for the overall system is depicted in Fig. 1. The output sampling rate of the sigma-delta modulator is M times the final sampling rate f_s. We state the amplitude requirements for the decimator in the form

$$1+\delta_p \le |H(e^{j2\pi f/(Mf_s)})| \le 1-\delta_p \quad \text{for} \quad 0 \le f \le \alpha\frac{f_s}{2} \tag{1a}$$

$$|H(e^{j2\pi f/(Mf_s)})| \le \delta_s \quad \text{for} \quad (2-\alpha)\frac{f_s}{2} \le f \le M\frac{f_s}{2}. \tag{1b}$$

When these specifications are satisfied, then the signal components aliasing into the passband $[0, \alpha f_s/2]$ are attenuated at least by $1/\delta_s$. We consider the following criteria:

$$\delta_p = 0.0001, \quad \delta_s = 0.000001, \quad \alpha = 0.907.$$

In this case, the stopband attenuation is at least 120 dB. α has been selected such that the passband edge becomes 20 kHz for the final output sampling rate of $f_s = 44.1$ kHz.

III. Proposed Class of Decimators

To reduce the arithmetic complexity of the decimator, it is preferred to construct it using several low-order stages, instead of one high-order stage (see, e.g., [7]). A multistage implementation of the proposed decimator is given in Fig. 2. The transfer function of the single-stage equivalent can be written in the form

$$H(z) = H_4(z^M)H_3(z^{M/2})H_2(z^{M/4})H_1(z^{M/8})G(z), \tag{2a}$$

where

$$G(z) = 2^{-P}\left[\frac{1-z^{-K}}{1-z^{-1}}\right]^L, \tag{2b}$$

$$K = M/8, \tag{2c}$$

and

$$P = L\log_2 K. \tag{2d}$$

In Fig. 2, $L = 5$. By changing K, the same decimator structure can be used for many decimation ratios M. $G(z)$ as given by Eq. (2b) is the transfer function from the overall filter input to the input of $H_1(z)$ in the case where the sampling rate reduction by K is not performed. The term in the parentheses in Eq. (2b) can be rewritten in the form

$$\frac{1-z^{-K}}{1-z^{-1}} = \sum_{r=0}^{K-1} z^{-r}.$$

CH2868-8/90/0000-3275$1.00 © 1990 IEEE

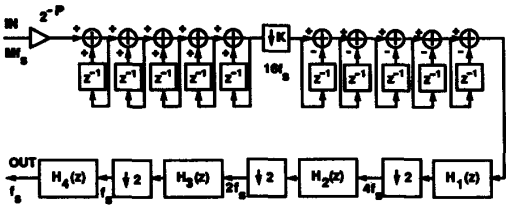


Fig. 2. *Implementation of the proposed decimator.*

Thus this term corresponds to linear-phase FIR filters. Linear-phase filters with transfer functions consisting of the above recursive terms have been used for sampling rate alteration in [8] and [9]. Using the techniques proposed in these papers, we can implement $G(z)$ using the substructures shown in Fig. 2. We note that when the feedforward term $1-z^{-K}$ is transfered after the sampling rate reduction by a factor of K, it becomes $1-z^{-1}$. It should be noted also that if 1's or 2's complement arithmetic (or modulo arithmetic in general) and the worst-case scaling are used, the output values of the filter $G(z)$ implemeted as shown in Fig. 2 are correct even though there may occur internal overflows in the feedback loops realizing the term $1/(1-z^{-1})$. The proofs of this fact can be found in [9] and [10]. Also, under the above conditions, the effect of temporary miscalculations vanishes from the output in finite time and initial resetting is not necessarily needed. The scaling constant 2^{-P} has been selected according to the worst-case scaling.

The design criteria stated in the previous section can be optimally met by selecting five terms in $G(z)$. The explanation to this will be given in the next section. In the following section, we shall show how the filter parts $H_1(z)$, $H_2(z)$, $H_3(z)$, and $H_4(z)$ can be properly designed in such a way that they contain no general multipliers and the overall filter meets the given criteria.

IV. Design of Subfilters

To avoid the use of general multipliers, the subfilters $H_1(z)$, $H_2(z)$, $H_3(z)$, and $H_4(z)$ have been designed to be special tailored filters. The transfer functions are

$$H_1(z) = 2^{-1}z^{-9} + \hat{H}_1(z^2), \tag{3a}$$

where

$$\hat{H}_1(z) = F_1(z)[(2^{-1}+2^{-2})z^{-3} + (-2^{-2}+2^{-20})[F_1(z)]^2] \tag{3b}$$

with

$$\begin{aligned} F_1(z) =& (-2^{-4}-2^{-7}-2^{-11})(1+z^{-3}) \\ &+ (2^{-1}+2^{-4}+2^{-7})(z^{-1}+z^{-2}). \end{aligned} \tag{3c}$$

$$H_2(z) = 2^{-1}z^{-21} + \hat{H}_2(z^2), \tag{4a}$$

where

$$\begin{aligned} \hat{H}_2(z) =& F_2(z)[(2^0+2^{-4}+z^{-5})z^{-9} \\ &+ (-2^0-2^{-4}-2^{-5})z^{-6}[F_2(z)]^2 \\ &+ (2^{-1}+2^{-3}+2^{-5})z^{-3}[F_2(z)]^4 \\ &+ (-2^{-3}-2^{-5}+2^{-20})[F_2(z)]^6] \end{aligned} \tag{4b}$$

with

$$\begin{aligned} F_2(z) =& (-2^{-4}-2^{-5})(1+z^{-3}) \\ &+ (2^{-1}+2^{-4}+2^{-5})(z^{-1}+z^{-2}). \end{aligned} \tag{4c}$$

$$H_3(z) = 2^{-1}z^{-147} + \hat{H}_3(z^2), \tag{5a}$$

where

$$\begin{aligned} \hat{H}_3(z) =& F_3(z)[(2^0+2^{-4}+z^{-5})z^{-63} \\ &+ (-2^0-2^{-4}-2^{-5})z^{-42}[F_3(z)]^2 \\ &+ (2^{-1}+2^{-3}+2^{-5})z^{-21}[F_3(z)]^4 \\ &+ (-2^{-3}-2^{-5}+2^{-20})[F_3(z)]^6] \end{aligned} \tag{5b}$$

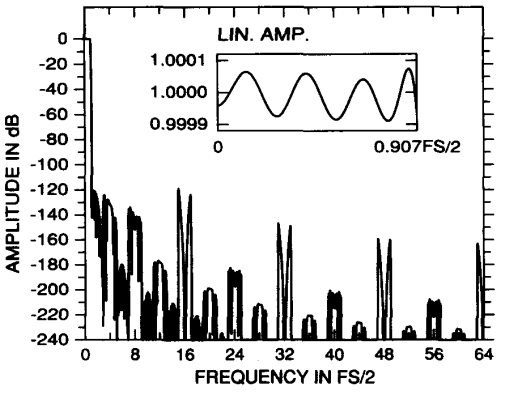


Fig. 3. *Amplitude response for the overall decimator.*

with

$$\begin{aligned} F_3(z) =& 2^{-6}(1+z^{-21}) + (-2^{-7}-2^{-8})(z^{-1}+z^{-20}) \\ &+ 2^{-6}(z^{-2}+z^{-19}) + (-2^{-6}-2^{-7})(z^{-3}+z^{-18}) \\ &+ 2^{-5}(z^{-4}+z^{-17}) + (-2^{-5}-2^{-7}-2^{-8})(z^{-5}+z^{-16}) \\ &+ (2^{-4}-2^{-8})(z^{-6}+z^{-15}) + (-2^{-4}-2^{-6}-2^{-8})(z^{-7}+z^{-14}) \\ &+ (2^{-3}-2^{-8})(z^{-8}+z^{-13}) + (-2^{-2}+2^{-5}+2^{-7})(z^{-9}+z^{-12}) \\ &+ (2^{-1}+2^{-3}+z^{-7})(z^{-10}+z^{-11}). \end{aligned} \tag{5c}$$

$$H_4(z) = z^{-7} + (2^{-6}+2^{-10})\hat{H}_3(z), \tag{6a}$$

where

$$\begin{aligned} \hat{H}_3(z) =& (-2^{-8}-2^{-11})(1+z^{-14}) \\ &+ (2^{-8}+2^{-9}+2^{-11})(z^{-1}+z^{-13}) \\ &+ (-2^{-7}-2^{-8}+2^{-11})(z^{-2}+z^{-12}) \\ &+ (2^{-6}+2^{-8}+2^{-9})(z^{-3}+z^{-11}) \\ &+ (-2^{-5}-2^{-7}-2^{-8})(z^{-4}+z^{-10}) \\ &+ (2^{-3}-2^{-6}-2^{-8})(z^{-5}+z^{-9}) \\ &+ (-2^{-1}+2^{-4}+2^{-9})(z^{-6}+z^{-8}) \\ &+ (2^{-1}+2^{-2}-2^{-5})z^{-7}. \end{aligned} \tag{6b}$$

The amplitude response of the overall design is depicted in Fig. 3 for $M = 64$. The subfilters $H_1(z)$, $H_2(z)$, and $H_3(z)$ are special half-band filters which can be implemeted effectively using a polyphase structure based on the commutative model [11]. The structures resulting by properly sharing the delays between the two branches are shown in Fig. 4. Fig. 4(a) gives the structure for $H_1(z)$. Fig. 4(b) for $H_2(z)$, and Fig. 4(c) for $H_3(z)$. The tap coefficients are $a_1 = 2^{-1}+2^{-2}$, $a_2 = -2^{-2}+2^{-20}$, $b_1 = 2^0+2^{-4}+2^{-5}$, $b_2 = -2^0-2^{-4}-2^{-5}$, $b_3 = 2^{-1}+2^{-3}+2^{-5}$, and $b_4 = 2^{-3}+2^{-5}+2^{-20}$. Note that the tap coefficients are the same for $H_2(z)$ and $H_3(z)$. One of the branches for all the three filters is a pure delay term. For $H_1(z)$, the other branch is a tapped cascaded interconnection of three identical subfilters of order 3, for $H_2(z)$, the other branch consists of seven identical subfilters of order 3, and, for $H_3(z)$, seven identical subfilters of order 21.

$H_3(z)$ has been designed to provide for the overall response at least a 120-dB attenuation on $[(2-\alpha)f_s/2, f_s]$. Because of the periodicity of the response of $H_3(z^{M/2})$ [cf. Eq. (2a)], the desired attenuation is simultaneously achieved also elsewhere on $[0, Mf_s/2]$ except for the "extra" passbands of $H_3(z^{M/2})$ centered at the frequencies $4kf_s/2$ for $k = 1,2,\ldots,M/4$. This is illustrated in Fig. 5, where the solid line gives the response of $H_3(z^{M/2})$ in the interval $[0, 32f_s/2]$. The second periodic half-band filter $H_2(z^{M/4})$ attenuates the first extra passband and some other passbands. The response of this filter is depicted by the dashed line in Fig. 5. It leaves the passbands centered at $8kf_s/2$ for $k = 1,2,\ldots,M/8$. The role of the third periodic filter $H_1(z^{M/8})$ (the response is given by a dash-dotted line in Fig. 5) is to attenuate some of the remaining extra passbands. It leaves the

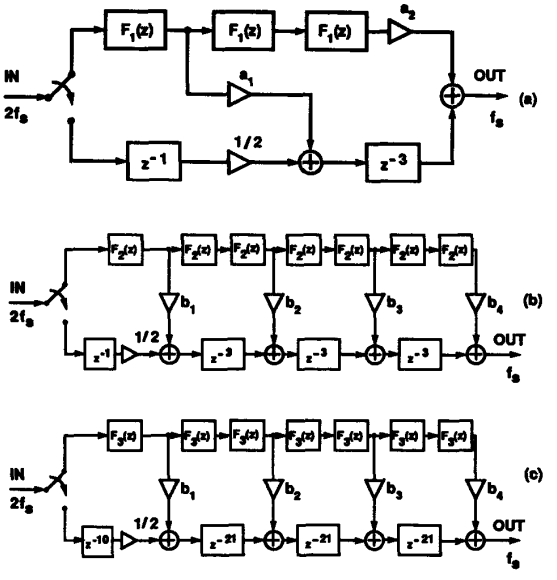


Fig. 4. *Efficient implementations for the half-band subfilters. (a) $H_1(z)$. (b) $H_2(z)$. (c) $H_3(z)$.*

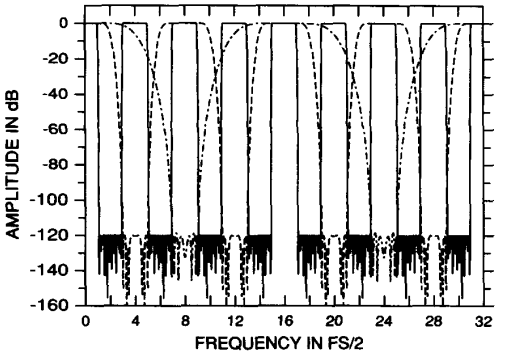


Fig. 5. *Amplitude responses for $H_3(z^{M/2})$ (solid line), $H_2(z^{M/4})$ (dahed line). and $H_1(z^{M/8})$ (dash-dotted line) in the interval $[0, 32f_s/2]$.*

passband centered at $16kf_s/2$ for $k = 1, 2, \ldots, M/16$. The remaining passbands are then attenuated by $G(z)$. This is illustrated in Fig. 6, where the thicker and thinner lines give the responses of $G(z)$ and $H_4(z^M)H_3(z^{M/2})H_2(z^{M/4})H_1(z^{M/8})$, respectively. When five terms are used in $G(z)$, the lowest attenuation of the peak just before the frequency $f = 16f_s/2$ is just 120 dB (see Fig. 3).

The last filter $H_4(z)$ has been designed to equalize the passband response within the given limits. Since the passband ripples of the half-band subfilters are very small, it can concentrate on equalizing the distortion caused by $G(z)$. Figure 7 gives the responses for $G(z)$ (dash-dotted line), $H_4(z^M)$ (dashed line), and $G(z)H_4(z^M)$ (solid line) in the passband $[0, 0.907f_s/2]$. It should be noted that the performance of the overall filter remains in the low frequencies practically the same as M is varied, enabling us to use the same four fixed filters for various values of M. Only the decimation ratio K needs to be changed.

The actual design of $H_1(z)$, $H_2(z)$, and $H_3(z)$ has been accomplished by properly modifying the methods proposed in [12] for optimally designing FIR filters as a tapped cascaded interconnection of identical subfilters. The decription of the resulting synthesis technique falls outside of the scope of this paper (this will be a subject of another paper).

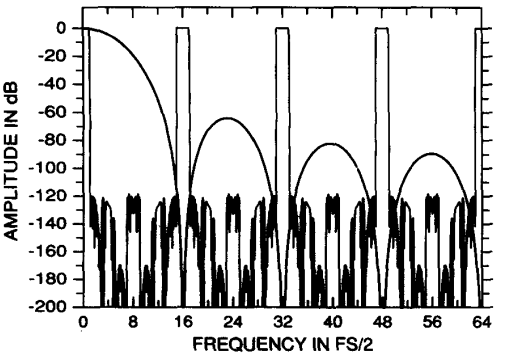


Fig. 6. *Amplitude responses for $G(z)$ (thicker line) and for $H_4(z^M)H_3(z^{M/2})H_2(z^{M/4})H_1(z^{M/8})$ (thinner line).*

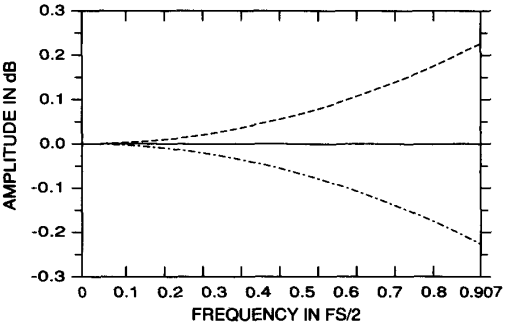


Fig. 7. *Amplitude responses for for $G(z)$ (dash-dotted line), $H_4(z^M)$ (dashed line), and $G(z)H_4(z^M)$ (solid line) in the passband $[0, 0.907f_s/2]$.*

The main advantage of using identical subfilters lies in the fact that it enables us to find an overall filter in such a way that there are no general multipliers. If conventional half-band filters are used for implementing $H_1(z)$, $H_2(z)$, and $H_3(z)$, the required orders of the direct-form filters $\hat{H}_1(z)$, $\hat{H}_2(z)$, and $\hat{H}_3(z)$ [cf. Eqs. (3b), (4b), and (5b)] are 7, 15, and 81, respectively. The delay terms in Eqs. (3a), (4a), and (5a) are in theses cases z^{-7}, z^{-15}, and z^{-81}, respectively. If direct rounding is used, the minimum number of decimal bits required for the coefficient representations are for these designs 22, 21, and 24, respectively. Because of a large silicon area required for implemeting a long general multiplier, the overall area for the VLSI-implementation of these filters is significantly more than that for the proposed design.

V. Filter Architecture

The filter structures described above lead directly to a very efficient VLSI implementation for the overall decimator. If a fourth or fifth order sigma-delta modulator is used, then an overall decimation ratio of $M = 64$ is enough to achieve a 20-bit resolution. This is illustrated in Fig. 8, where the dot-dashed and solid lines give the simulated baseband spectra at the output of a fifth-order modulator [3] and after filtering and decimation, respectively. As seen from this figure, the contribution of the aliased components to the overall noise is negligible.

For the decimation ratio of $M = 64$, the first filter stage $G(z)$ reduces the sampling rate by $K = 8$. This filter stage can be realized by using a mixed bit-parallel and bit-serial architecture for maximum layout compactness and speed. The layout is generated automatically from the system level specifications using parametrized layout generators (for details, see [13]). In this case, a 16-bit wordlength is required and the serial FIR modules must operate at the rate which is two times the modulator's sampling rate. For the remaining filters, an efficient circuit implementation can be achived if a dynamic one-transistor RAM can be used for data memory locations. Since com-

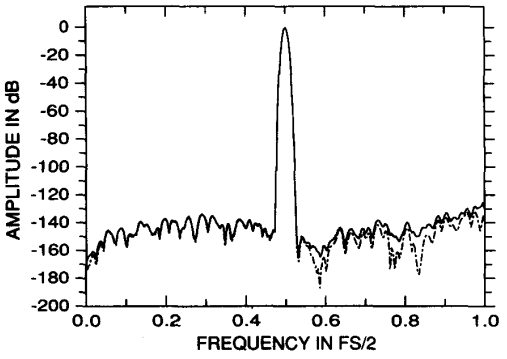


Fig. 8. *Noise spectra in the baseband $[0, f_s/2]$ for a fifth-order sigma-delta modulator [3]. The solid line gives the spectrum at the overall system consisting the modulator and the proposed decimator and the dot-dashed line gives the spectrum at the output of the modulator.*

plicated multiplication operations have been eleminated by the filter design, a simple dedicated filter processor [13] can be adapted with one-transistor RAM structures. Because of the constant data flow in RAM, no special refreshing circuitry is needed. The multiplication operations in the proposed filter algorithm can be implemented using three parallel shifters and one three-input adder/subtracter. A four-word FIFO buffer is needed at the processor's input in order to get the maximum throughput. The processor must operate at speed being 8 times higher (22.5 MHz) than that of the modulator. The processor's memory is addressed by three register: The of1 and of2 registers form a pointer inside the subfilter and the seg register determines the base address which this pointer is added to. The physical RAM location to be addressed is:

$$\text{address} = \text{seg} + [(\text{of1} - \lfloor \text{of2}/k_1 \rfloor) \bmod k_2],$$

where

$$k_1 = \begin{cases} 1 & \text{for } H_1(z) \\ 2 & \text{for } H_2(z) \\ 4 & \text{for } H_3(z) \text{ and } H_4(z), \end{cases}$$

$$k_2 = \begin{cases} 4 & \text{for } F_1(z) \text{ and } F_2(z) \\ 22 & \text{for } F_3(z) \\ 15 & \text{for } H_4(z), \end{cases}$$

and $\lfloor x \rfloor$ stands for integer part of x. In this way, the of2 register creates a ring buffer that is incremented correctly in spite of the decimation. The proposed addressing mechanism provides an efficient usage of RAM because only 3 additional memory locations are needed for the stack. The ROM is also very compact because the decimators $H_1(z)$, $H_2(z)$ and $H_3(z)$ repeat calls to the same subfilter subprogram. Moreover, the ALU is very small compared to a conventional multiplier that would require at least 24-bits operating at the same rate. As a whole, the area for the filter processor is dominated by the RAM and ROM modules. For the proposed algorithm, the RAM is 224 20-bit words, the ROM is about 100 20-bit words and the ALU has a 32-bit internal accuracy.

VI. Conclusion

An efficient linear-phase FIR filter structure has been proposed for eliminating the out-of-band noise generated by a high-order sigma-delta analog-to-digital converter. The main advantages of the proposed filter structure are:

1. It can be easily implemented in CMOS VLSI.
2. The quantization noise generated by the sigma-delta modulator is effectively attenuated.
3. Input signal components, such as possible sinusoidal components, aliasing into the passband are highly attenuated, thus relaxing the anti-aliasing prefilter requirements.
4. The same structure can be used for many oversampling ratios.
5. The overall filter structure contains no general multipliers. This enables us to implement the filter using a simple and small-area processor architecture.

Acknowledgements

This work has been supported by the National Microelectronics Program of Finland. The authors wish also to thank the Median-Free Group International for excellent working atmosphere and fruitful discussions during the course of this work.

References

[1] T. Karema, T. Ritoniemi H. Tenhunen, "Fourth order sigma-delta modulator circuit for digital audio and ISDN applications", in *Proc. IEE European Circuit Theory and Design*, Sep 1989, pp. 223-227.

[2] K. Uchimura, T. Hayashi, T. Kimura and A. Iwata, "Oversampling A-to-D and D-to-A converters with multistage noise shaping modulators," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 36, pp. 1899–1905, Dec. 1988.

[3] T. Ritoniemi T. Karema, and H. Tenhunen, "Design of stable high order 1-bit sigma-delta modulators," in *Proc. 1990 Int. Symp. Circuits Syst.* (New Orleans, Louisiana), this conference.

[4] T. Saramäki and H. Tenhunen, "Efficient VLSI-realizable decimators for a sigma-delta analog-to-digital converter," in *Proc. IEEE Int. Symp. Circuits Syst.* (Espoo, Finland), pp. 1525–1528, June 1988.

[5] T. Saramäki, H. Palomäki, and H. Tenhunen, "Multiplier-free decimators with efficient VLSI implementation for Sigma-Delta A/D converters," presented in *IEEE Workshop on VLSI Signal Processing* (Monterey, CA), Nov. 1988; included in *VLSI Signal Processing III*, edited by R. W. Brodersen and H. S. Moscovitz, New York: IEEE Press, 1988, pp. 523–534.

[6] T. Saramäki, T. Karema, T. Ritoniemi, J. Isoaho, and H. Tenhunen, "VLSI-realizable multiplier-free interpolators for sigma-delta D/A-converters" (invited paper), in *Proc. International Conference on Circuits and Systems* (Nanjing, China), pp. 60–63, July 1989.

[7] T. Saramäki, "Design of optimal multistage IIR and FIR filters for sampling rate alteration", in *Proc. IEEE Int. Symp. Circuits Syst.* (San Jose, CA), pp. 227–230, May 1986.

[8] T. Saramäki, "Efficient recursive digital filters for sampling rate conversion," in *Proc. IEEE Int. Symp. Circuits Syst.* (Newport Beach, CA), pp. 1322-1326, May 1983.

[9] S. Chu and C. S. Burrus, "Multirate filter design using comb filters," *IEEE Trans. Circuits Syst.*, vol. CAS-31, pp. 913–924, Nov. 1984.

[10] T. Saramäki, Y. Neuvo, and S. K. Mitra, "Design of computationally efficient interpolated FIR filters," *IEEE Trans. Circuits Syst.*, pp. 70–88, vol. CAS-35, Jan. 1988.

[11] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[12] T. Saramäki, "Design of FIR filters as a tapped cascaded interconnection of identical subfilters," *IEEE Trans. Circuits Syst.*, vol. CAS-34, pp. 1011–1029, Sept. 1987.

[13] T. Karema, T. Ritoniemi, and H. Tenhunen, "An oversampled sigma-delta A/D converter circuit using two-stage fourth order modulator", in *Proc. 1990 Int. Symp. Circuits Syst.* (New Orleans, Louisiana), this conference.