MODIFIED BUTTERWORTH AND CHEBYSHEV FUNCTIONS:
DIGITAL FILTER ROUNDOFF NOISE
AND BIT REQUIREMENTS

By

KHALIL ELIA MASSAD

Bachelor of Engineering
American University of Beirut
Beirut, Lebanon
1970

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1972

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
July, 1975

MODIFIED BUTTERWORTH AND CHEBYSHEV FUNCTIONS:

DIGITAL FILTER ROUNDOFF NOISE

AND BIT REQUIREMENTS

Thesis Approved:

Rao Yarlagadda
Thesis Adviser

Robert J. Mulholland

Gerald K. Goff

Edward L. Shreve

N N Durham
Dean of the Graduate College

ACKNOWLEDGEMENTS

Numerous consultations with and suggestions from Dr. Rao Yarlagadda, my thesis adviser and chairman of my doctoral committee, contributed significantly to this dissertation.

The interest and comments offered by Dr. Robert Mulholland, Dr. Edward Shreve, and Dr. Gerald Goff, members of my doctoral committee, are sincerely appreciated.

I thank Dr. John Chandler, Computer Science Department, Oklahoma State University, for his advice and for providing some of the computer programs used in the course of this thesis.

The financial support from the National Science Foundation under NSF Grant GU-3160, and the teaching assistantship from the School of Electrical Engineering, Oklahoma State University, is gratefully acknowledged.

For typing this thesis and for suggestions concerning style, I express my thanks to Dixie Jennings.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

β_i	Damping ratio of modified functions
γ_1	Inner multiplier of second order digital section
γ_2	Outer multiplier of second order digital section
γ_r	Multiplier of first order digital section
δ_i	Damping ratio of second order section
ε	Pass-band ripple factor
ρ	Reflection coefficient
σ_n^2	Output noise variance
ω_A	Analog frequency
ω_{bi}	Break point frequency of second order section
ω_c	Cutoff frequency of filter function
ω_D	Digital frequency
ω_d	Dominant poles location
ω_{pi}	Frequency of second order section magnitude peak
ω_r	Lowest specified stop-band frequency
ω_{tm}	Cross over frequency of original and modified functions
a	Real part of s-plane pole location ($s = -a \pm jb$)
b	Imaginary part of s-plane pole location
C_h	Magnitude of dominant pole section at break point frequency
c	Multiplicity of the dominant poles
$H_n(j\omega)$	Transfer function of order n
$L_m(j\omega)$	Modified transfer function of order n
MCF	Modified Chebyshev function

P_i	Peak magnitude of a second order function
Q	Pole quality factor
Q_d	Q of dominant poles
Q_{dc}	Q of dominant poles with multiplicity c
Q_r	Coefficient bits of first order section
Q_1	Coefficient bits of inner multiplier
Q_2	Coefficient bits of outer multiplier
$S(i)$	Scaling factors
s_x^T	Sensitivity of T with respect to x
$T_n(\ )$	Chebyshev polynomials of degree n
$v_i(k)$	ith section branch node output
$y_i(k)$	ith section output
$Z(s)$	Driving point impedance

CHAPTER I

INTRODUCTION

1.1 Statement of the Problem

In the past, minimum order transfer functions were considered optimal for filter designs. This is due to the fact that in the passive filter design the number of reactive elements is closely tied to the degree, n, of the transfer function, and that in the active and digital filter designs, in terms of first and second order cascade sections, the number of components depends upon the degree n. For a given class of functions such as Chebyshev, the poles will be closer to the imaginary axis in the s-domain for higher order functions. In the z-domain, the poles will be closer to the unit circle for higher n. Therefore, any change in the parameter values in the active or digital filters may move the poles into the region of instability [RA 1]. In addition, z-domain transfer functions with poles close to the unit circle will exhibit high output roundoff noise variance [GO 3], and the tolerance limits on the components of the active and passive filters will decrease with increasing n. Since the cost of the circuit components in the passive and active filter designs, in general, is inversely proportional to the tolerance limits, it is important to consider transfer functions which have poles away from the imaginary axis [GE 2].

In the digital filter design, the coefficient bit requirement increases as the poles approach the unit circle. Considering the cost aspect of filter designs, it is appropriate to find sub-optimal transfer functions which have poles away from the imaginary axis in the s-domain and away from the unit circle in the z-domain.

The worst case tolerance limits on the component values in the cascaded active filter design and coefficient bit requirements in cascaded digital filter design can best be estimated by considering the second order section corresponding to the dominant poles. For the dominant pole pair located at $s = -a \pm jb$, the corresponding quality factor Q_d is usually defined as $Q_d = \sqrt{a^2 + b^2} /2a$ [TE 1]. It is clear that the closer the poles are to the imaginary axis, the larger Q_d becomes.

In this thesis, the notion of modified transfer functions with dominant pole multiplicity greater than one is used in order to reduce Q_d and hence move the dominant poles away from the imaginary axis. These modified filter functions must always satisfy the pass-band and stop-band specifications, and they are here suggested as alternatives to the classical filter functions. The modified functions considered are the modified Butterworth and modified Chebyshev functions and are basically derived from classical Butterworth and Chebyshev functions, respectively.

Ideally a large number of bits is required to realize the coefficients (multipliers) of the discrete function, but due to the limited arithmetic word-length, the coefficients are rounded to the nearest quantization step. This change in coefficient values will result in an undesirable change in the pole location; therefore, it is important to use the minimum number of bits which can satisfy the pole tolerance limits. In this thesis a method is developed to estimate the coefficient bit requirement.

Several methods for calculating the digital filter output noise variance have been suggested [MI 1]; all of these methods rely on the z-domain transfer function for noise calculation. This requires that the s-domain filter function be transformed first into the z-domain prior to noise calculation. In this thesis, a method is given which relates the output noise variance computation directly to the s-domain filter transfer function and the filter driving point impedance.

1.2 Review of the Literature

Earlier, it was pointed out that there is a definite need for deriving sub-optimal transfer functions. This created an interest in deriving transfer functions with higher degree and lower dominant pole quality factor Q_d than is given by the original minimum order transfer functions [GO 1, KO 1]. These higher degree functions were obtained numerically. Kaiser pointed out that there is a definite need for developing analytical results in this area [KA 2]. Budak and Aronhime [BU 1] introduced the transitional Butterworth Chebyshev filters with reduced Q_d; this function is a combination of both the Butterworth and Chebyshev functions.

Recently, Premoli [PR 1] used the notion of multiplicity in the dominant poles to obtain analytically the multiple critical root maximally flat (MUCROMAF) polynomials for low-pass filters with lower Q_d and higher degree than the Butterworth functions. In addition, Premoli [PR 2] recently suggested a new class of multiple critical root pair equal ripple (MUCROER) filtering functions which possess reduced Q_d and a higher degree than the corresponding Chebyshev functions.

In considering the digital filter realization of transfer functions, the word-length requirements and output noise variance have been investigated [KN 1, GO 3]. Several methods for computing the output roundoff noise have been suggested [JU 1, JU 2, KN 1, GE 1, AS 1, MI 1]. An approach was also proposed for calculating the output roundoff noise variance by transforming the z-domain function into the s-domain [KI 1, GR 1].

It is well known that a digital filter, in general, has a low output noise variance when it is realized in terms of first and second order cascade sections [KU 1]. In addition, the ordering of the second order section for minimum output noise is important. Several numerical methods for reducing the digital filter output roundoff noise by selecting the optimum section ordering have been presented [JA 1, LE 1, GO 2]. A realization of the second order section with reduced quantization noise at low frequencies was also introduced [GO 3, KI 2].

In the area of coefficient word-length requirement, Mitra and Sherwood have presented a method for word-length estimation by evaluating the pole sensitivity with respect to coefficient changes due to quantization [MI 2]. Crochier approached the same problem by using a statistical method for word-length estimation [CR 1]. In order to satisfy the s-domain pole tolerance limits, White suggested a word-length estimation procedure that depends on the impulse invariant transformation [WH 1].

Cardwell [CA 1] attempted numerically to reduce the coefficient word-length by using higher order transfer functions; he succeeded in reducing the word-length at the cost of higher output noise variance. Avenhaus [AV 2, RA 1] investigated numerically word-length reduction using coefficient optimization techniques and higher degree functions; the word-length requirement was reduced but a higher output noise variance resulted. In this thesis, an attempt is made to reduce the output noise variance without increasing the word-length requirement (in many cases, lower word-length requirement results).

1.3 Technical Approach

The derivation of the modified Butterworth and Chebyshev functions is based on the notion of multiple dominant poles. The reduction in Q_d is possible because the multiple dominant poles will complement each other in giving the total high magnitude peak required originally by one second order high Q_d section. Since the dominant poles with multiplicity equal to two give the maximum percentage Q_d reduction [PR 1], further study is directed to this case.

The approach used in deriving the coefficients of the mth order low-pass multiple dominant pole modified Butterworth function $L_m(s)$ with reduced Q_d, given the nth order Butterworth function $H_n(s)$ with $m > n$, is explained in the following. Since $|H_n(j\omega)|$ is maximally flat at the origin, $|L_m(j\omega)|$ must also be maximally flat. This condition requires that the first $(n - 1)$ derivatives of $|L_m(j\omega)|^2$ with respect to ω^2 be equal to zero at $\omega = 0$. Due to the dominant pole multiplicity of c, the first $c - 1$ derivatives of the denominator of $|L_m(j\omega)|^2$ with respect to ω^2 must be zero when evaluated at the dominant pole location. In addition, the denominator of $|L_m(j\omega)|$ must equal zero when evaluated at the dominant pole location. To satisfy the pass-band specifications it is required that $|H_n(j1)| = |L_m(j1)|$. At $\omega = 0$, the magnitude $|L_m(j\omega)|^2$ must also be equal to 1 or $1/(1 + \varepsilon^2)$ for odd and even m, respectively, where ε is the ripple factor. The pass-band specifications will always be satisfied, and a root locus approach is used to increase the transition region attenuation.

The coefficients of the low-pass non-equal-ripple mth order modified Chebyshev functions (MCF's) with multiple dominant poles having significantly reduced Q_d are derived numerically by employing a new algorithm called the physical method. In this algorithm the dominant poles of the original Chebyshev function of order n $(n < m)$ are replaced by multiple dominant poles; the magnitude of every second order section is iteratively adjusted until the pass-band specifications are met. Since the MCF's are not unique, the classical least squares error algorithm is used to derive the MCF's and the two methods are compared. By increasing the dominant pole break frequency, intermediate modified Chebyshev functions with higher transition region attenuation can be obtained at the cost of increasing Q_d.

Having derived the MCF's, a comparison of the digital filter output noise variance and coefficient word-length requirement between the low-pass nth order Chebyshev functions and the low-pass double dominant pole MCF's of order $(n + 2)$ is investigated. The output noise variance is obtained after scaling and optimum section ordering as discussed by Cardwell [CA 1], while the estimation of the coefficient word-length follows a method derived in this study for the cases where the bilinear transformation is employed. In the suggested method for noise computation, the bilinear transform is also used in relating the output noise variance to the s-domain driving point impedance. The digital filter realization considered in this study is in terms of first and second order cascaded canonical sections [GO 3]. Fixed point arithmetic and rounding of products before summation is assumed.

1.4 Organization of the Thesis

Chapter II presents an analytical method for deriving the coefficients of the modified low-pass maximally flat Butterworth polynomial with low Q_d and multiplicity of the dominant pole pair greater than one. A root locus method is also presented to determine the modified Butterworth coefficients such that the attenuation of the transition region is increased.

Chapter III presents a new numerical algorithm (the physical method) which determines the coefficients of a low-pass non equal-ripple MCF function with multiple dominant poles and significantly lower Q_d than the corresponding Chebyshev function. For higher transition region attenuation the physical method can also generate intermediate MCF's. The results are compared with those obtained using the least squares error algorithm.

Chapter IV presents the digital filter output roundoff noise comparison between the Chebyshev and MCF functions. The suggested method for output roundoff noise calculation using the s-domain transfer function and the filter driving point impedance is also given.

Chapter V presents a method for estimating the coefficient word-length such that the s-domain pole tolerance limits are satisfied. The digital filter coefficient word-length comparison for the Chebyshev and MCF realization is also given.

Chapter VI presents a summary and suggestions for further study.

Appendices A, B, and C present the algorithms for the physical method, the least squares error method, and for roundoff noise computation.

CHAPTER II

MODIFIED BUTTERWORTH FUNCTIONS WITH LOW Q-FACTOR

2.1 Introduction

In active RC and digital filter designs, the precision requirement of each second order section might dictate a constraint on the maximum value of the dominant pole quality-factor Q_d [HU 1, KA 1, TE 1].

In this chapter, a method is given to determine the coefficients of a modified low pass maximally flat (at the origin) Butterworth polynomial, with a lower dominant pole pair Q-factor using a higher order polynomial with multiplicity of the dominant pole pair greater than one.

There has been some interest in deriving higher order transfer functions with low Q dominant poles [GO 1, LO 1]. Most of these are based upon numerical optimization techniques. It has been pointed out that there is a definite need in developing analytical results in this area [KA 2]. In a recent paper, the Multiple Critical Root Maximally Flat (MUCROMAF) polynomials for low pass filters were proposed where higher order polynomials with multiple critical roots were developed [PR 1]. The coefficients were obtained by solving two polynomial equations and a set of n - 2 simultaneous linear equations in (n - 2) unknowns, where n is the degree of the transfer function with no root multiplicity. Thus as n increases, the number of simultaneous equations to be solved increases.

Using this same notion of multiplicity in the dominant pole pair, this chapter proposes an alternate method which specifies the coefficients needed with fewer number of equations; the number of equations depends upon the dominant root multiplicity rather than the degree of the transfer function [MA 1]。 The same results that were obtained with the MUCROMAF polynomials are obtained here, but with fewer equations。 Computationally, the method presented here is superior to that presented in an earlier paper [PR 1] when the multiplicity of the dominant poles equal to two。 A modification of this method to fit more stringent frequency domain specifications is also presented。

2。2 Problem Statement

Let

$$|H_n(j\omega)|^2 = \frac{1}{1 + \varepsilon^2\omega^{2n}} \tag{2-1}$$

be the Butterworth function satisfying the pass and stop band requirements in the frequency domain。 It is required to find a modified Butterworth function

$$|L_m(j\omega)|^2 = \frac{1}{1 + \sum_{i=1}^{m} d_{2i}\omega^{2i}} \tag{2-2}$$

which satisfies the frequency domain specifications with the constraint that the magnitude function

$$F(\omega^2) = \left| \frac{H_n(j\omega)}{L_m(j\omega)} \right|^2 \tag{2-3}$$

is required to deviate the least amount from unity for frequencies close to $\omega = 0$ [HS 1]。 Furthermore, the Q of the dominant poles obtained from Equation (2-2) must be less than the specified Q。 Let the specified Q be

Q_d. It is evident that this condition implies that m > n.

In the following section, modified Butterworth functions are derived with the idea that the Q's of the dominant poles can be minimized by having the dominant roots with multiplicity c greater than one. It is clear that the magnitude of the dominant second order section of $H_n(s)$ will have a large overshoot for a low damping ratio. The reduction in Q_d is possible since the multiple critical poles complement each other in giving the total high peak required originally by one second order section in the Butterworth function. This results in identical second order sections each with low Q_d replacing the original high Q_d second order section. The resulting function is called modified Butterworth function, since it is maximally flat at the origin.

2.3 Modified Butterworth Polynomials

The coefficients d_{2i}, i = 1,...,m, in Equation (2-2) can be determined by observing the following interesting aspect of $F(\omega^2)$. The function $|H_n(j\omega)|^2$ is to be approximated by a higher order function $|L_m(j\omega)|^2$. Furthermore, $|H_n(j\omega)|^2$ is a maximally flat function and it is required that $|L_m(j\omega)|^2$ also be a maximally flat function. Therefore, $F(\omega^2)$ must also be a maximally flat function. Since $F(\omega)$ is a function of $\omega^2 = x$, it can be expressed in terms of its Taylor's series about $\omega = 0$, in the form [HS 1],

$$F(x) = f(0) + F'(0)\frac{x}{1!} + \cdots + F^{(i)}(0)\frac{x^i}{1!} + \cdots$$

where

$$F^{(i)}(0) = \frac{d^i}{dx^i} F(x)\Big|_{x=0} .$$

The maximally flat property of F(x) and its value at $\omega = 0$ implies that

$$F(0) = 1$$

and that $F^{(i)}(0) = 0$ for $i = 1,\ldots,n-1$. This results in

$$\left(\frac{d}{d\omega^2}\right)^i |L_m(j\omega)|^2 = 0 \qquad \text{for } i = 1,2,\cdots,n-1$$

which implies that

$$d_2 = d_4 = \cdots = d_{2n-2} = 0$$

so that Equation (2-2) can be written as

$$|L_m(j\omega)|^2 = \frac{1}{1 + \sum_{i=n}^{m} d_{2i}\ \omega^{2i}} = \frac{1}{1 + \varepsilon^2\ \omega^{2n}\ (\sum_{i=0}^{2(c-1)} a_{i+1}\ \omega^{2i})} \tag{2-4}$$

where the a_i's remain to be determined. Let $\omega_d = (R + jI)$ be the location of the multiple poles. Due to the pole multiplicity, the first $c - 1$ derivatives of the denominator in Equation (2-4) will have a zero at $\omega = \omega_d$. These result in the following equations. The first derivative results in

$$\sum_{i=0}^{2(c-1)} (n + i)\ a_{i+1}\ \omega_d^{2i} = 0 \quad . \tag{2-5}$$

The remaining derivatives result in

$$\begin{aligned} &(n + i - 1)(i - 1)!\ a_i + (n + i)i!\ a_{i+1}\ \omega_d^2 + \\ . &\sum_{k=2}^{2c-i-1} (n + i + k - 1)\ a_{i+k}\ \omega_d^{2k} \prod_{j=k+1-i}^{k-1} (i + j) = 0 \quad . \\ &\qquad\qquad i = 2,3,\cdots,c-1 \end{aligned} \tag{2-6}$$

In addition to these derivatives, the denominator in Equation (2-4) has a zero at $\omega = \omega_d$. This results in the equation

$$1 + \varepsilon^2 \omega_d^{2n} \left[\sum_{i=0}^{2(c-1)} a_{i+1} \omega_d^{2i} \right] = 0 \qquad . \text{(2-7)}$$

At the cut off frequency, $\omega = 1$ (normalized), it is required that $|H_n(j\omega)|^2 = |L_m(j\omega)|^2$; hence

$$a_1 + a_2 + a_3 \cdots + a_{2c-1} = 1 \qquad . \text{(2-8)}$$

By substituting a_1 from (2-8) into Equations (2-5) and (2-7), and by equating the real and imaginary parts of Equations (2-5), (2-6), and (2-7) to zero, a set of 2c simultaneous nonlinear equations in the 2c unknowns R, I, a_2, a_3, ..., a_{2c-1} results. There exists $\frac{n}{2}$ solutions ($\frac{n+1}{2}$ if n is odd) to these equations, each of which corresponds to a multiple pole on one complex conjugate pole pair of $H_n(s)$. In order that the dominant poles correspond to the multiple poles, one has to initialize the search subroutine used to solve the 2c equations with R and I values close to the dominant pole values of $H_n(s)$.

Next, let us examine the pass and stop band requirements.

Theorem 2.3.1

$|L_m(j\omega)|^2$ satisfies the pass-band specifications; that is

$$\frac{1}{1 + \varepsilon^2} \leq |L_m(j\omega)|^2 \leq 1 \qquad \text{for} \qquad 0 \leq \omega \leq 1 \qquad .$$

Proof

The constraints at the terminal points are evident from Equations (2-4) and (2-8). Therefore, one needs to show that for all ω, $0 < \omega < 1$,

$$0 < f(\omega) = \omega^{2n} \sum_{i=0}^{2(c-1)} a_{i+1} \omega^{2i} < 1 \qquad .$$

Thus it is sufficient to show that there exists no real ω_p such that

$0 < \omega_p < 1$ and $f'(\omega_p) = \frac{d}{d\omega^2} f(\omega)|_{\omega=\omega_p} = 0$. Differentiating $f(\omega)$ with respect to ω^2 and equating it to zero, it follows that

$$\omega^{2n-2} \left[\sum_{i=0}^{2(c-1)} (n + i)\, a_{i+1}\, \omega^{2i} \right] = 0 \tag{2-9}$$

which has $2n - 2$ roots at $\omega = 0$ and $4(c - 1)$ roots at $\omega = \omega_d = R + j\, I$ which is complex (see Equation 2-5)). Thus there exists no real ω_p such that $\omega_p > 0$ and $f'(\omega_p) = 0$. Therefore, the proof follows.

The above discussion indicates that the pass-band requirements will always be satisfied. The stop-band requirements will be discussed later. First, a special case is considered.

2.3.1 Double Pole (c = 2) Case

It can be observed that the rate of Q_{dc} drop is largest in the case of $c = 2$ [PR 1], where Q_{dc} is the Q-factor of the dominant pole pair. It is therefore necessary to investigate this case of $c = 2$ further. It will be demonstrated that the $2c = 4$ equations needed to solve for the coefficients a_2, a_3, R, and I can be reduced to two equations in two unknowns R and I.

The $2c = 4$ equations are the real and imaginary parts of Equations (2-5) and (2-7). Solving Equation (2-5) for ω_d^2 and using $\omega_d = R + j\, I$ and $a_1 = 1 - a_3 - a_2$ (Equation (2-8)), the following equations result.

$$(I^2 + R^2)^2 = \frac{n(1 - a_3 - a_2)}{(n + 2)a_3} \tag{2-10}$$

$$I^2 - R^2 = \frac{(n + 1)a_2}{2(n + 2)a_3} \tag{2-11}$$

Now, from (2-10) and (2-11), a_2 and a_3 can be expressed as

$$a_2 = \frac{2(I^2 - R^2)(n + 2)a_3}{n + 1} \qquad (2\text{-}12)$$

$$a_3 = \frac{n(n + 1)}{(n + 1)(n + 2)(I^2 + R^2)^2 + n(n + 1) + 2n(I^2 - R^2)(n + 2)} \qquad (2\text{-}13)$$

Substituting a_2, a_3, and a_1 (from Equation (2-8)) in Equation (2-7), the following equation results.

$$1 + \varepsilon^2\, \omega_d^{2n} \left[1 + \frac{n(\omega_d^4-1)(n+1) + 2n(n+2)(I^2-R^2)(\omega_d^2-1)}{(n+1)(n+2)(I^2+R^2)^2 + n(n+1) + 2n(n+2)(I^2-R^2)} \right] = 0 \;. \qquad (2\text{-}14)$$

The real and imaginary parts of Equation (2-14) result in two equations and two unknowns. These can be solved for the double pole location irrespective of the polynomial degree n; a_1, a_2, a_3, can then be determined from Equations (2-8), (2-12), and (2-13). In Table I, Q_{d_1}, Q_{d_2}, and a_i are given, where Q_{d_1} corresponds to the Q of the dominant pole pair of the Butterworth function; Q_{d_2} corresponds to the Q of the double dominant pair. Computationally, the above two equations are simpler to solve on a computer than a set of linear equations and two polynomials presented in [PR 1]. This is due to the fact the initial guess of the solution (dominant poles of the nth order Butterworth function) is very close to the solution itself. However, the story is different when c is greater than two. Convergency problems do arise, and the approach in [PR 1] is more appropriate.

In Table II, the poles of the modified Butterworth function for n = 3,...,15 are given.

TABLE I

Q_{dc} AND COEFFICIENT VALUES FOR c = 2 AND n = 3,...,15
$|L_m(j\omega)|^2 = 1/(1 + \varepsilon^2\omega^{2n}(a_1 + a_2\omega^2 + a_3\omega^4))$

c	1	2			
n	Q_{d1}	a_1	a_2	a_3	Q_{d2}
3	1.000000	1.139381	- 0.521951	0.382570	0.919211
4	1.306563	1.555091	- 1.222708	0.667616	1.135737
5	1.618034	2.059002	- 2.088475	1.029472	1.354353
6	1.931852	2.646374	- 3.115571	1.469197	1.574127
7	2.246980	3.315329	- 4.302638	1.987309	1.794610
8	2.562916	4.064973	- 5.649071	2.584098	2.015560
9	2.879385	4.894824	- 7.154565	3.259741	2.236834
10	3.196227	5.804601	- 8.818953	4.014351	2.458343
11	3.513337	6.794128	-10.642135	4.848007	2.680026
12	3.830649	7.863287	-12.624051	5.760763	2.901844
13	4.148114	9.012001	-14.764661	6.752659	3.123766
14	4.465702	10.240212	-17.063938	7.823725	3.345773
15	4.783385	11.547880	-19.521864	8.973984	3.567847

TABLE II

POLES OF MODIFIED BUTTERWORTH TRANSFER FUNCTION $L_m(s)$ FOR $c = 2$, $m = n + 2(c - 1)$

n	Double Pole							
3	-0.628901 ±J0.970179	-0.904764						
4	-0.491447 ±J1.002310	-0.836246 ±J0.298033						
5	-0.403614 ±J1.016042	-0.756913 ±J0.472797	-0.866199					
6	-0.342443 ±J1.022266	-0.684078 ±J0.586162	-0.843611 ±J0.202071					
7	-0.297358 ±J1.025022	-0.620842 ±J0.664403	-0.802678 ±J0.346784	-0.864777				
8	-0.262745 ±J1.026048	-0.566708 ±J0.720905	-0.756746 ±J0.454058	-0.854780 ±J0.155442				
9	-0.235335 ±J1.026173	-0.520374 ±J0.763175	-0.711259 ±J0.535792	-0.829941 ±J0.277152	-0.870318			
10	-0.213094 ±J1.025818	-0.480524 ±J0.795713	-0.668377 ±J0.599517	-0.798489 ±J0.373947	-0.865289 ±J0.127149			
11	-0.194687 ±J1.025208	-0.446018 ±J0.821357	-0.628833 ±J0.650191	-0.764691 ±J0.452043	-0.848742 ±J0.231922	-0.877209		
12	-0.179201 ±J1.024471	-0.415925 ±J0.841971	-0.592731 ±J0.691179	-0.730777 ±J0.515894	-0.825941 ±J0.318961	-0.874532 ±J0.107938		
13	-0.165994 ±J1.023679	-0.389493 ±J0.858824	-0.559904 ±J0.724831	-0.697893 ±J0.568735	-0.800005 ±J0.391874	-0.862804 ±J0.199882	-0.884001	
14	-0.154596 ±J1.022872	-0.366121 ±J0.872804	-0.530081 ±J0.752823	-0.666597 ±J0.612950	-0.772798 ±J0.453462	-0.845566 ±J0.278563	-0.882564 ±J0.093957	
15	-0.144662 ±J1.022075	-0.345325 ±J0.884550	-0.502965 ±J0.776379	-0.637124 ±J0.650319	-0.745435 ±J0.505898	-0.825107 ±J0.346247	-0.873873 ±J0.175880	-0.890293

2.3.2 Stop Band Specifications

Earlier, it was shown that the pass-band specifications are always satisfied. Now, the stop-band requirements will be examined for $|L_m(j\omega)|^2$ in Equation (2-4). Since, m > n, there exists an ω_t, such that

$$|L_m(j\omega)|^2 \leq |H_n(j\omega)|^2 \qquad \text{for} \qquad \omega \geq \omega_{tm} \quad .$$

Stop-band requirements can be examined by finding the frequencies at which $|L_m(j\omega)|^2 = |H_n(j\omega)|^2$. These frequencies can be found by solving the equation

$$\omega^{2n} \left(\sum_{i=0}^{2(c-1)} a_{i+1} \, \omega^{2i} - 1 \right) = 0 \quad .$$

For the special case (c = 2), the above equation reduces to

$$\omega^{2n}(a_3\omega^4 + a_2\omega^2 + a_1 - 1) = 0$$

which has 2n roots at the origin and the remaining four roots are located at (see Equation (2-8))

$$\omega_{1,2} = \pm 1 \quad , \quad \omega_{3,4} = \pm \sqrt{-1 - \frac{a_2}{a_3}} \tag{2-15}$$

where $\omega_{tm} = \omega_3$.

From Table II, one can see that $\omega_{tm} = \sqrt{-1 - \frac{a_2}{a_3}} < 1$ for n = 3, 4. For n > 5, $\omega_{tm} > 1$. This implies that for n = 3, 4, the stop-band requirements will always be satisfied. However, for n > 5, the stop band requirements are met if $\omega_r \geq \omega_{tm}$ where ω_r is the lowest specified stop-band frequency. On the other hand, if $\omega_r < \omega_{tm}$, then the specifications are not met. The procedure needs to be modified and the multiple poles must be separated into single poles, which is discussed below.

Now, a_2 can be expressed in terms of a_3 and ω_{tm} and is

$$a_2 = - a_3(\omega_{tm}^2 + 1) \qquad .(2\text{-}16)$$

From (2-8) and (2-16)

$$a_1 = 1 + a_3 \omega_{tm}^2 \qquad .(2\text{-}17)$$

Using these expressions in (2-4), $|L_m(j\omega)|^2$ can be reduced to

$$|L_m(j\omega)|^2 = \frac{1}{1 + \varepsilon^2 \omega^{2n}[(\omega_{tm}^2 - \omega^2)(1 - \omega^2)a_3 + 1]} \qquad .(2\text{-}18)$$

A root locus plot of the denominator in (2-18) in terms of the one variable a_3 with ω_{tm} set equal to ω_r, gives the value of a_3 such that the dominant poles are at their maximum distance from the $j\omega$ axis. a_2 and a_1 can then be calculated using (2-16) and (2-17).

2.3.3 Example

Given the eighth order low-pass Butterworth transfer function $|H_8(j\omega)|^2 = 1/(1 + \omega^8)$ with $\varepsilon = 1$ and $Q_{d_1} = 2.562916$ satisfying the magnitude specifications in the normalized frequency domain with $\omega_r = 1.2$, it is required to reduce Q_{d_1} using multiplicity of the dominant pole pair equal to two (c = 2). From Table I,

$$|L_{10}(j\omega)|^2 = \frac{1}{1 + \omega^8(4.065 - 5.649\,\omega^2 + 2.584\,\omega^4)}$$

$$Q_{d_2} = 2.01556$$

and from (2-15) $\omega_{tm} = 1.089$ which is less than ω_r. Therefore the specifications are met. In Figure 1, $|H_8(j\omega)|^2$ and $|L_{10}(j\omega)|^2$ are plotted to show the difference in the two approximations. If $\omega_r = 1.075 < \omega_{tm}$, then

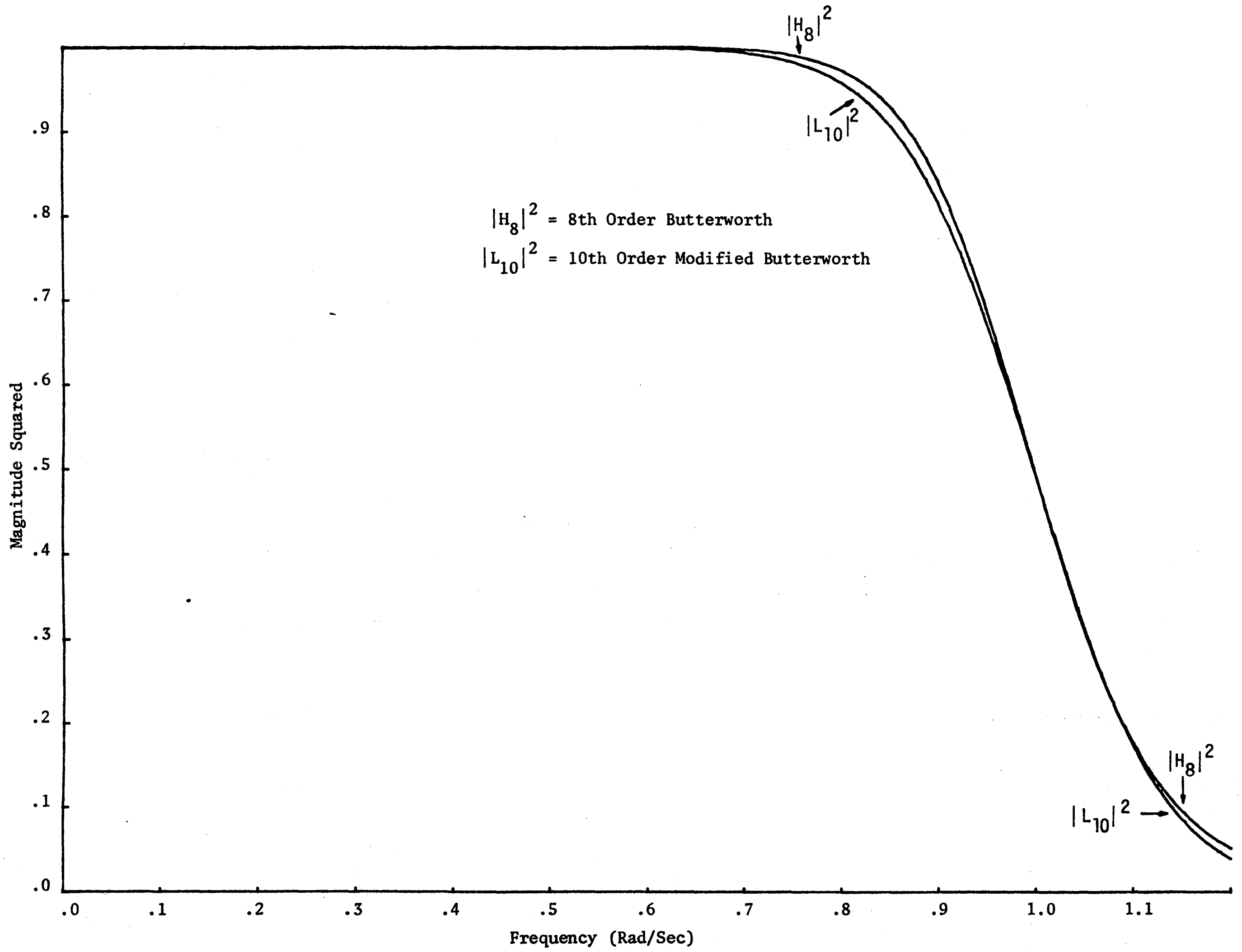

Figure 1. Magnitude Comparison Between Eighth Order Butterworth and Tenth Order Modified Butterworth Functions

in order to meet this specification the double poles need to be separated by applying the root locus technique discussed above. This results in $a_3 = 2.6285$, and $a_2 = -5.6661$ and $a_1 = 4.03756$. The Q of the dominant pole is given by $Q_d = 2.2258$.

2.4 Summary

A method that reduces the Q of the dominant pole pair of a Butterworth function $H_n(s)$ using a higher order function $L_m(s)$ is presented. It is assumed that $L_m(s)$ has dominant poles of multiplicity greater than one. Furthermore, the transfer function $L_m(s)$ is derived using the assumption that

$$\left(\frac{d}{d\omega^2}\right)^i |L_m(j\omega)|^2 = 0 \qquad \text{for} \qquad i = 1,2,\cdots,n-1 \quad .$$

It is shown that $|L_m(j\omega)|^2$ satisfies the pass-band specifications, and a method is given to fit $|L_m(j\omega)|^2$ to the stop-band specifications.

CHAPTER III

MODIFIED CHEBYSHEV FILTERING FUNCTIONS WITH LOW Q- FACTOR

3.1 Introduction

An important factor in the design of RC active filters and digital filters is the quality factor Q of the dominant poles. The precision requirements of each second order section realization of an RC or digital filter might dictate a constraint on the maximum value of Q [HU 1, KA 1, TE 1]. Modified Butterworth functions were presented in the previous chapter; here a method for modified Chebyshev function derivation is developed.

In this chapter, a new numerical algorithm is presented which determines the coefficients of a low-pass non-equal-ripple modified Chebyshev function with multiplicity of the dominant root pair greater than one; as a result its degree is higher than the corresponding Chebyshev polynomial but a much lower dominant root Q-factor Q_d is obtained. Intermediate modified Chebyshev functions with higher transition region attenuation and therefore increased Q_d are also discussed.

The concept of multiplicity in the dominant poles has been recently used and a substantial reduction in the critical Q-factor, Q_d, of the dominant poles resulted [PR 1, MA 1, PR 2]. In a recent paper [PR 2], a new class of multiple critical root pair, equal ripple (MUCROER) filtering functions, having a higher degree than the Chebyshev filtering functions, but with a much reduced Q_d and an improved time delay characteristic has been developed using the Remez algorithm; however, a reduction in the attenuation in the transition-band resulted.

Using this same notion of multiplicity in the dominant pole pair, a new numerical algorithm called the physical method is presented. The modified Chebyshev filtering function (MCF) obtained here is of a higher order than the original Chebyshev function. By relaxing the equal-ripple condition, a degree of freedom is obtained which results in a significantly lower critical quality factor Q_d and a better time delay characteristic than that achieved by either the MUCROER or Chebyshev polynomials; however, the transition region attenuation is further reduced. The algorithm generates a MCF function for every Chebyshev polynomial. It can also generate intermediate modified Chebyshev filtering functions (IMCF's) satisfying the pass band specifications and which have Q_d and transition region attenuation anywhere between the Q_d and transition region attenuation of MCF and MUCROER.

An example is given where for a low-pass filter with pass-band reflection coefficient of 50%, the Q_d of a tenth order Chebyshev polynomial is reduced 71.27% using a twelfth order MCF; whereas, the twelfth order MUCROER polynomial gives 58.89% reduction. Due to the reduction in Q_d, the tolerance limits on the realized components is reduced. The coefficient sensitivities are compared, and the output noise variance due to roundoff in the digital filter realization of each of these functions is given.

3.2 Problem Statement

Let

$$|H_n(j\omega)|^2 = \frac{1}{1 + \varepsilon^2 T_n^2(\omega)} \tag{3-1}$$

be the Chebyshev function satisfying the pass and stop-band requirements in the frequency domain, where $T_n(\omega)$ corresponds to the Chebyshev polynomial of degree n. It is required to find a MCF function

$$|L_m(j\omega)|^2 = \frac{1}{1 + \sum_{i=1}^{m} d_{2i}\ \omega^{2i}} \tag{3-2}$$

with a reduced Q_d of the dominant poles which also satisfies the frequency domain specifications. It is evident that this implies that $m > n$.

In the following section, modified Chebyshev functions are derived with the idea that the Q's of the dominant poles can be minimized by having the dominant roots with multiplicity $c > 1$.

3.3 Modified Chebyshev Functions (MCF's)

The numerical methods presented here are applicable to low-pass Chebyshev functions; however, by applying the classical frequency transformations, other modified Chebyshev low Q filter functions can be obtained.

First, a new numerical algorithm called Physical Method is developed. Second, the ideas are extended to the least squares error criterion. The results obtained by the two methods are in close agreement; however, the first method has more advantages and requires less computational time because it is developed for the present problem at hand, while the second method is more general and is presented here for purposes of convenience and comparison. Third, the results are extended to obtain the IMCF's.

3.3.1 Physical Method

A Chebyshev transfer function can be written, for n even and odd, respectively, as

$$H_n(s) = \prod_{i=1}^{n/2} \frac{\omega_{bi}^2}{s^2 + 2\delta_i\omega_{bi}s + \omega_{bi}^2} = \prod_{i=1}^{n/2} h_i(s) \tag{3-3a}$$

$$H_n(s) = K \prod_{i=1}^{(n-1)/2} \frac{\omega_{bi}^2}{s^2 + 2\delta_i\omega_{bi}s + \omega_{bi}^2} \cdot \frac{\omega_{bv}}{s + \omega_{bv}} = K \prod_{i=1}^{(n+1)/2} h_i(s) \tag{3-3b}$$

with

$$K = \sqrt{1 + \varepsilon^2} \qquad \text{and} \qquad v = \frac{(n + 1)}{2} \quad .$$

In the above equations, let $h_1(s)$ represent the section with dominant poles, δ_i is the damping ratio, ω_{bi} the break frequency, and K adjusts the maximum passband ripple to $\sqrt{1 + \varepsilon^2}$ for the design purposes; this value of K is selected for the convenience of the algorithm. Let the quality factor of the second order function corresponding to the dominant poles be represented by Q_d. It is clear that the magnitude function $|h_i(j\omega)|$ will have a large overshoot for a low δ_i. As pointed out in the last chapter, the reduction in Q_d is possible since the multiple critical poles complement each other in giving the total high peak required originally by one second order section in the Chebyshev function. This results in identical second order sections each with low Q_d replacing the original high Q_d second order section. The resulting function is called a modified Chebyshev function (MCF) since it is developed from the roots of the Chebyshev polynomial. This can be written for n even and odd, respectively, as

$$L_m(s) = \left[\frac{\omega_{b1}^2}{s^2 + 2\beta_1\omega_{b1}s + \omega_{b1}^2}\right]^c \prod_{i=2}^{n/2} \frac{\omega_{bi}^2}{s^2 + 2\beta_i\omega_{bi}s + \omega_{bi}^2} = (\ell_1(s))^c \prod_{i=2}^{n/2} \ell_i(s)) \tag{3-4a}$$

$$L_m(s) = K\left[\frac{\omega_{b1}^2}{s^2 + 2\beta_1\omega_{b1}s + \omega_{b1}^2}\right]^c \prod_{i=2}^{(n-1)/2} \frac{\omega_{bi}^2}{s^2 + 2\beta_i\omega_{bi}s + \omega_{bi}^2} \; \frac{\omega_{bv}}{s + \omega_{bv}} \tag{3-4b}$$

$$= (\ell_1(s))^c \prod_{i=2}^{(n+1)/2} \ell_i(s)$$

with $m = n + 2(c - 1)$ and $v = (n + 1)/2$, where c is the multiplicity of the dominant pole pair, and $\ell_1(s)$ represents the section with the dominant poles. The only unknowns in Equation (3-4a) or (3-4b) are the damping ratios β_i, and ω_{b1}; whereas, the break frequencies $\omega_{bi} i \neq 1$ are the same as in the Chebyshev case. Due to critical pole modification into multiple poles, the pass-band specifications are met by simply modifying ω_{b1} of the critical poles and β_i which controls the peak values of every second order section; by also fixing ω_{bi}, $i \neq 1$, the result will not be the MUCROER equiripple polynomial and the number of variables is reduced.

The equations in (3-4) provide a basis upon which the physical method is developed and is presented in terms of the following steps. Figure 2 gives the flow chart and Appendix A shows the program listing The first step is to solve for β_1 such that $(\max|\ell_1(j\omega)|)^c = \max|h_1(j\omega)|$.

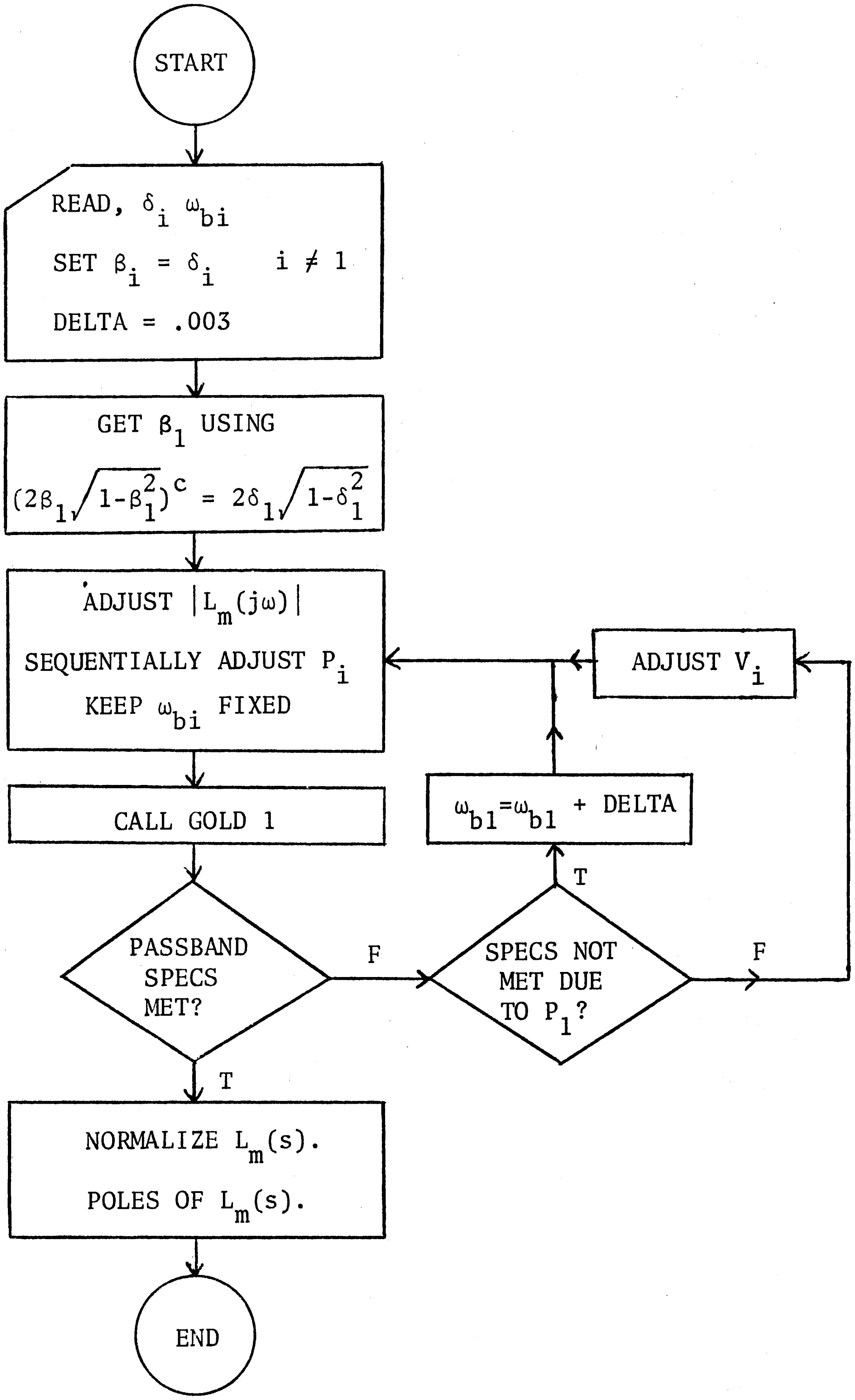
START
READ, δ_i ω_{bi}
SET $\beta_i = \delta_i$ $i \neq 1$
DELTA = .003
GET β_1 USING
$(2\beta_1\sqrt{1-\beta_1^2})^c = 2\delta_1\sqrt{1-\delta_1^2}$
ADJUST $|L_m(j\omega)|$
SEQUENTIALLY ADJUST P_i
KEEP ω_{bi} FIXED
CALL GOLD 1
PASSBAND SPECS MET?
T
F
SPECS NOT MET DUE TO P_1?
T
F
$\omega_{b1}=\omega_{b1}$ + DELTA
ADJUST V_i
NORMALIZE $L_m(s)$.
POLES OF $L_m(s)$.
END

Figure 2. Flow Chart Using the Physical Method

This implies that

$$\left[\frac{1}{2\beta_1\sqrt{1-\beta_1^2}}\right]^c = \left[\frac{1}{2\delta_1\sqrt{1-\delta_1^2}}\right] \tag{3-5}$$

which is obtained by first setting $d|h_1(j\omega)|/d\omega]_{\omega=\omega_{max}} = 0$ and then computing $|h_1(j\omega_{max})|$. In a similar manner, $\max|\ell_1(j\omega)|$ can be computed [ME 1]. The remaining β_i's are selected such that $\beta_i = \delta_i$. The resulting $|L_m(j\omega)|$ will not satisfy the pass-band specifications. The second step involves an iterative technique to modify the parameters in $L_m(j\omega)$ such that the pass-band specifications are satisfied. Let ω_{pi} be the frequency at which the peak of $|\ell_i(j\omega)|$ appear (see Figure 3), and let

$$P_i = \max_{\omega} |\ell_i(j\omega)| \quad , \quad V_i = |H_n(j\omega_{pi})| \qquad .$$

In the following, even n will be considered. However, the same procedure applies for odd n also. The iteration procedures is as follows:

1) Calculate $P_{(\frac{n}{2}-i)new}$ and $\beta_{(\frac{n}{2}-i)new}$, $i = 0,\ldots,\frac{n}{2}-2$, successively using [ME 1]

$$|\ell_1(j\omega)|^c \prod_{k=\frac{n}{2}-i+1}^{n/2} |\ell_{k\ new}(j\omega)|^2 \prod_{k=2}^{\frac{n}{2}-i-1} |\ell_k(j\omega)| P_{(\frac{n}{2}-i)new} \Bigg|_{\omega=\omega_{P(\frac{n}{2}-i)}} = V_{\frac{n}{2}-i} \tag{3-6a}$$

$$\beta_{(\frac{n}{2}-i)new} = \sqrt{.5 - .5\sqrt{1 - \left[1/P_{(\frac{n}{2}-i)new}\right]^2}} \quad . \tag{3-6b}$$

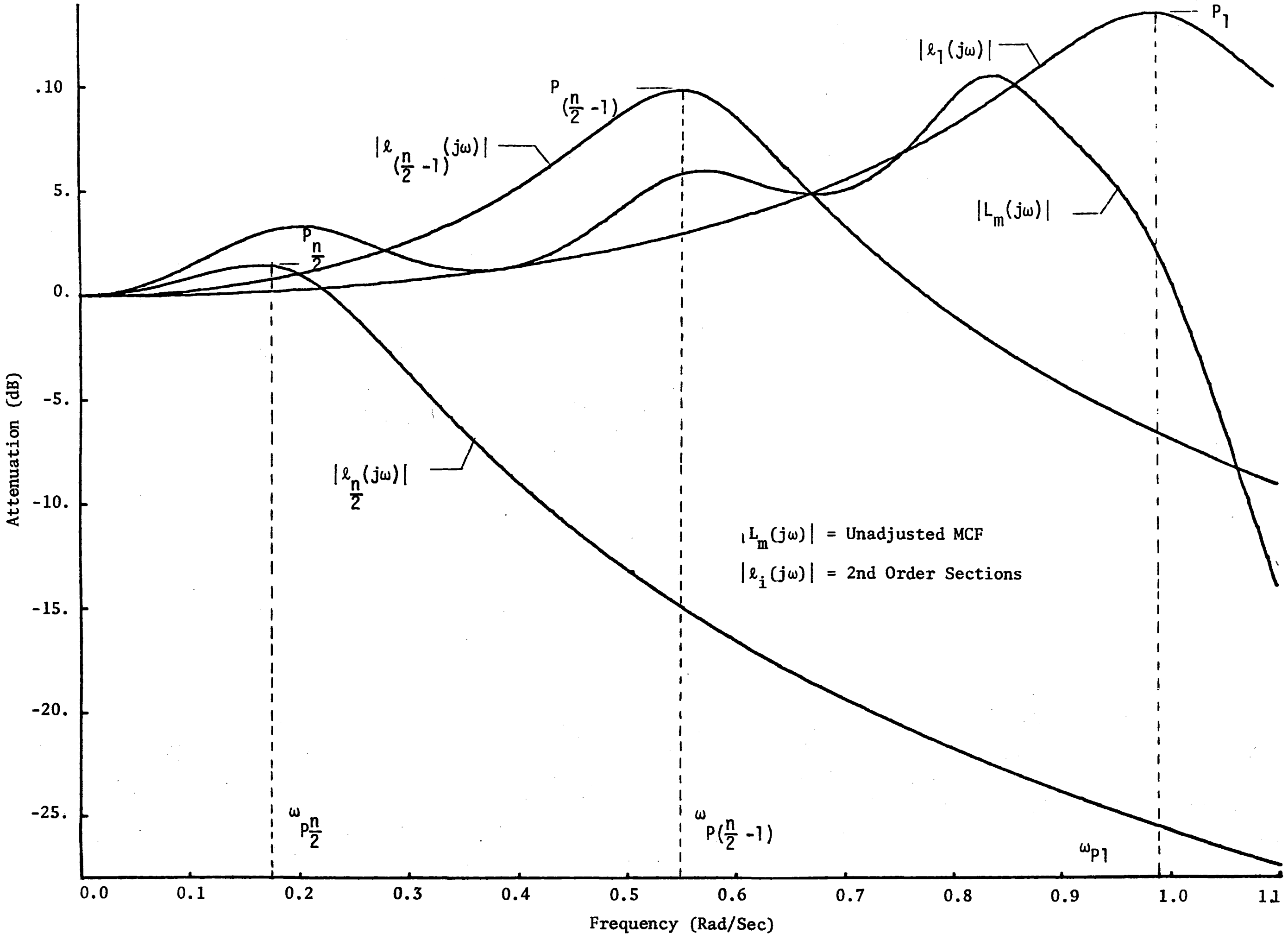

Figure 3. Plot of $|L_m(j\omega)|$ and Each of its Second Order Sections

Each step in the above computation reduces the peak of $\left|\ell_{(\frac{n}{2}-i)\text{new}}(j\omega)\right|$ such that $|L_{m\ \text{new}}(j\omega)|$ fits the specification at $\omega = \omega_{p(\frac{n}{2}-i)}$ (see Figure 3). Next, $P_{1\ \text{new}}$ can be computed using

$$(P_{1\ \text{new}})^c \cdot \prod_{k=2}^{n/2} \left|\ell_{k\ \text{new}}(j\omega)\right|_{\omega=\omega_{p1}} = V_1 \qquad (3\text{-}7)$$

and $\beta_{1\ \text{new}}$ can be computed using Equation (3-6b).

2) Keeping the break frequencies $\omega_{b1}, \omega_{b2}, \ldots, \omega_{b\ n/2}$ fixed, calculate the new location $\omega_{pi\ \text{new}}$ of the new peaks using $\omega_{pi\ \text{new}} = \omega_{bi}\sqrt{1 - 2\beta^2_{i\ \text{new}}}$. (This is obtained by setting $d|\ell_k(j\omega)|/d\omega]_{\omega=\omega_{\max}} = 0.0$, [ME 1]).

3) Repeat steps 1) and 2) until

$$\left|\beta_{(\frac{n}{2}-i)\text{new}} - \beta_{(\frac{n}{2}-i)\text{new-1}}\right| < \alpha \qquad i = 0,1,\cdots,\frac{n}{2}-1$$

where α is a specified small constant. Numerically, it was observed that the convergency rate is fast. For example, for $n = 10$, $c = 2$, 1/2 a dB ripple, and $\alpha = 10^{-3}$ only 5 iterations were required.

4) Call Golden section univariate search [WI 1] to calculate the minima and maxima of $|L_{m\ \text{new}}(j\omega)|$ in the pass-band. If the pass-band specifications are met, then $L_{m\ \text{new}}$ is normalized to a cutoff frequency $\omega_c = 1$. Otherwise, find the peak P_i which causes $|L_{m\ \text{new}}(j\omega)|$ to violate the specifications. If it is P_1, then set $\omega_{b1} = \omega_{b1}$ + delta, where delta is a small positive increment, and repeat steps (1) and (2); otherwise, adjust the corresponding V_i and repeat steps (1) and (2).

It should be pointed out that by increasing the break frequency ω_{b1}

of the multiple pole, the cutoff frequency of the $|L_{m\ new}(j\omega)|$ is increased. This obviously extends the range for the pass-band specifications allowing for the solution. However, no generality is lost since the function is normalized with respect to the cutoff frequency $\omega_c = 1$ at the end of the iteration.

The mth order modified Chebyshev function obtained will have a much lower Q_d than the nth order Chebyshev function and with an improved delay characteristics (see Table III). In Table IV the poles of the MCF functions are listed. The pass-band specifications are met; however, a lower transition region attenuation is obtained. Since, m > n it follows that for some ω_{tm}, $|L_m(j\omega)| < |H_n(j\omega)|$ for all $\omega > \omega_{tm}$. In section 3.3.3, a method is given to increase the attenuation in the stop-band.

3.3.2 Least Squares Error Algorithm

The results obtained using the physical method algorithm are not unique. For comparison, the MCF's were derived using the least squares error criterion. The results obtained are similar but not identical. The least squares error expression is given by [TE 2]

$$E = \sum_i (R_i(|L_m^r(j\omega_i)| - D))^2 \quad , \quad 0 \leq \omega_i \leq \omega_c^r \tag{3-8}$$

$$D = \frac{(1 + \sqrt{1 + \varepsilon^2})}{2}$$

where D and $|L_m^r(j\omega_i)|$ are the desired response and the calculated response after r adjustments, R_i is a weighting factor taken here to be one, and ω_c^r is the rth adjustment cutoff frequency. $L_m^r(s)$ is a function of two sets of variables: 1) the parameters denoted by the vector $\overline{\beta} = (\beta_1,\beta_2,\ldots,\beta_{n/2})^T$ for even n, or $\overline{\beta} = (\beta_1,\beta_2,\ldots,\beta_{(n-1)/2},\omega_{bv})^T$ for

TABLE III

CRITICAL QUALITY FACTORS Q_d OF CHEBYSHEV (CHEB.) $H_n(s)$ AND MODIFIED CHEBYSHEV (MCF) $L_m(s)$ c = 2 FUNCTIONS, m = n + 2(c - 1)

n	3 dB		2 dB		1 dB		1/2 dB	
	Cheb. Q_c	MCF Q_c	Cheb. Q_c	MCF Q_c	Cheb. Q_c	MCF Q_c	Cheb. Q_c	MCF Q_c
2	1.304694	1.031882	0.992736	0.955873	0.956520	0.863402	0.863721	0.809005
3	3.067657	1.592415	2.551637	1.524761	2.017720	1.386328	1.706190	1.266850
4	5.578868	2.150385	4.593878	1.996962	3.559044	1.760761	2.940554	1.661111
5	8.818828	2.529244	7.232256	2.409009	5.556439	2.246412	4.544964	2.109320
6	12.780106	3.359719	10.461586	3.173590	8.003696	2.930414	6.512843	2.747629
7	17.464518	3.859683	14.284086	3.736328	10.898676	3.528096	8.841798	3.346720
8	22.870358	4.974274	18.687274	4.749682	14.240465	4.055670	11.530788	4.114909
9	28.998422	5.599986	23.682711	5.435100	18.028681	5.116562	14.579336	4.818551
10	35.845802	6.935707	29.266127	6.636376	22.263082	6.155976	17.987144	5.710818

TABLE IV

POLES OF MODIFIED CHEBYSHEV TRANSFER FUNCTION $L_m(s)$ FOR $c = 2,\ m = n + 2(c - 1)$

n	Double Pole Column				
			Pass Band Ripple 3 dB		
2	-0.470452 ±J0.849309				
3	-0.307628 ±J0.930194	-0.285550			
4	-0.230871 ±J0.965711	-0.287986 ±J0.361960			
5	-0.197961 ±J0.981621	-0.328308 ±J0.544151	-0.241331		
6	-0.149769 ±J0.995156	-0.213053 ±J0.712803	-0.164728 ±J0.258955		
7	-0.130177 ±J0.996418	-0.227185 ±J0.777018	-0.189912 ±J0.421118	-0.166579	
8	-0.100878 ±J0.998510	-0.176052 ±J0.834525	-0.130437 ±J0.561031	-0.119919 ±J0.193954	
9	-0.089428 ±J0.997588	-0.193814 ±J0.860497	-0.143916 ±J0.642139	-0.149903 ±J0.327453	-0.134301
10	-0.072212 ±J0.999082	-0.148846 ±J0.892125	-0.110397 ±J0.711320	-0.100634 ±J0.456241	-0.094107 ±J0.155214
			Pass Band Ripple 2 dB		
2	-0.549688 ±J0.895632				
3	-0.332505 ±J0.957913	-0.327741			
4	-0.255443 ±J0.987724	-0.347855 ±J0.356831			
5	-0.211822 ±J0.998336	-0.373497 ±J0.539030	-0.289599		
6	-0.160832 ±J1.008079	-0.249319 ±J0.716091	-0.197987 ±J0.261923		

TABLE IV (Continued)

n	Double Pole Column				
7	-0.136018 ±J1.007275	-0.248081 ±J0.782287	-0.208585 ±J0.426160	-0.190717	
8	-0.106568 ±J1.006705	-0.201389 ±J0.837492	-0.152970 ±J0.564379	-0.144823 ±J0.195161	
9	-0.092856 ±J1.005082	-0.208897 ±J0.864663	-0.156681 ±J0.646631	-0.165264 ±J0.329872	-0.153658
10	-0.075945 ±J1.005136	-0.166927 ±J0.895139	-0.126406 ±J0.714637	-0.120074 ±J0.458053	-0.113539 ±J0.156245
		Pass Band Ripple 1 dB			
2	-0.713612 ±J1.004610				
3	-0.393775 ±J1.018318	-0.423073			
4	-0.308687 ±J1.042297	-0.426159 ±J0.375837			
5	-0.236329 ±J1.035148	-0.412406 ±J0.549875	-0.354125		
6	-0.179409 ±J1.036068	-0.294947 ±J0.719068	-0.254285 ±J0.265342		
7	-0.147484 ±J1.030173	-0.270280 ±J0.789302	-0.241153 ±J0.432801	-0.230998	
8	-0.127981 ±J1.030177	-0.229144 ±J0.884035	-0.181407 ±J0.570710	-0.187608 ±J0.196787	
9	-0.100117 ±J1.019604	-0.224024 ±J0.870529	-0.180628 ±J0.651450	-0.192646 ±J0.333528	-0.186209
10	-0.082806 ±J1.016140	-0.190242 ±J0.899211	-0.152606 ±J0.718776	-0.153264 ±J0.459993	-0.147676 ±J0.157705
		Pass Band Ripple 1/2 dB			
2	-0.899508 ±J1.144166				
3	-0.471142 ±J1.096823	-0.539845			

TABLE IV (Continued)

n	Double Pole Column				
4	-0.345099 ±J1.093325	-0.515889 ±J0.362793			
5	-0.263978 ±J1.081888	-0.452096 ±J0.559706	-0.429882		
6	-0.198689 ±J1.073619	-0.327491 ±J0.725646	-0.311706 ±J0.267381		
7	-0.160418 ±J1.061697	-0.282444 ±J0.797624	-0.278494 ±J0.437015	-0.274012	
8	-0.128411 ±J1.048966	-0.242252 ±J0.846921	-0.228218 ±J0.568612	-0.231069 ±J0.196907	
9	-0.108664 ±J1.041552	-0.220919 ±J0.877144	-0.207498 ±J0.652693	-0.221179 ±J0.336349	-0.219234
10	-0.090911 ±J1.034364	-0.191322 ±J0.904101	-0.179017 ±J0.719715	-0.184855 ±J0.460324	-0.182019 ±J0.158358

odd n; 2) the response sample points ω_i, denoted $\overline{\omega} = (\omega_1,\omega_2,\ldots,\omega_N)^T$ with $0 \leq \omega_i \leq \omega_c^r$. D is taken as a straight line in the middle of the pass band.

The unknown parameters in Equations (3-4a) and (3-4b) corresponding to the even and odd n, respectively, are β_i, $i = 1,\ldots,n/2$, and ω_{bv} and β_i, $i = 1,\ldots,(n-1)/2$. The object is to solve for these parameters subject to the condition that the pass-band specifications are met; this involves the examination of the pass band maxima and minima. Numerically, it can be seen that it is sufficient to examine only the maxima $M_1 = \max|L_m(j\omega)|$ closest to the cutoff frequency. This is due to the fact that M_1 is most affected by the multiple dominant pole action. The Golden Section [WI 1] univariate search technique is used to evaluate M_1, and the multidimensional pattern search [WI 1] is used to evaluate $\overline{\beta}$ parameters subject to minimization of the error term.

Figure 4 shows the flow chart and in Appendix B the program listing is given; the main inputs are: 1) XLO(I) and XHI(I), I = 1,...,n/2 for n even, or I = 1,...,(n+1)/2 for n odd. These are the lower and upper bounds of $\overline{\beta}$; 2) the break point frequency $\omega_b(I)$, I = 1,...,n/2 for n even, or I = 1,...,(n-1)/2 for n odd; 3) detal is the increment used to augment the value of the largest break point frequency ω_{b1}. As shown in the flow chart, obtain the $\overline{\beta}$ parameters by using the pattern search; next from Gold 1 search get the value of M_1. If M_1 is within the pass-band specifications then normalize $L_m(s)$ to a cutoff frequency $\omega_c = 1$ and end the program. Otherwise, set $\omega_{b1} = \omega_{b1}$ + delta and repeat the iteration process.

The mth order modified Chebyshev function obtained is similar to that obtained using the physical method. Q_d is reduced appreciably and

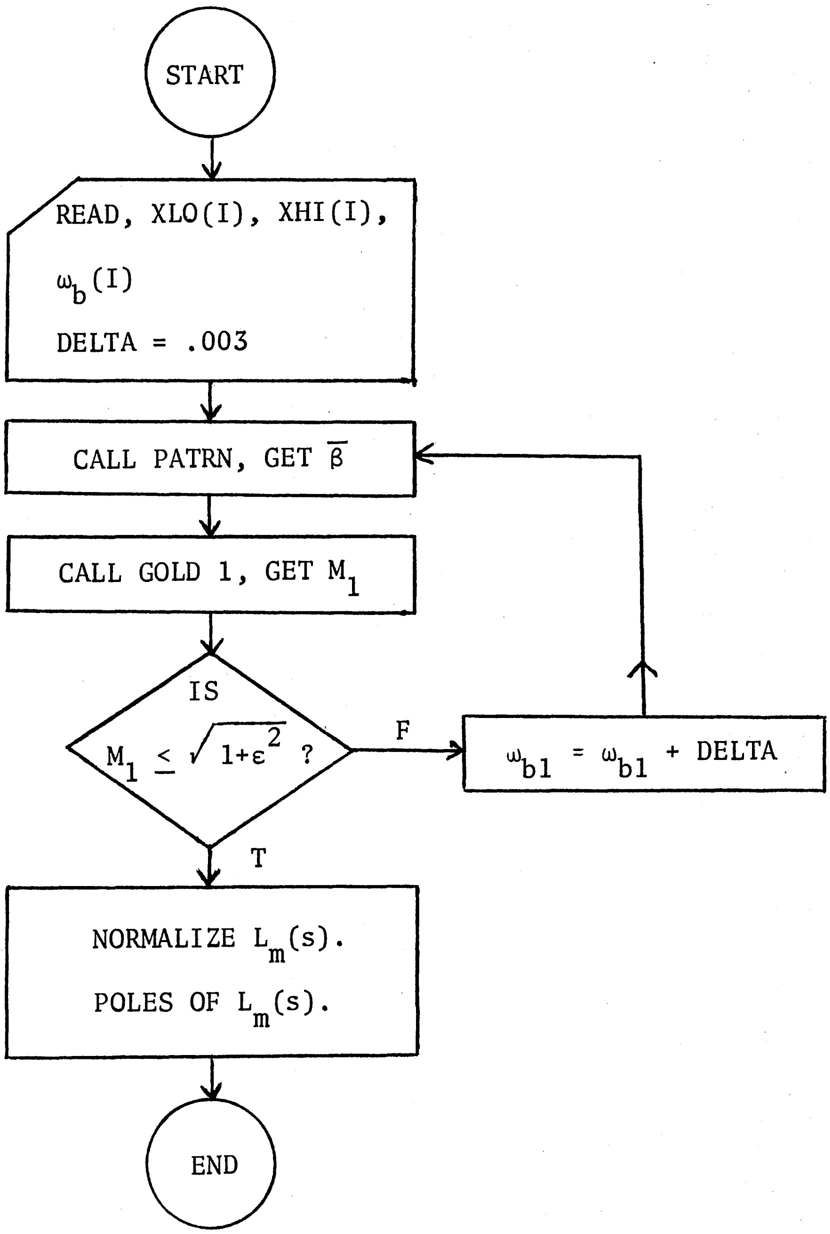

Figure 4. Flow Chart Using the Least Squares Error Method

the time delay is improved. Though the attenuation of the transition region is less steep, since $m > n$, we will have $|L_m(j\omega)| < |H_n(j\omega)|$ for all $\omega > \omega_{tm}$. A method is given below to satisfy the attenuation specifications in the stop band.

As pointed out earlier, the physical method is more advantageous as it is capable of adjusting $|L_m(j\omega)|$ at any frequency in the pass-band by adjusting each second order term. This also makes it an effective method for use in all pole filter functions other than the Chebyshev type where the break frequencies are spread out in the pass-band.

The rate of Q_d drop is largest for the case $c = 2$. Tables III and IV are given for $c = 2$, $2 \leq n \leq 10$, and a pass band ripple of 3, 2, 1, and 1/2 dB. In Table III the Q_d values of the modified Chebyshev and of the original Chebyshev functions are compared. In Table IV, the poles of the modified Chebyshev polynomials are listed.

3.3.3 Intermediate Modified Chebyshev Functions (IMCF's)

By relaxing the equal-ripple condition, a degree of freedom is obtained which makes such low Q_d values possible but at the cost of reduced transition band attenuation. In order to increase the attenuation in the transition region, the physical method or the least squares algorithm can generate IMCF functions which always satisfy the pass-band requirement and have a maximum Q_d and transition band attenuation approaching that of the MUCROER function, or have a minimum Q_d and transition band attenuation corresponding to the MCF. IMCF's are generated by further increasing ω_{b1} in the iterative method of Figures 2 or 4, beyond the value obtained by MCF. Each incremental increase in ω_{b1} gives rise to a new

IMCF function with larger Q_d value and higher transition region attenuation. For additional transition band attenuation the double poles may be slightly separated [MA 1], and/or imaginary axis zeros added [DU 1]. $|H_n(j\omega)|$ is equal ripple in the interval $0 \leq \omega \leq 1$ and has n half cycles; however, $|L_m(j\omega)|$ is not equal ripple and has n - 2 half cycles for $n \geq 5$ (see Figure 5a).

3.4 Examples

In the following, the tolerances for the Chebyshev, MUCROER, MCF and IMCF in terms of their coefficient sensitivities are computed. Let

$$T(s) = \prod_{i=1}^{k} \left[\frac{D_i}{(s^2 + A_i s + D_i)} \right] .$$

Then, the worst case tolerance is given by [HU 1],

$$\frac{\Delta T}{T} \simeq \sum_{i=1}^{k} \left(\left[\left| R_e s^T_{A_i} \frac{\Delta A_i}{A_i} \right| + \left| R_e s^T_{D_i} \frac{\Delta D_i}{D_i} \right| \right] + j \left[\left| I_m s^T_{A_i} \frac{\Delta A_i}{A_i} \right| + \left| I_m s^T_{D_i} \frac{\Delta D_i}{D_i} \right| \right] \right)$$

$$\simeq \frac{d|T|}{|T| + j \; d \arg T}$$

where the sensitivity $s^T_x = d\ell nT/d\ell nx$. For simplicity let $|d\,A_i/A_i| = |d\,D_i/D_i| = .05$. In the example below, $\Delta T/T$ is evaluated at the corner frequency $\omega = 1$. In addition to the tolerances, output noise variance (ONV) comparisons due to roundoff for cascaded second order canonical digital filter realization of functions is given. The bilinear transformation approach is used to find the digital functions. In computing the ONV, quantization step is taken as unity and scaling and section permutation for minimum ONV is performed [CA 1]. A more detailed explanation of ONV calculation and a comparison table will be given in

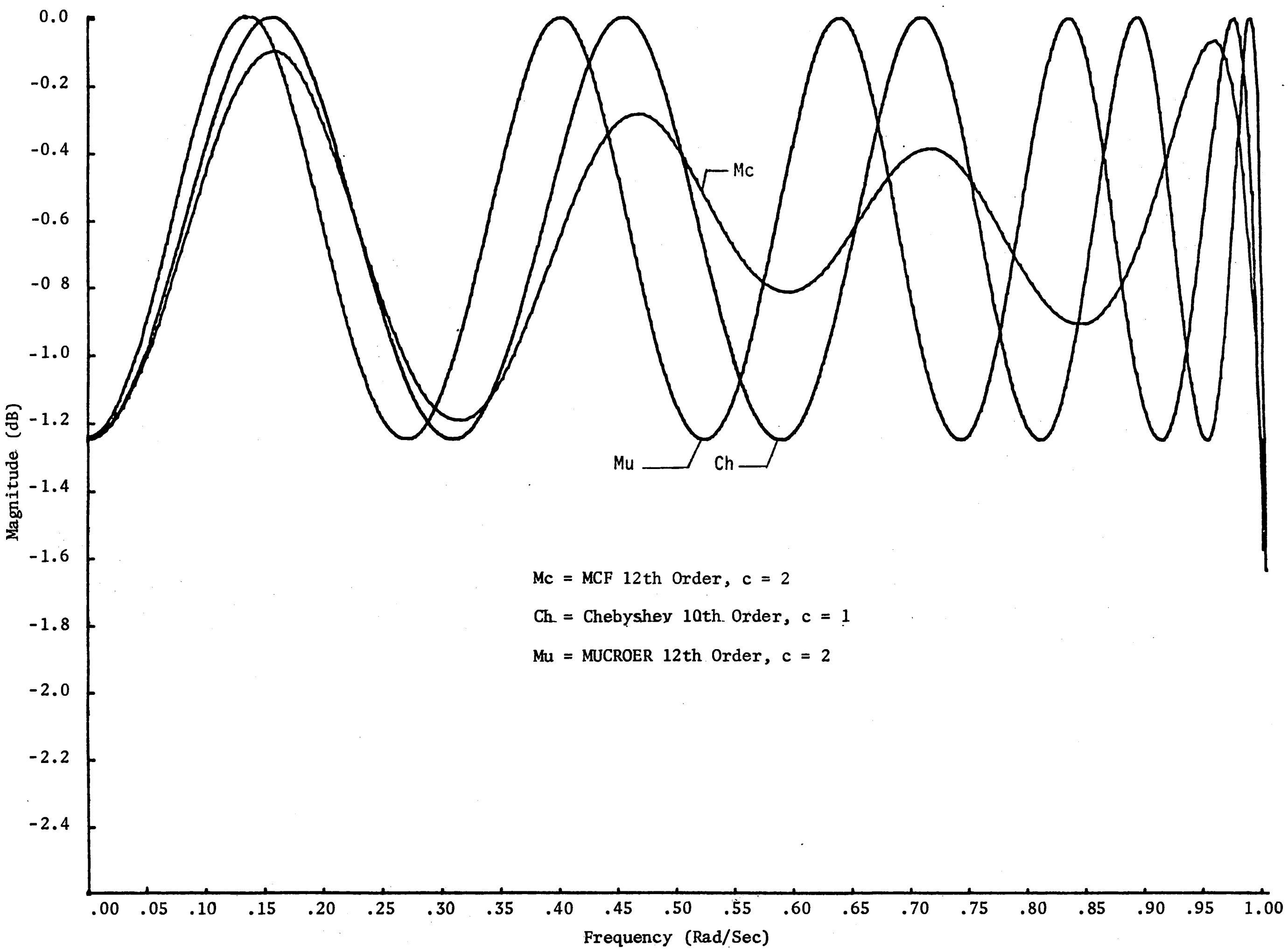

Figure 5a. Low-Pass Filter ρ = 50%, Pass-Band

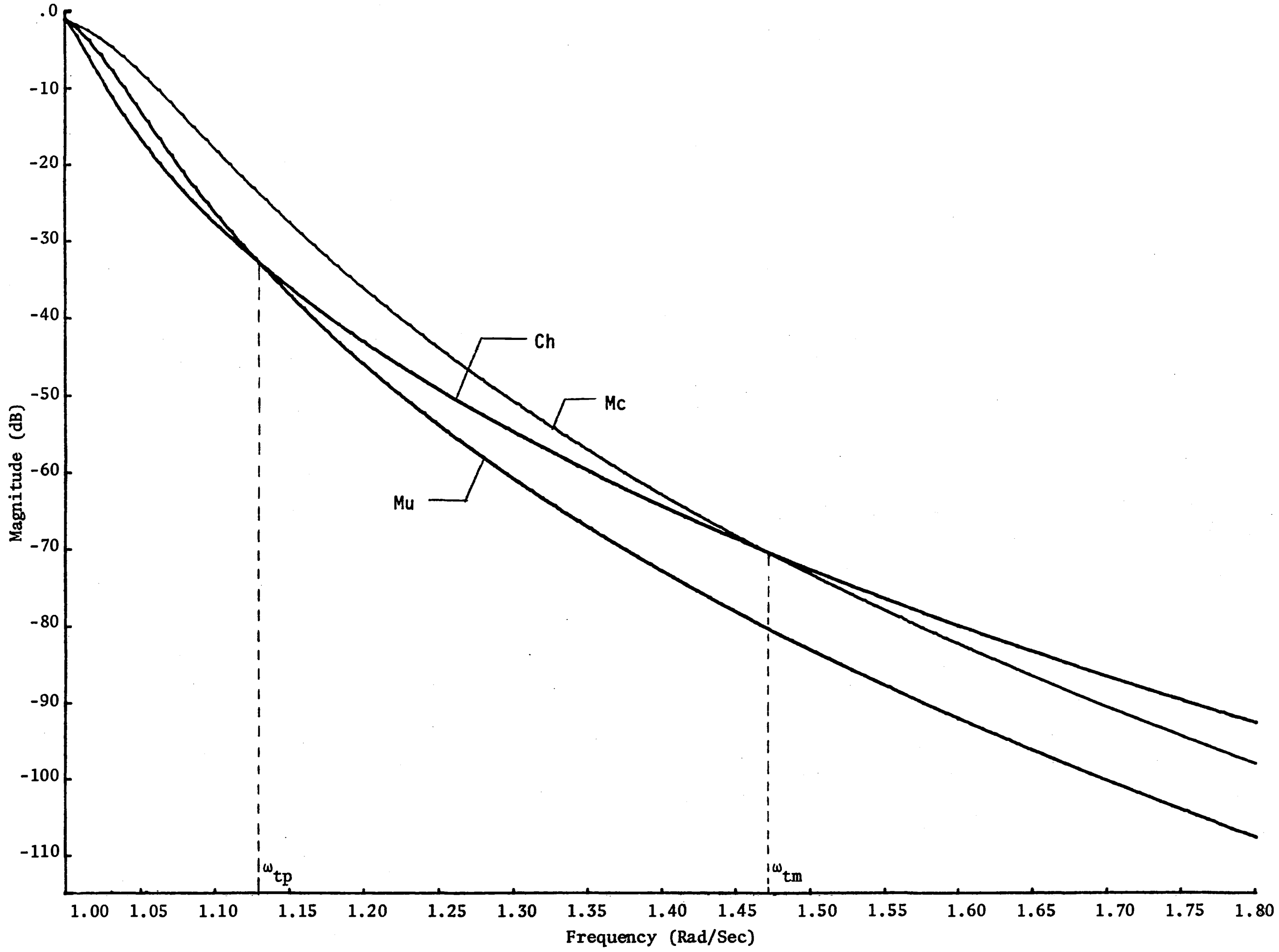

Figure 5b. Low-Pass Filter ρ = 50%, Stop-Band

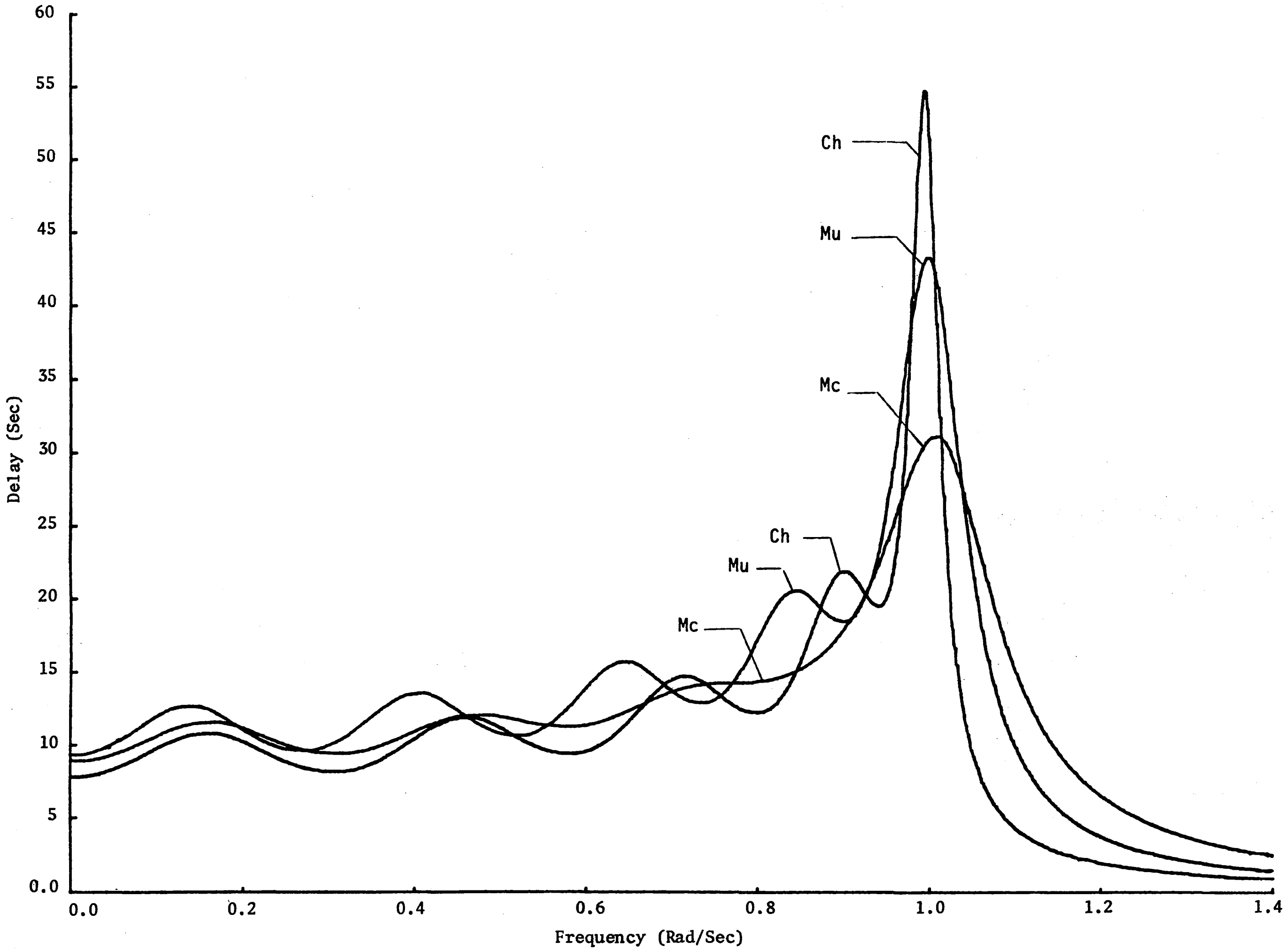

Figure 5c. Low-Pass Filter ρ = 50%, Time Delay

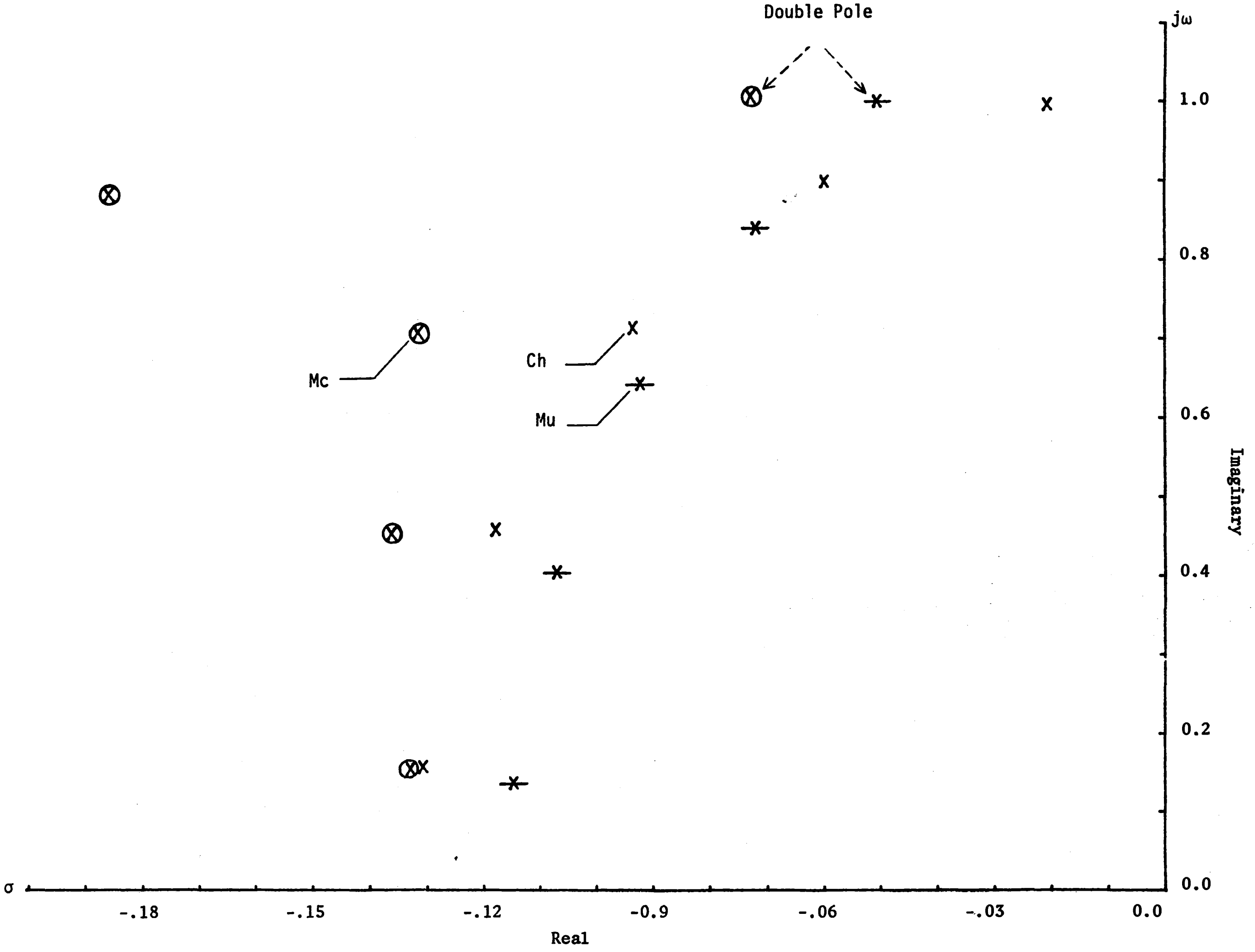

Figure 5d. Low-Pass Filter ρ = 50%, Pole Location

Chapter IV.

3.4.1 Example 1

Given a tenth order low-pass Chebyshev transfer function with a reflection coefficient $\rho = 50\%$ $(\rho^2 = \varepsilon^2/(1 + \varepsilon^2))$, $Q_d = 24.114576$, dominant pole location at $-0.020665 \pm J0.996267$, with $\Delta T/T \approx .7407 + J1.3619$, and ONV = 278.665, satisfying the magnitude specifications in the normalized frequency domain with $\omega_r = 1.5$, where ω_r is the lowest specified stop-band frequency. It is required to reduce Q_d using multiple dominant poles with multiplicity c = 2.

From the above specifications, we have $\overline{\delta} = (\delta_1, \delta_2, \ldots, \delta_5)^T = (.020738, .066571, .129831, .248896, .637172)^T$, and $\overline{\omega}_b = (\omega_{b1}, \omega_{b2}, \ldots, \omega_{b5})^T = (.996482, .900741, .719336, .472812, .204734)$. Using $\overline{\delta}$ and $\overline{\omega}_b$ in the physical method, the twelfth order modified Chebyshev function is obtained which has poles at: $-.132666 \pm J.155679$, $-.135916 \pm J.452385$, $-.131384 \pm J.706538$, $-.185692 \pm J.880515$, and dominant double poles at $-.072812 \pm J1.006276$ with $Q_d = 6.928174$, $\Delta T/T \approx .4742 + J.8842$, ONV = 195.161, and $\omega_{tm} = 1.448$ which is less than ω_r. Therefore the specifications are met and a reduction in Q_d, $\Delta T/T$, and ONV resulted (ω_{tm} is the frequency such that $|L_m(j\omega)| < |H_n(j\omega)|$ for all $\omega > \omega_{tm}$).

Comparing these results to the twelfth order $\rho = 50\%$ MUCROER function [PR 2] with c = 2 which gave $Q_d = 9.90919$, $\omega_{tp} = 1.132$, $\Delta T/T \approx .5294 + J1.115$, and ONV = 338.34 ($\omega_{tp}$ is the frequency such that |MUCROER Function| $< |H_n(j\omega)|$ for all $\omega > \omega_{tp}$), one can see that a 71.27% reduction in Q_d is obtained using the 12th order MCF function; whereas, the twelfth order MUCROER function gives a 58.89% reduction. An additional improvement in the time delay characteristic is also obtained but at the cost of reduced transition region attenuation. Figures 5a-5d show the pass-band ripple, stop-band attenuation, time delay, and pole location of the tenth order Chebyshev, twelfth order MUCROER, and twelfth order MCF with $\rho = 50\%$. However, if more transition region attenuation is desired, e.g., $\omega_r = 1.18 < \omega_{tm}$ then the specifications are not met; therefore ω_{b1} has to be increased as indicated above. This results in an IMCF function having poles at $-.124205 \pm J.143964$, $-.124136 \pm J.421191$, $-.124957 \pm J.65626$, $-.093207 \pm J.831314$, and dominant double poles at $-.055165 \pm J1.001483$ with $Q_d = 9.090909$, $\Delta T/T \approx .5102 + J1.053$, ONV = 262.944, and with $\omega_{tm\ new} = 1.178 < \omega_r$; therefore the specifications are met; note that Q_d, $\omega_{tm\ new}$, $\Delta T/T$, and ONV of IMCF approached those of the MUCROER function for increased attenuation in the transition region. Thus one can see the flexibility of the physical method (and the least squares method) in adjusting the filter function to meet steeper transition region attenuation; this results in a higher Q_d value. It should be noted that over 50% reduction in Q_d (see Table III) must be achieved by MCF or IMCF functions in order that $\Delta T/T$ and ONV are reduced; e.g., from Table III at 3 dB ripple MCF has more than 50% Q_d reduction over Chebyshev of 4th and higher orders.

3.4.2 Example 2

Given an eighth order low-pass Chebyshev transfer function with $\rho = 10\%$ (0.0436 dB pass-band ripple), $Q_d = 7.046669$, and dominant pole location at $-.074709 \pm J1.05024$. Using the physical method the corresponding tenth order MCF with dominant pole multiplicity $c = 2$ has poles at $-.411372 \pm J.212504$, $-.389079 \pm J.614068$, $-.358352 \pm J.918814$, and dominant double poles at -.184186 ± J1.163321 with Q_d = 3.197339 (54.63% Q_d reduction). It is to be noted that lower dB ripples give less percent reduction in Q_d as seen in Table III.

3.5 Conclusion

A modified non-equal-ripple Chebyshev function $L_m(s)$ with higher degree but much reduced dominant pole pair Q-factor Q_d than that of the corresponding Chebyshev function is presented. $L_m(s)$ is derived using a new numerical algorithm called the physical method. The pass-band specifications are satisfied; however, less transition region attenuation resulted. Intermediate modified Chebyshev functions with higher transition region attenuation and therefore larger Q_d are introduced. Improvement in the worst case sensitivity measure, and output noise variance of a digital filter realization required more than 50% Q_d reduction (see Table III). The physical method could also be effective in deriving modified filter functions other than the Chebyshev type where the break frequencies are spread out in the pass-band. Computer programs are given in Appendices A and B.

CHAPTER IV

ROUNDOFF NOISE: COMPARISON OF CHEBYSHEV AND MODIFIED CHEBYSHEV DIGITAL FILTERS

4.1 Introduction

The reduction of roundoff noise is of interest to many designers in the field of digital filter design. In this chapter the noise reduction capability of the new modified Chebyshev functions (MCF's) is demonstrated. In addition, a method for calculating the roundoff noise in terms of the driving point impedance is presented.

Several methods for the reduction of roundoff noise by optimum section ordering have been presented [JA 1, LE 1, GO 2, CH 1]. In addition, higher order functions were reported to give lower coefficient bit requirement, but in these cases larger roundoff noise resulted [CA 1, RA 1].

A digital filter with poles close to the unit circle in the z-plane (i.e., high Q_d filter) will have a high roundoff noise due to the rounding of products [GO 3]. This has prompted the development of the higher order MCF's with lower Q_d as a substitute to the Chebyshev functions. A reduction in the roundoff noise is achieved for the cases where MCF has over 50% Q_d reduction (see Table III). Coefficient bit comparison will be discussed in the next chapter where it will be shown that in many cases the MCF would require a lower number of bits.

Roundoff noise variance comparison of Butterworth and modified Butterworth functions are not presented here due to the fact that high degree functions have to be considered for a substantial reduction Q_d. However, substantial reduction in Q_d can be obtained for a lower order n using MCF's which are used in this chapter.

4.2 Roundoff Noise Calculation

In this section, a procedure for computing the output noise variance due to product rounding is given. First, the realization in terms of second order cascaded canonical sections is given. Second, the roundoff noise inputs are introduced in the realization. Third, the scaling and the optimal section ordering for the minimum output noise variance is discussed. In addition, a roundoff noise comparison table is given and a computer program listing for output roundoff noise calculation is included in Appendix C.

4.2.1 Realization

The output noise variance of a digital filter is a function of the realization used, circuit topology employed, the type of quantization used, and the location of product quantizations. For example, the realization in terms of cascaded second order sections has in general a lower output noise variance when compared to the output noise variance of a direct realization [KU 1]. In addition, the output noise variance will depend upon whether the products are quantized before or after summing.

In this section, the filter is realized in terms of first and second order canonical cascade sections as shown in Figure 6. This will not introduce any new problems even for functions such as MCF's which have multiple poles. Fixed point arithmetic and rounding of products prior to summing is used.

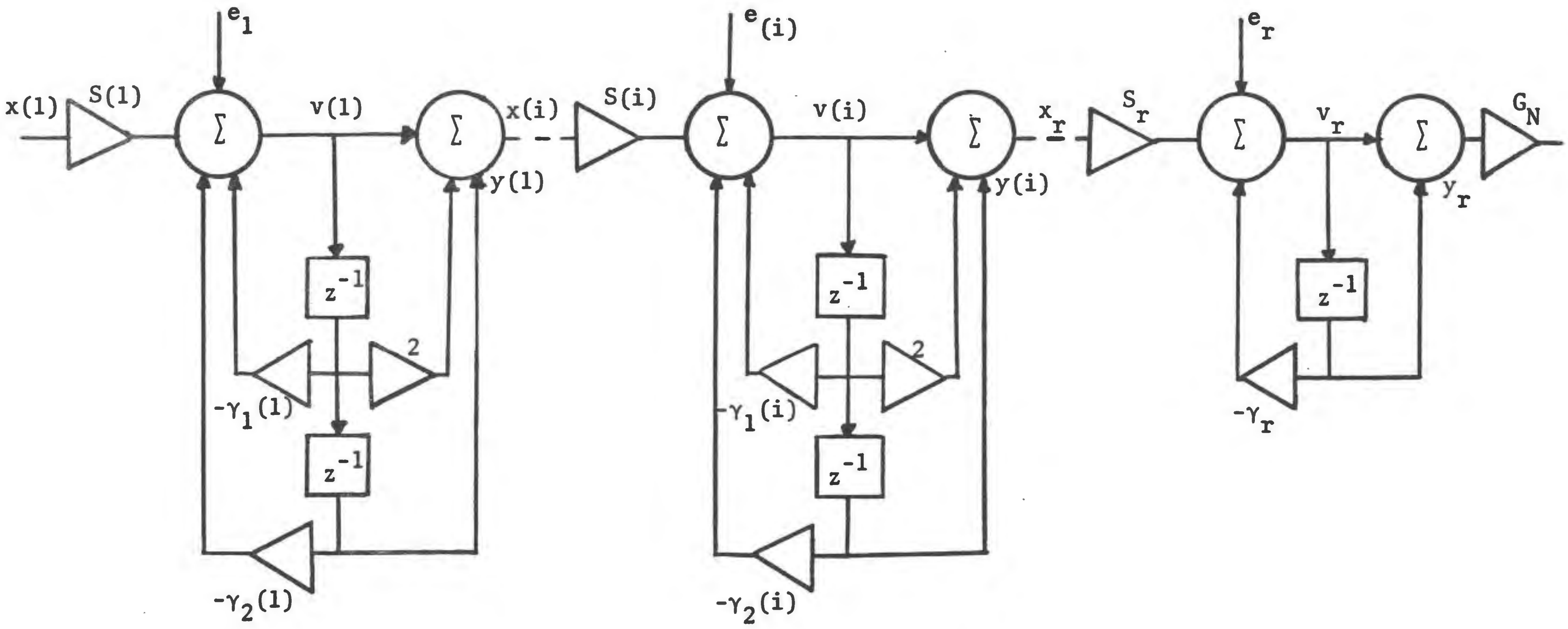

Figure 6. Cascade Filter Realization With Noise Inputs

The MCF's have been derived in the previous chapter and the discrete function can be obtained by making use of classical bilinear transformation (s → z-1/z+1). Let this function be written in the form

$$H(z) = \prod_{i=1}^{d} S(i) \frac{z^2 + 2z + 1}{z^2 + \gamma_1(i)z + \gamma_2(i)} \cdot S_r \frac{z + 1}{z + \gamma_r} \cdot G_N \qquad (4\text{-}1)$$

where S(i) and S_r are scaling factors used to prevent overflow. In section 4.2.3 the computation of these scaling factors is presented. The constants γ_1, γ_2, and γ_r are the multiplier coefficients, and G_N is introduced for normalizing the dc (z = 1) gain to unity. That is, G_N can be expressed, respectively, for even and odd functions by

$$G_N = \left[\begin{array}{ll} \displaystyle\prod_{i=1}^{d} \frac{1 + \gamma_1(i) + \gamma_2(i)}{4S(i)} & \text{for even functions} \\ & \\ \displaystyle\frac{1 + \gamma_r}{2S_r} \prod_{i=1}^{d} \frac{1 + \gamma_1(i) + \gamma_2(i)}{4S(i)} & \text{for odd functions} \quad . \end{array}\right. \qquad (4\text{-}2)$$

The explicit realization in terms of cascaded second order canonical sections along with the scaling factors is shown in Figure 6.

4.2.2 Noise Due to Product Rounding

One source of noise at the output of a digital filter is due to product rounding. The product of an m bit multiplicand and an n bit multiplier is an m + n bit product. Due to the finite register length of the hardware realization, or due to the finite filter word-length, the m + n bit word will be rounded to m bits. This quantization introduces an error e_i which can be represented as noise sources after each multiplier as shown in Figure 6 [GO 3]. Furthermore, the errors are assumed to be statistically independent and have a uniform probability density with zero mean (for rounding). If E_0 is the quantization step, the variance can be expressed by

$$\sigma^2 = \frac{E_0^2}{12} \quad . \tag{4-3}$$

Noting the statistical independence of the error sources, the total output noise of the filter can be expressed by

$$\sigma_0^2 = \frac{E_0^2}{12} \left[\sum_{i=1}^{d} \frac{M_i}{2\pi j} \oint G_i(z) G_i(\tfrac{1}{z}) \frac{1}{z} dz \right] \tag{4-4}$$

where d = number of sections, M_i = number of input error sources to the ith section, and $G_i(z)$ = transfer function between the input to the ith section and the filter output. Here, E_0 is taken as unity and the integration path is taken around the unit circle [GO 3]. The subroutine SALOSS given by Åström, et al. [AS 1] is used to evaluate Equation (4-4). A listing of SALOSS is given in Appendix C.

4.2.3 Scaling and Section Permutation

As pointed out earlier, scaling factors must be introduced at the input to every section in order to avoid overflow. Next, the computation of these scaling factors is discussed. Referring to Figure 6, the following transfer functions of interest expressed in the Z transform, are given below.

From the filter input to the ith section output

$$F_i(z) = \sum_{k=0}^{\infty} f_i(k) z^{-k} \quad .$$

From the filter input to the ith branch node

$$T_i(z) = \sum_{k=0}^{\infty} t_i(k)z^{-k} \quad .$$

These transfer functions can be expressed in terms of

$$X(z) = \sum_{k=0}^{\infty} x(k)z^{-n}$$

by

$$Y_i(z) = F_i(z)X(z)$$

$$V_i(z) = T_i(z)X(z) \quad .$$

For a filter input $|x(k)| \leq 1$ for all k, Jackson shows that

$$|y_i(k)| \leq \sum_{k=0}^{\infty} |f_i(k)|$$

and

$$|v_i(k)| \leq \sum_{k=0}^{\infty} |t_i(k)| \quad . \tag{4-5}$$

If the scaling factors are selected such that

$$\sum_{k=0}^{N} |f_i(k)| \leq 1$$

and

$$\sum_{k=0}^{N} |t_i(k)| \leq 1 \tag{4-6}$$

where N is chosen to be large with respect to the time constants of the filter, then $|y_i(k)| \leq 1$ and $|v_i(k)| \leq 1$ for all k [JA 2].

Cardwell's approach for computing the scaling factors will be used here. This approach insures that $|y_i(k)| \leq 1$ [CA 1]. In addition the requirement that $|v_i(k)| \leq 1$ will be taken into consideration. The first scaling factor S(1) is chosen such that

$$|y_1(k)| \leq 1$$

and

$$|v_1(k)| \leq 1 \qquad k = 0,1,2,\cdots \qquad . \quad (4\text{-}7)$$

This is satisfied provided that

$$\sum_{k=0}^{N} |f_1(k)| \leq 1$$

and

$$\sum_{k=0}^{N} |t_1(k)| \leq 1 \qquad . \quad (4\text{-}8)$$

The expressions $\sum_{k=0}^{N} |f_1(k)|$ and $\sum_{k=0}^{N} |t_1(k)|$ can be evaluated by solving the difference equations for the first and second order sections (see Figure 6). These equations are given by

$$v_1(k) = x(k) - \gamma_1(1)v_1(k - 1) - \gamma_2(1)v_1(k - 2) \qquad (4\text{-}9a)$$

$$y_1(k) = v_1(k) + 2v_1(k - 1) + v_1(k - 2) \qquad (4\text{-}9b)$$

where $v_1(-1) = v_1(-2) = 0$, $S(1) = 1$, and $x(k)$ is a unit impulse input applied to the digital filter where

$$x(k) = \begin{cases} 1 & k = 0 \\ 0 & \text{otherwise} \end{cases} \quad .$$

It is clear that $\sum_{k=0}^{N} |f_1(k)| = \sum_{k=0}^{N} |y_1(k)|$ and $\sum_{k=0}^{N} |t_1(k)| = \sum_{k=0}^{N} |v_1(k)|$ which can be evaluated from Equation (4-9). To insure the condition given by Equation (4-7), the scale factor S(1) is evaluated by

$$S(1) = \min \left[\frac{1}{\sum_{k=0}^{N} |f_1(k)|} , \frac{1}{\sum_{k=0}^{N} |t_1(k)|} \right] . \quad (4-10)$$

Similarly each S(i) can be evaluated by using the output of the (i - 1)th section multiplied by S(i - 1) as input to the ith section.

The subroutine SCALE that evaluates S(i) is given by Cardwell [CA 12]. A listing of a modified version of SCALE that takes $\sum_{k=0}^{N} |t_i(k)|$ into consideration is included in Appendix C.

Earlier, it was pointed out that the noise variance is a function of the realization used and the circuit topology. It is important to find an optimal ordering of first and second order sections for minimal output noise variance. The comparison of output noise variance of each filter is made on the basis of minimum noise output per filter configuration. This is achieved by permuting all the first and second order sections for a minimal output noise variance.

4.2.4 Noise Comparison

In Table V the output roundoff noise of an nth order low-pass Chebyshev function and low-pass MCF's of order (n + 2), are compared for 1/2 dB and 3 dB pass band ripples. Note that the MCF functions or order (n + 2) will give a lower noise variance than the corresponding nth order Chebyshev functions for cases where the MCF's have over 50% Q_d reduction (see Table III). Furthermore, for low dB ripples, the MCF's will give substantial lower noise variance when compared to Chebyshev functions, for higher order n. For example, from Table V, for 3 dB ripple, n is

TABLE V

ROUNDOFF NOISE σ_0^2: COMPARISON OF CHEBYSHEV $H_n(z)$ AND MCF $L_m(z)$ FOR $c = 2$ AND $m = n + 2$

n	3 dB		1/2 dB	
	Cheb. σ_0^2	MCF σ_0^2	Cheb. σ_0^2	MCF σ_0^2
2	4.351	6.487	2.268	4.208
3	9.564	10.710	4.980	7.930
4	29.259	19.957	15.307	15.137
5	43.781	21.435	23.959	18.457
6	104.267	54.861	49.245	40.853
7	110.273	50.553	61.835	43.594
8	222.835	118.417	118.902	77.895
9	205.532	94.582	121.430	83.605
10	387.803	219.876	195.102	144.645

four and for a 1/2 dB ripple n is five. Note also that higher order odd functions tend to have low output noise variance; this is due to the noise attenuation of the first order section.

4.3 On Calculating Roundoff Noise From the Driving Point Impedance

A new approach for the computation of the steady state output quantization noise variance of a digital filter is presented here. This method makes use of the s-domain transfer functions and their relation to the driving point impedance in the classical filter design [VA 1]. On the other hand, in the traditional method, the transfer function is transformed into the z-domain prior to noise calculation. In addition, the method presented here will not require the prior knowledge of the pole locations. Furthermore, the method can be applied to functions with multiple complex conjugate poles. In the following, a summary of the previous techniques of computing the output noise variance is given.

The steady state value of the output noise variance is given by

$$\sigma_n^2 = \frac{E_0^2}{12} \frac{1}{2\pi j} \oint H(z)H(\frac{1}{z}) \frac{1}{z} dz \qquad (4.11)$$

where H(z) is the transfer function from the noise sample input to the filter output [GO 3]. Evaluation of Equation (4-11) is important in communications and control problems. A tabulated solution of (4-11) for low order H(z) can be found in Jury [JU 1]; whereas, for high order H(z), evaluation of Equation (4-11) is difficult and the following methods have been used.

1) Using Cauchy's residue theorem and partial fractions expansion of H(z)H(1/z)1/z,

2) Using the form

$$\sigma_n^2 = \frac{E_0^2}{12} \frac{1}{\omega_s} \int_0^{\omega_s} |H(\omega_D)|^2 \, d\omega_D \tag{4-12}$$

where ω_s = sampling frequency and ω_D represents the digital frequency [KN 1],

3) Using the time series representation

$$\sigma_{(K)}^2 = \frac{E_0^2}{12} [\sum_{k=0}^{K} h^2(kT)]$$

where $h(kT) = Z^{-1}[H(z)]$, found in [GE 1],

4) Using the numerical formula of Åström, et al. [AS 1],

5) Using partial fraction expansion of H(z) has also been suggested [MI 1], and

6) Using the inners approach [JU 2].

4.3.1 Proposed Method

The proposed method is based upon the use of the bilinear transformation

$$z = \frac{1 + s}{1 - s} \tag{4-13}$$

in Equation (4-12) which permits analysis in the s-domain. Replacing z by $e^{j\omega_D T}$ and s by $j\omega$ in Equation (4-13), the analog frequency variable ω and the digital frequency variable $\omega_D T$ are related by

$$\omega = \tan \frac{\omega_D T}{2} \qquad .(4\text{-}14)$$

Using this expression in (4-12) with

$$d\omega_D = \frac{2}{T}\frac{1}{1+\omega^2}\,d\omega \qquad ,(4\text{-}15)$$

Equation (4-12) can be rewritten as [KI 1]

$$\sigma_n^2 = \frac{E_0^2}{6\pi}\int_0^\infty |H(\omega)|^2 \frac{1}{1+\omega^2}\,d\omega \qquad (4\text{-}16)$$

$$= \frac{E_0^2}{12}\frac{1}{\pi}\int_{-\infty}^{\infty} |F(\omega)|^2\,d\omega \qquad (4\text{-}17)$$

where the relation F(s) = H(s)/(s + 1) has been used. Greaves, et al. [GR 1] developed a method for obtaining the s-domain coefficients from the z-domain transfer function and then evaluating σ_n^2 from the tables given by Newton, et al. [NE 1]. In this section the procedure will be carried out one step further. Papoulis has pointed out the relationship for the energy E of a signal given by

$$E = \frac{1}{2\pi}\int_{-\infty}^{\infty} |F(\omega)|^2\,d\omega = \frac{1}{2}\lim_{s\to\infty} s\,Z(s) \qquad .(4\text{-}18)$$

where Z(s) is the driving point impedance [PA 1]. By using (4-18) in (4-17), it follows that

$$\sigma_n^2 = \frac{E_0^2}{12}\lim_{s\to\infty} s\,Z(s) \qquad .(4\text{-}19)$$

Equation (4-19) gives the output noise variance of a digital filter due to A/D noise or due to product quantization; it does not require root calculation since Z(s) could be obtained using Gewertz or Mitra's method [KA 3, MI 3] and it can be applied to filter functions with multiple conjugate poles. Equation (4-19) can be used directly on the s-domain filter function H(s) without transforming H(s) into the z-domain. This is true if the bilinear transformation is used to obtain H(z). However, if a method other than the bilinear transformation is employed to obtain

H(z), or if scaling is used, then H(z) must be obtained first; next $|F(\omega)|^2$ must be evaluated by applying Equation (4-13) to H(z), and finally σ_n^2 can be calculated from (4-19).

4.3.2 Example

Given the second order filter

$$H(s) = \frac{\omega_n^2}{s^2 + 2as + \omega_n^2} \tag{4-20}$$

which is to be realized digitally by using the bilinear transformation. It is required to find the output noise variance σ_n^2 of the digital filter due to A/D quantization.

From Equation (4-17),

$$F(s) = \frac{\omega_n^2}{(s^2 + 2as + \omega_n^2)(s + 1)} \tag{4-21}$$

where upon using Gewertz's method [KA 3],

$$Z(s) = \frac{a_2 s^2 + a_1 s + a_0}{s^3 + b_2 s^2 + b_1 s + b_0} . \tag{4-22}$$

By using (4-19), σ_n^2 can be obtained as

$$\begin{aligned} \sigma_n^2 &= \frac{E_0^2}{12} a_2 \\ &= \frac{E_0^2}{12} \frac{\omega_n^2(2a + 1)}{-\omega_n^2 + (2a + 1)(2a + \omega_n^2)} \end{aligned} \tag{4-23}$$

where a_2 is determined by using Gewertz's method. In [NE 1] a method for evaluating a_2 in matrix form is given. The output noise variance given in Equation (4-23) is in agreement with the answer obtained by using any of the previously mentioned methods. For example, by transforming H(s) into the z-domain by using the bilinear transformation and then applying Cauchy's residue theorem, results in an expression for σ_n^2 identical to that given in Equation (4-23).

4.4 Conclusion

In this chapter the following has been presented. First, the noise reduction capability of the modified Chebyshev functions has been demonstrated in a table comparing the output roundoff noise variance of a digital filter due to product rounding for the nth order Chebyshev and the n + 2 order modified Chebyshev functions including the 3 dB and 1/2 dB cases. The method used for noise calculation and scaling has also been discussed. Second, a method for calculating the roundoff noise in terms of the driving point impedance has been derived. This enables the designer, in the cases where the bilinear transformation has been used, to calculate the A/D quantization noise and the product quantization noise (if no scaling used) directly from the s-domain filter function H(s) without following the traditional method of transforming H(s) to the z-domain.

CHAPTER V

COEFFICIENT WORD-LENGTH: ESTIMATION AND COMPARISON OF CHEBYSHEV AND MODIFIED CHEBYSHEV DIGITAL FILTERS

5.1 Introduction

In this chapter a procedure is given for the estimation of multiplier word-length (coefficient bits) for the first and second order digital filter sections given the transfer function's s-domain poles and their tolerance specifications. This method is used to compare the multiplier bit requirement for digital filter realizations using Chebyshev and modified Chebyshev (MCF) functions.

Due to the finite arithmetic precision, the multiplier values have to be rounded to the nearest quantization step. This change in the multiplier accuracy will result in a corresponding change in the pole location. It is therefore required to obtain the minimum number of multiplier bits such that the corresponding pole shift satisfies the given tolerance limits.

In the literature, coefficient sensitivity and statistical approaches have been proposed for estimating the multiplier word-lengths [MI 2, CR 1]. In addition, for the cases where the impulse invariance transformation is used to obtain the discrete function, a method for coefficient bit estimation as a function of the s-domain poles and their tolerances has been suggested [WH 1]. For the cases where the bilinear transformation is used to obtain the discrete function, the method presented here gives a simple procedure for coefficient bit estimation as a function of the s-domain poles and their tolerances; this method will be used for the bit comparison. In addition, the coefficient word-length comparison of nth order low-pass Chebyshev functions and low-pass double dominant pole MCF's of order (n + 2) for the 1/2 dB and 3 dB ripple cases are tabulated. The word-length requirement is estimated for the dominant pole second order section such that a specified tolerance on both the break frequency and the magnitude of the dominant pole section evaluated at the break frequency is satisfied. From the table, it can be observed that in many cases the MCF will require a lower number of bits.

5.2 Coefficient Word-Length Estimation

In the following a procedure for estimating the word-length for first and second order sections is presented.

5.2.1 First Order Case

The first order transfer function considered here is given by

$$H(s) = \frac{p}{s + p} \quad . \tag{5-1}$$

By applying the bilinear transformation s→(z - 1)/(z + 1) to Equation (5-1), the following discrete function results

$$H(z) = \frac{R(z)}{z + \gamma_r}$$

where $\gamma_r = (p - 1)/(p + 1)$ is the multiplier for which the word-length requirements need to be estimated given that the pole located at $s = -p$ has a tolerance of Δp. Corresponding to this change in pole location, let $\gamma_r' = \gamma_r + \Delta\gamma_r$ be the new pole location in the z-domain, where $\Delta\gamma_r$ is half of the maximum quantization step size allowable for rounding. Expressing γ_r' in terms of Taylor's series around the nominal value of p and keeping only the first term, $\Delta\gamma_r$ can be expressed by

$$\Delta\gamma_r \simeq \frac{2\Delta p}{(p + 1)^2} \quad . \quad (5\text{-}4)$$

Assuming that rounding is used, the estimate for the number of bits Q_r required to keep the pole within the tolerance limits is obtained from $2^{-(Q+1)} = |\Delta\gamma_r|$. This results in

$$Q_r \geq - 1 - \log_2(\frac{2\Delta p}{(p + 1)^2}) \quad . \quad (5\text{-}5)$$

5.2.2 Example

Determine the coefficient bit requirement for a first order Butterworth digital filter which has a cutoff frequency of 1 rad/sec, pole tolerance $\Delta p/p$ = 10 per cent, and a sampling rate of 1 K.Hz.

In order to obtain the transfer function in the s-domain, prewarping must first be performed. Using Equation (4-14),

$$\omega_A = \tan \frac{\omega_D T}{2}$$

with $\omega_D = 1$ and $T = 1/1000$ the analog cutoff frequency is given by $\omega_A = .0005$. The first order filter transfer function is expressed as

$$H(s) = \frac{1}{s + 1} \quad . \quad (5\text{-}6)$$

Denormalizing (5-6) with respect to ω_A yields

$$H(s) = \frac{.0005}{s + .0005} \quad . \quad (5\text{-}7)$$

From Equation (5-5) Q_r is found to be 12 bits. This agrees with the result obtained when the impulse invariance transform was used for the filter design [WH 1].

5.2.3 Second Order Case

Let the second order transfer function in the s-domain be given by

$$H(s) = \frac{a^2 + b^2}{s^2 + 2as + a^2 + b^2} \quad . \quad (5\text{-}8)$$

The discrete function is obtained by applying the bilinear transform to Equation (5-8). This yields

$$H(z) = \frac{R_1(z)}{z^2 + \gamma_1 z + \gamma_2} \quad (5\text{-}9)$$

where

$$\gamma_1 = \frac{2(a^2 + b^2 - 1)}{a^2 + b^2 + 2a + 1} \quad (5\text{-}10a)$$

and

$$\gamma_2 = \frac{a^2 + b^2 - 2a + 1}{a^2 + b^2 + 2a + 1} \quad (5\text{-}10b)$$

are the multipliers for which the word-lengths need to be estimated given that the pole located at $s = -a \pm jb$ has a maximum allowable tolerance on a and b of Δa and Δb, respectively. Corresponding to this change in pole location, let $\gamma_i' = \gamma_i + \Delta\gamma_i$, $i = 1,2$ be the new multiplier values in the z-domain, where $\Delta\gamma_i$ is half of the maximum quantization step size allowable for rounding. Expressing γ_i' in terms of Taylor's series around the nominal pole location and keeping only the first derivative terms, $\Delta\gamma_i$

can be expressed as

$$\Delta\gamma_i = (\frac{\partial\gamma_i}{\partial a})\Delta a + (\frac{\partial\gamma_i}{\partial b})\Delta b \qquad (5\text{-}11)$$

which gives

$$\Delta\gamma_1 \approx \frac{4(a^2 + 2a - b^2 + 1)\Delta a + 8b(a + 1)\Delta b}{(a^2 + b^2 + 2a + 1)^2} \qquad (5\text{-}12a)$$

and

$$\Delta\gamma_2 \approx \frac{4(a^2 - b^2 - 1)\Delta a + 8ab\Delta b}{(a^2 + b^2 + 2a + 1)^2} \qquad .(5\text{-}12b)$$

If rounding is considered, then the estimate of the number of bits Q_i required to keep the poles within the tolerance limits is obtained by equating $2^{-(Q_i+1)} = |\Delta\gamma_i|$, i = 1,2. This gives

$$Q_1 \geq - 1 - \log_2|\Delta\gamma_1| \qquad (5\text{-}13a)$$

$$Q_2 \geq - 1 - \log_2|\Delta\gamma_2| \qquad .(5\text{-}13b)$$

Unlike the single coefficient first order case, each s-domain pole location of the second order section is determined by the value of the coefficients γ_1 and γ_2. In considering second order sections, Equation (5-13) gives an estimate of the coefficient bits Q_1 and Q_2, and therefore defines the maximum quantization step. Due to the independent rounding of γ_1 and γ_2 to a value within the maximum allowable quantization step, few cases might arise where the specified pole tolerance limits are slightly exceeded. For these cases the estimated bits Q_1 and Q_2 need to be further increased. Usually one bit more than the computed value would be adequate. An example for computing Q_1 and Q_2 is included in the next section illustrating some of the above ideas.

The computation of the multiplier bits requirement for setting the zeros within specified tolerance limits is similar to the above discussion and therefore omitted.

5.3 Coefficient Word-Length Comparison

In this section the coefficient word-length requirements are compared for the Chebyshev and the MCF functions. In this comparison, bit requirements for the dominant pole second order section are considered as it requires a large number of coefficient bits.

The dominant pole section for a Chebyshev function was given in Equation (3-3) and is

$$h_1(s) = \frac{\omega_{n1}^2}{s^2 + 2\delta_1\omega_{n1}s + \omega_{n1}^2} \qquad (5\text{-}14)$$

where ω_{n1} and $(\delta_1\omega_{n1})$ can be expressed in terms of the pole location as can be seen from Equation (5-8). These are given by

$$\omega_{n1}^2 = a^2 + b^2 \qquad (5\text{-}15a)$$

$$\delta_1\omega_{n1} = a \qquad .(5\text{-}15b)$$

Let the tolerance limits be given on the break frequency ω_{n1} and on $|h_1(j\omega_{n1})| \equiv C_h$. It is required to calculate the number of multiplier bits for the corresponding second order digital filter section such that the specified tolerance limits are met. The bit requirements are given in terms of a, b, Δa, and Δb in Equations (5-13a) and (5-13b). Therefore, the tolerance limits on ω_{n1} and C_h must be related to the tolerance limits on the pole locations. This aspect is discussed in the following.

First, C_h can be expressed in terms of a and b and is given by

$$C_h = |h_1(j\omega_{n1})| = \frac{1}{2\delta} = \frac{\omega_{n1}}{a} \quad . \quad (5\text{-}16)$$

Solving for a and b from Equations (5-15a) and (5-16), it follows that

$$a = \frac{\omega_{n1}}{2C_h} \quad (5\text{-}17a)$$

$$b = \sqrt{\omega_{n1}^2 - a^2} \quad . \quad (5\text{-}17b)$$

Using the incremental variations and keeping only the first order terms, Δa and Δb can be expressed as

$$\Delta a \simeq (\frac{\partial a}{\partial C_h})\Delta C_h + (\frac{\partial a}{\partial \omega_{n1}})\Delta\omega_{n1}$$

$$\Delta b \simeq (\frac{\partial b}{\partial a})\Delta a + (\frac{\partial b}{\partial \omega_{n1}})\Delta\omega_{n1} \quad .$$

Using Equations (5-17a) and (5-17b) in the above expressions, the following results

$$\Delta a \simeq \frac{(\Delta\omega_{n1} - \omega_{n1}\frac{\Delta C_h}{C_h})}{2a} \quad (5\text{-}18a)$$

$$\Delta b \simeq \frac{(\omega_{n1}\Delta\omega_{n1} - a\Delta a)}{b} \quad (5\text{-}18b)$$

which relates the pole tolerance limits to the tolerance limits on ω_{n1} and C_h. Equation (5-13) can then be used to give an estimate of the coefficient bit requirement such that the tolerance limits on ω_{n1} and C are satisfied.

In the following a step by step procedure for calculating Q_1 and Q_2 given the tolerance limits on ω_{n1} and C_h is outlined.

1) Obtain the pole tolerance limits Δa and Δb from the specified $\Delta\omega_{n1}$ and ΔC_h using Equations (5-18a) and (5-18b).

2) Evaluate the exact non-rounded values of the multipliers γ_1 and γ_2 using Equation (5-10).

3) Use Equation (5-13) to give an estimate of the required bits Q_1 and Q_2.

4) Calculate γ_{1q} and γ_{2q} the rounded values of γ_1 and γ_2 corresponding to Q_1 and Q_2 bits, respectively.

5) Calculate ω_{n1q} and C_{hq} the new values of ω_{n1} and C_h corresponding to γ_{1q} and γ_{2q}.

6) Check if ω_{n1q} and C_{hq} satisfy the specified tolerance limits; if the tolerance limits are satisfied, an attempt must be made to minimize Q_1 and Q_2. Hence, reduce Q_1 and Q_2 by one bit, respectively, and repeat steps 3 to 5. If the tolerance limits are not satisfied, perform coefficient rounding to the higher or lower quantization step (coefficient optimization [RA 1, AV 1]) and repeat steps 4 to 5. If the tolerance limits are not met after coefficient optimization, increase the estimated Q_1 and Q_2 by one bit, respectively, and repeat steps 4 to 5. Terminate the procedure when a minimum value of Q_1 and Q_2 is found such that the tolerance limits are satisfied.

From the various examples attempted, it can be stated that in the majority of cases, Equations (5-13a) and (5-13b) give directly the minimum bit requirement such that the given tolerance limits are satisfied.

Next, the coefficient bit requirements for double dominant poles are discussed. Earlier, C_h was defined to be the magnitude of the dominant pole section at $\omega = \omega_{n1}$. In Equation (3-4), the term corresponding to double dominant poles is given by

$$\ell_1^2(s) = \left[\frac{\omega_{b1}^2}{s^2 + 2\beta_1\omega_{b1}s + \omega_{b1}^2} \right]^2 \quad .$$

Let the magnitude of this function at $\omega = \omega_{b1}$ be identified by

$$C_t = |\ell_1(j\omega_{b1})|^2 \quad .$$

To relate this to the earlier work, let C_t be equal to C_ℓ^2 where $C_\ell = |\ell_1(j\omega_{b1})|$. Using the incremental variations, the tolerance limit ΔC_ℓ on C_ℓ can be expressed in terms of the given tolerance limit ΔC_t on C_t and is

$$\Delta C_\ell = C_\ell \left[\sqrt{1 + (\Delta C_t/C_t)} - 1 \right] \quad .$$

Hence, due to the presence of a double pole, a given tolerance on C_t will result in a lower tolerance on C_ℓ which is to be used in Equation (5-18) in order to evaluate Δa and Δb, and finally evaluate Q_1 and Q_2.

In the coefficient bit comparison study, the following binary coefficient representation is used. Since the filter coefficients γ_i lie in the range $-2 < \gamma_i < 2$, by assuming fixed point arithmetic and letting the most significant bit represent $2^0 = 1$, the coefficients γ_i are expressed in the form

$$\gamma_i = \pm \sum_{k=0}^{Q_i} d_k 2^{-k}$$

where

$$d_k = 0 \text{ or } 1 \qquad \text{for} \qquad \text{each } k \quad .$$

Following the earlier step by step procedure for minimum coefficient bit calculation, the coefficient bit comparison of the dominant pole section for the nth order low-pass Chebyshev functions and low-pass double dominant pole MCF's of order (n + 2) can be obtained. An example illustrating the above suggested step by step procedure for a minimum number of bits calculation is given below.

5.3.1 Example

The critical second order section of an eighth order 1/2 dB ripple low-pass Chebyshev function has a = 0.043620, b = 1.005002, ω_{n1} = 1.005948, and C_h = 11.530788. Find the coefficient bit requirement Q_1 and Q_2 such that $|\Delta\omega_{n1}/\omega_{n1}| \leq 5\%$ and $|\Delta C_h/C_h| \leq 5\%$.

From Equation (5-18), Δa = 0.0 and Δb = 0.050345. Substituting Δa and Δb in Equation (5-13) gives Q_1 = 3 and Q_2 = 7 bits. By including the sign and integer bits, Q_1 = 5 and Q_2 = 9 bits. Rounding the coefficients γ_1 to five bits and γ_2 to nine bits results in γ_{1q} = 0.0 and γ_{2q} = .914063, where γ_{1q} and γ_{2q} are the rounded coefficients. In order to verify whether the given tolerance limits are satisfied, the values of a_q, b_q, ω_{n1q}, and C_{hq} must be calculated, where a_q, b_q, ω_{n1q}, and C_{hq} are the values of a, b, ω_{n1}, and C_h after coefficient rounding. From Equation (5-10), a_q = .044899 and b_q = .99899. By substituting a_q and b_q in Equation (5-15), ω_{n1q} = 1.0 and C_{hq} = 11.1360. Therefore, the resulting percentage change in C_h and ω_{n1} is given by

$$\frac{\Delta C_h}{C_h} = \frac{C_h - C_{hq}}{C_h} = 3.42\%$$

and

$$\frac{\Delta\omega_{n1}}{\omega_{n1}} = \frac{\omega_{n1} - \omega_{n1q}}{\omega_{n1}} = 0.59\% \qquad .$$

These satisfy the specified tolerance limits. Coefficient bit minimization by setting $Q_1 = 5$ bits and $Q_2 = 8$ bits, or by setting $Q_1 = Q_2 = 8$ bits fails in satisfying the specified tolerance limits; therefore, it is necessary to use $Q_1 = 5$ bits and $Q_2 = 9$ bits.

Next, the above step by step procedure is used to compare the coefficient bit requirements of the dominant pole sections of the nth order low-pass Chebyshev functions with respect to the low-pass double dominant pole MCF's of order (n + 2). These results are given in Table VI. In Table VI the number of bits are calculated to satisfy specifications which set a maximum of five per cent ω_{n1} and C_h variations for the Chebyshev case, and a maximum of five per cent ω_{b1} and C_t variations for the MCF case. The bits given in Table VI include the integer bit and the sign bit. From Table VI it can be seen that the coefficient word-length requirement is approximately the same for both functions, and in many cases the MCF's would require a lower coefficient word-length.

5.4 Conclusion

A method for computing the coefficient word-length estimation for first and second order digital filter sections is presented in this chapter. This method can be used for the cases where the bilinear transformation method is employed to obtain the discrete equation. In the proposed method the coefficient bits are obtained from the s-domain poles and their tolerance specifications. In addition, coefficient bit word-length comparison of nth order low-pass Chebyshev functions and low-pass double dominant pole MCF's of order (n + 2) is tabluated (Table VI).

TABLE VI

COEFFICIENT BIT REQUIREMENT OF THE DOMINANT POLE SECTION FOR 5% TOLERANCE LIMIT

n	3 dB				1/2 dB			
	Cheby. $H_n(z)$		MCF $L_m(z)$ m=n+2; c=2		Cheby. $H_n(z)$		MCF $L_m(z)$ m=n+2; c=2	
	γ_1 Bits	γ_2 Bits	γ_1 Bits	γ_2 Bits	γ_1 Bits	γ_2 Bits	γ_1 Bits	γ_2 Bits
2	5	6	5	7	5	7	6	7
3	5	7	5	8	5	6	5	6
4	5	8	5	5	5	7	5	8
5	5	8	5	8	5	8	5	5
6	5	8	5	9	5	8	5	6
7	5	9	5	8	5	8	5	9
8	5	10	5	9	5	9	5	7
9	5	10	5	9	5	10	5	6
10	5	10	5	9	5	9	5	9

This comparison is given on the basis of five percent tolerance specification on both the break frequency and the magnitude of the dominant pole section. From the table it can be seen that in many cases the MCF's due to their low dominant Q-factor will require fewer number of coefficient bits than the Chebyshev functions. Examples illustrating these ideas are included.

CHAPTER VI

SUMMARY AND SUGGESTIONS FOR FURTHER STUDY

6.1 Summary

This thesis approaches from a new perspective the reduction of the digital filter output noise variance due to product rounding. This approach is developed for the Chebyshev and the Butterworth filter functions, and it consists of replacing the designed nth order filter by an (n + 2) double dominant pole modified filter function whose dominant pole quality-factor Q_d is significantly less than the Q_d of the original filter function.

An analytical approach for obtaining the coefficients of a modified low-pass maximally flat Butterworth function with multiple dominant pole and reduced Q_d is given. In addition, a new algorithm is presented which determines the coefficients of a low-pass non equal-ripple modified Chebyshev function (MCF) with multiple dominant poles and notably reduced Q_d. These modified filter functions will always satisfy the pass-band specifications; however, their transition region attenuation is reduced. Alternate methods are pointed out in order to increase the transition region attenuation of the modified functions at the cost of increasing the low Q_d.

The output noise variance and the coefficient word-length comparison of the nth order low-pass Chebyshev functions and the low-pass double dominant pole MCF's of order (n + 2) for 1/2 dB and 3 dB cases is drawn.

In this study a reduction in the output roundoff noise is achieved for the cases where Q_d reduction is more than 50%. This includes all high Q_d Chebyshev functions. In addition, the word-length requirements are approximately the same for both functions (in many cases the MCF's would require a lower coefficient word-length). For the modified Butterworth case, high order functions must be considered in order that Q_d reduction becomes substantial. Therefore, no comparison tables are given for the modified Butterworth functions.

A new approach is given for computing the output noise variance and for coefficient word-length estimation for the cases where the bilinear transform is used. The output noise variance is computed using the s-domain transfer function and the driving point impedance. The coefficient word-length estimation for the first and second order digital filter sections such that the s-domain pole tolerance limits are satisfied is presented. If the digital filter is designed based on other than the bilinear transformation then the suggested methods for coefficient word-length estimation and output noise variance calculation will require additional computation steps; in this case, the previously suggested methds in the referenced literature are more appropriate to employ.

6.2 Suggestions for Further Study

In the following, some extensions to the present study are given. Appropriate references are indicated.

6.2.1 Modified Functions

The modified Butterworth and modified Chebyshev functions with low dominant pole quality factor (Q_d) will always satisfy the pass-band specifications. However, in general, the stop band specifications may not be met, as the low Q_d is obtained at the cost of low attenuation in the transition region. A procedure is given in this thesis to increase the attenuation of the transition region at the cost of increasing the low Q_d. As an extension of this present research, the attenuation of the transition region might be increased without sacrificing the low Q_d by including a pair of complex conjugate zeros on the $j\omega$ axis. Some work has already been done in this area [DU 1].

The multiple dominant pole notion has been used in this thesis to develop alternate filter functions for two very common filter types, the Butterworth and the Chebyshev functions. However, this same notion of dominant pole multiplicity can be used to derive alternate filter functions for other common filter types such as the Bessel, Chebyshev type II, and the elliptic filters. A suggested approach for the Chebyshev type II function would be to derive an analytical method for obtaining the modified Chebyshev type II functions by making use of its maximally flat property. For deriving the modified elliptic filter function from the given elliptic filter function, a possible approach would be to derive a numerical algorithm similar to that used in obtaining the MCF's. It is anticipated that due to the elliptic filter's high Q_d property, a modified elliptic filter with multiple dominant poles will result in a substantial Q_d reduction.

6.2.2 Coefficient Bit Estimation

The coefficient bit estimation procedure given in this thesis will in many cases give the minimum number of bits required to meet the given pole tolerance limits in the s-domain. For the case where the minimum number of bits is not directly obtained from the equations given, a step by step iterative numerical algorithm that results in a minimum number of bits is presented. This problem of obtaining the minimum number of bits and not simply a close estimate exists in other methods that have already been suggested [MI 1, CR 1, WH 1]. It is therefore desirable to obtain an analytical method that will give the minimum bit requirement directly without the need of an iterative numerical minimization procedure.

6.2.3 Output Noise Variance

The output noise variance comparison conducted in this thesis is based on the minimum output noise variance of a cascaded first and second order section. This involves optimum digital filter section ordering that results in a minimum output noise variance. The present research in this area including this thesis relies on iterative numerical algorithms for optimum section ordering [CH 1, LE 1, JA 1]. An analytical approach to this problem is desired.

Finally, the digital filter output noise variance computation using the s-domain transfer function and its driving point impedance concept is used in this thesis. Further study in this area may involve these concepts in the z-domain. This may include a new notion of z-domain driving point impedance and its relation to the output noise variance of the digital filter and to the s-domain driving point impedance. A suggested approach would be to apply the bilinear transformation to every circuit element of the s-domain filter realization and to the s-domain driving point impedance. A good reference in this area would be the work done by Crochiere [CR 2].

SELECTED BIBLIOGRAPHY

[AS 1] Åström, K., E. Jury, and R. Agniel. "A Numerical Method for the Evaluation of Complex Integrals." IEEE Trans. Aut. Contr., Vol. AC-15 (August, 1970), 468-471.

[AV 1] Avenhaus, E., and H. W. Schuessler. "On the Approximation Problem in the Design of Digital Filters With Limited Word-Length." Arch. Elek. Ubertragung, Vol. 24 (December, 1970), 571-575.

[AV 2] Avenhaus, E. "On the Design of Digital Filters With Coefficients of Limited Word Length." IEEE Trans. Audio Electroacoust., Vol. AU-20 (August, 1972), 206-212.

[BU 1] Budak, A., and P. Aronhime. "Transitional Butterworth Chebyshev Filters." IEEE Trans. Circuit Theory, Vol. CT-18 (May, 1971), 413-415.

[CA 1] Cardwell, R. E. "Synthesis of Recursive Digital Filters." (Unpub. Ph.D. thesis, Polytechnic Institute of Brooklyn, June, 1973).

[CH 1] Chan, D. S. K., and L. R. Rabiner. "An Algorithm for Minimizing Roundoff Noise in Cascade Realizations of Finite Impulse Response Digital Filters." Bell System Technical Journal, Vol. 52 (March, 1973), 347-385.

[CR 1] Crochiere, R. E. "A New Statistical Approach to the Coefficient Word-Length Problem for Digital Filters." IEEE Trans. Circuits and Systems, Vol. CAS-22 (March, 1975), 190-196.

[CR 2] Crochiere, R. E. "On the Location of Zeros and a Reduction in the Number of Adds in a Digital Ladder Structure." IEEE Trans. Audio Electroacoust., Vol. AU-21 (December, 1973), 551-552.

[DU 1] Dutta, Roy, S. C. "On Pole Locations of Sharp Cutoff Low-Pass Filters." IEEE Proceedings (Lett.), Vol. 62 (April, 1974), 520-521.

[GE 1] Gersch, W. "Computation of the Mean-Square Response of a Stationary Time-Discrete System." IEEE Trans. Aut. Contr. (Corresp.), Vol. AC-16 (June, 1971), 277.

[GE 2] Géher, K. Theory of Network Tolerances. Budapest, Hungary: Akadémiai Kiadó, 1973.

[GO 1] Gorsky-Popiel, J. "Reduction of Network Sensitivity Through the use of Higher Order Approximating Functions." Electronics Letters, Vol. 3, No. 8 (August, 1967), 365-366.

[GO 2] Gowdi, J., and J. Hadstate. "Design of Optimum-Configurations of Digital Filters." 1973 IEEE Southeast. Con. Region 3 Conference, Louisville, Ky., (April, 1973), R-5-1, R-5-5.

[GO 3] Gold, B., and C. M. Rader. Digital Processing of Signals. New York: McGraw-Hill, 1969, Chapter 4.

[GR 1] Greaves, C. J., G. A. Gagne, and G. W. Bordner. "Evaluation of Integrals Appearing in Minimization Problems of Discrete-Data Systems." IEEE Trans. Aut. Cont., Vol. AC-11 (January, 1966), 145-148.

[HS 1] Hsia, T. C. "On the Simplification of Linear Systems." IEEE Trans. Aut. Contr., Vol. AC-17 (June, 1972), 372-374.

[HU 1] Huelsman, L. P. Theory and Design of Active RC Circuits. New York: McGraw-Hill, 1968, 13-32, 52-54.

[JA 1] Jackson, L. B. "Roundoff-Noise Analysis for Fixed-Point Digital Filters Realized in Cascade or Parallel Form." IEEE Trans. Audio Electroacoust., Vol. AU-18, No. 2 (June, 1970), 107-122.

[JA 2] Jackson, L. B. "On the Interaction of Roundoff Noise and Dynamic Range in Digital Filters." Bell System Technical Journal, Vol. 49, No. 2 (February, 1970), 159-184.

[JU 1] Jury, E. I. Theory and Applications of the Z-Transform Method. New York: John Wiley & Sons, 1964.

[JU 2] Jury, E. I., and S. Gutman. "The Inner Formulation for the Total Square Integral (SUM)." Proceedings of the IEEE (Lett.), Vol. 61 (March, 1973), 395-397.

[KA 1] Kaiser, J. F. "Some Practical Considerations in the Realization of Linear Digital Filters." Proc. Third Annual Allerton Conference on Circuit and System Theory, (October, 1965), 621-633.

[KA 2] Kaiser, J. "Digital Filters." IEEE NEREM Part 2, Signal Processing, (1973), 1-10.

[KA 3] Karni, S. Network Theory Analysis and Synthesis. Boston, Mass.: Allyn and Bacon, 1966, 190-205.

[KI 1] King, R. E., and W. A. Brown. "Stochastic Analysis of a Class of Linear Time-Quantized Control Systems." Journal of Electronics & Control, Vol. 17 (September, 1964), 319-344.

[KI 2] Kingsbury, N. G. "Digital Filter 2nd-Order Element With Low Quantization Noise for Poles and Zeros at Low Frequencies." *Electronics Letters*, Vol. 9, No. 12 (June, 1973), 271-273.

[KN 1] Knowles, J. B., and R. Edwards. "Effects of a Finite-Word-Length Computer in a Sampled Data Feedback System." *Proc. Inst. Elec. Eng.*, Vol. 112 (June, 1965), 1197-1207.

[KN 2] Knowles, J. B., and E. M. Olcayto. "Coefficient Accuracy and Digital Filter Response." *IEEE Trans. Circuit Theory*, Vol. 15 (March, 1968), 31-41.

[KO 1] Kominek, Z. "Low Pass Approximation Convenient for Circuits With Low Q." *Proc. of the Fourth Colloq. Microwave Communication*, Akadémiai Kiadó, Budapest, Hungary, Vol. 2 (1970), CT-15/1-CT-15/5.

[KU 1] Kuo, F. F., and J. F. Kaiser. *System Analysis by Digital Computer.* New York: John Wiley & Sons, 1966, Chapter 7.

[LE 1] Lee, W. "Optimization of Digital Filters for Low Roundoff Noise." *IEEE Trans. Circuits and Systems*, Vol. CAS-21, No. 3 (May, 1974), 424-431.

[LE 2] Ledbetter, J. D., and R. Yarlagadda. "Digital Filter Synthesis - A Low Sensitivity System Matrix." *IEEE Trans. Circuit Theory*, Vol. CT-20 (May, 1973), 322-324.

[LI 1] Liu, B. "Effect of Finite Word Length on the Accuracy of Digital Filters - A Review." *IEEE Trans. Circuit Theory*, Vol. CT-18 (November, 1971), 670-677.

[MA 1] Massad, K., and R. Yarlagadda. "Modified Butterworth Functions With Low Q-Factor." *Proc. Inst. Elec. Eng.*, Vol. 122 (February, 1975), 135-136.

[ME 1] Melsa, J., and D. Schultz. *Linear Control Systems.* New York: McGraw-Hill, 1969, 202-205.

[MI 1] Mitra, S., K. Hirano, and H. Sakaguchi. "A Simple Method of Computing the Input Quantization and Multiplication Roundoff Errors in a Digital Filter." *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-22 (October, 1974), 326-329.

[MI 2] Mitra, S. K., and R. J. Sherwood. "Estimation of Pole Zero Displacements of a Digital Filter due to Coefficient Quantization." *IEEE Trans. Circuits and Systems*, Vol. CAS-21 (January, 1974), 116-124.

[MI 3] Mitra, S. K. "On the Construction of a Positive Real Impedance From a Given Even Part." *IEEE Proceedings*, Vol. 51 (September, 1963), 1267.

[NE 1] Newton, G., L. Gould, and J. F. Kaiser. *Analytical Design of Linear Feedback Controls*. New York: John Wiley & Sons, 1957, 366-381.

[OP 1] Oppenheim, A. V., and C. W. Weinstein. "Effects of Finite Register Length in Digital Filters and the Fast Fourier Transform." *IEEE Proceedings*, Vol. 60 (August, 1972), 957-976.

[PA 1] Papoulis, A. *The Fourier Integral and its Applications*. New York: McGraw-Hill, 1962, 212-213.

[PR 1] Premoli, A. "The MUCROMAF Polynomials: An Approach to the Maximally Flat Approximation of RC Active Filters With Low Sensitivity." *IEEE Trans. on Circuit Theory*, Vol. CT-20 (January, 1973), 77-80.

[PR 2] Premoli, A. "A New Class of Equal-Ripple Filtering Functions With Low Q-Factors: The MUCROER Polynomials." *IEEE Trans. Circuits and Systems*, Vol. CAS-21 (September, 1974), 609-613.

[RA 1] Rabiner, L. R., and B. Gold. *Theory and Applications of Digital Signal Processing*. Englewood Cliffs, N.J.: Prentice-Hall, 1975, 336-344.

[TE 1] Temes, G. C., and S. K. Mitra. *Modern Filter Theory and Design*. New York: John Wiley & Sons, 1973, 334-344, 531-534.

[TE 2] Temes, G. C., and D. A. Calahan. "Computer Aided Network Optimization: the State-of-the Art." *IEEE Proceedings*, Vol. 55 (November, 1967), 1832-1863.

[VA 1] Van Valkenburg, M. E. *Modern Network Synthesis*. New York: John Wiley & Sons, 1967.

[WH 1] White, S. A. "Coefficient-Word-Length Requirement for Accurate Pole Location for Digital Filters." *Proc. Second Asilomar Conference on Circuits and Systems*, (November, 1968), 209-213.

[WI 1] Wilde, D. J. *Optimum Seeking Methods*. Englewood Cliff, N.J.: Prentice-Hall, 1964, 32-35, 145-157.

APPENDIX A

PHYSICAL METHOD ALGORITHM

This algorithm calculates the poles of the mth order low-pass double dominant pole modified Chebyshev function (MCF) with low dominant pole quality factor (Q_d). The algorithm is initiated with the break point frequency and the damping ratio of the nth order Chebyshev function. The relationship between the degree of the MCF, m, and the degree of the Chebyshev function, n, is given by $m = n + 2(c - 1)$, where c corresponds to the dominant pole multiplicity and is taken here as 2. In the algorithm, the double dominant poles replace the dominant poles of the nth order Chebyshev function, and an iterative procedure is used to fit the resultant function to the pass-band specifications. After meeting the pass-band specification and normalizing the poles to a cutoff frequency of one, the MCF is obtained. The algorithm output includes the following: print-out of the data, subroutine Gold 1 [ME 1] convergence monitor, the adjusted parameters in the iterative procedure, poles and Q_d of the MCF, and the MCF magnitude print-out.

```
      IMPLICIT REAL*8(A-H,O-Z)
      DOUBLE PRECISION DSQRT
C
C     THIS PROGRAM USES THE PHYSICAL METHOD TO CALCULATE THE COEFFICIENTS
C     OF THE MTH ORDER LOW-PASS MODIFIED CHEBYSHEV FUNCTION WITH LOW QUALITY
C     FACTOR Q, STARTING FROM THE POLES OF THE NTH ORDER ORIGINAL CHEBYSHEV
C     FUNCTION. M=N+2*(C-1)
C   INPUT QUANTITIES
C     CASCADAD EVEN TRANSFER FUNCTIONS OF THE FORM
C     WN**2/(S**2+2*ZETA*WN*S+WN**2)  IS CONSIDERED HERE.
C     A(I)=W(I)= THE BREAK FREQUENCY OF THE ORIGINAL CHEBYSHEV FUNCTION
C       ARRANGED IN ASCENDING SEQUENCE.
C     B(I)= THE DAMPING RATIO OF THE ORIGINAL CHEBYSHEV FUNCTION
C       ARRANGED IN DESCENDING SEQUENCE.
C     R(I)= ESTIMATE OF THE CAMPING RATIO OF THE MODIFIED CHEBYSHEV
C       FUNCTION. SET R(I)=B(I), I=1,....,N-1 AND SET R(N)=5*B(N)
C     L= ORDER OF THE ORIGINAL CHEBYCHEV FUNCTION
C     EPSI= PASS BAND RIPPLE FACTOR I.E FOR 1DB EPSI=.5088471
C     FRED= FRACTIONAL REDUCTION FOR SUBROUTINE GOLD1 SET FRED=.001 TO .00001
C     DATA REQUIRED: L; EPSI; FRED; A(I); B(I); AND R(I).
C
      DIMENSION V(10),X(120),W(15),R(9),H(10),P(10),A(10),B(10),F(10),Y(
     1120),Z(101),WNSQ(10),AX(601),AY(601),REAL(10),AEMAJ(10),XLJ(10),XH
     1J(10)
      SQRT(X)=DSQRT(X)
      READ(5,52)L,EPSI,FRED
      WRITE(6,77)L,EPSI,FRED
      L=L/2
      READ(5,12)(A(I),I=1,L)
      READ(5,12)(B(I),I=1,L)
      READ(5,12)(R(I),I=1,L)
      WRITE(6,78)(A(I),I=1,L)
      WRITE(6,79)(B(I),I=1,L)
      WRITE(6,91)(R(I),I=1,L)
      Q=SQRT(1+EPSI**2)
      NSKIP=0
      MJUMP=0
      NTOGL=L-1
      MUP=1
      MULTP=0
      Z(1)=0.
      K=L-1
      N=1
      M=0
      KPK=0
      LCK1=0
      DELSS=1.1/50.
C     SETTING W(I)=A(I)= BREAK FREQUENCY
      DO 1 J=1,L
    1 W(J)=A(J)
C     COMPUTING X(I) THE FREQUENCY WHERE THE PEAK OF EVERY 2ND ORDER
C       SECTION IN THE ORIGINAL CHEBYSHEV FUNCTION OCCURS.
      DO 3 J=1,L
      UND=1.-2.*B(J)**2
      IF(UND.GT.0.0)GO TO 2
      X(J)=A(J)
      GO TO 3
    2 X(J)=A(J)*SQRT(UND)
```

```
      3 CONTINUE
C       CALCULATING THE VALUE V(I) OF ORIGINAL CHEBY AT FREQ=X(I)
        DO 22 J=1,L
        V(J)=1.
        DO 22 I=1,L
        DUM=A(I)**2
        F(I)=DUM/SQRT((DUM-X(J)**2)**2+(2.*B(I)*A(I)*X(J))**2)
        V(J)=V(J)*F(I)
     22 CONTINUE
C       CALCULATING (DEC) THE INCREMENT IN V(L)
        DEC=(Q-1.)/5
        LAD=1
        NTK=0
        GO TO 29
     27 MJUMP=J
     29 D=0.
C       ADJUSTING THE PEAK OF EVERY 2ND ORDER SECTION TO MEET THE PASS BAND
C       SPECIFICATION
      6 DO 7 J=MUP,L
        Y(J)=1.
        DO 8 I=1,K
        DUM=W(I)**2
        H(I)=DUM/SQRT((DUM-X(J)**2)**2+(2.*R(I)*W(I)*X(J))**2)
        Y(J)=Y(J)*H(I)
      8 CONTINUE
        H(L)=W(L)**4/((W(L)**2-X(J)**2)**2+(2.*R(L)*W(L)*X(J))**2)
        Y(J)=Y(J)*H(L)
        P(J)=(V(J)*H(J))/Y(J)
        IF(J-L)15,14,14
     14 P(J)=SQRT(P(J))
     15 IF(P(J)-1.)10,10,9
      9 R(J)=SQRT(.5-.5*SQRT(1.-1./(P(J)**2)))
        GO TO 7
     10 R(J)=SQRT(((W(J)**4/P(J)**2)-(W(J)**2-X(J)**2)**2)/(4.*W(J)**2*X(J
       1)**2))
      7 CONTINUE
C       CALCULATING THE NEW X(I) & W(L)
        DO 17 I=MUP,K
        IF(R(I)-.707107)16,17,17
     16 X(I)=W(I)*SQRT(1.-2.*R(I)**2)
     17 CONTINUE
        W(L)=X(L)/SQRT(1.-2.*R(L)**2)
        IF(NTOGL.EQ.K)GO TO 49
        WRITE(6,75)(V(I),I=1,L),(P(I),I=1,L),(R(I),I=1,L),(X(I),I=1,L)
C       FIND THE NEW VALUE OF V(I)
     49 DO 19 J=MUP,K
        V(J)=1.
        DO 19 I=1,L
        DUM=A(I)**2
        F(I)=DUM/SQRT((DUM-X(J)**2)**2+(2.*B(I)*A(I)*X(J))**2)
        V(J)=V(J)*F(I)
     19 CONTINUE
C       FIX V(K) SUCH THAT THE LAST VALLEY IN THE PASS BAND IS GREATER THAN 1.
C         INCREASE V(K) BY A SMALL AMOUNT IF VALLEY IS NOT G.T. 1.
        V(K)=1.0018
        IF(NTOGL.EQ.K)GO TO 4
C       SUBTRACT A FIXED AMOUNT FROM V(I) IN ORDER TO ADJUST THE FINAL
C         PEAKS AT THE END OF THE ITERATIONS I.E. WHEN NTOGL = 1.
```

```
      V(MJUMP)=V(MJUMP)-.001*MULTP
    4 CONTINUE
      Q=Q+1.
      IF(Z(1)-1.)39,40,40
   39 IF(Q-12.)6,20,20
   40 IF(Q-6.)6,20,20
   20 C=1.
   46 CONTINUE
      IF(NSKIP.EQ.1)GO TO 47
      MJUMP=NTOGL
   47 Z(1)=2.
C     FOLLOWING IS GOLD1 DATA
C     NORM= 1 ALLOWS US TO EVALUATE THE PEAK OF THE MODIFIED CHEBYSHEV
      I1=0
      NORM=1
      PIKDEL=1./(2.*L)
      IF(KPK.GT.2)GO TO 68
      KPK=KPK+1
      SS=0.
      D1=1.
      D2=0.
      JF=1
C     COMPUTING XLOW & XHI OF EACH 2ND ORDER FOR GOLD1
      DO 66 J1=1,50
      DO 61 J2=1,K
      DUM=W(J2)**2
      H(J2)=DUM/SQRT((DUM-SS**2)**2+(2.*R(J2)*W(J2)*SS)**2)
      D1=D1*H(J2)
   61 CONTINUE
      H(L)=W(L)**4/((W(L)**2-SS**2)**2+(2.*R(L)*W(L)*SS)**2)
      D1=D1*H(L)
      IF(D1.GT.D2)NCH=0
      IF(D2.LT.D1)GO TO 62
      IF(NCH.EQ.1)GO TO 62
      NCH=1
      XLJ(JF)=SS-.2
      XHJ(JF)=SS+.1
      JF=JF+1
   62 CONTINUE
      D2=D1
      SS=SS+DELSS
   66 CONTINUE
      JF=JF-1
      WRITE(6,76)(XLJ(J3),J3=1,JF),(XHJ(J3),J3=1,JF)
   68 CONTINUE
C     CHECK IF PEAKS OF MOD CHEB SATISFY THE PASS BAND SPECIFICATION
      DO 21 J=MJUMP,K
   43 WRITE(6,53)J
      XLOW=XLJ(J)
      XHI=XHJ(J)
      IF(XLOW.LT.0.)XLOW=0.
      CALL GOLD1(I1,XLOW,XHI,FREQ,YBIG,XBIG,B3,B4,J5,W,R,L,NORM,UPRIPL,Z
     1Z)
      IF(MULTP.EQ.-1.)GO TO 23
      IF(MULTP.EQ.1.)GO TO 23
      IF(YBIG.GT.Q)GO TO 30
      IF(NTK.EQ.1)GO TO 21
      IF(NTOGL.EQ.K)GO TO 30
      V(J)=1.
C     CALCULATING THE NEW V(I) AFTER INCREASING X(L)
      DO 41 I=1,L
      F(I)=A(I)**2/SQRT((A(I)**2-X(J)**2)**2+(2.*B(I)*A(I)*X(J))**2)
      V(J)=V(J)*F(I)
   41 CCNTINLE
      GO TU 29
   35 WRITE (6,50)
      GU TC 32
   26 IF(M-2)37,36,36
   36 WRITE (6,51)
      GU TC 60
   37 M=N
   60 CONTINUE
   32 CONTINUE
      IF(NSKIP.EQ.1)GO TO 48
      NTOGL=1
      GO TC 46
   48 CONTINUE
      WRITE(6,11)(X(I),I=1,L)
      WRITE(6,11)(W(I),I=1,L)
      WRITE(6,11)(F(I),I=1,L)
      WRITE(6,11)(V(I),I=1,L)
      WRITE(6,11)(Y(I),I=1,L)
      WRITE(6,11)C,D
      DO 67 J=1,L
   67 WNSQ(J)=W(J)**2
C     CALL GOLD1 S FIND THE NORMALIZING FREQUENCY. SET NORM= 2
      NORM=2
      XLOW=XBIG
      XHI=1.3
      WRITE(6,54)
      CALL GCLD1(I1,XLCW,XHI,FRED,YBIG,XBIG,B3,B4,J5,W,R,L,NOFM,UPRIPL,Z
     1Z)
      DW=.01
      WF=0.
      NF=1.
      NP=150
      AX(1)=WF
C     PLOT THE MODIFIED CHEBYSHEV FUNCTION OBTAINED
      DO 85 J=1,NP
      WA=WF*XBIG
      WSQ=WA**2
      Y2=1.
      DO 80 I=NF,K
      Y1=WNSQ(I)/SQRT((WNSQ(I)-WSQ)**2+(2.*R(I)*W(I)*WA)**2)
   80 Y2=Y2*Y1
   81 CONTINUE
      YN=WNSQ(L)**2/((WNSQ(L)-WSQ)**2+(2.*R(L)*W(L)*WA)**2)
      AY(J)=Y2*YN
      AX(J+1)=AX(J)+DW
      WF=WF+DW
      WRITE(6,55)AX(J),AY(J)
   85 CONTINUE
      WRITE(6,56)
C     NORMALIZING THE POLES
      NH=1
      DO 90 J=NH,L
      NTK=0
      MULTP=-1.*LAD
      LAD=LAD+1
      MUP=J
      WRITE(6,74)J,V(J)
      Z(1)=2.
      NSKIP=1
      GO TO 27
   23 MULTP=0.
   21 CONTINUE
   30 CONTINUE
      IF(NTOGL.EQ.[illegible])GO TO 44
      IF(J.EQ.L)GO TO 46
C     SET (MULTP) IN ORDER TO REDUCE V(J)
      IF(MJUMP.EQ.J)GO TO 82
      MULTP=1
      GO TO 83
   82 MULTP=MULTP+1
   83 CONTINUE
      MJUMP=J
      MUP=J
      WRITE(6,74)J,V(J)
      Z(1)=2.
      NSKIP=1
      NTK=1
      GO TO 29
C     ADJUSTING THE CRITICAL 2ND ORDER SECTION TO FIT PASS BAND SPECS
   44 CONTINUE
      IF(NSKIP.EQ.1)GO TO 48
   45 WRITE(6,71)
      C=YBIG
      XP1=XBIG
      XLOP=XLOW+.01
      IF(LCK1.EQ.2)GO TO 26
      IF(XBIG.LT.XLOP)GO TO 24
      IF(LCK1.EQ.1)GO TO 24
      IF(YBIG.GT.Q)GO TO 28
   24 CONTINUE
      M=2.*M
      Z(1)=2.
      LCK1=1
      IF(YBIG.LE.Q)GO TO 33
      LCK1=2
      WRITE(6,72)V(L)
      V(L)=V(L)-DEC
      WRITE(6,72)V(L)
      GO TO 29
   33 WRITE(6,72)V(L)
      V(L)=V(L)+DEC
      WRITE(6,72)V(L)
      GO TO 29
   28 WRITE(6,73)X(L)
C     INCREASE X(L) IF PASS BAND SPECS CAN NOT BE MET
      X(L)=X(L)+.0031
      WRITE(6,73)X(L)
      IF(X(L)-1.1)38,38,35
   38 M=0
      J=L
      REAL(J)=R(J)*W(J)
      AEMAJ(J)=SQRT(W(J)**2-REAL(J)**2)
      REAL(J)=REAL(J)/XBIG
      AEMAJ(J)=AEMAJ(J)/XBIG
      WRITE(6,57)REAL(J),AEMAJ(J)
   90 CONTINUE
      QUALF=(SQRT(REAL(L)**2+AEMAJ(L)**2))/(2.*REAL(L))
      WRITE(6,58)QUALF
      WRITE(6,59)Q
C     WRITE(6,70)
C     CALL GRAPH(AX,AY,AX,NP,0,1)
   11 FORMAT (6F9.5)
   12 FORMAT (5F11.7)
   50 FORMAT (1X,27HFREQ RANGE X(L) IS EXCEEDED)
   51 FORMAT (1X,40HCAUGHT IN A LOOP CF OVER AND UNDER SPECS)
   52 FORMAT(I2,2F10.7)
   53 FORMAT(/1X,'GOING INTO GOLD1 FOR GETTING PEAK NUMBER',I2)
   54 FORMAT(/1X,'FINDING THE NORMALIZING FREQUENCY')
   55 FORMAT(1X,5E17.8)
   56 FORMAT(//1X,'POLE LOCATION',10X,'REAL',16X,'IMAGINARY',/)
   57 FORMAT(14X,E18.8,3X,E18.8)
   58 FORMAT(//1X,'CRITICAL QUALITY FACTOR Q =',E18.8)
   59 FORMAT(/1X,'MAX RIPPLE MAGNITUDE ABOVE 1 I.E SQRT(1+E**2) RIPPLE='
     1,E18.8)
   70 FORMAT(///11X,'----------NOW THE NORMALIZED GRAPH----------')
   71 FORMAT(/1X,'ADJUSTING CRITICAL 2ND ORDER NOW')
   72 FORMAT(/1X,'V(L) =',F10.6)
   73 FORMAT(/1X,'X(L) =',F10.6)
   74 FORMAT(/1X,'V(',I1,') =',F10.6)
   75 FORMAT(1X,5F6.4,2X,5F6.4,2X,5F6.4,2X,5F6.4)
   76 FORMAT(1X,'XLOW(I) & XHI(I) =',10F9.5)
   77 FORMAT(1X,'CHEBYSHEV DEGREE =',I2,' RIPPLE FACTOR =',F10.6,' GOLD1
     1 FRACTIONAL REDUCTION =',F11.8)
   78 FORMAT(1X,'CHEBYSHEV BREAK FREQUENCIES WN(I)=',6F11.7)
   79 FORMAT(1X,'CHEBYSHEV DAMPING RATIO =',6F11.7)
   91 FORMAT(1X,'DAMPING RATIO STARTING ESTIMATES=',6F11.7)
      STOP
      END
      SUBROUTINE GOLD1(K,XL,XR,F,YBIG,XBIG,XL1,XR1,N,WN,XZT,NRD,NORM,UPR
     1IPL,Z)
      IMPLICIT REAL*8(A-H,C-Z)
C
C     THIS SUBROUTINE WILL SEARCH OVER A ONE-DIMENSIONAL UNIMODAL FUNCTION
C     AND REPORT THE EXTREME ORDINATE FOUND, ITS ABSCISSA, FINAL ABSCISSAS
C     BOUNDING THE INTERVAL OF UNCERTAINTY, AND THE NUMBER OF FUNCTION
C     EVALUATIONS EXPENDED DURING THE SEARCH.
C     FOR REFERENCE SEE C.MISCHKE BOOK 'INTRO. TO COMPUTER-AIDED DESIGN'P.180
C
C     THE SUBROUTINE REQUIRES THE SPECIFICATION OF THE PRESENT INTERVAL OF
C     UNCERTAINTY, FRACTIONAL REDUCTION IN THE INTERVAL OF UNCERTAINTY,
C     AND WHETHER OR NOT A CONVERGENCE MONITOR PRINTOUT IS DESIRED.
C     PROVIDE A SUBROUTINE MERIT1(X,Y) WHICH RETURNS THE ORDINATE Y WHEN
C     THE ABSCISSA X IS TENDERED.
C     VARIABLES
C     K=0 CONVERGENCE MONITOR WILL NOT PRINT.
C     K=1 CONVERGENCE MONITOR WILL PRINT.
C     XL= ORIGINAL LEFTHAND ABSCISSA OF INTERVAL OF UNCERTAINTY.
C     XR= ORIGINAL RIGHTHAND ABSCISSA OF INTERVAL OF UNCERTAINTY.
C     F= FRACTIONAL REDUCTION IN INTERVAL OF UNCERTAINTY DESIRED.
C     YBIG= EXTREME ORDINATE DISCOVERED DURING SEARCH.
C     XBIG= ABSCISSA OF EXTREME ORDINATE.
C     XL1= FINAL LEFTHAND ABSCISSA OF INTERVAL OF UNCERTAINTY.
C     XR1= FINAL RIGHTHAND ABSCISSA OF INTERVAL OF UNCERTAINTY.
C     N= NUMBER OF FUNCTION EVALUATIONS EXPENDED DURING SEARCH.
C
      DIMENSION XZT(9),WN(15)
C     QABS(ARG)=ABS(ARG)
      QABS(ARG)=DABS(ARG)
C   FO REFERENCE SEE C.MISCHKE BOOK 'INTRO TO COMPUTER-AIDED DESIGN'PAGE 180
      GO TO 100
C   .... PRINT CONVERGENCE MONITOR HEADINGS IF REQUIRED ......
  111 IF(K)32,31,32
   32 WRITE(6,33)
   33 FORMAT(37H1CONVERGENCE MONITOR SUBROUTINE GOLD1,//,58H      N
     1 Y1              Y2               X1              X2,//)
   31 N = 0
      XLEFT = XL
      XRIGHT = XR
   13 SPAN = XR - XL
      DELTA=QABS(SPAN)
   14 X1 = XL + 0.381966*DELTA
      X2 = XL + 0.618034*DELTA
      CALL MERIT1(X1,Y1,WN,XZT,NRD,NORM,UPRIPL,Z)
      CALL MERIT1(X2,Y2,WN,XZT,NRD,NORM,UPRIPL,Z)
      N = N + 2
    3 IF(K)34,9,34
   34 WRITE(6,35)N,Y1,Y2,X1,X2
   35 FORMAT(I5,4(1X,E15.7))
    9 IF(QABS(XL-XR)-QABS(F*SPAN))4,4,8
    8 DELTA = 0.618034*DELTA
      IF(Y1 - Y2)1,10,2
    1 XL = X1
      X1 = X2
      Y1 = Y2
      X2 = XL + 0.618034*DELTA
      CALL MERIT1(X2,Y2,WN,XZT,NRD,NORM,UPRIPL,Z)
      N = N + 1
      GO TO 3
    2 XR = X2
      Y2 = Y1
      X2 = X1
      X1 = XL + 0.381966*DELTA
      CALL MERIT1(X1,Y1,WN,XZT,NRD,NORM,UPRIPL,Z)
      N = N + 1
      GO TO 3
    4 IF(Y2 - Y1)5,5,6
    5 YBIG = Y1
      XBIG = X1
      GO TO 7
    6 YBIG = Y2
      XBIG = X2
    7 XL1 = XL
      XR1 = XR
      GO TO 39
   10 XL = X1
      XR = X2
      DELTA = XR - XL
      GO TO 14
   39 IF(K)40,37,37
   37 WRITE(6,38)XLEFT,XRIGHT,F,YBIG,XBIG,XL1,XR1,N
   38 FORMAT(//,
     154H LEFTHAND ABSCISSA OF INTERVAL OF UNCERTAINTY ........,E15.7,/,
     254H RIGHTHAND ABSCISSA OF INTERVAL OF UNCERTAINTY .......,E15.7,/,
     354H FRACTIONAL REDUCTION OF INTERVAL OF UNCERTAINTY .....,E15.7,/,
     454H EXTREME ORDINATE DISCOVERED DURING SEARCH ...........,E15.7,/,
     554H ABSCISSA OF EXTREME ORDINATE ........................,E15.7,/,
     654H NEW LEFTHAND ABSCISSA OF INTERVAL OF UNCERTAINTY ....,E15.7,/,
     754H NEW RIGHTHAND ABSCISSA OF INTERVAL OF UNCERTAINTY ...,E15.7,/,
     854H NUMBER OF FUNCTION EVALUATIONS EXPENDED IN SEARCH ...,I15,//)
   40 XL = XLEFT
      XR = XRIGHT
      RETURN
  100 IF(K)102,101,101
  101 IF(K - 1)104,104,102
  102 WRITE(6,103)K
  103 FORMAT(41H *****ERROR MESSAGE SUBROUTINE GOLD1*****,/,9H       I1,,
     1I15,14H IS NOT 0 OR 1)
      RETURN
  104 IF(XR - XL)105,107,107
  105 WRITE(6,106)XL,XR
  106 FORMAT(41H *****ERROR MESSAGE SUBROUTINE GOLD1*****,/,9H       A2,,
     1E15.7,21H NOT SMALLER THAN A3,,E15.7)
      RETURN
  107 IF(F)109,109,108
  108 IF(F - 1.0)111,109,109
  109 WRITE(6,110)F
  110 FORMAT(41H *****ERROR MESSAGE SUBROUTINE GOLD1*****,/,9H       A4,,
     1E15.7,31H DOES NOT LIE BETWEEN 0. AND 1.)
      RETURN
      END
      SUBROUTINE MERIT1(WA,Y,W,R,L,NORM,UPRIPL,Z)
      IMPLICIT REAL*8(A-H,O-Z)
C
C     MERIT1 IS A SUBROUTINE TO GOLD1 SEARCH, AN ORDINATE Y IS RETURNED
C       WHEN COLUMN VECTOR OF ABSCISSA W IS TENDERED.
C
      DOUBLE PRECISION DSQRT
      DIMENSION R(9),W(15),WNSQ(10)
      SQRT(X)=DSQRT(X)
      Y2=1.
      K=L-1
      NG=1
      NF=1
      DO 5 J=NG,L
    5 WNSQ(J)=W(J)**2
      WSQ=WA**2
      DO 10 J=NF,K
      Y1=WNSQ(J)/SQRT((WNSQ(J)-WSQ)**2+(2.*R(J)*W(J)*WA)**2)
   10 Y2=Y2*Y1
   15 CONTINUE
      YN=WNSQ(L)**2/((WNSQ(L)-WSQ)**2+(2.*R(L)*W(L)*WA)**2)
      Y=Y2*YN
      IF(NORM.EQ.1)GO TO 80
      IF(NORM.EQ.2)GO TO 25
   25 CONTINUE
C     FOLLOWING IS TO GET XBIG TO NORMALIZE THE MODIFIED CHEBYSHEV
      IF(Y.GT.1.)GO TO 30
      Y=1.+(1.-Y)
   30 CONTINUE
C     NOW LET Y BE ALWAYS -VE EXCEPT AT Y=1 WHERE ITS EQUAL TO ZERO,
C     THUS WE CAN GET MAX Y WHERE IT INTERSECTS LINE 1.
      Y=1.-Y
      GO TO 80
   80 CONTINUE
      RETURN
      END
```

APPENDIX B

LEAST SQUARES ERROR ALGORITHM

This algorithm gives the poles of the mth order low-pass double dominant poles modified Chebyshev function (MCF) with low dominant pole quality factor (Q_d). The input data includes the break point frequency of the nth order Chebyshev function, and the upper and lower bounds for the estimated damping ratio of the MCF. The relationship between the orders m and n is given by m = n + 2(c - 1), where c corresponds to the dominant pole multiplicity and is taken here as 2. In the algorithm double dominant poles replace the dominant poles of the nth order Chebyshev function, and the pass-band specifications are met by adjusting the parameters of the new function. This is achieved by using Gold 1 and Pattern search [ME 1] to minimize the pass-band error function. The MCF is obtained after meeting the pass-band specifications, and normalizing the poles to a cutoff frequency of one. The algorithm output includes the following: print-out of the data, pattern and Gold 1 convergence monitor, poles and Q_d of the MCF, and the MCF magnitude print-out. Subroutine Gold 1 will not be listed here since it has been included in Appendix A.

```
C
C      THIS PROGRAM GIVES THE POLES OF THE MODEFIED CHEBYCHEV POLYNOMIAL
C      WITH REDUCED CRITICAL QUALITY FACTOR Q. USING MULTIPLE POLES
C      IT CAN BE USED FOR CHEBYCHEV POLYNOMIALS OF DEGREE 2 & GREATER.
C      SECOND ORDER POLYNOMIALS OF THE FORM S**2+2*ZETA*WN*S+WN**2 ARE
C      CONSIDERED  IN HERE
C    INPUT QUANTITIES
C      N= NUMBER OF UNKNOWN PARAMETERS= NOF DAMPING RATIOS + 1 FOR 1ST
C        ORDER SECTION IN ODD CHEBYSHEV FUNCTIONS.
C      EPSI= PASSBAND RIPPLE FACTOR I.E.FOR 2DB EPSI= .7647831
C      Z= ORDER OF ORIGINAL CHEBYSHEV POLYNOMIAL
C      F= MINIMUM STEP SIZE IN SUBROUTINE PATTERN BEFORE QUITTING I.E F=.0001
C      FRED= FRACTIONAL REDUCTION FOR SUBROUTINE GOLD1 I.E FRED=.0001
C      XLO(J)= LOWER BOUND OF ZETA REQUIRED FOR PATTERN SEARCH
C      XHI(J)= UPPER BOUND OF ZETA REQUIRED FOR PATTERN SEARCH
C      WN(J)= BREAK FREQUENCIES OF ORIGINAL CHEBYSHEV POLYNOMIAL
C      DATA REQUIRED: F; FRED; XLO(I); XHI(I); WN(I).
       IMPLICIT REAL*8(A-H,O-Z)
       DIMENSION XHI(9),XLO(9),AY(150),AX(150),WN(15),X(9),WNSQ(10)
      1,REAL(9),AEMAJ(9)
C      FOLLOWING IS FOR DOUBLE PRECISION
C      QSQRT(ARG)=SQRT(ARG)
       QSQRT(ARG)=DSQRT(ARG)
C      JINT(ARG)=INT(ARG)
       JINT(ARG)=IDINT(ARG)
       READ(5,21)N,Z,EPSI,F,FRED
   21  FORMAT(I2,F3.0,3F11.9)
C      THE LOWER AND UPPER BOUND FOR ZETAS FOLLOWS
C      THE BREAK FREQUENCIES FOLLOWS
    1  READ(5,22)(XLO(J),J=1,N)
       IF(XLO(1).EQ.0.0)GO TO 99
       READ(5,22)(XHI(J),J=1,N)
       READ(5,23)(WN(J),J=1,N)
       WRITE(6,86)Z
   22  FORMAT(5F4.2)
   23  FORMAT(5F11.8)
       WRITE(6,24)(WN(J),J=1,N)
   24  FORMAT(/1X,'INITIAL CORNER FREQ WN =',5E16.8,/)
C      UPPIPL= THE HIGHIEST RIPPLE= SQRT(1+EPSI**2).NOTE THAT THE
C      LOWER RIPPLE LIMIT IS ALWAYS = 1.
C      DELWN= INCREMENT FOR THE LAST BREAK FREQ
       DELWN=.003
       UPRIPL=QSQRT(1+EPSI**2)
C      FOLLOWING IS PATRN DATA
C      ULTWT= MAX FREQ INCLUDED IN ERROR COST TERM
       ULTWT=1.
       LP=2
       DELTA=.001
C        NUM= 1 IS TO READ IN THE LATEST X(I) TO PATRN
       NUM=0
C      THE GOLD1 VALUES FOLLOWS
C      NORM= 0 GIVES US THE MAX OF FILTER FUNCTION
C      NORM= 1 ALLOWS US TO EVALUATE THE NORMALIZING FREQ
       NORM=0
       I1=1
       ODDEQ=Z/2
       ODCHK=JINT(ODDEQ)
C      ARRANGING XLOW FOR GOLD1 TO GET LAST PEAK
```

```
      IF(ODDEQ.EQ.ODCHK)GO TO 2
      IF(Z.NE.3)GO TO 2
      XLOW=.35
      GO TO 4
    2 CONTINUE
      IF(N.GT.1)GO TO 3
      XLOW=0.0
      GO TO 4
    3 CONTINUE
      XLOW=WN(N-1)
    4 CONTINUE
      XHIG=1.25
      XLCHK=XLOW+.01
    6 CONTINUE
C     CALL PATTERN SEARCH TO OBTAIN NEW ZETA VALUE
      CALL PATRN(N,LP,XHI,XLO,DELTA,F,X,WN,NUM,UPRIPL,Z,ULTWT)
    5 CONTINUE
C     CALL GOLD1 TO OBTAIN THE VALUE OF THE LAST PEAK.IF IT IS WITHIN
C     THE PASS BAND SPECIFICATION THE ITERATION STOPS;OTHERWISE,WN(N)
C     IS INCREMENTED AND PATTERN IS CALLED TO GIVE THE NEW ZETAS
      NORM=0
      CALL GOLD1(I1,XLOW,XHIG,FREQ,YBIG,XBIG,B3,B4,J5,WN,X,N,NORM,UPRIPL
     1,Z)
      IF(XBIG.GT.XLCHK)GO TO 20
   15 CONTINUE
      XLOW=XLOW+.02
      XLCHK=XLOW+.01
      WRITE(6,70)XLOW
      GO TO 5
   20 CONTINUE
      IF(YBIG.LE.UPRIPL)GO TO 50
      NUM=1
   25 CONTINUE
      WN(N)=WN(N)+DELWN
      WRITE(6,71)WN(N)
      DELTA=.001
      CALL PATRN(N,LP,XHI,XLO,DELTA,F,X,WN,NUM,UPRIPL,Z,ULTWT)
      CALL MERIT1(XBIG,YBIG,WN,X,N,NORM,UPRIPL,Z)
      IF(YBIG.GT.UPRIPL)GO TO 25
      GO TO 5
   50 WRITE(6,60)(X(I),I=1,N)
   60 FORMAT(//1X,'THE OPTIMUM VALUES OF ZETA ARE X(I)=',9E16.8,/)
      AX(1)=0.
      W=0.
      DW=1.5/100.
      NP=100
      NM=N-1
      NF=1
      NG=1
      IF(ODDEQ.EQ.ODCHK)GO TO 26
      NG=2
   26 CONTINUE
      DO 35 J=NG,N
   35 WNSQ(J)=WN(J)**2
C     FOLLOWING IS TO PLOT THE NORMALIZED MAGNITUDE OF MOD CHEBYSHEV
C     FIRST OBTAIN THE NORMALIZING FREQUENCY FROM GOLD1
   44 CONTINUE
      NORM=1
      XLOW=XBIG
      XHIG=1.3
      CALL GOLD1(I1,XLOW,XHIG,FRED,YBIG,XBIG,B3,B4,J5,WN,X,N,NORM,UPRIPL
     1,Z)
      DW=1./100.
      WF=0.
      NF=1
      NP=NP+50
      DO 85 J=1,NP
      W=WF*XBIG
      WSQ=W**2
      Y2=1.
      IF(ODDEQ.EQ.CDCHK)GO TO 82
      Y2=X(1)*UPRIPL/QSQRT(X(1)**2+W**2)
      NF=2
      IF(N.EQ.2)GO TO 81
   82 CONTINUE
      IF(N.EQ.1)GO TO 81
      DO 80 I=NF,NM
      Y1=WNSQ(I)/QSQRT((WNSQ(I)-WSQ)**2+(2.*X(I)*WN(I)*W)**2)
   80 Y2=Y2*Y1
   81 CONTINUE
      YN=WN(N)**4/((WN(N)**2-WSQ)**2+(2.*X(N)*WN(N)*W)**2)
      AY(J)=Y2*YN
      AX(J+1)=AX(J)+DW
      WF=WF+DW
      WRITE(6,65)AX(J),AY(J)
   85 CONTINUE
      WRITE(6,74)
      NH=1
      IF(ODDEQ.EQ.CDCHK)GO TO 89
      NH=2
      WRITE(6,79)X(1)
   89 CONTINUE
      DO 90 J=NH,N
      REAL(J)=X(J)*WN(J)
      AEMAJ(J)=QSQRT(WN(J)**2-REAL(J)**2)
C     NORMALIZING THE POLES
      REAL(J)=REAL(J)/XBIG
      AEMAJ(J)=AEMAJ(J)/XBIG
      WRITE(6,75)REAL(J),AEMAJ(J)
   90 CONTINUE
      QUALF=(QSQRT(REAL(N)**2+AEMAJ(N)**2))/(2.*REAL(N))
      WRITE(6,77)QUALF
      WRITE(6,78)UPRIPL
      WRITE(6,76)
C     CALL GRAPH(AX,AY,AX,NP,0,1)
C     NORM=2 GETS THE STOP FREQ WHERE MOD CHEB & ORIGINAL CHEB INTERSECT
      NORM=2
      XLOO=1.01
      XHII=1.75
      CALL GOLD1(I1,XLOO,XHII,FRED,YB,XB,B3,B4,J5,WN,X,N,NORM,UPRIPL,Z)
      WRITE(6,87)WN(N),XB
   99 CONTINUE
   65 FORMAT(1X,5E17.8)
   70 FORMAT(1X,'NEW INCREASED XLO= ',E17.8)
   71 FORMAT(1X,'NEW INCREASED WN(N) =',E17.8,/)
   72 FORMAT(1X,'GOING INTO GOLD1',/)
```

```
73 FORMAT(/,/,/,/,11X,'------ NOW THE  NON NORMALIZED GRAPH ---')
74 FORMAT(//1X,'POLE LOCATION',10X,'REAL',16X,'IMAGINARY',/)
75 FORMAT(14X,E18.9,3X,E18.9)
76 FORMAT(/,/,/,11X,'------ NOW THE NORMALIZED GRAPH -----')
77 FORMAT(//1X,'CRITICAL QUALITY FACTOR Q =',E19.8)
78 FORMAT(/1X,'MAX RIPPLE MAGNITUDE ABOVE 1 I.E. SQRT(1+E**2). RIPPLE
  1 =',E18.9)
79 FORMAT(/1X,'VALUE FOR ODD POLYNOMIAL X(1) =',E18.9,/)
86 FORMAT(///////,'******** FOLLOWING IS FOR CHEBY DEGREE N =',F3.0,'
  1TO N+2 ******************',//)
87 FORMAT(/1X,'CHEB & MOD CHEB WITH WN(N)=',E17.8,'INTERSECT AT STOP
  1BANB AT FREQ W=',E17.8/)
   STOP
   END
```

```
      SUBROUTINE PATRN(N,NP,XHI,XLO,DELTA,F,X2,WN,NUM,UPRIPL,Z,ULTWT)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION XHI(9),XLO(9),X1(9),X11(9),X12(9),AM(9),X2(9),XSAVE(9),X
     13(9),AM1(9),X4(9),COMNT(20),WN(15)
C*********************************************************************C
C                                                                     C
C     THIS SUBROUTINE CONDUCTS A PATTERN SEARCH WITHIN REGIONAL       C
C     CONSTRAINTS IN A HYPERSPACE OF UP TO EIGHT INDEPENDENT          C
C     VARIABLES.  THIS PROCEDURE WAS DEVISED BY HOOKE AND JEEVES      C
C     (REF: OPTIMUM SEEKING METHODS BY WILDE, PG. 145)                C
C                                                                     C
C     PROVIDE A DIMENSION DECLARATION AS FOLLOWS:                     C
C     DIMENSION XHI(9),XLO(9)                                         C
C                                                                     C
C     PROVIDE A SUBROUTINE MERIT(X,Y) FROM WHICH AN ORDINATE Y IS     C
C     RETURNED WHEN COLUMN VECTOR OF ABSCISSA X IS TENDERED.          C
C                                                                     C
C     PROVIDE THE MERIT FUNCTION ON A DATA CARD                       C
C                                                                     C
C     NP=0 CONVERGENCE MONITOR WILL NOT PRINT                         C
C     ..........NOMENCLATURE..........                                C
C     N=NUMBER OF INDEPENDENT VARIABLES IN SEARCH (8 OR LESS)         C
C     NP=1 CONVERGENCE MONITOR WILL PRINT EVERY ITERATION             C
C     NP=2 CONVERGENCE MONITOR WILL PRINT EVERY 2ND ITERATION         C
C     DELTA=INITIAL STEP SIZE                                         C
C     F=MINIMUM STEP SIZE BEFORE QUITING
C     XLO=LOWER BOUND OF SEARCH DOMAIN, COLUMN VECTOR                 C
C     XHI=UPPER BOUND OF SEARCH DOMAIN, COLUMN VECTOR                 C
C                                                                     C
C*********************************************************************C
 1000 FORMAT (///' CONVERGENCE MONITOR - PATTERN SEARCH SUBROUTINE',//'
     1  NN',4X,'DELTA',7X,'Y',7X,'X(1)',6X,'X(2)',6X,'X(3)',6X,'X(4)',6X
     2,'X(5)',6X,'X(6)',6X,'X(7)',6X,'X(8)'/)
 1001 FORMAT (1X,I4,10E15.7)
 1002 FORMAT ('1')
 1003 FORMAT (///' LARGEST MERIT ORDINATE FOUND DURING SEARCH ..........
     1',E15.8/' NUMBER OF FUNCTION EVALUATIONS USED DURING SEARCH ...',I
     215/' FINAL SEARCH STEPSIZE ..............................',E15.8/
     3//)
 1004 FORMAT (' X(',I1,') =',E15.8/)
 1005 FORMAT (' XLO(',I1,') =',E15.8,5X,'XHI(',I1,') =',E15.8)
 1006 FORMAT ('1','THE MERIT FUNCTION EVALUATED'//1X,20A4////' REGIONAL
     1CONSTRAINTS'/)
 1007 FORMAT (20A4)
      DO 1 I=1,N
    1 WRITE (6,1005) I,XLO(I),I,XHI(I)
C
C     ------> INITIALIZE <------
C
      D15=DELTA
      IF(NUM.EQ.0)GO TO 3
      DO 2 J=1,N
      X1(J)=X2(J)
      X11(J)=X1(J)
      X12(J)=X1(J)
    2 XSAVE(J)=X1(J)
      GO TO 6
    3 CONTINUE
```

```
      DO 5 I=1,N
      X1(I)=XHI(I)-((XHI(I)-XLO(I))/2.0)
      X11(I)=X1(I)
      X12(I)=X1(I)
    5 XSAVE(I)=X1(I)
    6 CONTINUE
      CALL MERIT (X1,Y1,WN,N,UPRIPL,Z,ULTWT)
      ITER=0
      IF (NP) 9,10,9
    9 WRITE (6,1000)
C
      WRITE (6,1001) ITER,DELTA,Y1,(X1(I),I=1,N)
C     ------> EVALUATE THE STAR PATTERN      <------
C     ------> TO DETERMINE BASE POINT, B(I) <------
C
   10 AM(1)=Y1
      DO 55 I=1,N
      J=I+1
      X11(I)=X1(I)+DELTA
      IF (X11(I)-XHI(I)) 20,20,15
   15 X11(I)=XHI(I)
   20 CALL MERIT (X11,YA,WN,N,UPRIPL,Z,ULTWT)
      X12(I)=X1(I)-DELTA
      IF (X12(I)-XLO(I)) 25,30,30
   25 X12(I)=XLO(I)
   30 CALL MERIT (X12,YB,WN,N,UPRIPL,Z,ULTWT)
      IF (AM(I)-YA) 40,35,35
   35 IF (AM(I)-YB) 45,50,50
   40 AM(J)=YA
      X2(I)=X11(I)
      X12(I)=X11(I)
      GO TO 55
   45 AM(J)=YB
      X2(I)=X12(I)
      X11(I)=X12(I)
      GO TO 55
   50 AM(J)=AM(I)
      X2(I)=X1(I)
      X11(I)=X1(I)
      X12(I)=X1(I)
   55 CONTINUE
      Y2=AM(J)
      ITER=ITER+1
      IF (NP-1) 59,58,57
   57 D=ITER/2.0
      K=D
      D1=K
      IF (D1-D) 59,58,59
   58 WRITE (6,1001) ITER,DELTA,Y2,(X2(I),I=1,N)
   59 IF (Y2-Y1) 65,60,65
   60 DELTA=DELTA/8.
      DEL=DELTA-F/8.
      IF (DEL) 175,65,65
C
C     ------> EVALUATE PROJECTED TRIAL POINT  <------
C     ------> AS TEMPORARY HEAD POINT, T(I,0) <------
C
   65 DO 85 I=1,N
```

```
      X3(I)=2.0*X2(I)-XSAVE(I)
      IF (X3(I)-XLO(I)) 75,85,70
   70 IF (X3(I)-XHI(I)) 85,85,80
   75 X3(I)=XLO(I)
      GO TO 85
   80 X3(I)=XHI(I)
   85 CONTINUE
      CALL MERIT (X3,Y3,WN,N,UPRIPL,Z,ULTWT)
      IF (Y3-Y2) 90,90,165
C     ------> EVALUATE THE STAR PATTERN    <------
C     ------> AROUND PROJECTED TRIAL POINT <------
   90 DO 95 I=1,N
      X11(I)=X3(I)
   95 X12(I)=X3(I)
      AM1(1)=Y2
      DO 140 I=1,N
      J=I+1
      X11(I)=X3(I)+DELTA
      IF (X11(I)-XHI(I)) 105,105,100
  100 X11(I)=XHI(I)
  105 CALL MERIT (X11,YA,WN,N,UPRIPL,Z,ULTWT)
      X12(I)=X3(I)-DELTA
      IF (X12(I)-XLO(I)) 110,115,115
  110 X12(I)=XLO(I)
  115 CALL MERIT (X12,YB,WN,N,UPRIPL,Z,ULTWT)
      IF (AM1(I)-YA) 120,125,125
  120 AM1(J)=YA
      X12(I)=X11(I)
      X4(I)=X11(I)
      GO TO 140
  125 IF (AM1(I)-YB) 135,130,130
  130 AM1(J)=AM1(I)
      X11(I)=X3(I)
      X12(I)=X3(I)
      X4(I)=X3(I)
      GO TO 140
  135 AM1(J)=YB
      X11(I)=X12(I)
      X4(I)=X12(I)
  140 CONTINUE
      IF (AM1(J)-Y2) 155,155,145
C
C     ------> ESTABLISH A STAR PATTERN POINT  <------
C     ------> AS TEMPORARY HEAD POINT, T(I,0) <------
C
  145 DO 150 I=1,N
      XSAVE(I)=X2(I)
      AMSAV=Y2
      X1(I)=X4(I)
      X11(I)=X4(I)
      X12(I)=X4(I)
  150 Y1=AM1(J)
      GO TO 10
C
C     ------> ESTABLISH PREVIOUS BASE POINT   <------
C
C     ------> AS TEMPORARY HEAD POINT, T(I,0) <------
  155 DO 160 I=1,N
```

```
      XSAVE(I)=X1(I)
      AMSAV=Y1
      X1(I)=X2(I)
      X11(I)=X2(I)
      X12(I)=X2(I)
  160 Y1=Y2
      GO TO 10
C
C     ------> ESTABLISH PROJECTED TRIAL POINT <------
C     ------> AS TEMPORARY HEAD POINT, T(I,0) <------
C
  165 DO 170 I=1,N
      XSAVE(I)=X2(I)
      AMSAV=Y2
      X1(I)=X3(I)
      X11(I)=X3(I)
      X12(I)=X3(I)
  170 Y1=Y3
      DELTA=D15
      GO TC 10
  175 DELTA=DELTA*8.0
      WRITE (6,1003) Y2,ITER,DELTA
      DO 180 I=1,N
  180 WRITE (6,1004) I,X2(I)
      WRITE (6,1002)
      CONTINUE
      RETURN
      END
      SUBROUTINE MERIT(X,Y,WN,N,UPRIPL,Z,ULTWT)
      IMPLICIT REAL*8(A-H,O-Z)
C
C     MERIT IS A SUBROUTINE OF PATTERN SEARCH, AN ORDINATE Y IS RETURNED
C     WHEN COLUMN VECTOR OF ABSCISSA X IS TENDERED.
C
      DIMENSION WN(15),X(9),WNSQ(10)
C     QSQRT(ARG)=SQRT(ARG)
      QSQRT(ARG)=DSQRT(ARG)
      JINT(ARG)=IDINT(ARG)
C     NUMPIK=NUMBER OF CHEBY PEAKS
C     RESPID= THE IDEAL PASS BAND STRAIGHT LINE ABOVE ONE DESIRED
      Y=0.
C     FOLLOWING IS FOR 10TH ORDER CHEBY 3DB TO 12TH ORDER
      RESPID=1.+(UPRIPL-1.)/2.
      WT=1.
      NUMPIK=N
      WTMXST=3.
      W=0.
      MAXIT=NUMPIK*2*10
      DW=ULTWT/MAXIT
      NM=N-1
      NF=1
      NG=1
      ODDEQ=Z/2
      ODCHK=JINT(ODDEQ)
      IF(ODDEQ.EQ.ODCHK)GO TO 26
      NG=2
   26 CONTINUE
      DO 35 J=NG,N
   35 WNSQ(J)=WN(J)**2
      DO 10 J=1,MAXIT
      WSQ=W**2
      Y2=1.
      IF(ODDEQ.EQ.ODCHK)GO TO 39
      Y2=X(1)*UPRIPL/QSQRT(X(1)**2+W**2)
      NF=2
      IF(N.EQ.2)GO TO 41
   39 CONTINUE
      IF(N.EQ.1)GO TO 41
      DO 40 I=NF,NM
      Y1=WNSQ(I)/QSQRT((WNSQ(I)-WSQ)**2+(2.*X(I)*WN(I)*W)**2)
   40 Y2=Y2*Y1
   41 CONTINUE
      YN=WN(N)**4/((WN(N)**2-WSQ)**2+(2.*X(N)*WN(N)*W)**2)
      Y20=Y2*YN
      Y=Y+(WT*(Y20-RESPID))**2
      W=W+DW
      IF(W.GT.WTMXST)GO TO 3
      GO TO 5
    3 WT=7.
    5 CONTINUE
   10 CONTINUE
      Y=1./Y
      RETURN
      END
      SUBROUTINE MERIT1(W,Y,WN,X,N,NORM,UPPRIPL,Z)
      IMPLICIT REAL*8(A-H,O-Z)
C
C     MERIT1 IS A SUBROUTINE TO GOLD 1 SEARCH, AN ORDINATE Y IS RETURNED
C     WHEN COLUMN VECTOR OF ABSCISSA W IS TENDERED.
C
      DOUBLE PRECISION DLOG10
      DIMENSION X(9),WN(15),WNSQ(10)
C     QSQRT(ARG)=SQRT(ARG)
      QSQRT(ARG)=DSQRT(ARG)
      JINT(ARG)=IDINT(ARG)
      QABS(ARG)=DABS(ARG)
      Y2=1.
      NM=N-1
      NG=1
      NF=1
      ODDEQ=Z/2
      ODCHK=JINT(ODDEQ)
      IF(ODDEQ.EQ.ODCHK)GO TO 26
      NG=2
   26 CONTINUE
      DO 35 J=NG,N
   35 WNSQ(J)=WN(J)**2
      WSQ=W**2
      IF(ODDEQ.EQ.ODCHK)GO TO 39
      Y2=X(1)*UPPRIPL/QSQRT(X(1)**2+W**2)
      NF=2
      IF(N.EQ.2)GO TO 41
   39 CONTINUE
      IF(N.EQ.1)GO TO 41
      DO 40 I=NF,NM
      Y1=WNSQ(I)/QSQRT((WNSQ(I)-WSQ)**2+(2.*X(I)*WN(I)*W)**2)
   40 Y2=Y2*Y1
   41 CONTINUE
      YN=WN(N)**4/((WN(N)**2-WSQ)**2+(2.*X(N)*WN(N)*W)**2)
      Y=Y2*YN
      IF(NORM.EQ.0)GO TO 80
      IF(NORM.EQ.1)GO TO 42
      IF(NORM.EQ.2)GO TO 47
   42 CONTINUE
C     FOLLOWING IS TO GET XBIG TO NORMALIZE THE MOD CHEB
      IF(Y.GT.1.)GO TO 45
      Y=1.+(1.-Y)
   45 CONTINUE
C     NOW LET Y BE ALLWAS -VE EXCEPT AT Y=1. WHERE ITS EQUAL TO ZERO,
C     THUS WE CAN GET MAX Y WHERE IT INTERSECTS LINE 1.
      Y=1.-Y
      GO TO 80
   47 CONTINUE
C FINDING STOP BAND INTERSECTION FREQ BET MOD CHEB & ORIGINAL CHEB
      WN(11)=.63717164
      WN(12)=.24889627
      WN(13)=.12983136
      WN(14)=.066570948
      WN(15)=.020737564
      WN(6)=.99648154
      FN=1.
      DO 50 J=1,4
      FA=WNSQ(J)/QSQRT((WNSQ(J)-WSQ)**2+(2.*WN(J)*WN(10+J)*W)**2)
   50 FN=FN*FA
      FA=WN(6)**2/QSQRT((WN(6)**2-WSQ)**2+(2.*WN(6)*WN(15)*W)**2)
      FN=FN*FA
      Y=20*DLOG(Y)
      FN=20*DLOG10(FN)
      Y=-QABS(Y-FN)
      GO TO 80
   80 CONTINUE
      RETURN
      END
```

APPENDIX C

ROUNDOFF NOISE COMPUTATION

This program evaluates the output noise variance of a digital filter by computing the integral of $(1/(2\pi j))H(z)H(1/z)1/z$ around the unit circle in the z-plane. Realization in terms of first and second order cascaded canonic sections is considered, and the bilinear transformation is employed to transform from the s-domain function to the z-domain. The scaling factors are evaluated in subroutine SCALE, and the integral is evaluated by subroutine SALOSS [AS 1]. Input data includes ripple factor EPSI, desired section ordering, and s-plane poles. The algorithm output includes section ordering, and output noise variance due to A/D conversion and multiplier roundoff.

```
      IMPLICIT REAL*8(A-H,O-Z)
      DOUBLE PRECISION DSQRT
      DIMENSION A(15),B(15),AS(11),VA(8),KS(10),GNT(10),B1(10),B2(10),YF
     1(400),S(10)
     1,PP(7),QQ(7),LE(7),KN(7),KE(7)
C
C     THIS PROGRAM EVALUATES THE OUTPUT NOISE VARIANCE OF A DIGITAL FILTER
C     BY COMPUTING THE INTEGRAL OF (1/(2*PI*J))*B(Z)*B(1/Z)/(A(Z)*A(1/Z)*Z)
C     AROUND THE UNIT CIRCLE
C     THE SECTIONS ARE IN CANONIC CASCADE FORM OF 1ST & 2ND ORDER SECTIONS
C     THE BILINEAR TRANSFORMATION FROM S TO Z DOMAIN IS USED IN THIS PROGRAM
C     P1,P2,... ARE REAL POLE LOCATION IN S PLANE,P1=CRITICAL POLE,
C          P2=NEXT CRITICAL POLE, ETC.
C     Q1,Q2,... ARE IMAG POLE LOCATION IN S PLANE,Q1=CRITICAL POLE,
C          Q2=NEXT CRITICAL POLE, ETC.
C     PROGRAM FINDS NOISE DUE TO SECTION P1,Q1 CLOSEST TO OUTPUT, THEN
C     THE NOISE DUE TO P1,Q1,&P2,Q2 AND SO ON
C     ODD & EVEN FUNCTIONS CAN BE USED
C     G SETS H(Z)=ATTENUATION (ATT) WHEN Z=1
C     KS = NUMBER OF INPUT NOISE SOURCES
C     NSCT= NUMBER OF FIRST AND SECOND ORDER SECTIONS.
C     N= ORDER OF POLYNOMIALS A & B
C     GN1,GN2,...ARE TO SET H(Z)=ATTENUATION(ATT) AT Z=1 ATT=1 FOR ODD
C     FUNCTION H(Z)
C     DATA NEEDED ARE P1,Q1,P2,Q2,....,& NSCT AT PROGRAM END. EPSI & KS(I)
C       AT PROGRAM TOP. FOR ODD FUNCTIONS SET Q1 OF REAL POLE=0.0, &
C       EPSI=0.0
C     EPSI = RIPPLE FACTOR I.E., FOR 1 DB RIPPLE,THEN EPSI=.508847
C     LP & MP ARE THE DESIRED SECTION ORDERING,MP IS THE SECTION CLOSEST
C       TO OUTPUT,I.E.,FOR TWO 2ND ORDER SECTIONS IF LP=2 & MP=1,IT MEANS
C       THAT THE CRITICAL SECTION IS CLOSEST TO THE OUTPUT.
C     DATA REQUIRED: 1- S-PLANE 2ND QUADRANT POLE LOCATION P1,Q1,P2,Q2,...;
C     2- EPSI; 3- NSCT; 4- LP & MP.
C
C     READ POLES OF 4TH ORDER FUNCTION
      READ(5,83)P1,Q1,P2,Q2,EPSI,NSCT
C     READ DESIRED SECTION ORDERING FOR 2 SECTIONS(I.E.,3D OR 4TH ORDER)
      READ(5,80)LP,MP
      WRITE(6,84)P1,Q1,P2,Q2,EPSI,NSCT,LP,MP
      LZ=1
      NC=1
C     RIPPLE CALCULATION
      ATT=1./DSQRT(1.+EPSI**2)
      GN1=0.0
      GN2=0.0
      GN3=0.0
      GN4=0.0
      GN5=0.0
      GN6=0.0
      ERMX=10000.0
      JR=0
      JQ=0
      JP=0
      KP=0
      NOUS=0
      JINTR=0
    7 CONTINUE
C     ASSIGNING POLE LOCATION FOR 2 SECTIONS,3 SECTIONS, ETC.
```

```
      PP(MP)=P1
      QQ(MP)=Q1
      PP(LP)=P2
      QQ(LP)=Q2
      GO TO 100
  100 CONTINUE
C     OBTAINING THE CORRECT SECTION ORDERING FOR PRINT OUT
      LE(1)=MP
      LE(2)=LP
      LE(3)=KP
      LE(4)=JP
      LE(5)=JQ
      LE(6)=JR
      DO 159 J=1,6
      KE(J)=0
  159 KN(J)=0
      JA=1
      DO 161 J=1,6
      IF(LE(J).LT.KN(JA))GO TO 161
      KN(JA)=LE(J)
      KE(JA)=J
  161 CONTINUE
      JA=1
      JB=2
  162 DO 163 J=1,6
      IF(LE(J).GE.KN(JA).OR.LE(J).LT.KN(JB))GO TO 163
      KN(JB)=LE(J)
      KE(JB)=J
  163 CONTINUE
      JA=JA+1
      JB=JB+1
      IF(JB.LE.NSCT)GO TO 162
      WRITE(6,81)(KE(J),J=1,6)
      JN=1
      NODD=0
      KS(1)=3
      KS(2)=3
      KS(3)=3
      KS(4)=3
      KS(5)=3
      KS(6)=3
C     CALCULATING MULTIPLIER VALUES
      G=ATT
      IF(QQ(1).EQ.0.0)GO TO 111
      U1=1.
      F1=2.
      D1=1.
      E1=2.*PP(1)
      W1=PP(1)**2+QQ(1)**2
      GN1=ATT*W1/(1.+W1+E1)
      X1=2.*(W1-1.)/(1.+W1+E1)
      Y1=(1.+W1-E1)/(1.+W1+E1)
      B1(NSCT)=X1
      B2(NSCT)=Y1
      G=G*(1.+X1+Y1)/4.
      GO TO 112
  111 X1=1.
      Y1=(PP(1)-1.)/(PP(1)+1.)
      D1=0.0
      U1=0.0
      F1=1.
      B1(NSCT)=Y1
      G=G*(1.+Y1)/2.
      KS(1)=2
  112 CONTINUE
      NF=1
      IF(NSCT.EQ.NF)GO TO 20
      IF(QQ(2).EQ.0.0)GO TO 113
      U2=1.
      F2=2.
      D2=1.
      E2=2.*PP(2)
      W2=PP(2)**2+QQ(2)**2
      GN2=W2/(1.+W2+E2)
      X2=2.*(W2-1.)/(1.+W2+E2)
      Y2=(1.+W2-E2)/(1.+W2+E2)
      B1(NSCT-NF)=X2
      B2(NSCT-NF)=Y2
      G=G*(1.+X2+Y2)/4.
      GO TO 114
  113 X2=1.
      Y2=(PP(2)-1.)/(PP(2)+1.)
      D2=0.0
      U2=0.0
      F2=1.
      B1(NSCT-NF)=Y2
      G=G*(1.+Y2)/2.
      KS(2)=2
  114 CONTINUE
      NF=2
      IF(NSCT.EQ.NF)GO TO 20
   20 CONTINUE
C     OBTAIN THE SCALING MULTIPLIERS
      CALL SCALE(NSCT,B1,B2,G,S,GN,QQ)
      DO 10 J=1,15
   10 A(J)=0.0
      IF(NSCT.LT.1)GO TO 40
C     FOR FIRST 2ND ORDER SECTION CLOSEST TO OUTPUT
      IF(QQ(1).EQ.0.0)GO TO 123
      N=2
      GO TO 124
  123 N=1
      NODD=1
  124 CONTINUE
      VAT=0.0
      IN=15
      IF(QQ(1).EQ.0.0)GO TO 140
      A(N-1)=D1
  140 A(N)=X1
      A(N+1)=Y1
      DO 29 J=1,15
   29 YF(J)=A(J)
      GNT(1)=GN
    1 CONTINUE
      GT=GNT(JN)
      IF(QQ(1).EQ.0.0)GO TO 146
      B(N-1)=U1*GT
  146 B(N)=F1*GT
      B(N+1)=1.*GT
      LN=1
      RC=1.
      CALL SALOSS(A,B,N,IERR,V,IN)
      RC=2.
      IF(IERR.EQ.0)GO TO 50
      IF(JN.EQ.1)GO TO 30
      V=V/12.
      WRITE(6,76)V
      GO TO 40
   30 CONTINUE
      VA(1)=KS(1)*V
      VAT=VAT+VA(1)
      IF(NSCT.LT.2)GO TO 40
C     NOW FOR CASCADE OF TWO SECOND ORDER SECTIONS
      IF(QQ(2).EQ.0.0)GO TO 125
      IF(NODD.EQ.1)GO TO 125
      N=4
      GO TO 126
  125 N=3
      NODD=1
  126 CONTINUE
      R1=D1*X2+X1*D2
      R2=D1*Y2+D2*Y1+X1*X2
      R3=X1*Y2+X2*Y1
      R4=Y1*Y2
      IF(QQ(2).EQ.0.0)GO TO 141
      IF(NODD.EQ.1)GO TO 141
      A(N-3)=D1*D2
  141 A(N-2)=R1
      A(N-1)=R2
      A(N)=R3
      A(N+1)=R4
      DO 31 J=1,15
   31 YF(J)=A(J)
      GNT(1)=GN*S(NSCT)
    2 CONTINUE
      C1=U1*U2
      C2=U1*F2+F1*U2
      C3=U1+F1*F2+U2
      C4=F1+F2
      GT=GNT(JN)
      IF(QQ(2).EQ.0.0)GO TO 147
      IF(NODD.EQ.1)GO TO 147
      B(N-3)=C1*GT
  147 B(N-2)=C2*GT
      B(N-1)=C3*GT
      B(N)=C4*GT
      B(N+1)=1.*GT
      LN=2
      RC=3.
      CALL SALOSS(A,B,N,IERR,V,IN)
      RC=4.
      IF(IERR.EQ.0)GO TO 50
      IF(JN.EQ.1)GO TO 32
      V=V/12.
      WRITE(6,76)V
      GO TO 40
   32 CONTINUE
      VA(2)=KS(2)*V
      VAT=VAT+VA(2)
      IF(NSCT.LT.3)GO TO 40
   40 CONTINUE
      IF(JN.EQ.2)GO TO 45
      DO 42 J=1,15
   42 A(J)=YF(J)
C     NOW OBTAINING THE A/D NOISE VARIANCE
      JN=2
      GNT(2)=GNT(1)*S(1)
      IF(LN.EQ.1)GO TO 1
      IF(LN.EQ.2)GO TO 2
   45 CONTINUE
C     ADD ONE TO VAT TO ACCOUNT FOR OUTPUT MULTIPLIER IN SECTION CLOSE TO
C     OUTPUT
      VAT=VAT+1.
      VAT=VAT/12.
      WRITE(6,72)VAT
      IF(VAT.GE.ERMX)GO TO 175
      ERMX=VAT
      MJ1=KE(1)
      MJ2=KE(2)
      MJ3=KE(3)
      MJ4=KE(4)
      MJ5=KE(5)
      MJ6=KE(6)
  175 CONTINUE
  110 CONTINUE
      WRITE(6,82)ERMX,MJ1,MJ2,MJ3,MJ4,MJ5,MJ6
      ERMX=10000.0
      NC=1
   50 CONTINUE
      WRITE(6,70)IERR,N
   51 CONTINUE
   70 FORMAT(1X,'***ERROR-PCLES OUTSICE UNIT CIRCLE.IERR=',I1,' AT STAGE
     1 N=',I2,/)
   71 FORMAT(1X,'OUTPUT NOISE VARIANCE V(',I1,')=',F16.7,'*Q**2/12')
   72 FORMAT(1X,'TOTAL OUTPUT NOISE VARIANCE =',F16.7,'*Q**2')
   73 FORMAT(1X,'RC = ',F4.2)
   74 FORMAT(1X,'VALUE OF GN(',I1,')=',F16.7,'    & GNT=',F16.7)
   75 FORMAT(1X,'TOTAL OUTPUT NOISE VARIANCE =',F16.7,'*Q**2/12')
   76 FORMAT(1X,' OUTPUT NOISE VARIANCE DUE TO A/D =',F16.7,'*Q**2')
   77 FORMAT(1X,'INITIAL OUTPUT GAIN =',F16.7)
   80 FORMAT(6I2)
   81 FORMAT(/,6I2)
   82 FORMAT(/,10X,'***MINIMUM OUTPUT NOISE VARIANCE=',F16.7,'*Q**2
     1AT SECTION ORDERING',6I2,//)
   83 FORMAT(5F10.7,I2)
   84 FORMAT(1X,'P1=',F10.7,'Q1=',F10.7,'P2=',F10.7,'Q2=',F10.7,'EPSI=',
     1F10.7,'NSCT=',I2,'LP=',I2,'MP=',I2)
      STOP
      END
      SUBROUTINE SALCSS(A,B,N,IERR,V,IN)
C
C     PROGRAM FOR EVALUATING THE INTEGRAL OF THE RATIONAL FUNCTION
C       1/(2*PI*I)*B(Z)*B(1/Z)/(A(Z)*A(1/Z)*Z)
C     AROUND THE UNIT CIRCLE
C     REFERENCE: ASTROM,JURY,&AGNIEL 'A NUMERICAL METHOD FOR THE EVALUATION
C     OF COMPLEX INTEGRALS',IEEE TRANS ON AUTOMATIC CONTROL,AUG 1970,PP468-471
C     A- VECTOR WITH THE COEFFICIENTS OF THE POLYNOMIAL
C       A(1)*Z**N+A(2)*Z**(N-1)+...+A(N+1)
C     IT IS ASSUMED THAT A(1) IS GREATER THAN ZERO
C     B- VECTOR WITH THE COEFFICIENTS OF THE POLYNOMIAL
C       B(1)*Z**N+B(2)*Z**(N-1)+...+B(N+1)
C
C       THE VECTORS A AND B ARE DESTROYED
C
C     N- ORDER OF THE POLYNOMIALS A AND B (MAX 10)
C     IERR- WHEN RETURNING IERR = 1 IF A HAS ALL ZEROS INSIDE UNIT CIRCLE
C       IERR= 0 IF THE POLYNOMIAL A HAS ANY ROOT OUTSIDE OR ON
C       THE UNIT CIRCLE OR IF A(1) IS NOT POSITIVE
C     V- THE RETURNED LOSS I.E RETURNED VALUE OF THE COMPLEX INTEGRAL
C     IN- DIMENSION OF A AND B IN MAIN PROGRAM
C
C     SUBROUTINES REQUIRED: NONE
C
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION A(IN),B(IN),AS(12)
C
C     CRUDE STABILITY TEST
      NP=N+1
      IF(A(1))50,50,1
    1 R=A(1)
      DO 2 I=1,N
    2 R=R+A(I+1)
      IF(R)50,50,3
    3 R=A(1)
      N1=1
      DO 4 I=1,N
      N1=-N1
    4 R=-R+A(I+1)
      IF(N1)5,5,6
    5 R=-R
    6 CONTINUE
      IF(R)50,50,7
C
C     BEGIN MAIN LOOP
C
    7 A0=A(1)
      IERR=1
      V=0.0
      DO 10 K=1,N
      L=N+1-K
      L1=L+1
      ALFA=A(L1)/A(1)
      BETA=B(L1)/A(1)
      V=V+BETA*B(L1)
      DO 20 I=1,L
      M=L+2-I
      AS(I)=A(I)-ALFA*A(M)
```

```
20 B(I)=B(I)-BETA*A(M)
   IF(AS(1))50,50,30
30 DO 40 I=1,L
40 A(I)=AS(I)
10 CONTINUE
   V=V+B(1)**2/A(1)
   V=V/A0
   RETURN
50 IERR=0
70 FORMAT(1X,'A(I)=',10F12.6)
71 FORMAT(1X,'B(I)=',10F12.6)
72 FORMAT(/1X,'R=',F12.6)
73 FORMAT(/1X,'AS(1)=',F16.9)
   RETURN
   END
```

```
      SUBROUTINE SCALE(NSCT,B1,B2,G,S,GN,QQ)
      IMPLICIT REAL*8(A-H,O-Z)
      DOUBLE PRECISION DABS
      DIMENSION B1(10),B2(10),YF(400),S(10),QQ(7)
C     THIS SUBROUTINE CALCULATES THE SCALING FACTORS FOR THE CASCADE DIGITAL
C     FILTER SECTIONS SUCH THAT V(NT).LE.1 ,& Y(NT).LE.1
C     NSCT = NUMBER OF FILTER SECTIONS
C     B1 & B2 = INNER & OUTER FEEDBACK MULTIPLIERS
C     G= FILTER GAIN IT SETS H(Z)=1 AT Z=1
C     GN= NEW RETURNED FILTER GAIN = G/S(K)
      S(1)=1.
      DO 10 J=1,75
   10 YF(J)=0.0
      YF(1)=1.
      K=1
      JR=NSCT
   11 CONTINUE
      IF(JR.EQ.0)GO TO 58
      IF(K.GT.2)GO TO 15
      KM=1
      GO TO 20
   15 KM=K-1
   20 CONTINUE
      IF(QQ(JR).EQ.0.0)GO TO 51
      V1=0.0
      V2=0.0
      SUMV=0.0
      SUMY=0.0
      DO 40 LS=1,75
      LL=LS
      V=S(KM)*YF(LL)-B1(K)*V1-B2(K)*V2
      YF(LL)=V+2.*V1+V2
      SUMV=SUMV+DABS(V)
      SUMY=SUMY+DABS(YF(LL))
      V2=V1
      V1=V
   40 CONTINUE
      IF(SUMY.GT.SUMV)GO TC 45
      S(K)=1./SUMV
      GO TO 46
   45 S(K)=1./SUMY
   46 CONTINUE
   50 CONTINUE
      K=K+1
      JR=JR-1
      GO TO 11
   51 CONTINUE
      SUMV=0.0
      SUMY=0.0
      V1=0.0
      DO 53 LL=1,75
      V=S(KM)*YF(LL)-B1(K)*V1
      YF(LL)=V+V1
      SUMV=SUMV+DABS(V)
      SUMY=SUMY+DABS(YF(LL))
      V1=V
   53 CONTINUE
      IF(SUMY.GT.SUMV)GO TO 56
   S(K)=1./SUMV
   GO TO 57
56 S(K)=1./SUMY
57 CONTINUE
   K=K+1
   JR=JR-1
   GO TO 11
58 CONTINUE
   GN=G
   DO 55 K=1,NSCT
55 GN=GN/S(K)
70 FORMAT(/1X,'   SUMV(',I1,')=',F16.7,4X,' & SUMY(',I1,')=',F16.7)
71 FORMAT(1X,'SCALING FACTOR S(',I1,')=',F16.7)
72 FORMAT(1X,'ADJUSTED OUTPUT GAIN=',F16.7)
   RETURN
   END
```

VITA

Khalil Elia Massad

Candidate for the Degree of

Doctor of Philosophy

Thesis: MODIFIED BUTTERWORTH AND CHEBYSHEV FUNCTIONS: DIGITAL FILTER ROUNDOFF NOISE AND BIT REQUIREMENTS

Major Field: Electrical Engineering

Biographical:

Personal Data: Born in Beirut, Lebanon, October 20, 1946, the son of Mr. and Mrs. Elia K. Massad.

Education: Received the high school diploma (Lebanese Baccalaureate Part I) and the Lebanese Baccalaureate Part II (scientific) from National College Choueifat, Choueifat, Lebanon, in June, 1965, and June, 1966; received the Bachelor of Engineering degree in Electrical Engineering from the American University of Beirut, Beirut, Lebanon, June, 1970; received the Master of Science degree from Oklahoma State University, December, 1972, with a major in Electrical Engineering; completed requirements for the Doctor of Philosophy degree at Oklahoma State University with a major in Electrical Engineering, July, 1975.

Professional Experience: Engineer in Training, research center of Electricité De France, Clamart, France, summer of 1969; Maintenance Engineer, Faat Engineering International, Beirut, Lebanon, from July, 1970 to August, 1971; Teaching Assistant, Electrical Engineering, Oklahoma State University, 1972-1975; Research Assistant, Electrical Engineering, Oklahoma State University, 1973-1974; Electrical Engineer/Scientist, Collins Radio, Richardson, Texas, summer of 1974.

Professional Organizations: Member of Institute of Electrical and Electronics Engineers.