DESIGN OF DIGITAL FILTERS AND FILTER BANKS BY OPTIMIZATION: APPLICATIONS

Tapio Saramäki and Juha Yli-Kaakinen
Signal Processing Laboratory,
Tampere University of Technology,
Finland
e-mail: ts@cs.tut.fi

Tampere University of Technology
Signal Processing Laboratory

ABSTRACT

This paper emphasizes the usefulness and the flexibility of optimization for finding optimized digital signal processing algorithms for various constrained and unconstrained optimization problems. This is illustrated by optimizing algorithms in six different practical applications:

- Optimizing nearly perfect-reconstruction filter banks subject to the given allowable errors,
- minimizing the phase distortion of recursive filters subject to the given amplitude criteria,
- optimizing the amplitude response of pipelined recursive filters,
- optimizing the modified Farrow structure with an adjustable fractional delay,
- finding the optimum discrete values for coefficient representations for various classes of lattice wave digital filters, and
- finding the multiplierless coefficient representations for the linear-phase finite impulse response filters.

WHY THERE IS A NEED TO USE OPTIMIZATION?

Among others, there exist following three reasons:

1) Thanks to dramatic advances in VLSI circuit technology and signal processors, more complicated DSP algorithms can be implemented faster and faster.

- In order to generate effective DSP products, old algorithms have to be reoptimized or new ones should be generated subject to the implementation constraints.

2) All the subalgorithms in the overall DSP product should be of the same quality:

- In the case of lossy coding, nearly perfect-reconstruction filter banks are more beneficial: lower overall delay and shorter filters.

3) There are various problems where one response is desired to optimized subject to the given criteria for other responses:

- A typical example is to design recursive filters such that the phase is made as linear as possible subject to the given amplitude criteria.

TWO-STEP PROCEDURE

It has turned out that the following procedure is very efficient:

1) Find in a systematic simple manner a suboptimum solution.

2) Improve this solution using a general-purpose nonlinear optimization procedure:

- Dutta-Vidyasagar algorithm
- Sequential quadratic programming

Desired Form: Find the adjustable parameters included in the vector Φ to minimize

$$\rho(\Phi) = \max_{1 \leq i \leq I} f_i(\Phi) \tag{1}$$

subject to constraints

$$g_l(\Phi) \leq 0 \quad \text{for} \quad l = 1, 2, \ldots, L \tag{2}$$

and

$$h_m(\Phi) = 0 \quad \text{for} \quad m = 1, 2, \ldots, M. \tag{3}$$

COMMENTS

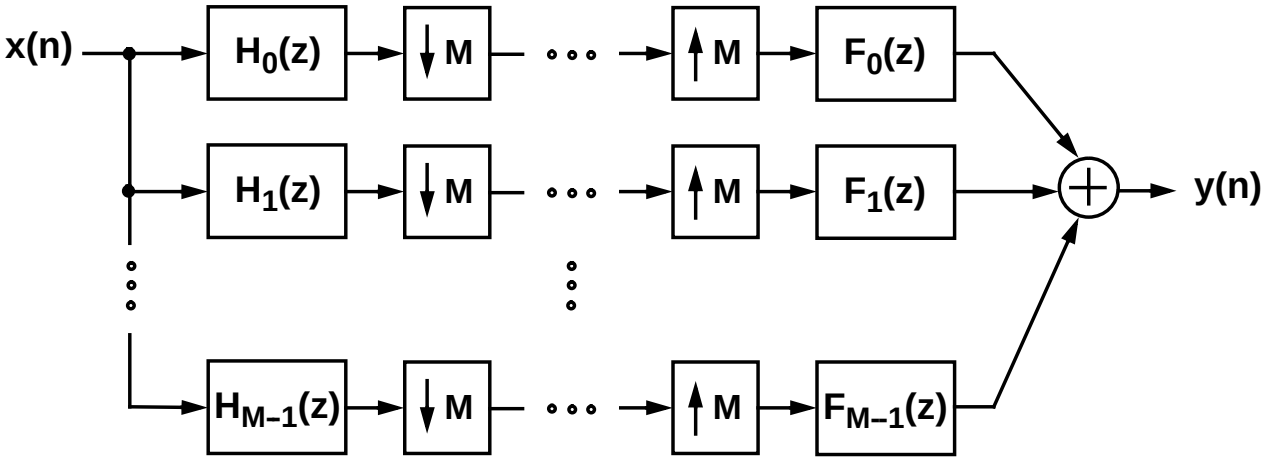
- The proposed two-step procedure is very efficient when a good start-up solution being rather close to the optimum solution can found.

- For each problem under consideration the way of generating this initial solution is very different.

- A good understanding of the problem at hand is needed.

- If a good enough start-up solution cannot be found or there are several local optima, then simulated annealing or genetic algorithms can be used.

NEARLY PERFECT-RECONSTRUCTION COSINE-MODULATED FILTER BANKS



Linear-phase prototype filter:

$$H_p(z) = \sum_{n=0}^{N} h_p(n) z^{-n}, \tag{4}$$

Filters in the bank:

$$h_k(n) = 2h_p(n)\cos\left[(2k+1)\frac{\pi}{2M}\left(n - \frac{N}{2}\right) + (-1)^k\frac{\pi}{4}\right] \tag{5}$$

$$f_k(n) = 2h_p(n)\cos\left[(2k+1)\frac{\pi}{2M}\left(n - \frac{N}{2}\right) - (-1)^k\frac{\pi}{4}\right]. \tag{6}$$

INPUT-OUTPUT RELATION

$$Y(z) = T_0(z)X(z) + \sum_{l=1}^{M-1} T_l(z)X(ze^{-j2\pi l/M}), \quad \text{(7a)}$$

where

$$T_0(z) = \frac{1}{M}\sum_{k=0}^{M-1} F_k(z)H_k(z) \quad \text{(7b)}$$

and for $l = 1, 2, \ldots, M-1$

$$T_l(z) = \frac{1}{M}\sum_{k=0}^{M-1} F_k(z)H_k(ze^{-j2\pi l/M}). \quad \text{(7c)}$$

Here, $T_0(z)$ is the reconstruction transfer function and the remaining ones are aliased transfer functions. It is desired that $T_0(z) = z^{-N}$ and the remaining transfer functions are zero.

STATEMENT OF THE PROBLEMS

Problem I: Given ρ, M, and N, find the coefficients of $H_p(z)$ to minimize

$$E_2 = \int_{\omega_s}^{\pi} |H_p(e^{j\omega})|^2 d\omega, \tag{8a}$$

where

$$\omega_s = (1+\rho)\pi/(2M) \tag{8b}$$

subject to

$$1-\delta_1 \leq |T_0(e^{j\omega})| \leq 1+\delta_1 \quad \text{for} \quad \omega \in [0,\ \pi] \tag{8c}$$

and for $l = 1, 2, \ldots, M-1$

$$|T_l(e^{j\omega})| \leq \delta_2 \quad \text{for} \quad \omega \in [0,\ \pi]. \tag{8d}$$

Problem II: Given ρ, M, and N, find the coefficients of $H_p(z)$ to minimize

$$E_\infty = \max_{\omega \in [\omega_s,\ \pi]} |H_p(e^{j\omega})| \tag{9}$$

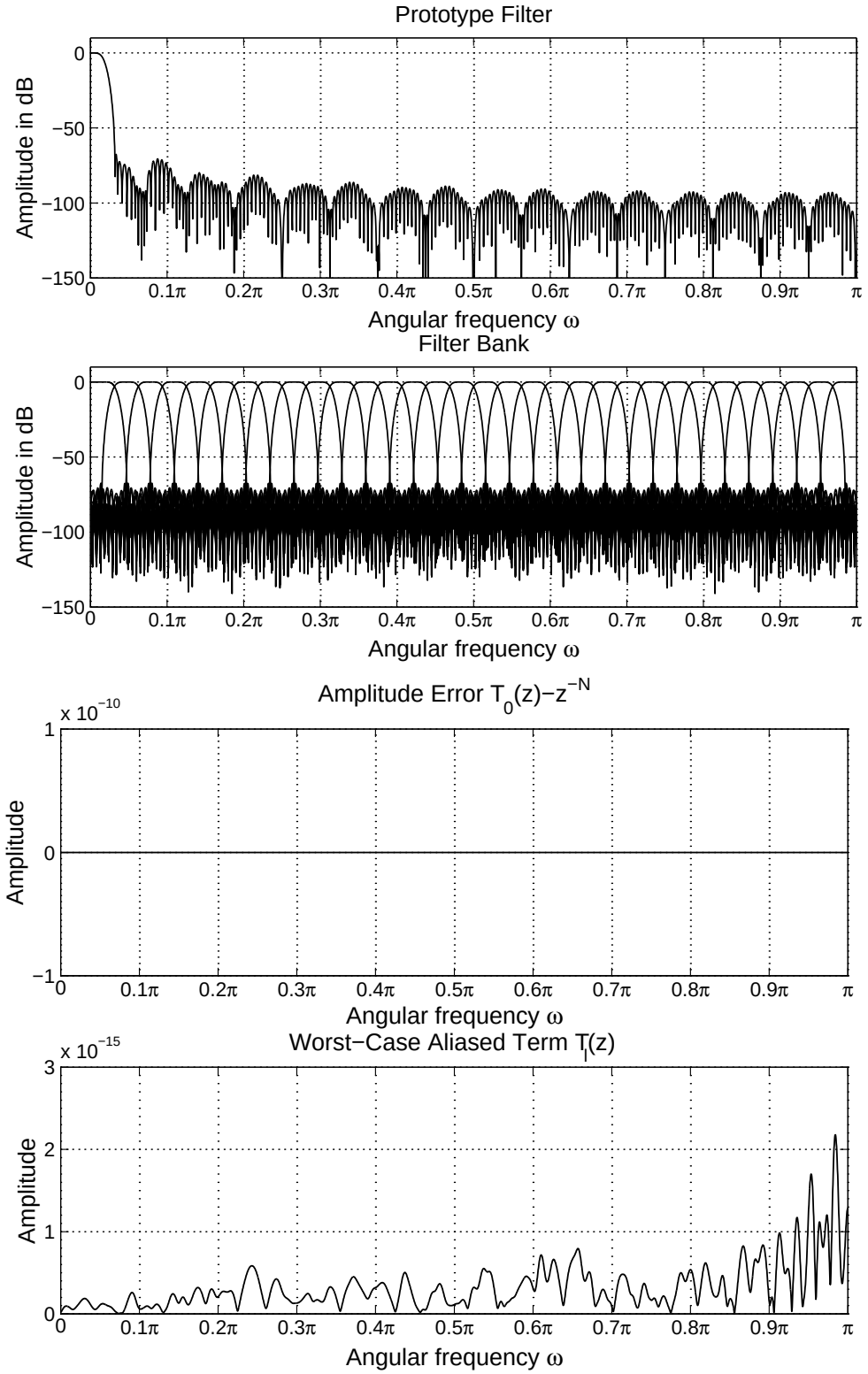
subject to the conditions of Eqs.(8c) and (8d).

COMPARISONS BETWEEN FILTER BANKS WITH $M = 32$ and $\rho = 1$

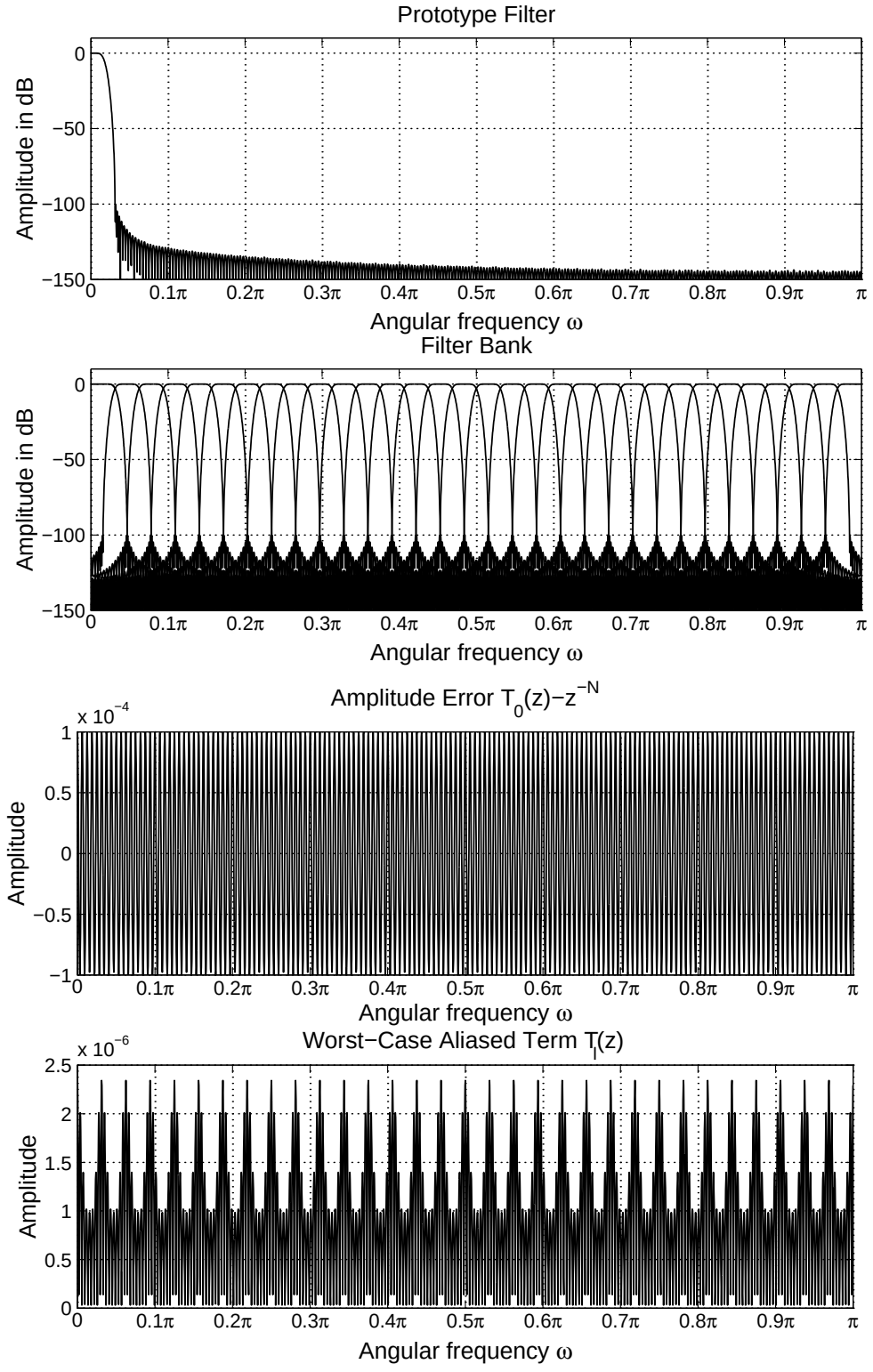
Boldface numbers indicate that these parameters have been fixed in the optimization.

Criterion	K	N	δ_1	δ_2	E_∞	E_2
Least Squared	**8**	**511**	**0**	**0** $-\infty$ dB	$1.2 \cdot 10^{-3}$ -58 dB	$7.4 \cdot 10^{-9}$
Minimax	**8**	**511**	**0**	**0** $-\infty$ dB	$2.3 \cdot 10^{-4}$ -73 dB	$7.5 \cdot 10^{-8}$
Least Squared	**8**	**511**	$\mathbf{10^{-4}}$	$2.3 \cdot 10^{-6}$ -113 dB	$1.0 \cdot 10^{-5}$ -100 dB	$5.6 \cdot 10^{-13}$
Minimax	**8**	**511**	$\mathbf{10^{-4}}$	$1.1 \cdot 10^{-5}$ -99 dB	$5.1 \cdot 10^{-6}$ -106 dB	$3.8 \cdot 10^{-11}$
Least Squared	**8**	**511**	**0**	$9.1 \cdot 10^{-5}$ -81 dB	$4.5 \cdot 10^{-4}$ -67 dB	$5.4 \cdot 10^{-10}$
Least Squared	**8**	**511**	$\mathbf{10^{-2}}$	$5.3 \cdot 10^{-7}$ -126 dB	$2.4 \cdot 10^{-6}$ -112 dB	$4.5 \cdot 10^{-14}$
Least Squared	**6**	**383**	$\mathbf{10^{-3}}$	**0.00001** **-100 dB**	$1.7 \cdot 10^{-4}$ -75 dB	$8.8 \cdot 10^{-10}$
Least Squared	**5**	**319**	$\mathbf{10^{-2}}$	**0.0001** **-80 dB**	$8.4 \cdot 10^{-4}$ -62 dB	$2.7 \cdot 10^{-9}$

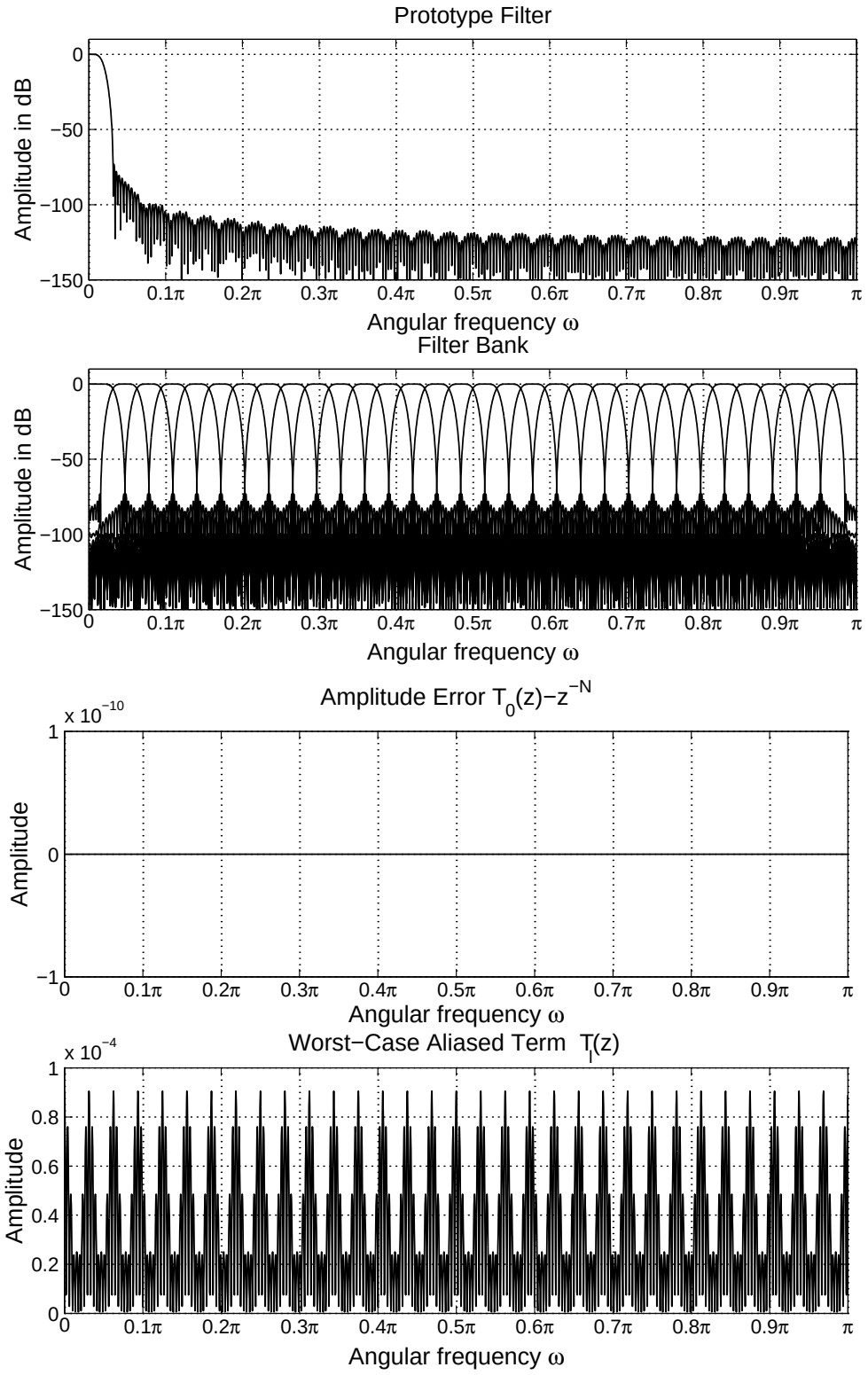
PERFECT-RECONSTRUCTION FILTER BANK with $N = 511$



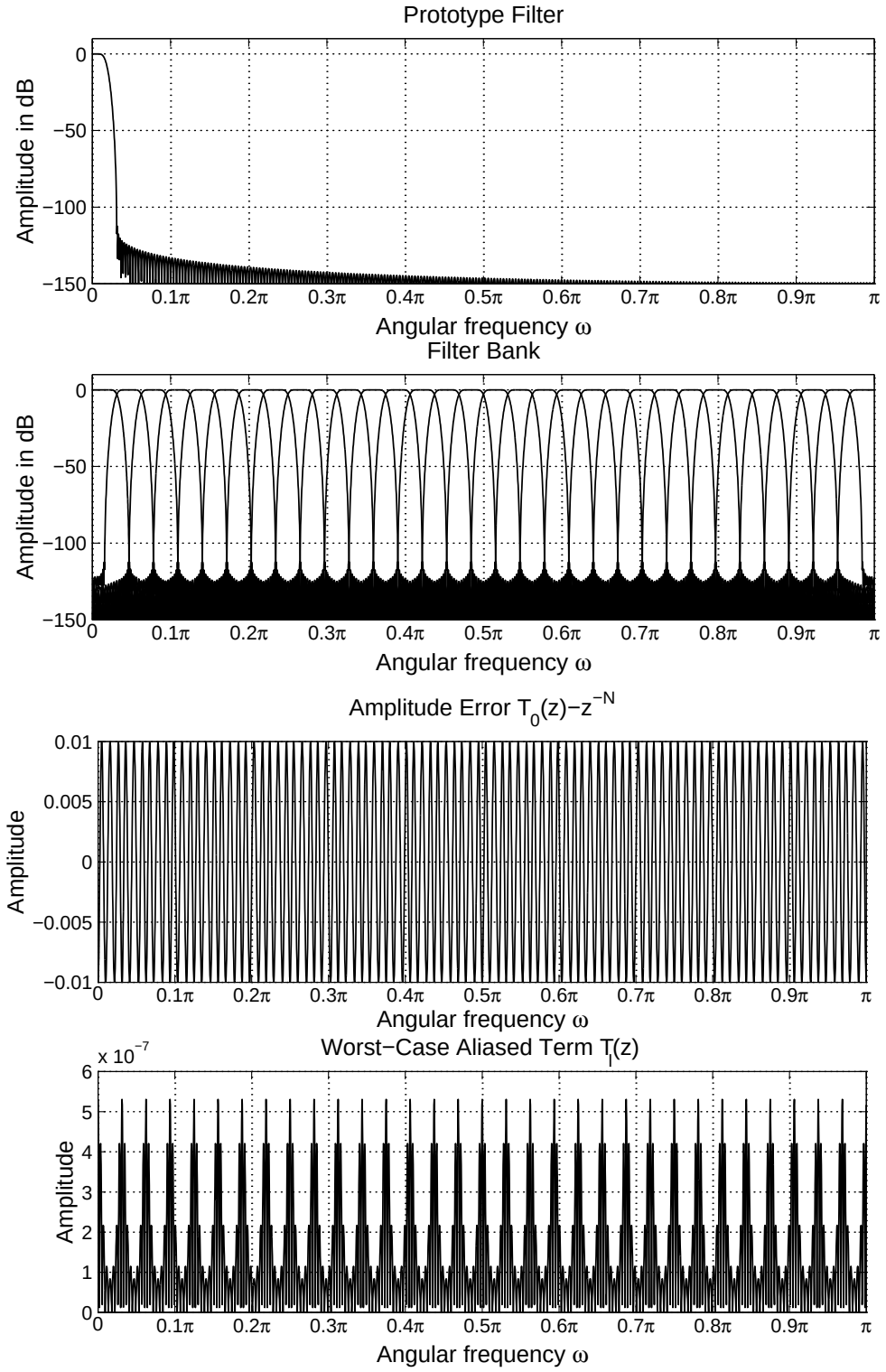
NEARLY PR FILTER BANK with $N = 511$ for $\delta_1 = 0.0001$



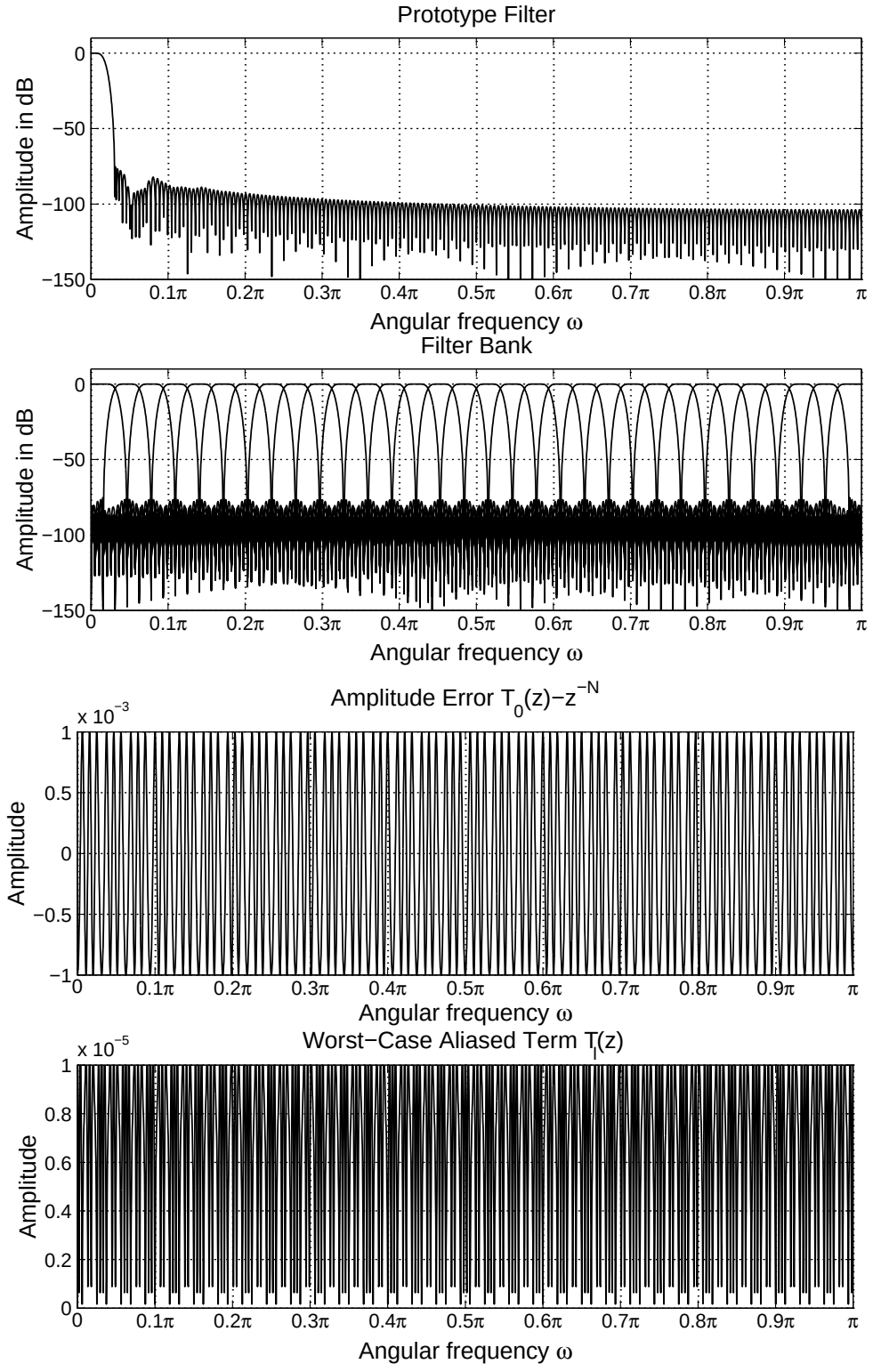
NEARLY PR FILTER BANK with $N = 511$ for $\delta_1 = 0$



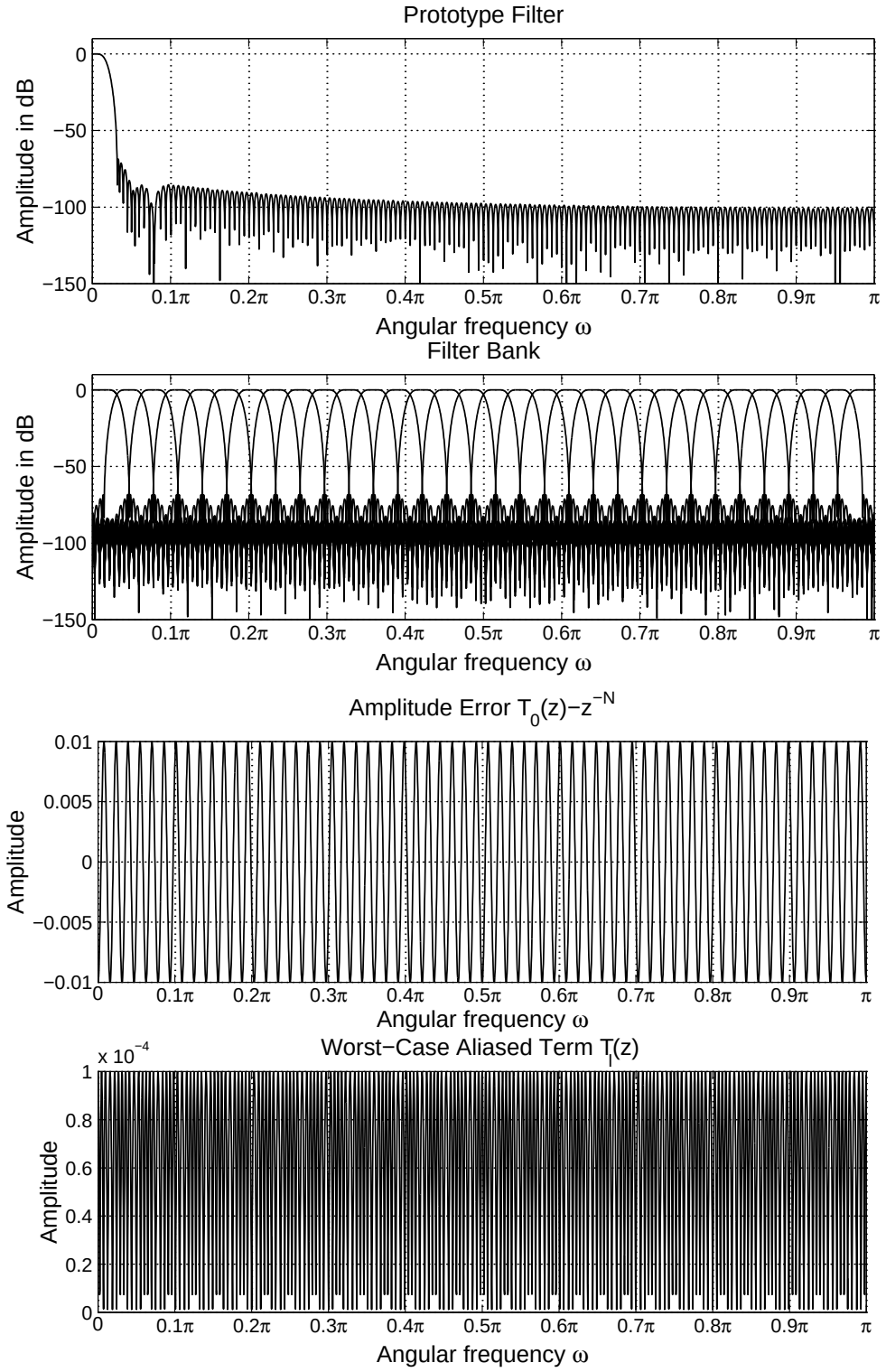
NEARLY PR FILTER BANK with $N = 511$ for $\delta_1 = 0.01$



NEARLY PR BANK with $N = 383$ for $\delta_1 = 0.001$ and $\delta_2 = 0.00001$



NEARLY PR BANK with $N = 319$ for $\delta_1 = 0.01$ and $\delta_2 = 0.0001$



DESIGN OF APPROXIMATELY LINEAR PHASE RECURSIVE DIGITAL FILTERS

- It is shown how the minimize the maximum deviation of the passband phase of a recursive digital filter subject to the given amplitude criteria.

- The filters under consideration are conventional cascade-form filters and lattice wave digital (LWD) filters (parallel connections of two all-pass filters).

- There exist very efficient schemes for designing initial filters in the lowpass case.

- Before stating the problems, we denote the overall transfer function by $H(\Phi, z)$, where Φ is the adjustable parameter vector.

- The unwrapped phase response of the filter is denoted by $\arg H(\Phi, e^{j\omega})$.

STATEMENT OF THE PROBLEMS

Approximation Problem I: Given ω_p, ω_s, δ_p, and δ_s, as well as the filter order N, find Φ and ψ, the slope of a linear phase response, to minimize

$$\Delta = \max_{0 \le \omega \le \omega_p} |\arg H(\Phi, e^{j\omega}) - \psi\omega| \tag{10a}$$

subject to

$$1 - \delta_p \le |H(\Phi, e^{j\omega})| \le 1 \quad \text{for} \quad \omega \in [0, \omega_p], \tag{10b}$$

$$|H(\Phi, e^{j\omega})| \le \delta_s \quad \text{for} \quad \omega \in [\omega_s, \pi], \tag{10c}$$

and

$$|H(\Phi, e^{j\omega})| \le 1 \quad \text{for} \quad \omega \in (\omega_p, \omega_s). \tag{10d}$$

Approximation Problem II: Given ω_p, ω_s, δ_p, and δ_s, as well as the filter order N, find Φ and ψ to minimize Δ as given by Eq. (10a) subject to the conditions of Eqs. (10b) and (10c) and

$$\frac{d|H(\Phi, e^{j\omega})|}{d\omega} \le 0 \quad \text{for } \omega \in (\omega_p, \omega_s). \tag{10e}$$

EXAMPLE: $\omega_p = 0.05\pi$, $\omega_s = 0.1\pi$, δ_p=0.0228 (0.2-dB passband ripple), and $\delta_s = 10^{-3}$ (60-dB attenuation)

Elliptic Filter: The minimum order is **five**.

Cascade-Form Filter for Problem I: For the filter of order **seven** the maximum deviation from $\phi_{\text{ave}}(\omega) = -47.06\omega$ is 0.29 degrees.

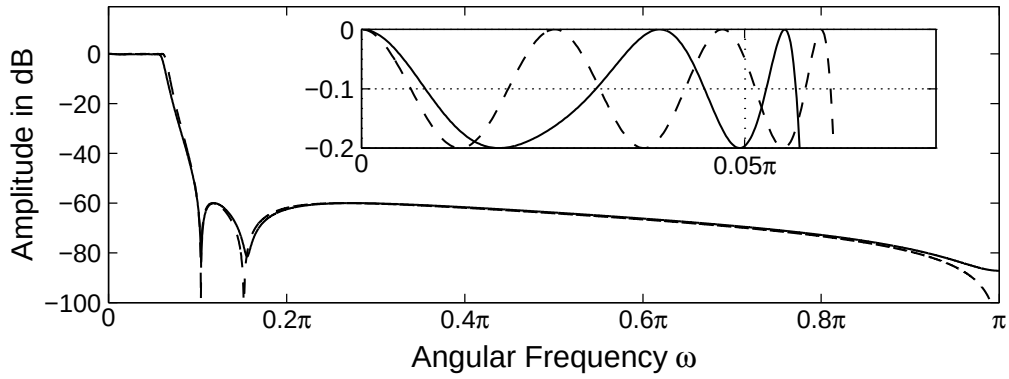
Cascade-Form Filter for Problem II: For the filter of order **seven** the maximum deviation from $\phi_{\text{ave}}(\omega) = -47.56\omega$ is 0.50 degrees.

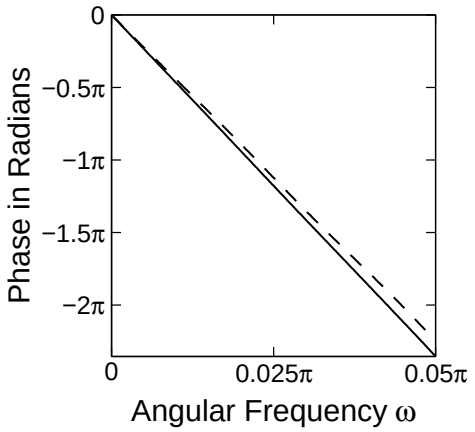
Lattice Wave Digital Filter for Problem I: For the filter of order **nine** the maximum deviation from $\phi_{\text{ave}}(\omega) = -40.38\omega$ is 0.094 degrees.

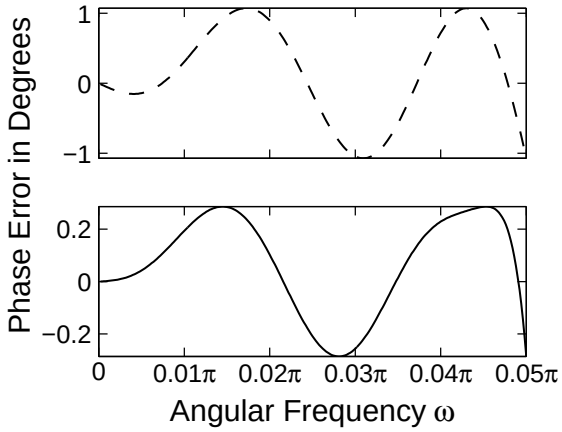
Lattice Wave Digital Filter for Problem II: For the filter of order **nine** the maximum deviation from $\phi_{\text{ave}}(\omega) = -42.92\omega$ is 0.27 degrees.

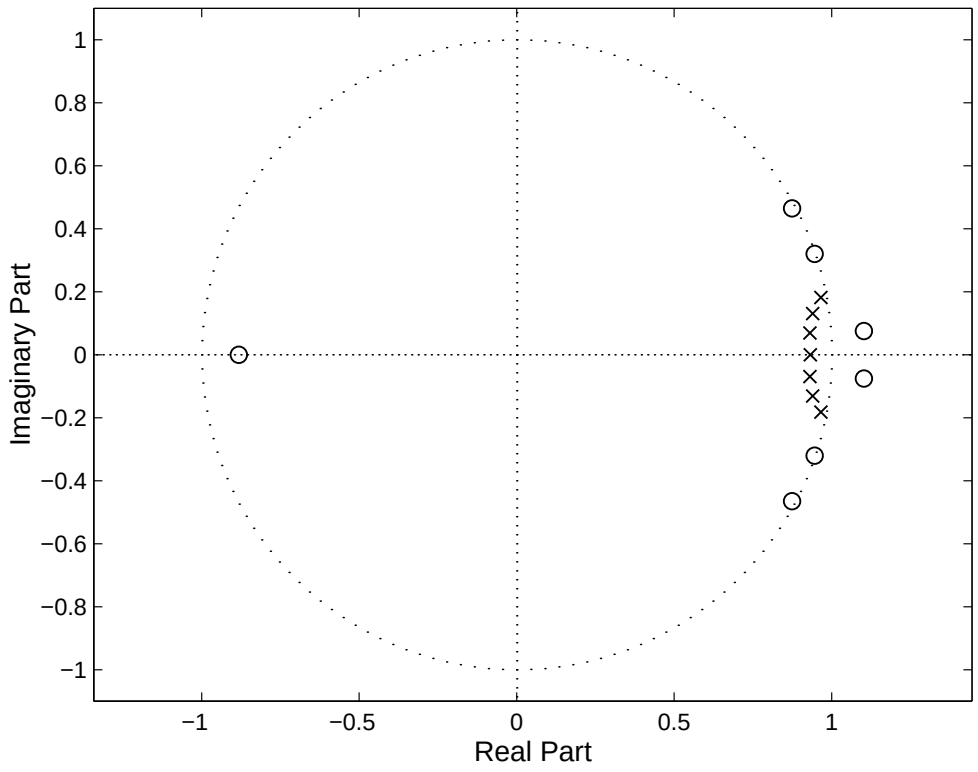
Linear-Phase FIR Filter: Minimum order is 107 and delay is 53.5 samples.

Cascade-Form Filter for Problem I

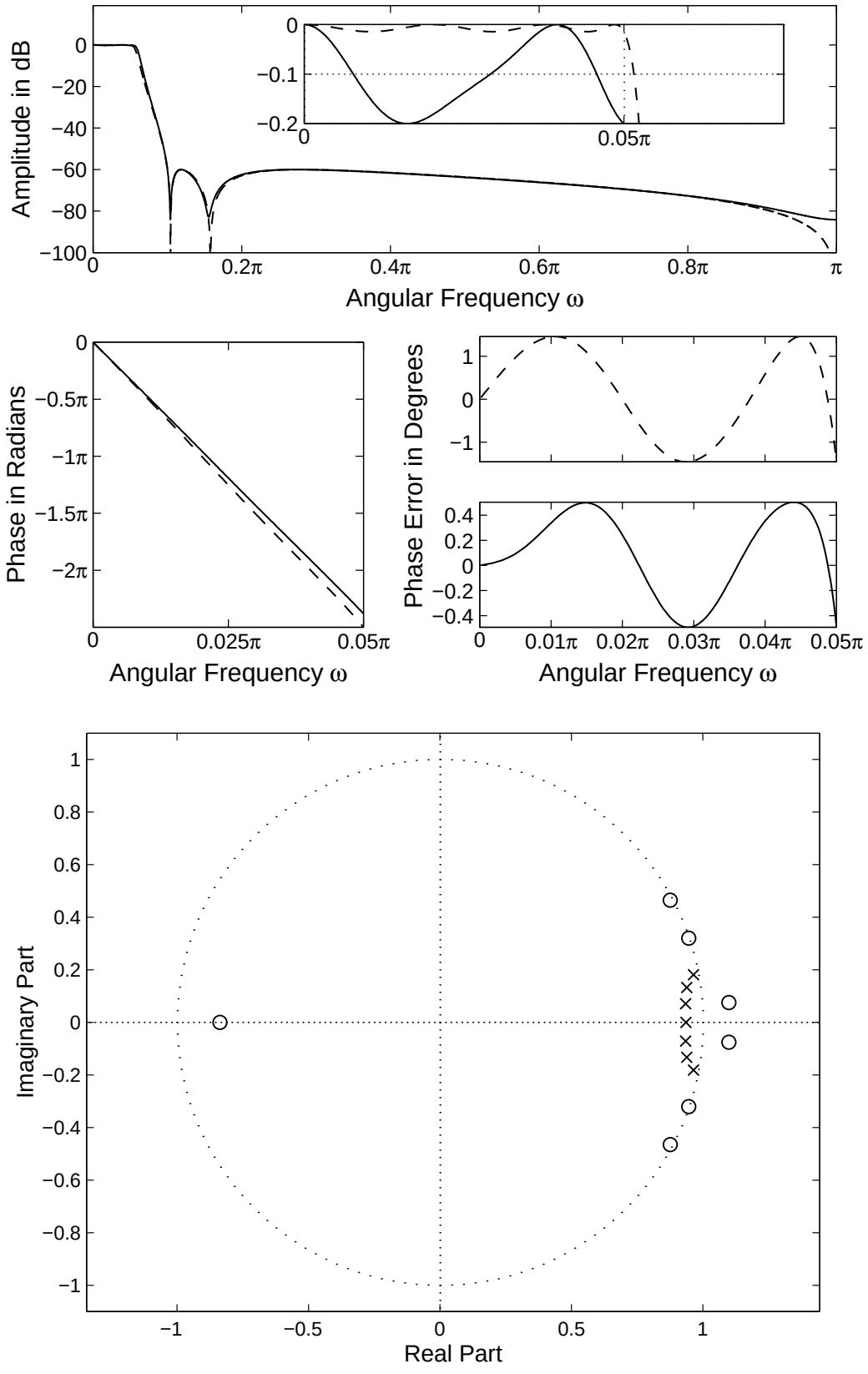




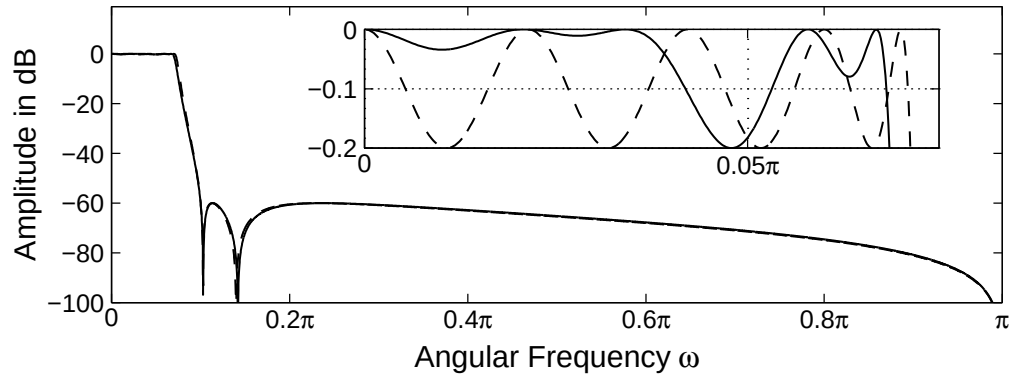


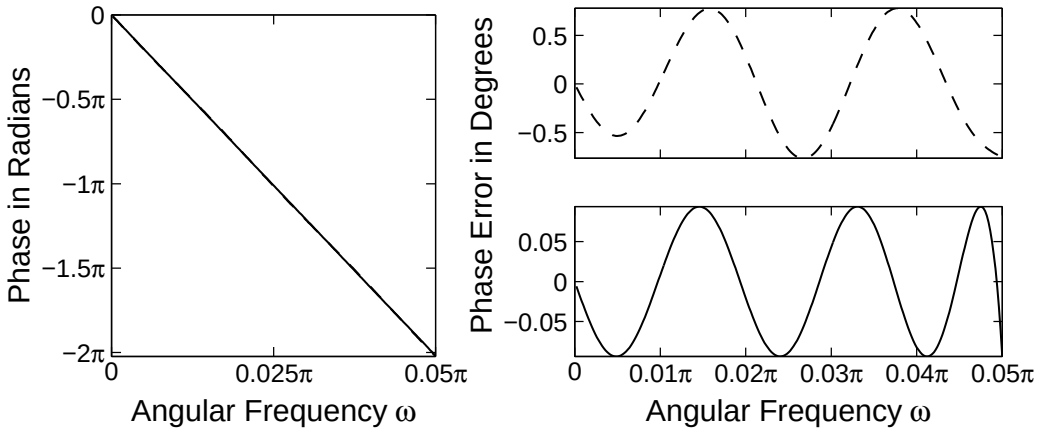


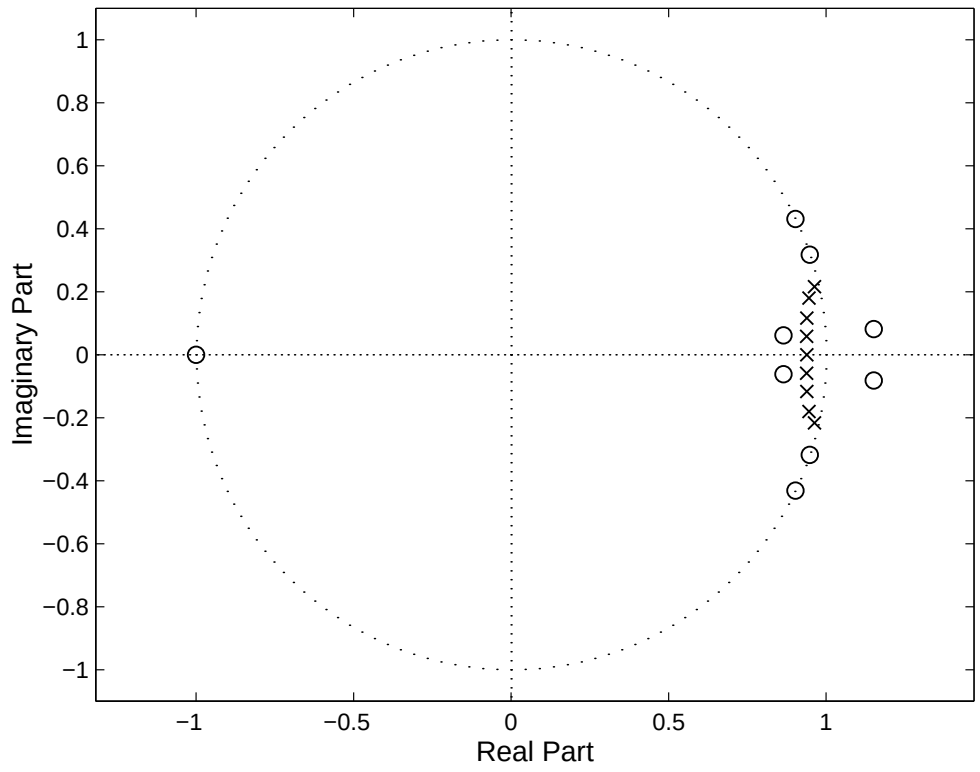
Cascade-Form Filter for Problem II



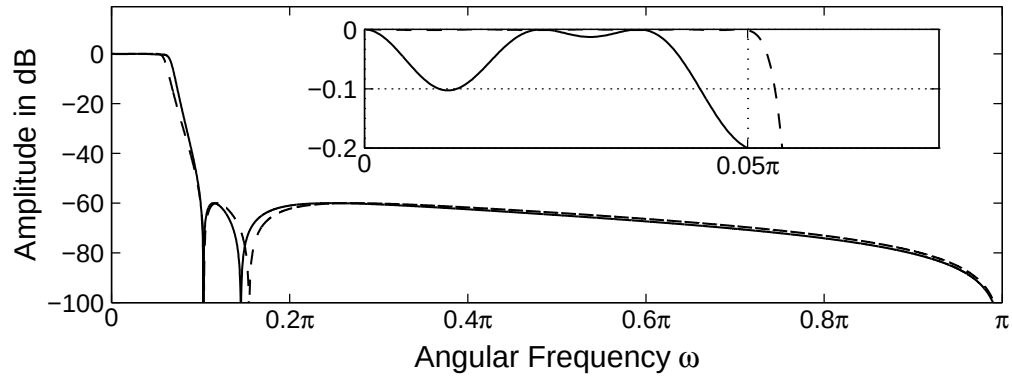
Lattice Wave Digital Filter for Problem I

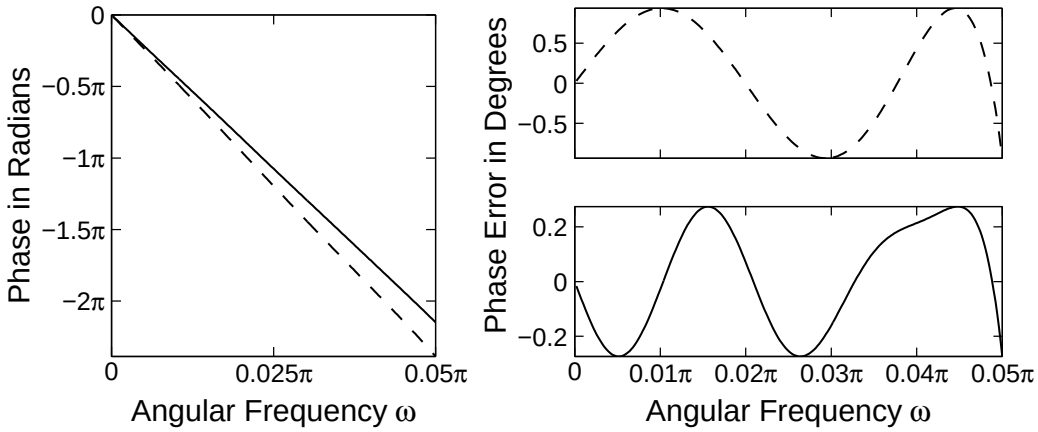


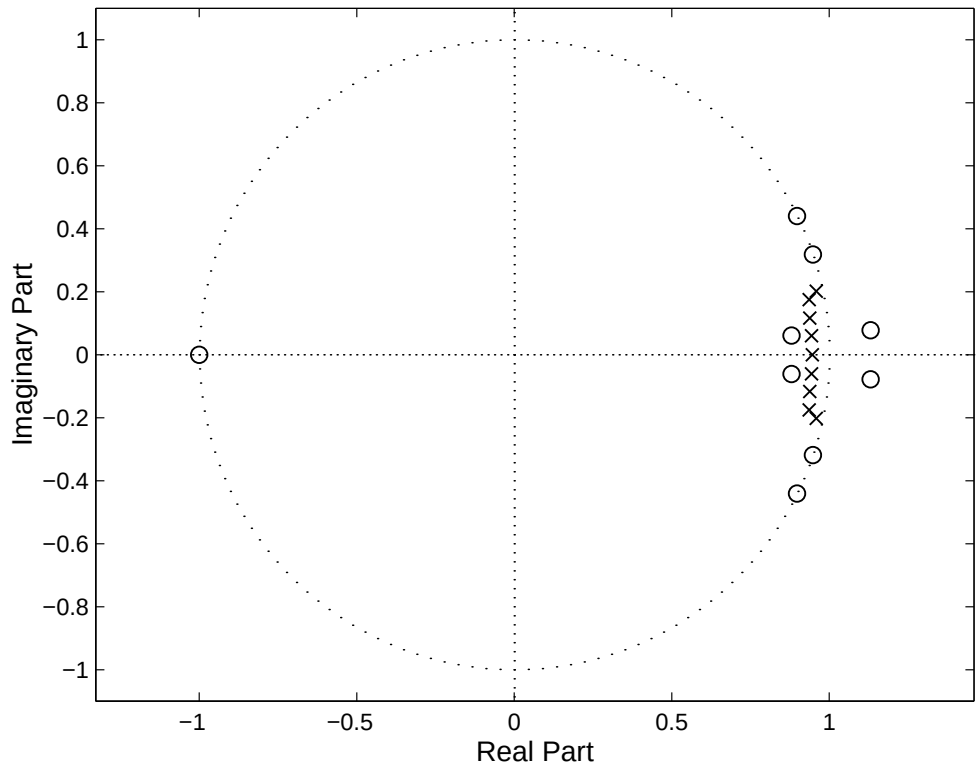




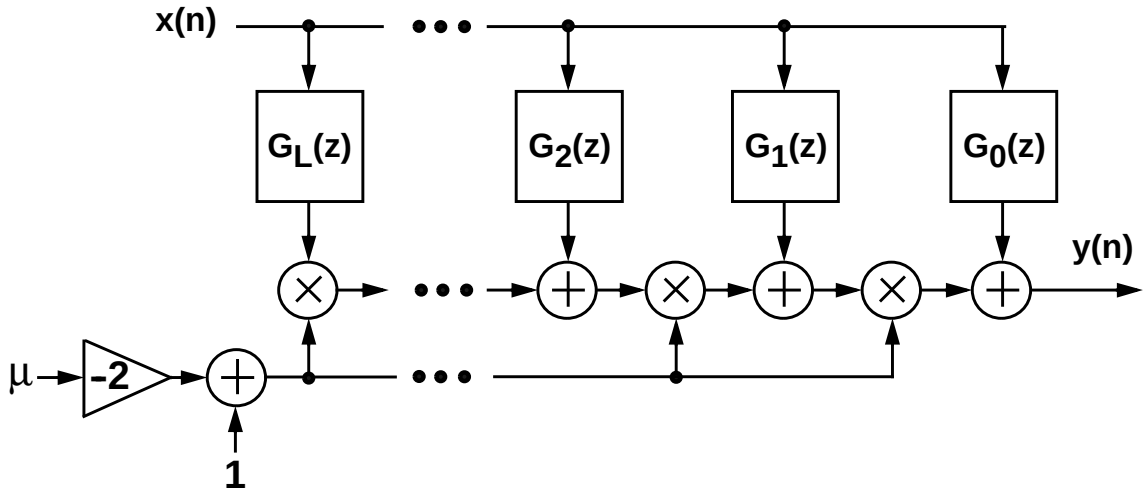
Lattice Wave Digital Filter for Problem II







MODIFIED FARROW STRUCTURE WITH ADJUSTABLE FRACTIONAL DELAY



Fixed linear-phase filters for $l = 0, 1, \ldots, L$:

$$G_l(z) = \sum_{n=0}^{N-1} g_l(n) z^{-n} \tag{11a}$$

where N is an even integer and

$$g_l(n) = \begin{cases} g_l(N-1-n) & \text{for } l \text{ even} \\ -g_l(N-1-n) & \text{for } l \text{ odd.} \end{cases} \tag{11b}$$

Delay: $N/2-1+\mu$, where the fractional delay $0 \leq \mu < 1$ is directly the adjustable parameter of the structure.

TRANSFER FUNCTION

The overall transfer function is given by

$$H(\Phi, z, \mu) = \sum_{n=0}^{N-1} h(n, \Phi, \mu) z^{-n}, \tag{12a}$$

where

$$h(n, \Phi, \mu) = \sum_{l=0}^{L} g_l(n)(1 - 2\mu)^l \tag{12b}$$

and Φ is the adjustable parameter vector

$$\begin{aligned}\Phi = \big[&g_0(0), g_0(1), \ldots, g_0(N/2-1), g_1(0), g_1(1), \ldots, \\ &g_1(N/2-1), \ldots, g_L(0), g_L(1), \ldots, g_L(N/2-1)\big].\end{aligned} \tag{12c}$$

AMPLITUDE AND PHASE DELAY RESPONSES

The frequency, amplitude, and phase delay responses of the proposed Farrow structure are given by

$$H(\Phi, e^{j\omega}, \mu) = \sum_{n=0}^{N-1} h(n, \Phi, \mu) e^{-j\omega n}, \tag{13a}$$

$$|H(\Phi, e^{j\omega}, \mu)| = \left| \sum_{n=0}^{N-1} h(n, \Phi, \mu) e^{-j\omega n} \right|, \tag{13b}$$

and

$$\tau_p(\Phi, \omega, \mu) = -\arg(H(\Phi, e^{j\omega}, \mu))/\omega, \tag{13c}$$

respectively.

STATEMENT OF THE PROBLEM

Optimization Problem: Given L, N, Ω_p, and ϵ, find the adjustable parameter vector Φ to minimize

$$\delta_p = \max_{0 \le \mu < 1} \left[\max_{\omega \in \Omega_p} |\tau_p(\Phi, \omega, \mu) - (N/2 - 1 + \mu)| \right] \tag{14a}$$

subject to

$$\delta_a = \max_{0 \le \mu < 1} \left[\max_{\omega \in \Omega_p} ||H(\Phi, e^{j\omega}, \mu)| - 1| \right] \le \epsilon. \tag{14b}$$

EXAMPLES

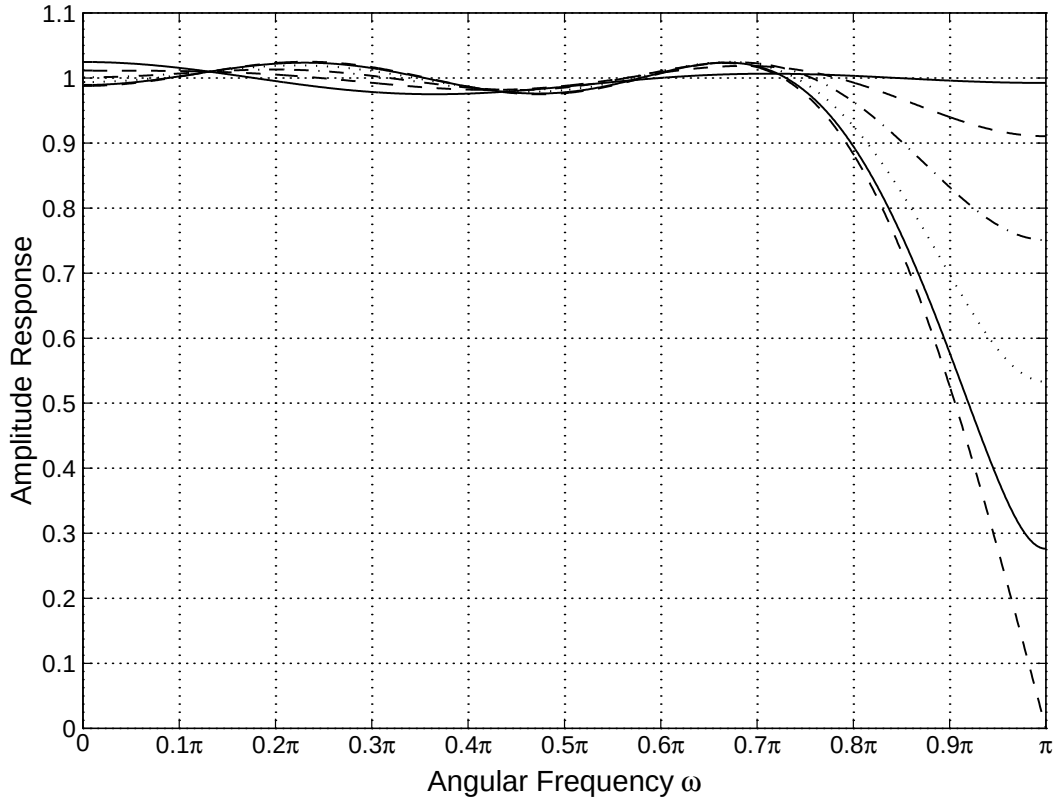
Example 1: $\Omega_p = [0,\ 0.75\pi]$, $\epsilon = 0.025$, and $\delta_p \leq 0.01$.

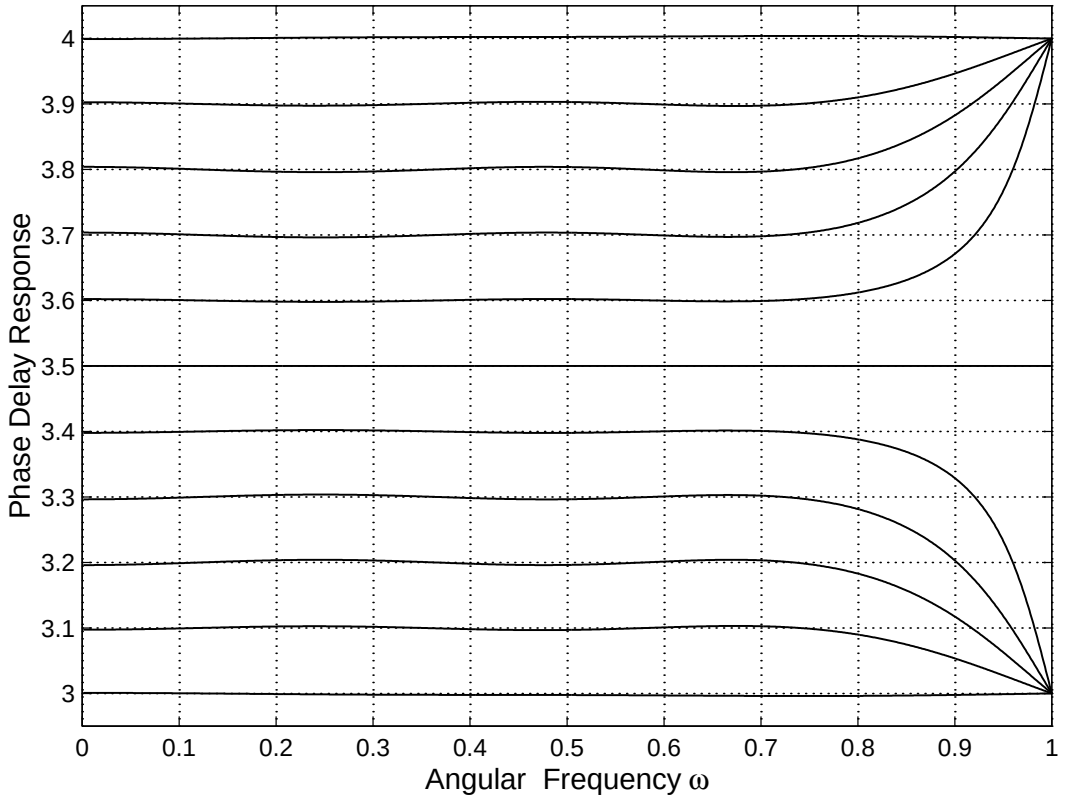
$\delta_p = 0.00402$ is achieved by $N = 8$ and $L = 3$.

Example 2: $\Omega_p = [0,\ 0.9\pi]$, $\epsilon = 0.01$, and $\delta_p \leq 0.001$.

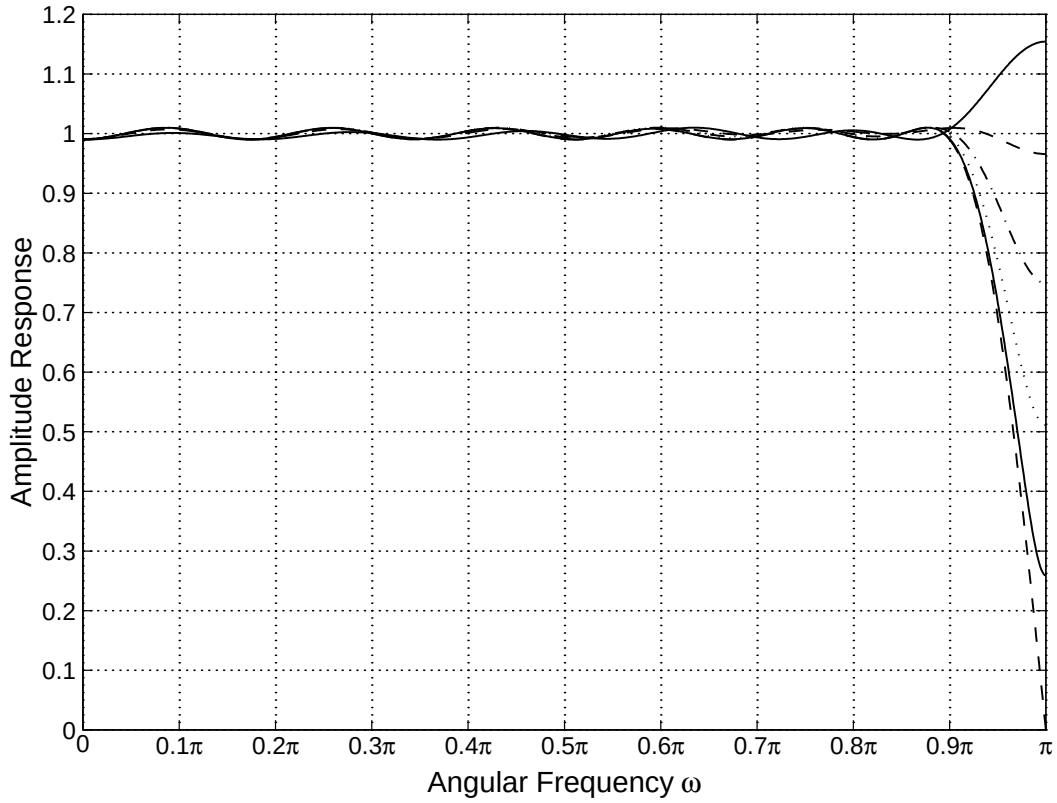
The criteria are met by $N = 26$ and $L = 4$.

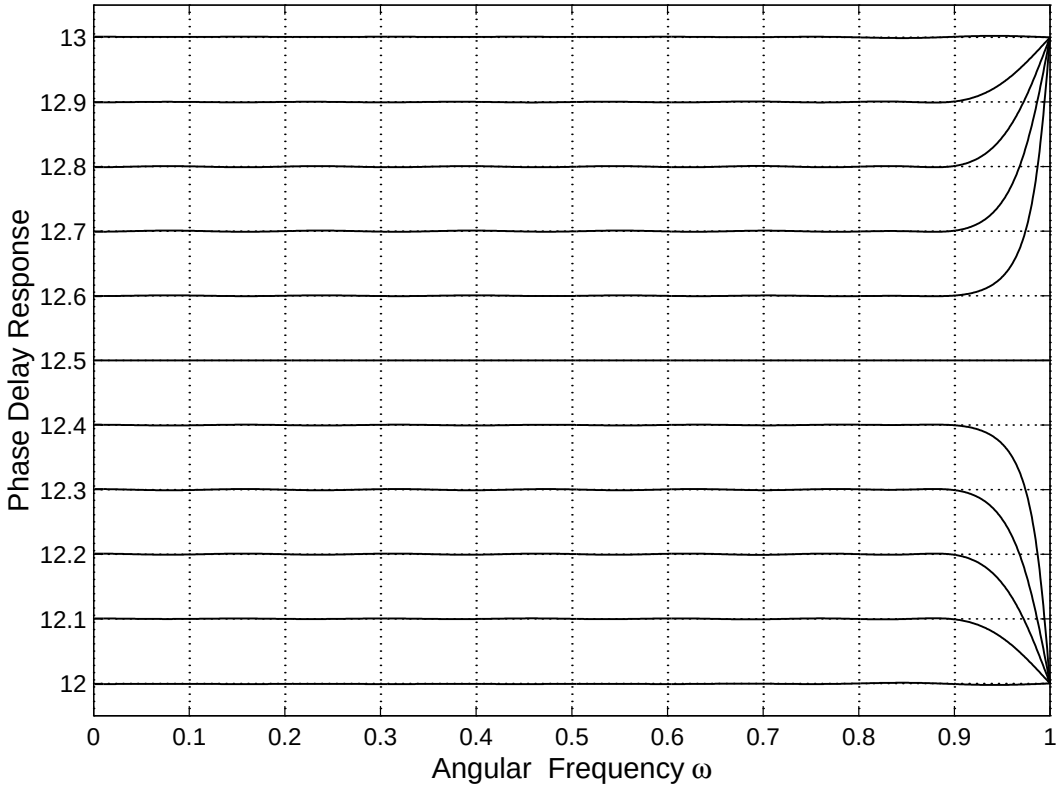
RESPONSES FOR EXAMPLE 1





RESPONSES FOR EXAMPLE 2





DESIGN OF LATTICE WAVE DIGITAL FILTERS WITH SHORT COEFFICIENT WORDLENGTH

- It is shown how the coefficients of the various classes of lattice wave digital (LWD) filters can be conveniently quantized.

- The filters under consideration are
  - the conventional LWD filters,
  - cascades of low-order LWD filters providing a very low sensitivity and roundoff noise, and
  - LWD filters with an approximately linear phase in the passband.

- There exist very efficient schemes for designing initial filters in the lowpass case.

- Before stating the problems, we denote the overall transfer function as $H(\Phi, z)$, where Φ is the adjustable parameter vector.

Overall Transfer Function

The most general form of the transfer function is given by

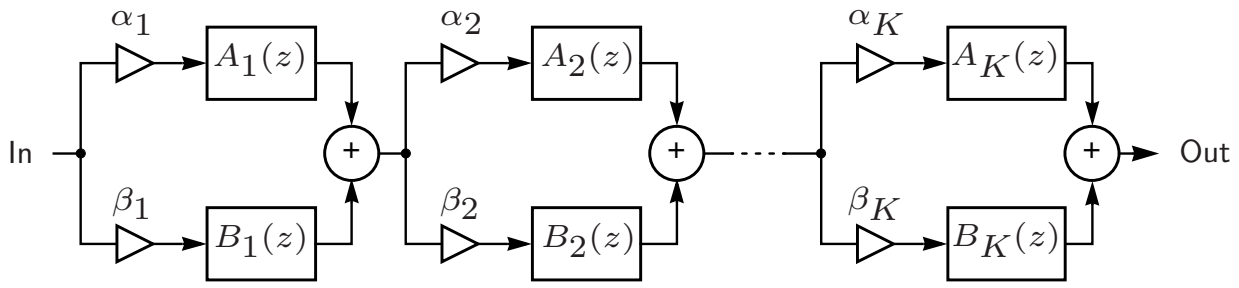
$$H(\Phi, z) = \prod_{k=1}^{K} H_k(\Phi, z), \tag{15a}$$

where

$$H_k(\Phi, z) = \alpha_k A_k(z) + \beta_k B_k(z). \tag{15b}$$

Here, $A_k(z)$'s and $B_k(z)$'s are stable all-pass filters of orders M_k and N_k, respectively.

For conventional and approximately linear-phase LWD filters $K = 1$.



Statement of the Problems

In VLSI applications it is desirable to express the coefficient values in the form

$$\sum_{r=1}^{R} a_r 2^{-P_r}, \tag{16}$$

where each of the a_r's is either 1 or -1 and the P_r's are positive integers in the increasing order.

The target is to find all the coefficient values included in Φ, in such a way that:

1. R, the number of powers of two, is made as small as possible and
2. P_R, the number of fractional bits, is made as small as possible.

Problem I: Find K, the number of subfilters, the M_k's and N_k's, as well as the adjustable parameter vector Φ in such a way that:

1. $H(\Phi, z)$ meets the criteria given by

$$1 - \delta_p \leq |H(\Phi, e^{j\omega})| \leq 1 \quad \text{for } \omega \in [0, \omega_p] \quad \text{(17a)}$$

$$|H(\Phi, e^{j\omega})| \leq \delta_s \quad \text{for } \omega \in [\omega_s, \pi]. \quad \text{(17b)}$$

2. The coefficients included in Φ are quantized to achieve the above-mentioned target for their representations.

Problem II: Find Φ as well as τ, the slope of the linear-phase response, in such a way that:

1. $H(\Phi, z)$ meets the criteria given by Eq. (17) and

$$|\arg[H(\Phi, e^{j\omega_i})] - \tau\omega| \leq \Delta \quad \text{for} \quad i = 1, 2, \ldots, L_p,$$

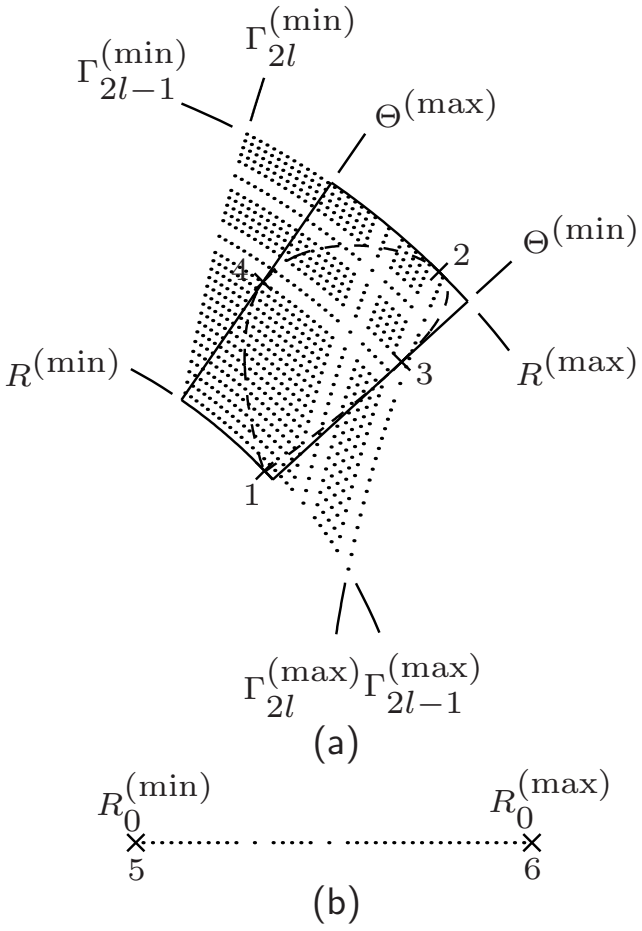
where $\arg[H(\Phi, e^{j\omega})]$ denotes the unwrapped phase response of the filter and Δ is the maximum allowable phase error from the linear-phase response.

2. The coefficients are quantized to achieve the above-mentioned target for their representations.

Quantization Algorithm

The coefficient optimization is performed in two stages:

1. A nonlinear optimization algorithm is used for determining a parameter space of the infinite-precision coefficients including the feasible space where the filter meets the criteria.



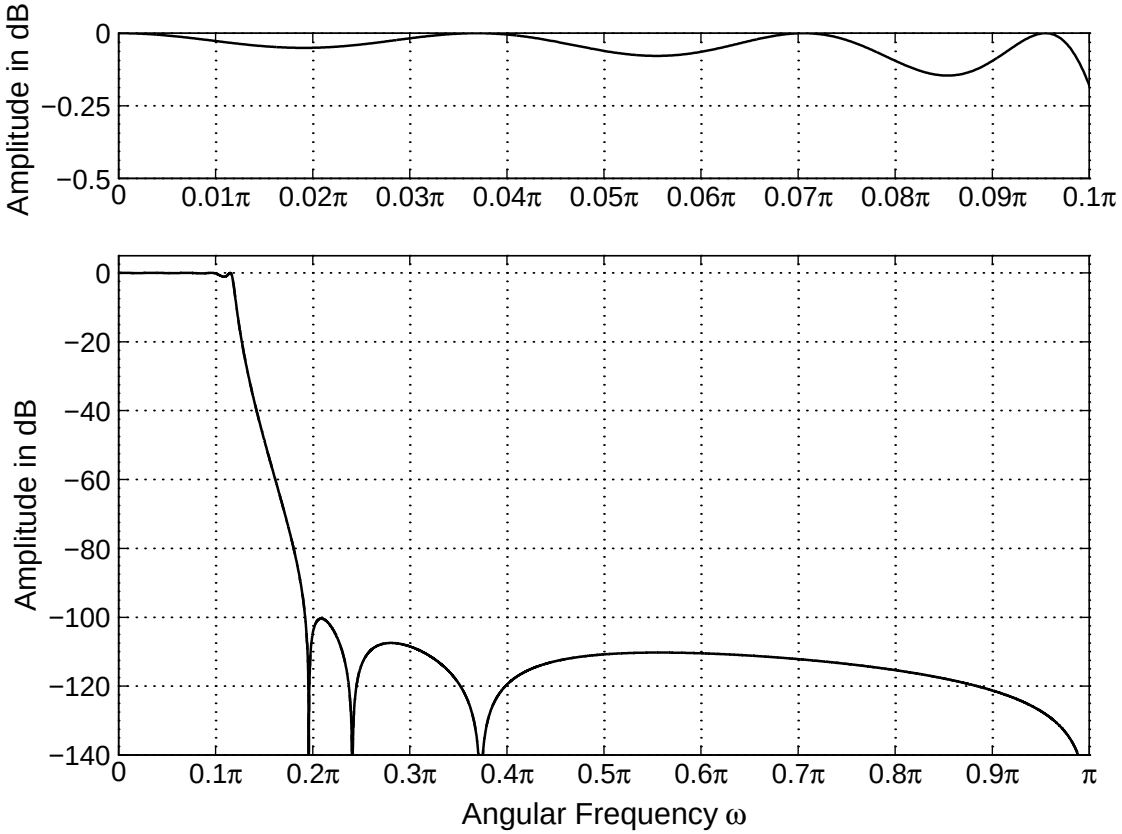
2. The filter parameters in this space are searched in such a manner that the resulting filter meets the given criteria with the simplest coefficient representation forms.

The algorithm guarantees that the optimum finite-wordlength solution can be found for both the fixed-point and the multiplierless coefficient representations.

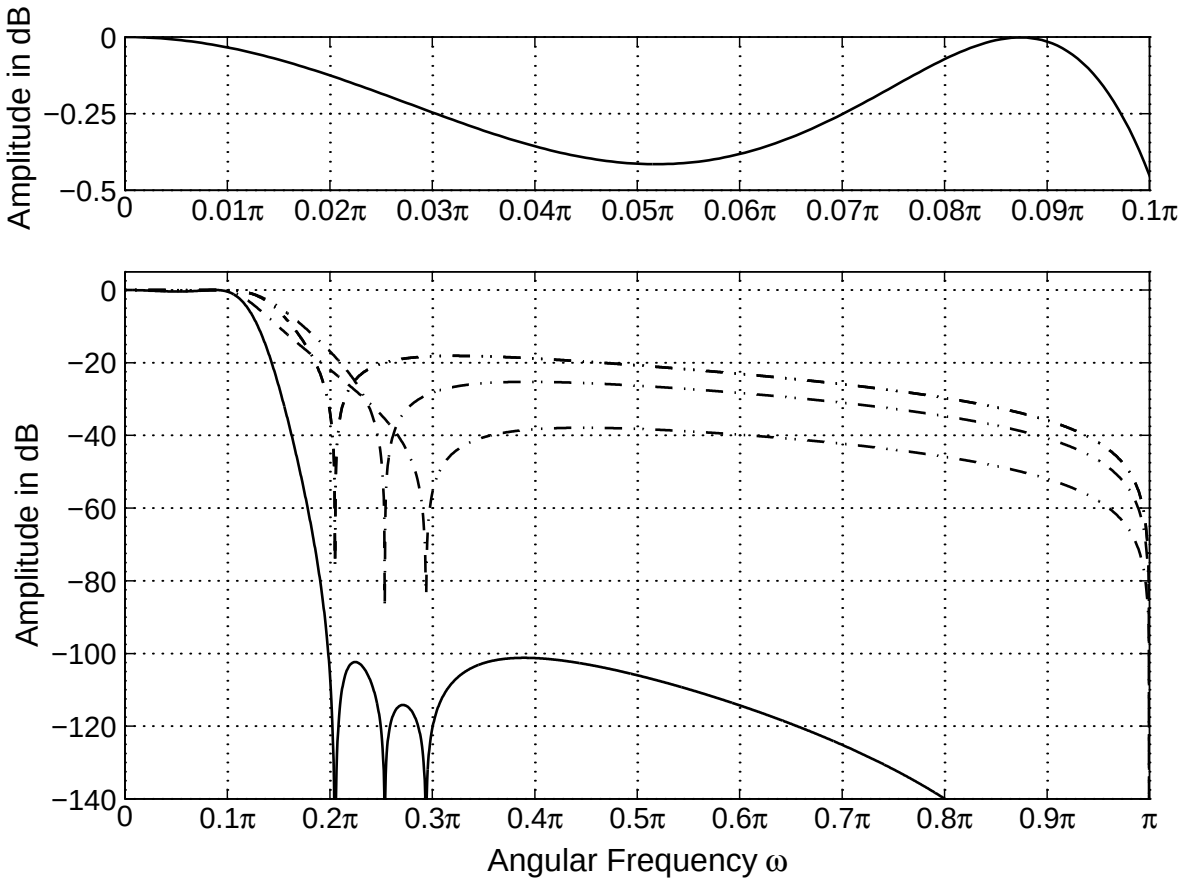
Numerical Examples

Filter specifications: $\delta_p = 0.0559$ (0.5-dB passband variation), $\delta_s = 10^{-5}$ (100-dB stopband attenuation), $\omega_p = 0.1\pi$, and $\omega_s = 0.2\pi$.

Ninth-order direct LWD filter is required to meet the criteria ($M_1 = 5$ and $N_1 = 4$).



Alternatively, cascade of **four 3rd-order** LWD filters is needed to satisfy the amplitude specifications.



For the cascade of four LWD filters, only **5** fractional bits are needed for coefficient implementation compared to **9** bits required by the direct LWD filter.

The number of adders required to implement all the coefficients are **12** and **6**, for the direct and cascade implementations, respectively.

The price paid for this is a slight increase in the overall filter order (from nine to twelve).

Optimized Finite-Precision Adaptor Coefficient Values for the Cascade of Four LWD Filters

$A(z)$	$B(z)$
$\gamma_0^{(1,2)} = 2^{-1} + 2^{-3}$	$\widehat{\gamma}_1^{(1,2)} = -1 + 2^{-2} - 2^{-5}$ $\widehat{\gamma}_2^{(1,2)} = 1 - 2^{-3} + 2^{-5}$
$\gamma_0^{(3)} = 2^{-1} + 2^{-3} + 2^{-5}$	$\widehat{\gamma}_1^{(3)} = -1 + 2^{-2}$ $\widehat{\gamma}_2^{(3)} = 1 - 2^{-3} + 2^{-5}$
$\gamma_0^{(4)} = 1 - 2^{-2} + 2^{-5}$	$\widehat{\gamma}_1^{(4)} = -1 + 2^{-2} - 2^{-4}$ $\widehat{\gamma}_2^{(4)} = 1 - 2^{-4}$

Approximately Linear-Phase LWD Filter

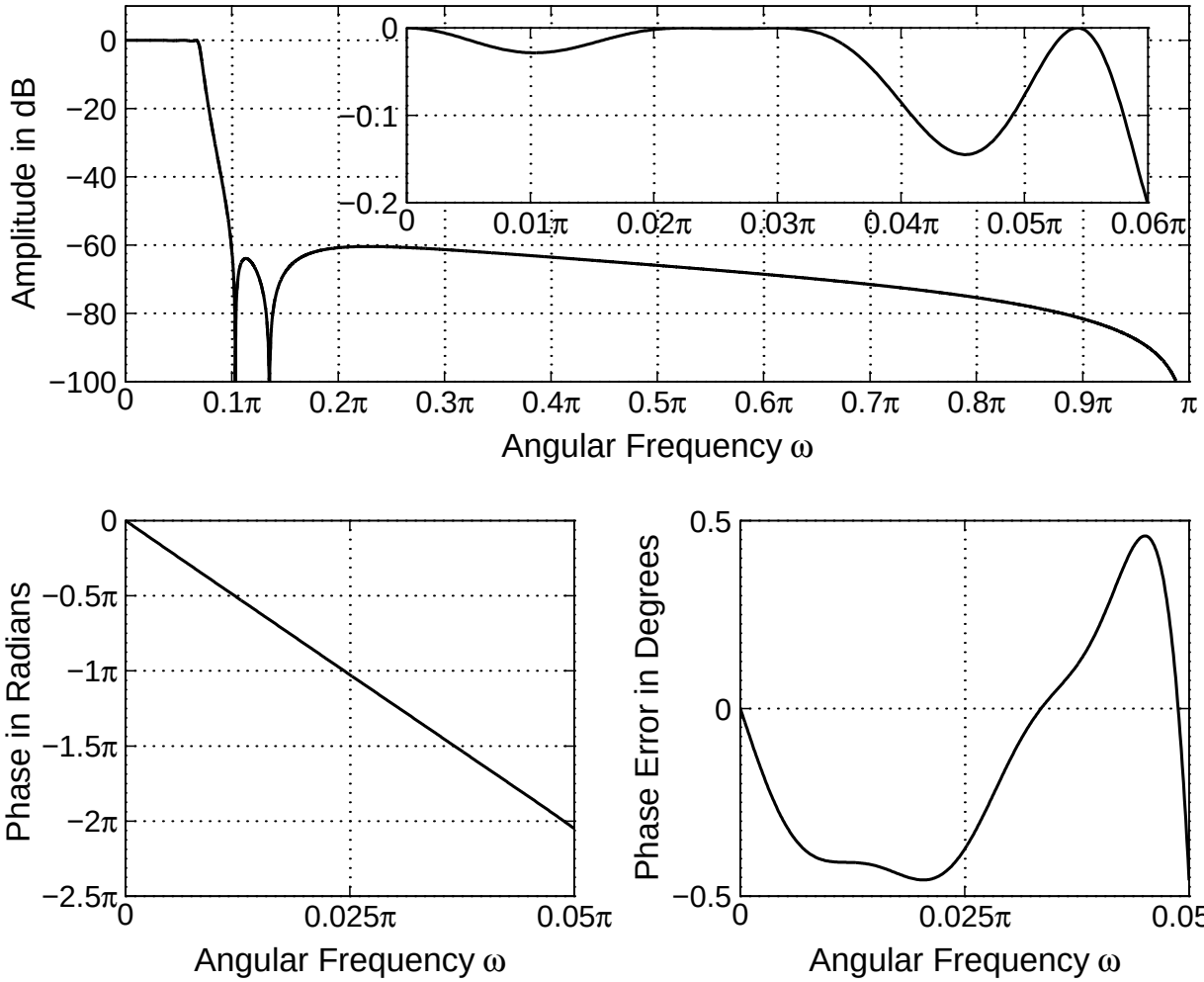
Filter specifications: $\delta_p = 0.0228$ (0.2-dB passband variation), $\delta_s = 10^{-3}$ (60-dB stopband attenuation), $\omega_p = 0.05\pi$, and $\omega_s = 0.1\pi$.

The minimum order of an elliptic filter to meet the amplitude specifications in **five**. An excellent phase performance is obtained by increasing the filter order to **nine**.

For the optimal infinite-precision filter the phase error is 0.09399 degrees.

To allow some tolerance for the quantization, the maximum allowable phase error is increased to 0.5 degrees.

Amplitude and Phase Responses for the Quantized Filter



For the optimized filter, only 10 adders with 11 fractional bits are required to implement all the adaptor coefficients.

The phase error for the optimized filter is $0.458\,55$ degrees.

A SYSTEMATIC ALGORITHM FOR THE DESIGN OF MULTIPLIERLESS FIR FILTERS

In this application, we show how the coefficients of the linear-phase FIR filters can be conveniently quantized using optimization techniques.

The zero-phase frequency response of a linear-phase Nth-order FIR filter can be expressed as

$$H(\omega) = \sum_{n=0}^{M} h(n)\,\mathrm{Trig}(\omega, n), \tag{18}$$

where the $h(n)$'s are the filter coefficients and $\mathrm{Trig}(\omega, n)$ is an appropriate trigonometric function depending on whether N is odd or even and whether the impulse response is symmetrical or antisymmetrical. Here, $M = N/2$ if N is even, and $M = (N+1)/2$ if N is odd.

Desired Coefficient Representation Form

The general form for expressing the FIR filter coefficient values as a sums of signed-powers-of-two (SPT) terms is given by

$$h(n) = \sum_{k=1}^{W_n+1} a_{k,n} 2^{-P_{k,n}} \quad \text{for} \quad n = 1, 2, \ldots, M, \quad (19)$$

where $a_{k,n} \in \{-1, 1\}$ and $P_{k,n} \in \{1, 2, \ldots, L\}$ for $k = 1, 2, \ldots, W_n + 1$. In this representation form, each coefficient $h(n)$ has W_n adders and the maximum allowable wordlength is L bits.

Statement of the Problem

When finding the optimized simple discrete-value representation forms for the coefficients of FIR filters, it is a common practise to accomplish the optimization in such a manner that the scaled response meets the given amplitude criteria.

In this case, the criteria for the filter can be expressed as

$$1 - \delta_p \leq H(\omega)/\beta \leq 1 + \delta_p \quad \text{for} \quad \omega \in [0, \omega_p] \tag{20a}$$

$$-\delta_s \leq H(\omega)/\beta \leq \delta_s \quad \text{for} \quad \omega \in [\omega_s, \pi], \tag{20b}$$

where

$$\beta = \frac{1}{2}\big[\max H(\omega) + \min H(\omega)\big] \quad \text{for} \quad \omega \in [0, \omega_p] \tag{20c}$$

is the average passband gain.

These criteria are preferred to be used when the filter coefficients are desired to be quantized on a highly nonuniform discrete grid as in the case of the power-of-two coefficients.

Optimization Problem

Given ω_p, ω_s, δ_p, and δ_s, as well as L, the number of fractional bits, and the maximum allowed number of SPT terms per coefficient, find the filter coefficients $h(n)$ for $n = 1, 2, \ldots, M$ as well as β to minimize implementation cost, in such a manner that:

1. The magnitude criteria, as given by Eq. (20), are met and
2. the normalized peak ripple (NPR), as given by

$$\delta_{\text{NPR}} = \max\{\Delta_p/W,\, \Delta_s\}, \tag{21}$$

   is minimized.

Here, Δ_p and Δ_s are the passband and stopband ripples of the finite-precision filter scaled by $1/\beta$ and $W = \delta_p/\delta_s$.

Coefficient Optimization

The coefficient optimization is performed in two stages:

1. For each of the filter coefficient $h(n)$ for $n = 0, 1, \ldots, M-1$ the largest and smallest values of the coefficient are determined in such a manner that the given amplitude criteria are met subject to $h(M) = 1$.

   This restriction, simplifying the overall procedure, can be stated without loss of generality since the scaling constant β, as defined by Eq. (20), can be used for achieving the desired passband amplitude level.

   These problems can be solved conveniently by using linear programming.

2. It has been experimentally observed that the parameter space defined above forms the feasible space where the filter specifications are satisfied. After finding this larger space, all what is needed is to check whether in this space there exists a combination of the discrete coefficient values with which the overall criteria are met.

Optimization of Infinite-Precision Coefficients

The goal is achieved by solving $2M$ problems of the following form. Find the filter coefficients $h(n)$ for $n = 0, 1, \ldots, M-1$ as well as β to minimize ψ subject to the conditions

$$\sum_{n=0}^{M-1} h(n)\,\mathrm{Trig}(\omega_i, n) - \beta(\delta_p + 1) \leq -\,\mathrm{Trig}(\omega_i, M),$$

$$-\sum_{n=0}^{M-1} h(n)\,\mathrm{Trig}(\omega_i, n) - \beta(\delta_p - 1) \leq -\,\mathrm{Trig}(\omega_i, M),$$

for $\omega_i \in [0, \omega_p]$ and

$$\sum_{n=0}^{M-1} h(n)\,\mathrm{Trig}(\omega_i, n) - \beta\delta_s \leq -\,\mathrm{Trig}(\omega_i, M),$$

$$-\sum_{n=0}^{M-1} h(n)\,\mathrm{Trig}(\omega_i, n) - \beta\delta_s \leq -\,\mathrm{Trig}(\omega_i, M),$$

for $\omega_i \in [\omega_s, \pi]$.

Here ψ is $-h(n)$ or $h(n)$ where $h(n)$ is one among the filter coefficients $h(n)$ for $n = 0, 1, \ldots, M-1$.

Optimization of Finite-Precision Coefficients

In the above procedure, $h(M)$ was fixed to be unity. The search for a proper combinations of discrete values can be conveniently accomplished by using a scaling constant $\alpha \equiv h(M)$.

For this constant, all the existing values between $1/3$ and $2/3$ are selected from the look-up table containing all the possible power-of-two numbers for a given wordlength and a given maximum number of SPT terms per coefficient.

Then for each value of $\alpha = h(M)$, the largest and smallest values of the infinite-precision coefficients are scaled in the look-up table as

$$\widehat{h}(n)^{(\text{min})} = \alpha h(n)^{(\text{min})} \quad \text{for} \quad n = 0, 1, \ldots, M \quad \text{(23a)}$$

$$\widehat{h}(n)^{(\text{max})} = \alpha h(n)^{(\text{max})} \quad \text{for} \quad n = 0, 1, \ldots, M \quad \text{(23b)}$$

and the magnitude response is evaluated for each combination of the power-of-two numbers in the ranges $[\widehat{h}(n)^{(\text{min})}, \widehat{h}(n)^{(\text{max})}]$ for $n = 0, 1, \ldots, M$ to check whether the filter meets the amplitude criteria.

Numerical Examples

Example 1: $N = 37$, $\delta_p = \delta_s = 10^{-3}$, $\omega_p = 0.3\pi$, and $\omega_s = 0.5\pi$.

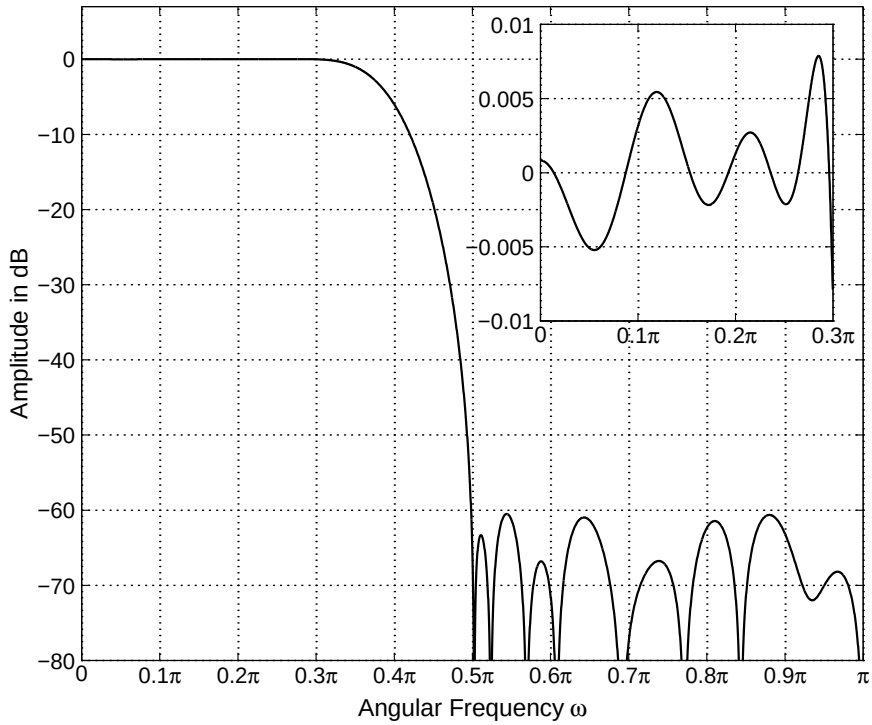
Method	δ_{NPR} (dB)	No. Powers of Two	No. Adders
Lim and Parker	-62.08	43	–
Chen and Willson	-60.87	40	–
Proposed	-60.48	34	48

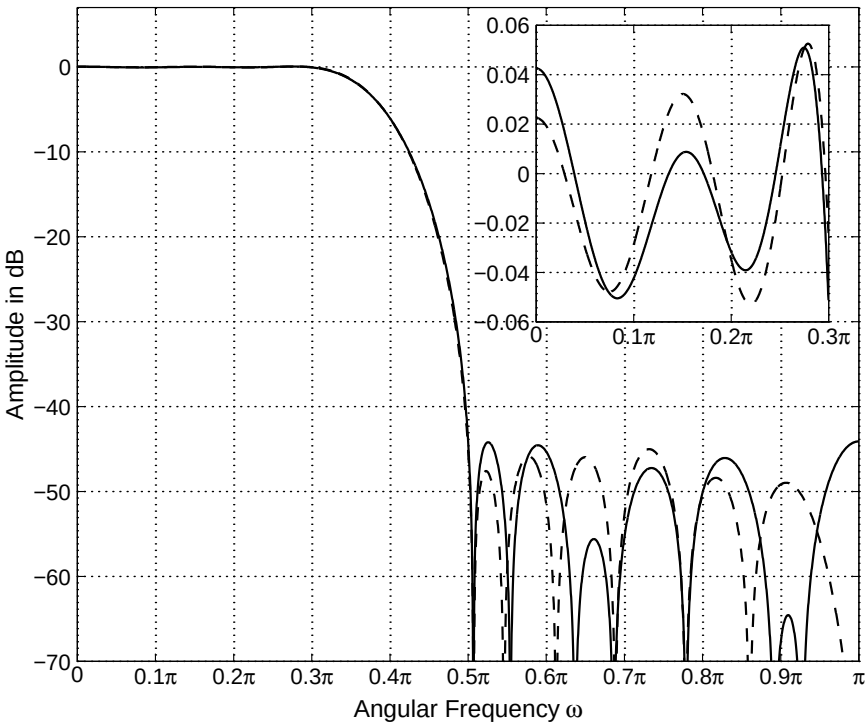
Example 2: $N = 24$, $\delta_p = \delta_s = 0.005$, $\omega_p = 0.3\pi$, and $\omega_s = 0.5\pi$.

Method	δ_{NPR} (dB)	No. Powers of Two	No. Adders
Samueli	-42.17	24	35
Li *et al.*	-43.33	24	–
Chen and Willson	-43.97	24	33
Proposed	-44.09	21	30

If redundancies within the coefficients are utilized only 26 adders are required to meet the specifications.

Responses for Examples 1 and 2





Optimized Finite-Precision Coefficient Values for the FIR Filter in Example 1

$$
\begin{aligned}
h(0) &= h(37) = -2^{-11} \\
h(1) &= h(36) = 0 \\
h(2) &= h(35) = +2^{-9} - 2^{-12} \\
h(3) &= h(34) = +2^{-9} \\
h(4) &= h(33) = -2^{-9} - 2^{-11} \\
h(5) &= h(32) = -2^{-7} + 2^{-9} - 2^{-11} \\
h(6) &= h(31) = 0 \\
h(7) &= h(30) = +2^{-6} - 2^{-8} \\
h(8) &= h(29) = +2^{-7} + 2^{-9} \\
h(9) &= h(28) = -2^{-6} + 2^{-8} - 2^{-10} \\
h(10) &= h(27) = -2^{-5} + 2^{-8} + 2^{-12} \\
h(11) &= h(26) = 0 \\
h(12) &= h(25) = +2^{-4} - 2^{-6} - 2^{-9} \\
h(13) &= h(24) = +2^{-5} + 2^{-8} + 2^{-10} \\
h(14) &= h(23) = -2^{-4} + 2^{-6} - 2^{-10} \\
h(15) &= h(22) = -2^{-3} + 2^{-6} + 2^{-8} \\
h(16) &= h(21) = 0 \\
h(17) &= h(20) = +2^{-2} + 2^{-6} \\
h(18) &= h(19) = +2^{-1}
\end{aligned}
$$

Effective Implementation Exploiting Coefficient Symmetry for the Multiplierless FIR Filter in Example 1