ROUNDOFF NOISE IN CASCADE REALIZATION OF FINITE IMPULSE RESPONSE DIGITAL FILTERS

by

David So Keung Chan

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREES OF

BACHELOR OF SCIENCE

and

MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1972

Signature of Author
Department of Electrical Engineering, August 14, 1972

Certified by
Thesis Supervisor (Academic)

Certified by
Thesis Supervisor (VI-A Cooperating Company)

Accepted by
Chairman, Departmental Committee on Graduate Students

ROUNDOFF NOISE IN CASCADE REALIZATION OF FINITE IMPULSE RESPONSE DIGITAL FILTERS

by

David So Keung Chan

Submitted to the Department of Electrical Engineering on August 14, 1972 in partial fulfillment of the requirements for the Degrees of Bachelor of Science and Master of Science

ABSTRACT

The implementation of digital filters using finite precision or fixed point arithmetic introduces several quantization problems, among which is roundoff noise. An investigation of the behavior of roundoff noise in cascade realizations of finite impulse response digital filters as a function of filter parameters as well as section ordering is carried out, with both theoretical bases as well as experimental results. It is shown that most orderings of a filter have relatively low noise. Linear phase filters are used as examples throughout, and a unified, rigorous treatment of the theory of these filters is provided. Furthermore, several methods for the scaling of cascade filters to meet dynamic range constraints are rigorously summarized, and two of these methods are treated in depth, including a comparison between their effects on roundoff noise.

Arguments are presented which demonstrate a correlation between roundoff noise and a parameter which is defined in terms of the peakedness of certain subfilter spectra. This result enables one to judge by inspection the relative merit of a filter ordering. Experimental evidence is presented to support this result.

Based on a simple procedure proposed by others, an algorithm which finds, for a cascade filter, an ordering which has very low noise is developed. Application of this algorithm to over 50 filters has in every case shown excellent results. For practical cascade filter orders of interest, viz. up to 50, the algorithm requires less than 20 seconds on the Honeywell 6070 computer. A filter of 128^{th} order has been ordered using this algorithm, yielding an ordering with rms noise of approximately 4Q to 6Q (Q = quantization step size), depending on the type of scaling performed, compared to a potentially possible value of over 10^{27}Q.

THESIS SUPERVISOR: Alan V. Oppenheim
TITLE: Associate Professor of Electrical Engineering

THESIS SUPERVISOR: Lawrence R. Rabiner
TITLE: Member of the Technical Staff at the Bell Telephone Laboratories

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my advisor at Bell Laboratories, Dr. Lawrence R. Rabiner, for his guidance, encouragement, and helpful suggestions. Most of all, his enthusiasm in my research work is greatly appreciated. I am also grateful to Dr. Ronald W. Schafer of Bell Laboratories for the stimulating discussions I have had with him, and to Professor Alan V. Oppenheim for agreeing to be my academic supervisor.

This thesis research was supported by Bell Laboratories under a cooperative program with M.I.T. and I am grateful for the use of their facilities. My special thanks go to Mrs. B. A. MaSaitis for a very able job of typing this thesis. Finally, I am deeply indebted to my parents, Mr. and Mrs. K. Chan, for their continuing aid, understanding, and encouragement.

TABLE OF CONTENTS

		Page
TITLE PAGE		1
ABSTRACT		2
ACKNOWLEDGEMENTS		4
TABLE OF CONTENTS		5
LIST OF FIGURES AND TABLES		7
SECTIONS		
1.0	Introduction	10
	1.1 Quantization Effects in Practical Filters	14
	1.2 Contributions of this Thesis Research	17
2.0	Linear Phase FIR Filters	21
	2.1 Criteria for Linear Phase	23
	2.2 Design Techniques and Realization Structures	35
3.0	Techniques for Roundoff Error Analysis	48
	3.1 Statistical Model for Roundoff Errors	50
	3.2 Roundoff Noise in Cascade Form FIR Filters	52
	3.3 Dynamic Range Constraints in the Cascade Form	58
4.0	Behavior of Roundoff Noise in Cascade Filters	79
	4.1 Dependence of Roundoff Noise on Section Ordering	85
	4.2 Comparison of Sum Scaling and Peak Scaling	115
	4.3 Dependence of Roundoff Noise on Other Filter Parameters	135

Page

4.4 Spectral Peaking and Roundoff Noise 144

5.0 An Algorithm for Obtaining a Low Noise Ordering for a Cascade Filter 164

6.0 Conclusions 186

BIBLIOGRAPHY 188

APPENDICES

A.1 - List of Orderings and Noise Variances of Filter of Fig. 4.6 193

A.2 - List of Orderings and Noise Variances of Filter of Fig. 4.12 195

A.3 - List of Orderings, Noise Variances, and Parameter Pk for Filter no. 14 197

A.4 - Comparison of Orderings and their Reverse for Filter no. 14 200

B.1 - Plots of Subfilter Spectra for a High Noise Ordering of Filter no. 14 203

B.2 - Plots of Subfilter Spectra for a Low Noise Ordering of Filter no. 14 210

LIST OF FIGURES AND TABLES

Figure	Page
2.1 - Direct Form Linear Phase Filter	43
2.2 - Cascade Form Filter Section	46
2.3 - Alternate Cascade Filter Section	46
2.4 - General Cascade Filter Section	47
3.1 - Model for Practical Cascade Filter	54
3.2 - Alternate Equivalent Model for Cascade Filter	54
4.1 - Definition of Filter Parameters	86
4.2 - Transpose Configuration of Fig. 2.2	91
4.3 - Alternate Configuration to Fig. 2.2	91
4.4 - Alternate Configuration to Fig. 2.3	92
4.5 - Flow Chart of Scaling and Noise Calculation Subroutine	94
4.6(a) - Impulse Response and Step Response of Typical 13-Point Filter	103
4.6(b) - Magnitude Frequency Response of Typical 13-Point Filter	104
4.7 - Zeros of Filter of Fig. 4.6	105
4.8 - Noise Distribution Histogram of Filter of Fig. 4.6	107
4.9 - Cumulative Noise Distribution of Filter of Fig. 4.6	108
4.10 - Detailed Noise Distribution Histogram of Filter of Fig. 4.6	110
4.11 - Noise Distribution Histogram of Typical 11-Point Filter	112

		Page
4.12 -	Noise Distribution Histogram of Another 13-Point Filter	113
4.13 -	Noise Distribution Histogram of a Third 13-Point Filter	114
4.14 -	Noise Distribution Histogram of 15-Point Filter Example 1	116
4.15 -	Noise Distribution Histogram of 15-Point Filter Example 2	117
4.16 -	Noise Distribution Histogram of 15-Point Filter Example 3	118
4.17 -	Peak Scaling Versus Sum Scaling Noise Output Comparison for Typical Filter	129
4.18 -	Noise Distribution Histogram of Filter of Fig. 4.17 Using Sum Scaling	131
4.19 -	Noise Distribution Histogram of Filter of Fig. 4.17 Using Peak Scaling	132
4.20 -	Peak Scaling Versus Sum Scaling Noise Output Comparison for Filter of Fig. 4.6	133
4.21 -	Noise Distribution Histogram of Filter of Fig. 4.6 Using Peak Scaling	134
4.22 -	Output Noise Variance as a Function of Filter Length	138
4.23 -	Output Noise Variance as a Function of Bandwidth	140
4.24 -	Output Noise Variance as a Function of Passband Approximation Error	141
4.25 -	Output Noise Variance as a Function of Stopband Approximation Error	142
4.26 -	Values of Parameter C for Different Filter Lengths	150
4.27(a) -	Zeros of Typical 33-Point Filter	153

Page

4.27(b) - Zeros of Typical 67-Point Filter 154

4.28 - Zeros of Filter of Fig. 4.12 158

4.29 - Zeros of Filter #14 of Table 4.1 161

5.1 - Flow Chart of Ordering Algorithm 171

5.2 - Computation Time Versus Filter Length for Ordering Algorithm 179

Table

4.1 - List of Filters and Their Noise Distribution Statistics 137

4.2 - Tabulation of Filters Used For Fig. 4.26 151

5.1 - List of Filters and the Results of Ordering Algorithm 174

5.2 - List of Filters and the Results of Ordering Algorithm 175

5.3 - List of Filters and the Results of Alg. 1' 184

1.0 Introduction

Within the past decade, great advances have been made in the field of digital filtering. Many efficient techniques have been developed for the design of filter transfer functions with specified frequency response characteristics. Digital filters have been successfully employed and found to be indispensable in many signal processing tasks, such as speech and picture processing and analysis. The advantages of digital systems over functionally-equivalent analog systems are clear - high reliability, arbitrarily high accuracy, stable and easily alterable parameter values, and straight-forward realization.

Arbitrarily high accuracy, however, is possible in any digital system only at the expense of arbitrarily long wordlength used to represent data. Clearly, increased wordlength implies increased system complexity and cost. Because special-purpose digital systems dedicated to the task of filtering have become feasible through the use of large-scale integrated circuits, system wordlength has become an important design parameter. Therefore, given a specified filtering task that a system is to perform, it is desirable to minimize the accuracy or wordlength required of that system, in order that size and cost may be held to a minimum.

While excellent filter transfer functions designed on the basis of infinite-precision arithmetic are readily available, it is as yet unclear what types of algorithms are most efficient for implementing any particular filter transfer function using finite-precision arithmetic. In fact the approximation (design of transfer function) and implementation phases of digital filter design are not really independent since given a filter wordlength and algorithm or configuration, one may find a better solution for a transfer function than that obtainable by quantizing the result of an "infinite precision" transfer function. In any event, in order to solve the implementation problem, it is necessary to understand as much as possible the effects, commonly referred to as quantization effects, that finite word-length has on the behavior of a practical filter. In this report we shall consider the behavior of one type of quantization effect in one type of practical digital filter.

Digital filters can be divided into two fundamental classes - those with impulse responses of infinite duration and those with impulse responses of finite duration. We shall refer to the former class as

"Infinite Impulse Response" (IIR) filters and to the latter class as "Finite Impulse Response" (FIR) filters. In the frequency domain these two classes are distinguished by the fact that transfer functions of IIR filters are rational functions in z^{-1}, hence are represented by both poles and zeros in the z-plane, whereas transfer functions of FIR filters are polynomial functions of z^{-1}, represented by zeros only in the finite z-plane. The reader may recognize that IIR and FIR filters are also referred to in the literature as "Recursive" and "Nonrecursive" filters. However, since both IIR and FIR filters can be realized recursively as well as nonrecursively[15] we shall reserve the terms "Recursive" and "Nonrecursive" to describe types of realizations of filters.

The study of IIR digital filters has had a longer history mainly because of the generality of IIR filters and the close resemblance between the form of their transfer function and that of traditional analog filters. By simple algebraic transformations it is possible to convert transfer functions of analog filters into transfer functions for IIR filters while preserving frequency response or time response characteristics of interest. In this way design specifications for a digital filter can be phrased in analog filter terms, so that the great body of knowledge already existing for continuous filter design can be put to advantage in digital filter design. However, IIR filters have some important inherent short-comings. First of all, limit cycles can occur in IIR filters, causing non-zero output even with zero input. Secondly, quantization of the coefficients of a stable IIR filter can lead to an unstable filter. Finally, IIR filters with stringent specifications on their magnitude frequency response have highly nonlinear phase response characteristics.

In order to find solutions to these problems where they are significant, and also as a result of important developments in discrete optimization theory, a great deal of interest has been turned to FIR filters in recent years. FIR filters have several important advantages over IIR filters. Most notable among these are the following:

1) With proper constraints on their coefficients FIR filters can be easily made to have exactly linear phase response. These filters can then be used to approximate any arbitrary magnitude frequency response.

2) When realized nonrecursively FIR filters are always stable, thus quantization of coefficients cannot lead to instability. Furthermore, limit cycles cannot occur in nonrecursive FIR filters.

In this report we shall restrict all experimental investigations to FIR filters with linear phase. However, most results obtained can be easily generalized for FIR filters in general. Sections 2.0 to 2.2 will present a unified discussion of linear phase filters. But meanwhile, we first consider quantization effects in general and then define our research problem area.

1.1 Quantization Effects in Practical Filters

A practical digital filter (i.e. one realized with finite precision arithmetic) introduces several quantization effects that are unexplained by a simple theoretical transfer function. These may be classified into three basic categories:

1) Quantization of the values of samples derived from a continuous input waveform causes inaccuracies in the representation of the waveform.
2) Finite precision representation of the filter coefficients alters the frequency response characteristics of the filter and may cause a stable filter to become unstable.
3) Finite precision arithmetic causes inaccuracies in the filter output, which, together with the finite dynamic range of the filter, limits the signal-to-noise ratio attainable.

The first type of quantization effect, commonly known as A-D (analog-to-digital) noise, is independent of the method of transfer function implementation, and can be easily analyzed. It will be shown to be negligible compared to roundoff errors in most filters realized in the cascade form of interest. The second type of quantization effect, known as the coefficient sensitivity problem, does depend in degree and character on the type of structure used to implement a filter. Much effort has been given to studying the nature of this effect in IIR filters[7-10,24]. However, though some of the findings on IIR filters can be specialized to FIR filters, others are not applicable. Herrmann and Schüssler[6] have provided some insights into the sensitivity of coefficients in the linear phase FIR cascade structure, but more work needs to be done before the full implications are clear.

In this report we will consider only the third type of quantization effect. Errors in this category are introduced into a filter by the quantization of results of arithmetic operations within the filter. The exact nature of these errors depends on the "mode" of arithmetic employed (fixed-point or floating-point), as well as the type of quantization used (rounding or truncation). For the same quantization step size truncation leads to a larger error variance than rounding. Therefore, in general rounding is preferred.

Extensive studies have been made on the statistical properties of roundoff errors in both floating-point and fixed-point arithmetic[27-30]. With regard to quantization errors the major difference between these two modes of arithmetic is that given the wordlength used, in the former case the maximum possible error committed when quantizing the result of a multiplication depends on the magnitude of the result whereas in the latter case it is independent of the data magnitude. Also, addition introduces quantization errors in floating-point arithmetic but not in fixed-point arithmetic.

Although, for a given wordlength, floating-point arithmetic generally results in less error than fixed-point arithmetic, for reasons of economy fixed-point arithmetic is generally employed in special-purpose digital equipment. In this report our analyses will be based upon the assumption of fixed-point arithmetic with rounding. However, the results obtained are essentially independent of the mode of arithmetic employed as well as the type of quantization performed. Only the formulas for the calculation of noise variances are different among the different cases.

1.2 Contributions of this Thesis Research

The major contributions of this thesis to the understanding of FIR digital filters can be summarized as follows:

1). Sections 2.0-2.1: The theory of linear phase filters is presented in mathematical rigor.

2). Section 3.2: A noise figure in number of bits is defined which relates an upper bound on the noise output magnitude of a filter to the number of bits required to represent noise so as to free signal bits from noise.

3). Section 3.3: A thorough, rigorous treatment of known scaling methods to meet dynamic range constraints in cascade filters is presented and optimal scaling methods are defined and proved for two classes of input signals to a filter.

4). Section 4.1: Considering all possible orderings of sections of relatively low order cascade filters, the distribution of output noise variance values over their range of possible occurrence was determined for a large number of filters. The shape of the distributions is found to be essentially independent of the characteristics of the transfer function of a filter. In particular, most orderings of a filter are found to have relatively very low noise. Also, certain orderings equivalent in terms of output noise are determined and their equivalence proved.

5). Section 4.2: A comparison of the output noise variances of a filter determined by two different scaling methods is presented. Analytical reasoning shows the results of the two methods to be very comparable, at least in order of magnitude. Experimental results then show that for almost all orderings the variances are approximately in a constant ratio independent of ordering for the filter.

6). Section 4.3: Experimental results are provided to show that roundoff noise tends to increase with all four parameters, viz. filter length, bandwidth, passband and stopband approximation errors, which characterize a filter transfer function. In particular, noise tends to increase exponentially with filter length.

7). Section 4.4: A correlation is established heuristically and experimentally between output noise level and a parameter defined in terms of the amount of peaking of certain subfilter spectra. The result enables one to judge by inspection the relative merit of an ordering for a filter. It also explains to a degree why most orderings for a filter have low noise and helps the designer in the absense of an ordering algorithm to sensibly choose, with minimal effort, a good ordering for a filter.

8). Section 5.0: A completely automatic machine algorithm is presented which finds, for a cascade filter, an ordering with very low noise. This algorithm is developed based on a simple procedure proposed by Avenhaus[16]. For practical cascade filter orders of interest, viz. up to 50, the algorithm requires less than 20 seconds of computation time on the Honeywell 6070 computer. Application of this algorithm to over 50 filters has, in every case, shown excellent results. Typically the resulting filters have only 2 to 3 bits of noise. A typical 128^{th} order (129 point) filter has been ordered by this algorithm, yielding an ordering with rms noise of approximately 4Q to 6Q (Q = quantization step size), depending on the type of scaling performed, or ~3 bits, compared to a potentially possible value of over 10^{27}Q, or 91 bits.

Because of the rigorous nature of the presentations in sections 2.1 and 3.3, the major results are stated in theorems followed by their proofs. Thus, the reader not interested in rigorous proofs may, without loss of continuity, read only the statements of the theorems. In fact, for an understanding of the discussions of sections 4.1 and higher, the reader need only be familiar with the definitions and some of the theorem statements of earlier sections.

2.0 Linear Phase FIR Filters

The general transfer function for an N point FIR filter can be written in the form

$$H(z) = \sum_{k=0}^{N-1} h(k)z^{-k} \tag{2.1}$$

where the real-valued sequence $\{h(k),\ k = 0,\ldots,N-1\}$ is the impulse response of the filter. Alternatively, H(z) can be expressed in the factored form

$$H(z) = \prod_{i=1}^{Ns} (b_{oi} + b_{1i}z^{-1} + b_{2i}z^{-2}) \tag{2.2}$$

where b_{ji}, $j = 0,1,2$, $i = 1,\ldots,Ns$ are real numbers and Ns, the number of factors, is defined as

$$Ns = \begin{cases} \frac{N-1}{2} & N \text{ odd} \\ \frac{N}{2} & N \text{ even} \end{cases}$$

and $b_{2Ns} = 0$ if N is even.

We shall define a linear phase filter to be a filter whose transfer function H(z) is expressible in the form

$$H(z)\Big|_{z=e^{j\omega}} = H(e^{j\omega}) = \pm\ |H(e^{j\omega})|e^{-j\alpha\omega} \qquad (2.3)$$

where α is a real positive constant with the physical significance of delay in number of samples. The factor $\pm$ is necessary since $H(e^{j\omega})$ actually is of the form

$$H(e^{j\omega}) = H^*(e^{j\omega})e^{-j\alpha\omega}$$

where $H^*(e^{j\omega})$ is a real function taking on both positive and negative values. We will also find it useful to define a mirror-image polynomial (MIP) of degree N to be a polynomial of the form $\sum_{k=0}^{N} a_k z^k$ whose coefficients satisfy the relation

$$a_k = a_{N-k} \qquad 0 \leq k \leq N$$

In the next section we shall derive necessary and sufficient conditions on the coefficients of H(z) such that a filter with transfer function H(z) will have an exactly linear phase response.

2.1 Criteria for Linear Phase

The conditions on $H(z)$ which are necessary and sufficient for linear phase are summarized in the statement of Theorem 2.1 below. All notations are as they were defined in the previous section. A rigorous proof of the theorem is provided. However, the reader who is either already familiar with the results of Theorem 2.1 or is not interested in a rigorous proof may skip to the end of the proof on page 34 without loss of continuity.

<u>Theorem 2.1</u>: $H(z)$ can be expressed in the form (2.3) if and only if one of the following equivalent conditions hold:

(a) $h(k) = h(N-1-k) \quad 0 \le k \le N-1$

(b) If z_i is a zero of $H(z)$, then z_i^{-1} is also a zero of $H(z)$. Also if $z_i = +1$ is a zero of $H(z)$ then it occurs in even multiplicity.

(c) Suppose z_i is a zero of the i^{th} factor in (2.2). Let $S = \{i: \ z_i \text{ is real}\}$ and $Q = \{i: \ i \notin S\}$. Then $f(z) = \prod_{i \varepsilon S} (b_{oi} + b_{1i}z^{-1} + b_{2i}z^{-2})$ is a mirror-image polynomial in z^{-1}, and for all $i \varepsilon Q$, either $b_{oi} = b_{2i}$ or there exists $j \neq i$, $j \varepsilon Q$, such that

$$\frac{b_{oi}}{b_{2j}} = \frac{b_{1i}}{b_{1j}} = \frac{b_{2i}}{b_{oj}}$$

Furthermore, the following is a sufficient condition for H(z) to be expressible in the form (2.3):

(d) In (2.2), for $1 \leq i \leq Ns$, either $b_{2i} = 0$ and $b_{oi} = b_{1i}$, or $b_{oi} = b_{2i}$, or there exists $j \neq i$, $1 \leq j \leq Ns$, such that

$$\frac{b_{oi}}{b_{2j}} = \frac{b_{1i}}{b_{1j}} = \frac{b_{2i}}{b_{oj}}$$

In all cases the value of α is $\alpha = \frac{N-1}{2}$.

Before proceeding with the proof of Theorem 2.1 we will need the following result on mirror-image polynomials.

Lemma: The product of mirror-image polynomials is a mirror-image polynomial.

Proof: We first show that f(z) is an MIP of degree N iff $f(z) = z^N f(z^{-1})$. To see this, let

$$f(z) = \sum_{k=0}^{N} a_k z^k$$

Then

$$z^N f(z^{-1}) = \sum_{k=0}^{N} a_k z^{N-k}$$

$$= \sum_{k=0}^{N} a_{N-k} z^k$$

Hence $f(z) = z^N f(z^{-1})$ iff $a_k = a_{N-k}$, $0 \leq k \leq N$. Now let f(z) and g(z) be any two MIP's of degrees N and M. Then

$$(f \cdot g)(z) = f(z)g(z)$$

$$= \left[z^N f(z^{-1})\right]\left[z^M g(z^{-1})\right]$$

$$= z^{N+M}(f \cdot g)(z^{-1})$$

Hence $(f \cdot g)$ is an MIP (of degree N+M). By the associativity of polynomial multiplication the lemma is proved.

Q.E.D.

Proof of Theorem 2.1:

We shall prove necessity and sufficiency for condition (a) and then (a) → (b) → (c) → (a), and finally (d) → (a).

Suppose there exists some α such that

$$H(e^{j\omega}) = \pm |H(e^{j\omega})| e^{-j\alpha\omega} \tag{2.4}$$

Then since from (2.1)

$$H(e^{j\omega}) = \sum_{k=0}^{N-1} h(k) e^{-j\omega k} \tag{2.5}$$

we have

$$\sum_{k=0}^{N-1} h(k) e^{-j\omega k} = \pm |H(e^{j\omega})| e^{-j\alpha\omega} \tag{2.6}$$

Equating real and imaginary parts of (2.6), we obtain

$$\sum_{k=0}^{N-1} h(k) \cos k\omega = \pm |H(e^{j\omega})| \cos \alpha\omega \tag{2.7}$$

$$\sum_{k=0}^{N-1} h(k) \sin k\omega = \pm |H(e^{j\omega})| \sin \alpha\omega \tag{2.8}$$

Dividing (2.8) by (2.7) yields

$$\frac{\sin \alpha\omega}{\cos \alpha\omega} = \frac{\sum_{k=0}^{N-1} h(k) \sin k\omega}{\sum_{k=0}^{N-1} h(k) \cos k\omega} \tag{2.9}$$

or

$$\tan \alpha\omega = \frac{\sum_{k=1}^{N-1} h(k)\sin k\omega}{h(o) + \sum_{k=1}^{N-1} h(k)\cos k\omega} \tag{2.10}$$

First, suppose $\alpha = 0$, then we must have $h(k) = 0$ for all $k > 0$, hence $N = 1$ and $H(z) = h(0)$. Clearly, (a) is satisfied with $\alpha = \frac{N-1}{2} = 0$.

Now suppose $\alpha \neq 0$, then we can rewrite (2.9) as

$$\sum_{k=0}^{N-1} h(k) \cos k\omega \sin \alpha\omega - \sum_{k=0}^{N-1} h(k)\sin k\omega \cos \alpha\omega = 0 \tag{2.11}$$

or

$$\sum_{k=0}^{N-1} h(k) \sin (\alpha-k)\omega = 0 \tag{2.12}$$

The only possible solution to (2.12) for all ω is

$$\alpha = \frac{N-1}{2}$$

$$h(k) = h(N-1-k) \quad 0 \le k \le N-1$$

Conversely, suppose condition (a) holds. Furthermore, for the time being suppose N is even. Then (2.5) can be rewritten as

$$H(e^{j\omega}) = \sum_{k=0}^{\frac{N}{2}-1} h(k)e^{-j\omega k} + \sum_{k=\frac{N}{2}}^{N-1} h(k)e^{-j\omega k}$$

$$= \sum_{k=0}^{\frac{N}{2}-1} h(k)e^{-j\omega k} + \sum_{k=0}^{\frac{N}{2}-1} h(k)e^{-j\omega(N-1-k)}$$

$$= \sum_{k=0}^{\frac{N}{2}-1} h(k)e^{-j\omega(\frac{N-1}{2})}\left[e^{j\omega(\frac{N-1}{2} - k)} + e^{-j\omega(\frac{N-1}{2} - k)}\right]$$

$$= \left[\sum_{k=0}^{\frac{N}{2}-1} 2h(k)\ \cos(\frac{N-1}{2} - k)\omega\right]e^{-j(\frac{N-1}{2})\omega} \qquad (2.13)$$

Similarly, if N is odd, we obtain

$$H(e^{j\omega}) = \left[\sum_{k=0}^{\frac{N-3}{2}} 2h(k)\ \cos(\frac{N-1}{2} - k)\omega + h(\frac{N-1}{2})\right]e^{-j(\frac{N-1}{2})\omega} \qquad (2.14)$$

Hence H(z) indeed satisfies (2.3) with $\alpha = \frac{N-1}{2}$.

(a) → (b):

Note from the proof of the lemma that (a) implies

$$H(z) = z^{-(N-1)}H(z^{-1}) \qquad (2.15)$$

Therefore if z_i is a zero of H(z), then z_i^{-1} is also a zero of H(z).

If $z_i = +1$ is a zero of $H(z)$ occurring with odd multiplicity, write $H(z)$ as

$$H(z) = g(z)(1-z^{-1}) \tag{2.16}$$

Now $g(z)$ satisfies condition (b) with $H(z)$ replaced by $g(z)$. In the course of the remainder of this proof it will be shown that condition (b) implies that $H(z)$ is a mirror-image polynomial in z^{-1}. Hence $g(z)$ is an MIP in z^{-1} and can be expressed as

$$g(z) = \sum_{k=0}^{N-2} a_k z^{-k} \tag{2.17}$$

where $a_k = a_{N-2-k}$ $0 \leq k \leq N-2$.

But

$$H(z) = (1-z^{-1}) \sum_{k=0}^{N-2} a_k z^{-k}$$

$$= \sum_{k=0}^{N-2} a_k z^{-k} - \sum_{k=0}^{N-2} a_k z^{-k-1}$$

$$= a_0 + \sum_{k=1}^{N-2} (a_k - a_{k-1}) z^{-k} - a_{N-2} z^{-(N-1)} \tag{2.18}$$

Identifying coefficients in (2.1) and (2.18) we see that

$$\left.\begin{aligned} h(0) &= -h(N-1) \\ h(k) &= a_k - a_{k-1} \\ h(N-1-k) &= a_{N-1-k} - a_{N-2-k} \\ &= a_{k-1} - a_k \end{aligned}\right\} \quad 1 \le k \le N-2 \tag{2.19}$$

Therefore h(k) = -h(N-1-k) for all k and condition (a) is contradicted unless $H(z) \equiv 0$.

(b) → (c):

Next, suppose (b) holds. Clearly (b) also holds with H(z) replaced by f(z). Suppose f has an even number of zeros. Then we can group these into reciprocal pairs and write

$$f(z) = \prod_{j=1}^{M} \beta_j (1-z^{-1}r_j)(1-z^{-1}r_j^{-1}) \tag{2.20}$$

where $\{r_i, r_i^{-1}, i = 1,\ldots,M\}$ are the real-valued zeros and β_j are real constants. Expanding (2.20) gives

$$f(z) = \prod_{j=1}^{M} \beta_j (1-(r_j+r_j^{-1})z^{-1}+z^{-2}) \tag{2.21}$$

Clearly each factor in (2.21) is an MIP in z^{-1}, hence by the lemma f(z) is also an MIP. If f has an odd number of zeros, we can write f as

$$f(z) = g(z)(1+z^{-1}) \tag{2.22}$$

The preceeding arguments for f(z) apply to g(z), and since $1+z^{-1}$ is an MIP in z^{-1}, so f(z) must be. Next if $i \varepsilon Q$, we can write

$$\begin{aligned} b_{oi} + b_{1i}z^{-1} + b_{2i}z^{-2} &= \beta_i\left(1-z^{-1}r_i e^{j\theta_i}\right)\left(1-z^{-1}r_i e^{-j\theta_i}\right) \\ &= \beta_i\left(1-2r_i \cos\theta_i z^{-1} + r_i^2 z^{-2}\right) \end{aligned} \tag{2.23}$$

If $r_i = 1$, then $b_{oi} = \beta_i = b_{2i}$. On the other hand if $r_i \neq 1$ there must be some $j \neq i$ such that $r_j = r_i{}^{-1}$ and $\theta_j = -\theta_i$, or

$$\begin{aligned} b_{oj} + b_{1j}z^{-1} + b_{2j}z^{-2} &= \beta_j\left(1-z^{-1}r_i^{-1}e^{-j\theta_i}\right)\left(1-z^{-1}r_i^{-1}e^{j\theta_i}\right) \\ &= \beta_j\left(1-2r_i^{-1} \cos\theta_i z^{-1} + r_i^{-2}z^{-2}\right) \end{aligned} \tag{2.24}$$

Identifying coefficients we obtain

$$\frac{b_{oi}}{b_{2j}} = \frac{\beta_i}{\beta_j r_i^{-2}} = \frac{\beta_i}{\beta_j} r_i^2$$

$$\frac{b_{1i}}{b_{1j}} = \frac{2\beta_i r_i \cos \theta_i}{2\beta_j r_i^{-1} \cos \theta_i} = \frac{\beta_i}{\beta_j} r_i^2$$

$$\frac{b_{2i}}{b_{oj}} = \frac{\beta_i r_i^2}{\beta_j}$$

Hence

$$\frac{b_{oi}}{b_{2j}} = \frac{b_{1i}}{b_{1j}} = \frac{b_{2i}}{b_{oj}} \tag{2.25}$$

(c) → (a):

Now let

$$g(z) = \prod_{i \varepsilon Q} \left(b_{oi} + b_{1i} z^{-1} + b_{2i} z^{-2}\right) \tag{2.26}$$

We can write

$$g(z) = g_1(z) g_2(z) \tag{2.27}$$

where $g_1(z)$ contains those factors in which $b_{oi} = b_{2i}$ and $g_2(z)$ contains the remainder. Clearly by the lemma $g_1(z)$ is an MIP in z^{-1}. Now for each factor of index i in $g_2(z)$ there is a factor of index $j \neq i$ such that (2.25) holds. Combining two such factors yields

$$\left(b_{oi} + b_{1i}z^{-1} + b_{2i}z^{-2}\right)\left(b_{oj} + b_{1j}z^{-1} + b_{2j}z^{-2}\right)$$

$$= \beta\left(b_{oi} + b_{1i}z^{-1} + b_{2i}z^{-2}\right)\left(b_{2i} + b_{1i}z^{-1} + b_{oi}z^{-2}\right)$$

$$= \beta\Big[b_{oi}b_{2i} + (b_{oi}b_{1i} + b_{1i}b_{2i})z^{-1} + \left(b_{oi}^2 + b_{1i}^2 + b_{2i}^2\right)z^{-2} + \left(b_{oi}b_{1i} + b_{1i}b_{2i}\right)z^{-3} + b_{oi}b_{2i}z^{-4}\Big] \quad (2.28)$$

where β is a proportionality constant. Clearly (2.28) is an MIP, and since $g_2(z)$ is a product of such factors, it is also an MIP. Thus g(z) is an MIP. Since

$$H(z) = f(z)g(z) \quad (2.29)$$

H(z) is an MIP of degree N-1 in z^{-1}, which means

$$h(k) = h(N-1-k) \quad 0 \leq k \leq N-1 \tag{2.30}$$

Thus we have proved (c) $\rightarrow$ (a).

Sufficiency of condition (d) is now clear from the fact that each condition stated for the b_{ji}'s leads to a factor for H(z) which is an MIP, hence H(z) must be an MIP, and (2.30) holds. This completes the proof of Theorem 2.1.

Q.E.D.

The definition (2.3) of a linear phase filter requires that the filter has both constant group delay and constant phase delay. However, if we are content with only constant group delay we can define a second type of "linear phase" filter in which the phase of $H(e^{j\omega})$ is a piecewise affine function of ω, i.e.,

$$H(e^{j\omega}) = \pm |H(e^{j\omega})| e^{j(\beta-\alpha\omega)} \tag{2.31}$$

By proceeding exactly as in the proof of Theorem 2.1 it is easily shown that with the constraint (2.1) on the form of H(z) the only possible solutions for $\beta\varepsilon[-\pi,\pi]$ is $\beta = \pm \frac{k\pi}{2}$, k = 0,1,2. If $\beta = 0, \pm\pi$ (2.31) reduces to (2.3). Thus the only new cases added are when $\beta = \pm \frac{\pi}{2}$. It can be seen from the proof of Theorem 2.1 that these cases arise exactly when $z_i = +1$ occurs as a zero of $H(z)$ in odd multiplicity, or equivalently when $\{h(k)\}$ satisfies

$$h(k) = -h(N-1-k) \qquad 0 \leq k \leq N-1$$

Filters of this special type are useful in the design of wide-band differentiators.[17] However, we shall not consider them further in this report, but shall restrict the term "linear phase filter" to refer to those satisfying (2.3).

2.2 Design Techniques and Realization Structures

There are three basic techniques for the design of FIR filters. These are the windowing, frequency-sampling, and optimal design techniques.[5] Although both the windowing and frequency-sampling techniques yield suboptimal filters, they are useful because of their simplicity and ease of design. However, we shall not consider further these techniques in this report, but instead will focus on the third design technique. Optimal design is attractive because the filters generated are optimum in a sense which we shall describe presently, and because efficient algorithms exist for its implementation. For simplicity we shall consider only filters with an odd number of points in their impulse responses.

A linear phase filter with an odd number of points has the nice property that with a simple translation of its impulse response samples in the time domain, its frequency response can be made purely real. Thus if $H(e^{j\omega})$ is the frequency response of an N-point, linear phase filter with impulse response $\{h(k), k=0,...,N-1\}$, where N is odd, define a new sequence $\{g(k)\}$ by

$$g(k) = h(\tfrac{N-1}{2} + k) \quad k = -\tfrac{N-1}{2}, \ldots, \tfrac{N-1}{2} \qquad (2.32)$$

Since $\{h(k)\}$ satisfies $h(k) = h(N-1-k)$, $0 \leq k \leq N-1$ we have $g(k) = g(-k)$, $0 \leq k \leq \frac{N-1}{2}$.

Hence

$$G(e^{j\omega}) = \sum_{k=-\frac{N-1}{2}}^{\frac{N-1}{2}} g(k)e^{-j\omega k}$$

$$= g(0) + \sum_{k=1}^{\frac{N-1}{2}} 2g(k) \cos k\omega \qquad (2.33)$$

which is the desired real frequency response. Now $G(e^{j\omega})$ is simply plus or minus the magnitude of $H(e^{j\omega})$, as

$$G(e^{j\omega}) = \sum_{k=-\frac{N-1}{2}}^{\frac{N-1}{2}} h(\tfrac{N-1}{2} + k)e^{-j\omega k}$$

$$= \sum_{k=0}^{N-1} h(k)e^{-j\omega(k - \frac{N-1}{2})}$$

$$= e^{j(\frac{N-1}{2})\omega} \sum_{k=0}^{N-1} h(k)e^{-j\omega k}$$

$$= e^{j(\frac{N-1}{2})\omega} H(e^{j\omega})$$

$$= \pm |H(e^{j\omega})| \tag{2.34}$$

where the last step is obvious from (2.14). Since in the design of linear phase filters we can shape only the magnitude of the frequency response, we will assume in the remainder of this section that $H(e^{j\omega})$ is real and of the form (2.33).

For the special case of low-pass filters, we can state simply that a filter designed via the optimal technique is optimum in the sense that given the order of the filter, the passband edge, and the maximum allowable approximation error in the passband and the stopband, it has the minimum attainable transition bandwidth.[2] In general, the optimality criterion can be stated as follows:

Let $D(\omega)$ be an ideal transfer characteristic which we wish to approximate with $H(e^{j\omega})$ for all $\omega \varepsilon S$, where S is any closed subset of $[0,\pi]$, not necessarily connected. Define a weighted error function on S as $E(\omega) = W(\omega)\,[D(\omega)-H(e^{j\omega})]$, and let $||E|| = \max_{\omega \varepsilon S} |E(\omega)|$. Then a filter $H(e^{j\omega})$ designed via the optimal technique is optimum in the sense that given $D(\omega)$, $W(\omega)$, S, and the number of points N of the filter, it results in the least possible $||E||$.

It can be proved[2] that an $H(e^{j\omega})$ which satisfies the above optimality criterion exhibits on S at least $\frac{N+3}{2}$ "alternations", i.e., if $Q = \{\omega_i \varepsilon S,\ i=1,\ldots,M \mid \omega_i < \omega_{i+1},\ E(\omega_i) = -E(\omega_{i+1}) = \pm||E||,\ i=1,\ldots,M-1\}$ then Q has at least $\frac{N+3}{2}$ elements, or $M \geq \frac{N+3}{2}$. In the case where $H(e^{j\omega})$ is a low-pass transfer function with passband edge ω_p and stopband edge ω_s, $S = [0,\omega_p]U[\omega_s,\pi]$. Since by definition $\omega_p \varepsilon Q$ and $\omega_s \varepsilon Q$, and all other elements of Q must be extrema of $H(e^{j\omega})$, the condition on the number of elements in Q implies that the optimum $H(e^{j\omega})$ must have at least $\frac{N-1}{2}$ points of extrema on **S**. We will show next that any $H(e^{j\omega})$ can have at most $\frac{N+1}{2}$ extrema on $[0,\pi]$, hence on S.

Recall that $H(e^{j\omega})$ is plus or minus the magnitude of the frequency response of some causal linear phase filter and has the form (2.33), i.e.,

$$H(e^{j\omega}) = \sum_{k=0}^{\frac{N-1}{2}} g_k \cos\omega k \tag{2.35}$$

for some sequence $\{g_k\}$.

Now for each k we have the trigonometric relation

$$\cos k\omega = \sum_{m=0}^{k} \alpha_{mk} (\cos\omega)^m \tag{2.36}$$

for some real sequence $\{\alpha_{mk}\}$. Therefore (2.35) can be written as

$$H(e^{j\omega}) = \sum_{k=0}^{\frac{N-1}{2}} g_k \Big(\sum_{m=0}^{k} \alpha_{mk} (\cos\omega)^m \Big)$$

$$= \sum_{k=0}^{\frac{N-1}{2}} d_k (\cos\omega)^k \tag{2.37}$$

where $\{d_k\}$ is some appropriate sequence. Differentiating (2.37), we obtain

$$\frac{d}{d\omega} H(e^{j\omega}) = \sum_{k=0}^{\frac{N-1}{2}} k d_k (\cos\omega)^{k-1} (-\sin\omega)$$

$$= -\sin\omega \sum_{k=0}^{\frac{N-3}{2}} (k+1)d_{k+1}(\cos\omega)^k \qquad (2.38)$$

Now consider the one-to-one mapping from $[0,\pi]$ onto $[-1,1]$ defined by $x = \cos\omega$. With this transformation we can define a new function $G(x)$ by

$$G(x) = \frac{d}{d\omega} H(e^{j\omega})\Big|_{\omega=\cos^{-1}x} = f_1(x)f_2(x) \qquad (2.39)$$

where

$$\left.\begin{aligned} f_1(x) &= -\sqrt{1-x^2} \\ f_2(x) &= \sum_{k=0}^{\frac{N-3}{2}} (k+1)d_{k+1}x^k \end{aligned}\right\} x\varepsilon[-1,1] \qquad (2.40)$$

Clearly, $f_1(x)$ has two zeros at $x = \pm 1$. Now $f_2(x)$ is the restriction of a polynomial of degree $\frac{N-3}{2}$ to the interval $[-1,1]$, hence can have at most $\frac{N-3}{2}$ zeroes on the open interval $(-1,1)$. Therefore $G(x)$ can have at most $\frac{N+1}{2}$ zeroes on $[-1,1]$. But this means $H(e^{j\omega})$ can have at most $\frac{N+1}{2}$ extrema on $[0,\pi]$.

Thus a low-pass filter designed via the optimal design technique, which we shall call an optimal low-pass filter, has an $H(e^{j\omega})$ which has either $\frac{N-1}{2}$ or $\frac{N+1}{2}$ extrema on $[0,\pi]$. Following conventional usage we shall call an extremum of $H(e^{j\omega})$ a "ripple" and refer to the value of $|E(\omega)|$ at an extremum as the height of the ripple. Now to achieve optimality an $H(e^{j\omega})$ need only have $\frac{N-1}{2}$ ripples of equal height on $[0,\pi]$. Hence, in general, an optimal filter is not necessarily equiripple, meaning that for instance if more than $\frac{N-1}{2}$ ripples are present, no more than $\frac{N-1}{2}$ of them need be of equal height.

Those low-pass filters which exhibit $\frac{N+1}{2}$ equal-height ripples constitute a special class of optimal low-pass filters, which we shall refer to as extraripple filters, following the usage by Parks and McClellan[2]. Because of the uniqueness of optimal filters, given the maximum allowable approximation error in the passband and the stopband, there are exactly $\frac{N-1}{2}$ possible extraripple filters of length N, which are uniquely determined once the number of ripples in the passband or the stopband is specified. Furthermore, given the approximation error and N there are exactly $\frac{N-1}{2}$ unique values of ω which are possible passband edges for an N-point extraripple filter. Finally, within the class of all optimal low-pass filters with identical impulse response length and approximation error, extraripple filters are shown to be locally optimum in the sense that if $F(\omega)$ denotes transition bandwidth as a function of passband edge and if ω_x is a passband edge for an extraripple filter, then $F(\omega)$ possesses a local minimum at $\omega = \omega_x$.[18]

Several methods for the optimal design of filters are currently available. The polynomial interpolation[3] and nonlinear optimization[19] methods are both only capable of designing extraripple filters. However, the former technique is considerably more efficient than the latter. The Chebyshev approximation[2] and linear programming[20] methods can both be used to design optimal filters in general. Though less flexible, the former technique is much more efficient. Given the same specifications, where applicable, all four techniques yield identical solutions. The extraripple filters used as examples in this report will be generated using the polynomial interpolation method.

Having obtained a desired transfer function, the next step in the design of a digital filter is choosing a method of implementation. There are several "structures" in which a given linear phase FIR transfer function can be realized. Perhaps the simplest of these is the direct form.

Figure 2.1 shows the block diagram of an N-point filter in direct form, where N is odd. This structure can be easily derived by writing H(z) in the form

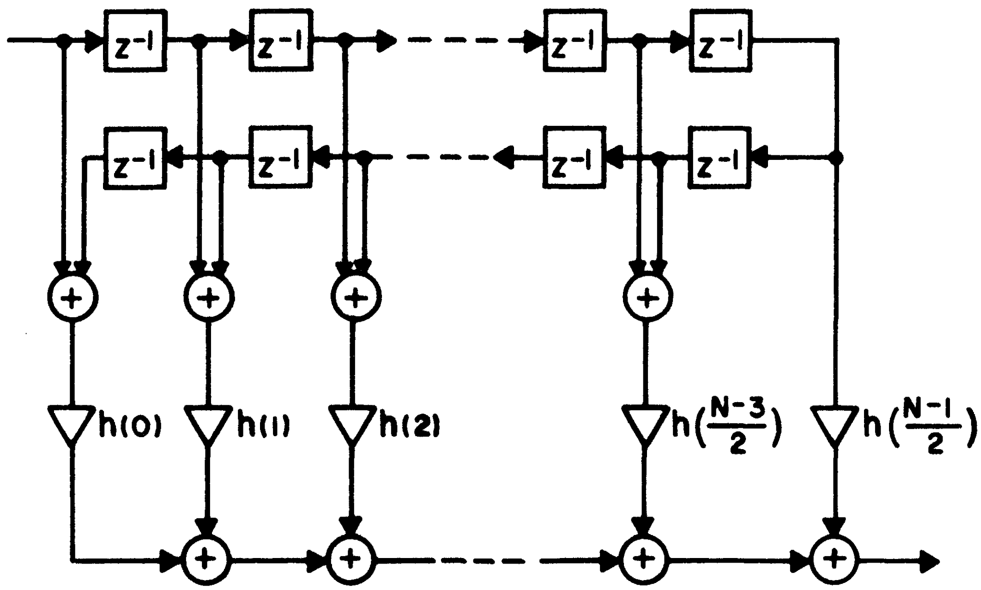


FIG. 2.1 DIRECT FORM LINEAR PHASE FILTER

$$H(z) = \sum_{k=0}^{\frac{N-3}{2}} h(k)z^{-k} + \sum_{k=\frac{N+1}{2}}^{N-1} h(k)z^{-k} + h(\tfrac{N-1}{2})z^{-(\frac{N-1}{2})}$$

$$= \sum_{k=0}^{\frac{N-3}{2}} \left[h(k)z^{-k} + h(N-1-k)z^{-(N-1-k)}\right] + h(\tfrac{N-1}{2})z^{-(\frac{N-1}{2})}$$

$$= \sum_{k=0}^{\frac{N-3}{2}} h(k)\left[z^{-k} + z^{-(N-1-k)}\right] + h(\tfrac{N-1}{2})z^{-(\frac{N-1}{2})}$$

(2.41)

A similar structure arises when N is even.

Alternatively, H(z) can be realized in cascade form. Because complex zeros of linear phase FIR filters may occur in quadruplets where the four zeros in each group are interdependent, it is natural to attempt a cascade structure using 4th order subfilters as building blocks. However, the results of this report will show that from the viewpoint of roundoff errors it is generally undesirable to group together reciprocal zeros in a cascade structure. Therefore we will consider only a cascade structure built upon 2nd order filter sections.

Condition (d) of Theorem 2.1 provides us a way to assign zeros to individual 2nd-order sections of a cascade filter so that linear phase is preserved. In particular, complex zeros are grouped by conjugate pairs, real zeros that are reciprocals of each other are paired together, while doubled or higher multiplicity zeros are grouped by pairs of the same kind. In this way the only zero that can occur by itself in a section is $z = -1$ (since by Theorem 2.1 $z = +1$ is not allowed as a zero of odd multiplicity). This strategy of zero assignment will be assumed in all cascade filters discussed in this report. Thus for a cascade filter we write H(z) in the form

$$H(z) = \prod_{i=1}^{N_s} (b_{oi}+b_{1i}z^{-1}+b_{2i}z^{-2}) \tag{2.42}$$

where $\{b_{ij}\}$ satisfies condition (d) of Theorem 2.1 and N_s is the number of sections. Figure 2.2 shows a general block diagram for the i^{th} section of H(z). However, when $b_{oi} = b_{2i}$ we will find it desirable to use instead the configuration in Figure 2.3, since it leads to reduced roundoff errors. Figure 2.4 shows that these two configurations can be readily accommodated in a more general subfilter structure, therefore in the remainder of this report we will assume that both configurations are used in a cascade structure. In particular, for the i^{th} section

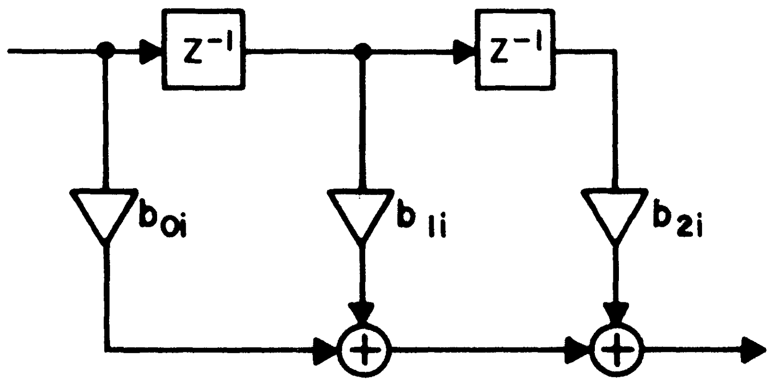


FIG. 2.2 — CASCADE FORM FILTER SECTION

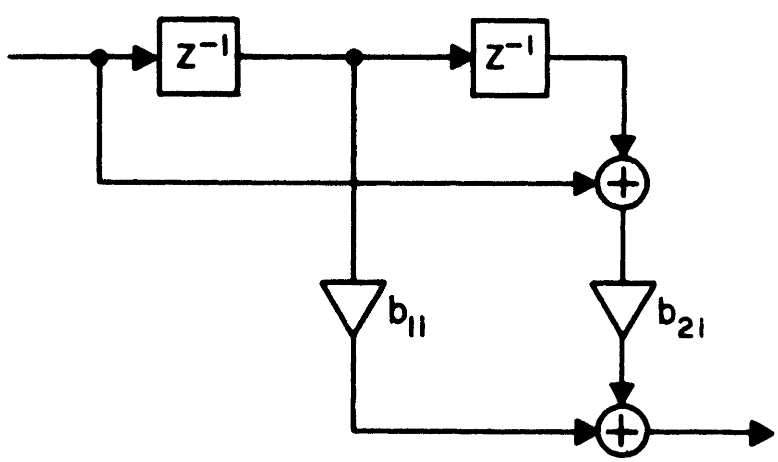


FIG. 2.3 — ALTERNATE CASCADE FILTER SECTION

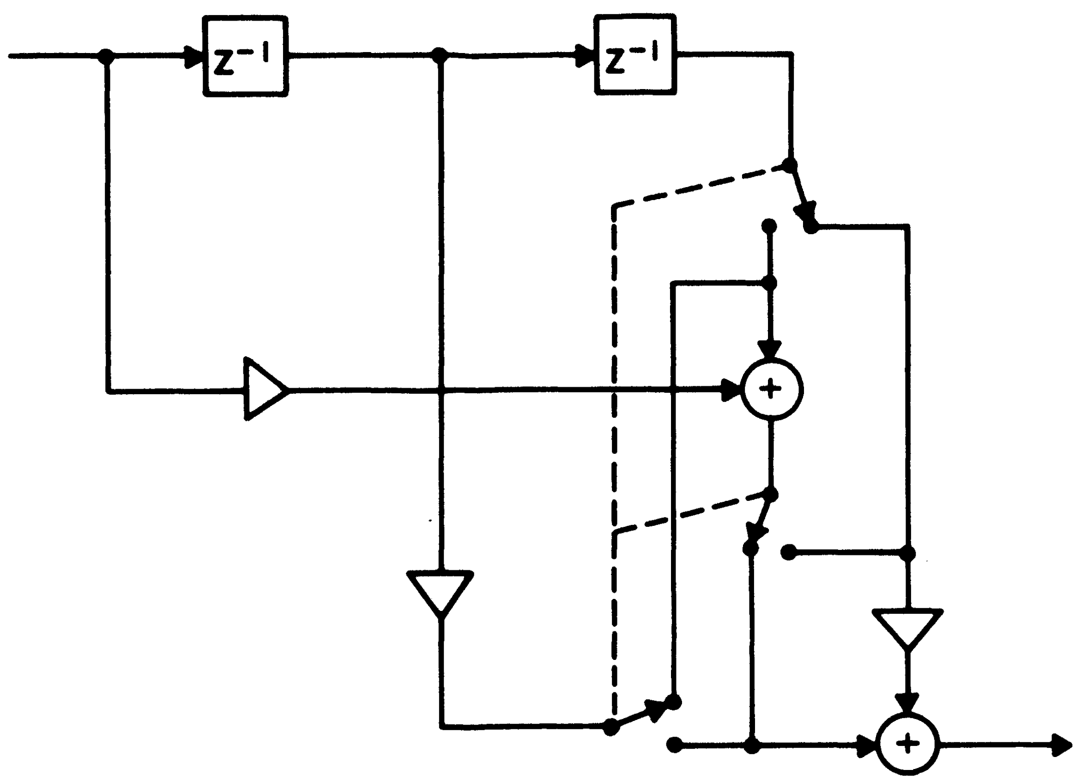


FIG. 2.4 GENERAL CASCADE FILTER SECTION

if $b_{oi} = b_{2i}$, Figure 2.3 is used and if $b_{oi} \neq b_{2i}$, Figure 2.2 is used. Furthermore, though minor variations to Figures 2.2 and 2.3 are possible as building blocks for a cascade form, we will assume unless otherwise stated that Figures 2.2 and 2.3 are meant whenever the term cascade form is used. For more on cascade form variations see Section 4.1.

Other structures are possible for the realization of linear phase FIR filters. A well known example is the frequency-sampling structure[15], which is particularly well adapted to the frequency-sampling design approach. Other less well-known structures based upon polynomial interpolation formulas have also been proposed. These include the Lagrange, Newton, Hermite, and Taylor structures[14]. We shall not consider any of these other structures in this report.

3.0 Techniques for Roundoff Error Analysis

Although digital filters are usually analyzed using linear system techniques, such as difference equations and z-transformations, any practical realization of a digital filter is necessarily nonlinear because of quantization effects. Nonlinearities are introduced when results of arithmetic operations are quantized. Thus a single input-output linear relation cannot accurately describe the behavior of a practical digital filter.

The powerful techniques of linear system theory, however, can still be applied if we take a slightly different approach and model a digital filter as a multi-input rather than a single-input system. The additional inputs are contrived in such a way that the overall system becomes linear. More specifically, each multiplier in a filter is modelled as an ideal multiplier followed by a summation node where an auxiliary input signal is added to the product. The samples of this extra input are devised in such a way that the summation result always equals some quantized level in the filter. Thus if the extra input sequence is chosen to have magnitudes less than or equal to half of a quantized step, our model is exactly equivalent to a multiplier which rounds its result.

Since fixed-point arithmetic is assumed, addition introduces no error. Thus with each multiplier modelled as in the above, we arrive at a multi-input linear system model for a practical digital filter. Quantization as a direct constraint disappears from our analysis since all data at physical points of interest in the filter model automatically take on quantized values. In effect we have shifted our problem from dealing with a nonlinear system transfer function to determining a set of auxiliary inputs.

Clearly, given a practical filter and its input, all the appropriate auxiliary input sequences can be exactly determined. However, the highly complex nature and lack of generality of any such attempt deems it totally unfeasible. Therefore, we shall not venture into any deterministic analysis of the extra inputs, but instead will adopt a statistical approach which has proven to be very fruitful.

In view of the statistical analysis, we shall refer to roundoff errors as "roundoff noise" and call each auxiliary input in our model a noise source. The next three sections will formulate the model for roundoff noise and apply it to cascade FIR filters with dynamic range constraints.

3.1 Statistical Model for Roundoff Errors

In the previous section, we have developed a model for a practical filter consisting of noise inputs as well as signal input. We will now formulate a statistical description for the noise sources so that using the linearity of our model the roundoff noise in a filter can be analyzed independent of the signal.

Clearly, each noise source is simply a sequence of samples each of which is an error term due to rounding. Therefore, it is reasonable to model each sample as a random variable with uniform probability density on the interval $(-\frac{Q}{2}, \frac{Q}{2})$ and zero density elsewhere, where Q is the quantization step size. Thus each sample is a zero mean random variable with a variance of $\frac{Q^2}{12}$. If all data were represented by fractions, then $Q = 2^{-(t-1)}$ where t is the number of bits in a data word (1 sign bit and t-1 numerical bits).

Furthermore we shall assume the following:

1) Any two different samples from the same noise source are uncorrelated.
2) Any two different noise sources, regarded as random processes, are uncorrelated.
3) Each noise source is uncorrelated with the input signal.

Thus each noise source is modelled as a discrete stationary white random process with a uniform power density spectrum of magnitude $\frac{Q^2}{12}$. Although the above assumptions can be shown to be invalid for many pathological cases, they have been supported by a great deal of experimental results for a large class of signals and quantization step sizes of interest.[11, 13, 27, 28]

Thus far, the vast majority of studies on roundoff noise has been carried out based upon the assumptions stated above. The results have been found to be useful and agree well with experimental evidence. Therefore, we shall do likewise in this report.

3.2 Roundoff Noise in Cascade Form FIR Filters

Having formulated a model for roundoff errors, we will now apply it to the analysis of roundoff noise in cascade form FIR filters. Block diagrams for general sections of a cascade FIR filter with quantization effects ignored are shown in Figures 2.2-2.4. As can be seen from these diagrams, adding a noise source to the output of any multiplier in any of these section configurations is equivalent to adding a noise source to the output of the section. Therefore, to model a section of a practical cascade filter we need simply add k_i noise sources to the output of the section, where k_i is the number of multipliers with non-integer coefficients in the section. Or equivalently, by assumption 2 of the previous section, we can instead add one noise source of variance $k_i \frac{Q^2}{12}$.

Before proceeding further, we shall need to develop some notations. Let $H_i(z)$ denote the transfer function of the i^{th} section of a filter H(z), i.e.,

$$H(z) = \prod_{i=1}^{Ns} H_i(z) \qquad (3.1)$$

where

$$H_i(z) = b_{oi} + b_{1i} z^{-1} + b_{2i} z^{-2}$$

Furthermore, define

$$G_i(z) = \begin{cases} \prod_{j=i+1}^{Ns} H_j(z) & 0 \leq i \leq Ns-1 \\ 1 & i = Ns \end{cases} \qquad (3.2)$$

and let $\{g_i(k)\}$ be the impulse response of $G_i(z)$, i.e.,

$$G_i(z) = \sum_k g_i(k) z^{-k} \qquad (3.3)$$

Then we can model a practical cascade filter as in Figure 3.1 or equivalently as in Figure 3.2. Letting $\{E_i(n)\}$ denote the noise sequence at the filter output due to the i^{th} noise source alone, we have

$$E_i(n) = \sum_k g_i(k) e_i(n-k) \qquad (3.4)$$

By the stationarity of $\{e_i(n)\}$ the variance of $E_i(n)$ is independent of n, hence denoting this variance by σ_i^2 , we obtain by assumption 1 of section 3.1,

$$\sigma_i^2 = \sum_k g_i^2(k) \overline{e_i(n-k)^2}$$

$$= k_i \frac{Q^2}{12} \sum_k g_i^2(k) \qquad (3.5)$$

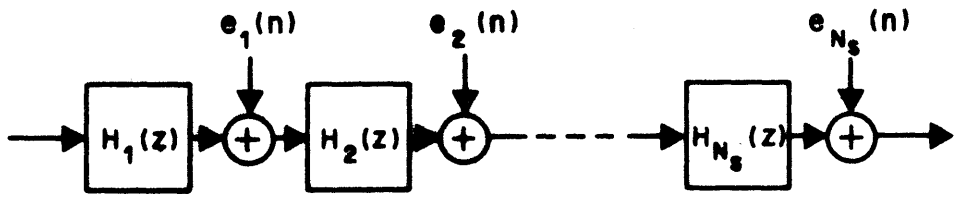


FIG. 3.1 – MODEL FOR PRACTICAL CASCADE FILTER

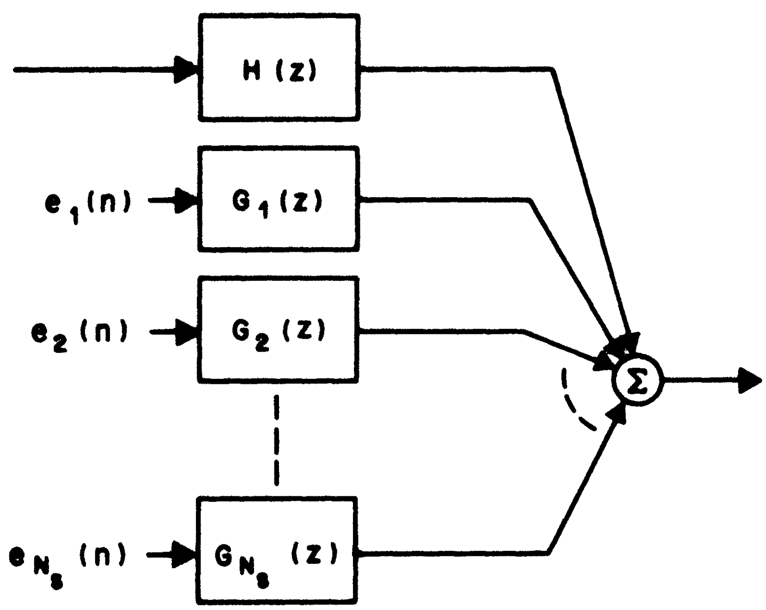


FIG. 3.2 – ALTERNATE EQUIVALENT MODEL FOR CASCADE FILTER

Now the total noise output is given by

$$E(n) = \sum_{i=1}^{Ns} E_i(n) = \sum_{i=1}^{Ns} \sum_{k} g_i(k)e_i(n-k) \qquad (3.6)$$

Therefore by assumptions 1 and 2 of the previous section,

$$\sigma^2 = \overline{E^2(n)} = \sum_{i=1}^{Ns} \sigma_i^2 \qquad (3.7)$$

It is instructive to re-derive (3.5) by a slightly different approach. Since different noise sources are uncorrelated white processes, their power density spectra add, therefore we can write the power spectrum of $\{e_i(n)\}$ as

$$S_i(\omega) = k_i \frac{Q^2}{12} \qquad (3.8)$$

Let $N_i(\omega)$ be the power spectrum of $\{E_i(n)\}$, then from linear system noise theory,

$$N_i(\omega) = |G_i(e^{j\omega})|^2 S_i(\omega)$$

$$= k_i \frac{Q^2}{12} |G_i(e^{j\omega})|^2 \qquad (3.9)$$

Therefore

$$\sigma_i^2 = \frac{1}{2\pi} \int_0^{2\pi} N_i(\omega) d\omega$$

$$= k_i \frac{Q^2}{12} \cdot \frac{1}{2\pi} \int_0^{2\pi} |G_i(e^{j\omega})|^2 d\omega \qquad (3.10)$$

But by Parseval's theorem for discrete signals

$$\frac{1}{2\pi} \int_0^{2\pi} |G_i(e^{j\omega})|^2 d\omega = \sum_k g_i^2(k) \qquad (3.11)$$

Therefore again we arrive at (3.5). In comparing different orderings of a given cascade filter we shall use the output noise variance σ^2 as a figure of merit. However, in terms of the actual deviation of an output sample from the value it would have if quantization effects were absent, the standard deviation σ is more applicable. σ is the rms noise value, or in some sense a measure of the expected magnitude of a noise sample. If a large number of noise sources were present in a filter, we can argue from the Central Limit Theorem of probability theory that the distribution of the output noise will be approximately Gaussian. In that case we can say that essentially (i.e., with high probability) all output errors are bounded in magnitude by 3σ.

In general some multiple of σ can be used as an essential upperbound on the noise magnitude. From (3.5) and (3.7) σ is directly proportional to Q. We will now show that if all output errors of a filter are bounded in magnitude by δQ, where δ is some positive constant and Q is the quantization step size, then t bits of accuracy in the output can be assured if all data are represented by $t+\ell$ bits, where $\ell \geq \log_2\delta+1$.

To show this, observe that with t bits of accuracy all data are represented to within an error of $\pm 2^{-t}$, since 2^{-t} is half of the quantization step size. Thus if the roundoff error magnitudes at the filter output are no more than 2^{-t}, then t bits of accuracy is preserved. Now if $t+\ell$ bits are used to represent all data in the filter, then $Q = 2^{-(t+\ell-1)}$. Therefore to assure t bits of accuracy we require

$$\delta Q = \delta \cdot 2^{-(t+\ell-1)} \leq 2^{-t} \qquad (3.12)$$

or

$$\ell \geq \log_2\delta+1 \qquad (3.13)$$

Because of the statistical approach which we adopted, we cannot set an absolute upperbound on the output errors. However, as an engineering criterion for choosing the number of extra bits required to compensate for roundoff noise in a filter, we can use

$$\delta Q = \sigma$$

or

$$\delta = \sigma/Q \tag{3.14}$$

Thus we can define a noise figure for a filter in number of bits as

$$\text{Noise in number of bits} = \log_2 \left(\frac{\sigma}{Q}\right) + 1 \tag{3.15}$$

Notice that this noise figure is independent of the quantization step size or the wordlength employed.

3.3 Dynamic-Range Constraints in the Cascade Form

A practical digital filter, necessarily implemented as a physical device, must have a finite dynamic range. Especially when fixed-point arithmetic is employed, this dynamic range sets a practical limit to the maximum range of signal levels representable in a filter and acts to constrain the signal-to-noise ratio attainable.

In some filter structures, such as the direct form, given the filter transfer function the designer has no control over the relative signal levels at points within the filter. Only the gain of the overall filter can be varied. However, in a cascade realization with Ns sections there are Ns-1 degrees of freedom available in addition to the overall filter gain and the ordering of sections.

To see this let us define a factorization for H(z) which is unique up to ordering of factors, in the form

$$H(z) = \beta \prod_{i=1}^{Ns} \hat{H}_i(z)$$

$$\hat{H}_i(z) = a_{oi} + a_{1i}z^{-1} + a_{2i}z^{-2} \qquad (3.16)$$

where $\{a_{ij}\}$ satisfies

$$a_{oi} \geq 0, \quad \sum_{j=0}^{2} |a_{ji}| = 1 \quad i = 1,\ldots,Ns \qquad (3.17)$$

Then the transfer function for the i^{th} section in a cascade realization can be written as

$$H_i(z) = S_i\hat{H}_i(z) \qquad (3.18)$$

where S_i is an arbitrary constant, subject only to the constraint that

$$\prod_{i=1}^{Ns} S_i = \beta \qquad (3.19)$$

Thus given β, Ns-1 of the S_i's can be chosen at will.

We shall refer to any rule for assigning values to $\{S_i\}$ as a scaling method. Obviously, some scaling method must be employed in the design of a cascade filter whether or not one is concerned with dynamic range constraints since numerical values must be assigned to the S_i's. When dynamic range is an issue, the constraints it imposes can be met in some best manner by choosing the proper scaling method. In this thesis we shall be concerned only with filters designed so that no arithmetic overflow in them can cause distortion in the filter output. Therefore, our investigation of scaling methods will be restricted to those methods which guarantee that for a given class of input signals no distortion-causing overflow occurs in the scaled filter.

It can be shown[21] that in an addition operation if two's complement arithmetic is used, as is usually the case, then as long as the final result is within the representable numerical range, individual partial sums can be allowed to overflow without causing inaccuracies in the result. We shall assume in this thesis that all additions in a filter are done using two's complement arithmetic. Then, to guarantee that no distortion caused by overflow occurs at a cascade filter's output, only the input and output of each filter section need be constrained not to overflow.

To simplify the discussion of scaling methods, we make the following definitions. Let

$$F_i(z) = \sum_{k=0}^{2i} f_i(k) z^{-k} = \prod_{j=1}^{i} H_j(z) \qquad (3.20)$$

and $\qquad 1 \leq i \leq Ns$

$$\hat{F}_i(z) = \sum_{k=0}^{2i} \hat{f}_i(k) z^{-k} = \prod_{j=1}^{i} \hat{H}_j(z) \qquad (3.21)$$

Also, let $\{v_i(n)\}$ be the output sequence of $F_i(z)$ or $H_i(z)$. Furthermore, assume that the maximum magnitude of numerical data representable in a filter is 1.0. Then the necessary overflow constraints on a cascade filter can be stated as

$$|v_i(n)| \leq 1 \qquad 1 \leq i \leq Ns, \qquad \text{all } n \qquad (3.22)$$

We now state and prove necessary and sufficient conditions for (3.22) to hold for two classes of input signals. Theorem 3.1 deals with the class of input sequences $\{x(n)\}$ which satisfy $|x(n)| \leq 1$ for all n. For simplicity we shall refer to this class as class 1. Theorem 3.2 deals with the class of inputs of the form $\{x(n)\}$ with transform $X(e^{j\omega})$ which satisfy

$$\frac{1}{2\pi}\int_0^{2\pi} |X(e^{j\omega})|\,d\omega \leq 1.$$

This class will be called class 2. By virtue of the fact that

$$x(n) = \frac{1}{2\pi}\int_0^{2\pi} X(e^{j\omega})e^{j\omega n}\,d\omega \tag{3.23}$$

and hence

$$|x(n)| \leq \frac{1}{2\pi}\int_0^{2\pi} |X(e^{j\omega})|\,d\omega \tag{3.24}$$

class 2 is a subset of class 1.

Theorem 3.1: Suppose $|x(n)| \leq 1$. Then condition (3.22) is satisfied if and only if

$$\sum_{k=0}^{2i} |f_i(k)| \le 1 \qquad i = 1,\ldots,Ns \tag{3.25}$$

Proof: If (3.25) holds, then since for $i = 1,\ldots,Ns$

$$v_i(n) = \sum_{k=0}^{2i} f_i(k)x(n-k)$$

we have

$$|v_i(n)| \le \sum_{k=0}^{2i} |f_i(k)||x(n-k)|$$

$$\le \max|x(n)| \sum_{k=0}^{2i} |f_i(k)|$$

$$\le \sum_{k=0}^{2i} |f_i(k)| \le 1$$

Hence (3.22) holds.

On the other hand if (3.25) does not hold, then for some i $\sum_{k=0}^{2i} |f_i(k)| = \delta$, where $\delta > 1$. Now let $\{x(n)\}$ be any sequence satisfying $|x(n)| \le 1$ such that for some n_o

$$x(k) = \frac{f_i(n_o-k)}{|f_i(n_o-k)|} \qquad n_o-2i \le k \le n_o$$

Clearly $\{x(n)\}$ can be chosen to be a causal sequence by letting $n_o \geq 2i$. But now

$$v_i(n_o) = \sum_{k=0}^{2i} f_i(k)x(n_o-k)$$

$$= \sum_{k=0}^{2i} f_i(k) \left[\frac{f_i(k)}{|f_i(k)|}\right]$$

$$= \sum_{k=0}^{2i} |f_i(k)| = \delta > 1$$

Hence (3.22) does not hold.

Q.E.D.

Theorem 3.2: Suppose $\frac{1}{2\pi}\int_0^{2\pi} |X(e^{j\omega})|d\omega \leq 1$. Then condition (3.22) is satisfied if and only if

$$|F_i(e^{j\omega})| \leq 1 \qquad i = 1,\ldots,Ns$$

$$0 \leq \omega \leq 2\pi \qquad (3.26)$$

Proof: In general, for $i = 1,\ldots,Ns$

$$v_i(n) = \frac{1}{2\pi}\int_0^{2\pi} F_i(e^{j\omega})X(e^{j\omega})e^{j\omega n}d\omega$$

If (3.26) holds, then

$$|v_i(n)| \leq \max |F_i(e^{j\omega})| \cdot \frac{1}{2\pi}\int_0^{2\pi} |X(e^{j\omega})|d\omega$$

$$\leq \frac{1}{2\pi}\int_0^{2\pi} |X(e^{j\omega})|d\omega \leq 1$$

However, if (3.26) does not hold, then for some i and some ω_o

$$|F_i(e^{j\omega_o})| = \varepsilon \quad \text{where} \quad \varepsilon > 1 \tag{3.27}$$

Let

$$F_i(e^{j\omega}) = |F_i(e^{j\omega})|e^{j\theta_i(\omega)} \tag{3.28}$$

and let the input sequence $\{x(n)\}$ be defined by

$$x(n) = \cos(\omega_o(n-\beta_i)) \tag{3.29}$$

where

$$\beta_i = \frac{\theta_i(\omega_o)}{\omega_o} \tag{3.30}$$

Then

$$X(e^{j\omega}) = \pi[\delta(\omega-\omega_o) + \delta(\omega+\omega_o)]e^{-j\omega\beta_i} \quad -\pi \leq \omega \leq \pi$$

and

$$\frac{1}{2\pi}\int_0^{2\pi} |X(e^{j\omega})|d\omega = 1$$

But

$$v_i(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |F_i(e^{j\omega})| e^{j\theta_i(\omega)} X(e^{j\omega}) e^{j\omega n} d\omega$$

$$= \frac{1}{2}\left[|F_i(e^{j\omega_o})| e^{j\theta_i(\omega_o)} e^{j\omega_o(n-\beta_i)} \right.$$

$$\left. + |F_i(e^{-j\omega_o})| e^{j\theta_i(-\omega_o)} e^{-j\omega_o(n-\beta_i)} \right]$$

Since $\{f_i(k)\}$ is real,

$$|F_i(e^{j\omega})| = |F_i(e^{-j\omega})|$$

and

$$\theta_i(\omega) = -\theta_i(-\omega)$$

Therefore

$$v_i(n) = |F_i(e^{j\omega_o})| \cos[\theta_i(\omega_o) + \omega_o n - \omega_o \beta_i]$$

$$= \varepsilon \cos \omega_o n$$

where we have used (3.27) and (3.30). Hence $v_i(0) = \varepsilon > 1$ which shows that (3.22) does not hold.

Q.E.D.

Conditions (3.25) and (3.26) of Theorems 3.1 and 3.2 can be re-stated to give conditions on $\{S_i\}$. Recall that the $\hat{H}_i(z)$'s are unique once H(z) is given, hence the $\hat{F}_i(z)$'s and $\{\hat{f}_i(k)\}$'s are also unique. From (3.18), (3.20) and (3.21) we have

$$f_i(k) = \Big(\prod_{j=1}^{i} S_j\Big)\hat{f}_i(k) \tag{3.31}$$

and

$$F_i(z) = \Big(\prod_{j=1}^{i} S_j\Big)\hat{F}_i(z) \tag{3.32}$$

Therefore, conditions (3.25) and (3.26) can be re-stated respectively as

$$\prod_{j=1}^{i} |S_j| \leq \left[\sum_{k=0}^{2i} |\hat{f}_i(k)|\right]^{-1} \tag{3.33}$$

and

$$\prod_{j=1}^{i} |S_j| \leq \left[\max_{0 \leq \omega \leq 2\pi} |\hat{F}_i(e^{j\omega})|\right]^{-1} \tag{3.34}$$

$i = 1,\ldots,Ns$ (for both (3.33) and (3.34))

These then are conditions which, for the class of inputs concerned, a scaling method must satisfy. We shall show next that in some sense optimum scaling methods are obtained when (3.33) and (3.34) are satisfied with equality. For ease of reference we shall first define and name two scaling methods.

Define sum scaling to be the rule

$$\prod_{j=1}^{i} S_j = \left[\sum_{k=0}^{2i} |\hat{f}_i(k)| \right]^{-1} \qquad i = 1,\ldots,Ns \quad (3.35)$$

or stated recursively,

$$S_i = \begin{cases} \left[\sum_{k=0}^{2} |\hat{f}_1(k)| \right]^{-1} & i = 1 \\ \left[\left(\prod_{j=1}^{i-1} S_j \right) \sum_{k=0}^{2i} |\hat{f}_i(k)| \right]^{-1} & i = 2,\ldots,Ns \end{cases} \quad (3.36)$$

Also, define peak scaling to be the rule

$$\prod_{j=1}^{i} S_j = \left[\max_{0 \le \omega \le 2\pi} |\hat{F}_i(e^{j\omega})| \right]^{-1} \quad i = 1,\ldots,Ns \ (3.37)$$

or

$$S_i = \begin{cases} \left[\max_{0 \le \omega \le 2\pi} |\hat{F}_1(e^{j\omega})| \right]^{-1} & i = 1 \\ \left[\left(\prod_{j=1}^{i-1} S_j \right) \max_{0 \le \omega \le 2\pi} |\hat{F}_i(e^{j\omega})| \right]^{-1} & i = 2,\ldots,Ns \end{cases} \quad (3.38)$$

<u>Theorem 3.3</u>: Given a transfer function to be realized in cascade form (as defined in figures 2.2, 2.3) using fixed-point arithmetic of a given word-length, and given the ordering of filter sections, assume that

a) the number of noise sources in each section (i.e. k_i) is independent of the scaling method.

b) all filter coefficients can be represented to arbitrary precision.

c) no overflow is allowed to occur at the input and output of each section.

d) the overall gain of the filter is maximized subject to no overflow at the filter output.

Then each of the following scaling methods is optimum for the class of input signals stated in the sense that it yields the minimum possible roundoff noise variance as defined in (3.7) among all scaling methods which satisfy conditions (c) and (d) above for the class of inputs considered.

1) Sum scaling for class 1* signals

2) Peak scaling for class 2* signals

Proof: From (3.7) and (3.10),

$$\sigma^2 = \sum_{i=1}^{Ns} k_i \frac{Q^2}{12} \cdot \frac{1}{2\pi} \int_0^{2\pi} |G_i(e^{j\omega})|^2 d\omega \qquad (3.39)$$

---

* See page 62 for definitions.

or using (3.2) and (3.18),

$$\sigma^2 = \sum_{i=1}^{Ns-1} \left(\prod_{j=i+1}^{Ns} S_j \right)^2 \cdot C_i + k_{Ns} \cdot \frac{Q^2}{12} \tag{3.40}$$

where

$$C_i = k_i \frac{Q^2}{12} \cdot \frac{1}{2\pi} \int_0^{2\pi} \left| \prod_{j=i+1}^{Ns} \hat{H}_j(e^{j\omega}) \right|^2 d\omega$$

$$1 \leq i \leq Ns-1$$

We can rewrite (3.40) as

$$\sigma^2 = \beta \sum_{i=1}^{Ns-1} \frac{C_i}{\left(\prod_{j=1}^{i} S_j \right)^2} + k_{Ns} \cdot \frac{Q^2}{12} \tag{3.41}$$

where $\beta = \left(\prod_{j=1}^{Ns} S_j \right)^2$ is by assumption independent of scaling. Also, the last term in (3.41) and all the C_i's are by the assumptions independent of scaling. Therefore σ^2 is minimized when the summation in (3.41) is minimized. But since each term in the summation is nonnegative, the sum is minimized by minimizing each term individually. This means we must maximize $\prod_{j=1}^{i} |S_j|$ for $i = 1,\ldots,Ns-1$. Referring to (3.33), (3.34), (3.35), and (3.37), clearly sum scaling and peak scaling satisfy conditions (c) and

(d) of the theorem. Also we see that the maximization of $\prod_{j=1}^{i} |S_j|$ is accomplished when sum scaling is used for class 1 signals or peak scaling is used for class 2 signals.

Q.E.D.

Thus optimal scaling methods are established for two classes of input signals. It is possible to define other classes of signals by considering the "Lp norm" of their transforms[11]. The Lp norm, or p-norm, of a function f(x) on an interval [a,b] is defined as[22]

$$||f(x)||_p = \left[\int_a^b |f(x)|^p dx\right]^{1/p} \qquad 1 \leq p < \infty \quad (3.42)$$

In general, the p-norm of a function f(x) is defined as long as $|f(x)|^p$ is Lebesque integrable over [a,b]. Also, the results which we shall obtain are applicable in this general case. However, we shall be concerned only with the case when f(x) is a continuous function.

For a sequence {x(n)} with transform $X(e^{j\omega})$, let us define the p-norm of $X(e^{j\omega})$ as

$$||X(e^{j\omega})||_p = ||g(x)||_p \quad 1 \leq p < \infty \qquad (3.43)$$

where g(x) is a function on [0,1] defined by

$$g(x) = X(e^{j2\pi x}) \qquad 0 \leq x \leq 1 \tag{3.44}$$

In other words

$$||X(e^{j\omega})||_p = \left[\frac{1}{2\pi}\int_0^{2\pi} |X(e^{j\omega})|^p d\omega\right]^{1/p} \qquad 1 \leq p < \infty \tag{3.45}$$

Next let us extend definitions (3.42) and (3.43) to values of p in the extended real number system by defining

$$||f(x)||_\infty = \lim_{p\to\infty} ||f(x)||_p \tag{3.46}$$

and

$$||X(e^{j\omega})||_\infty = ||g(x)||_\infty \tag{3.47}$$

For each p we can now define a class of signals consisting of those sequences whose transform satisfy

$$||X(e^{j\omega})||_p \leq 1 \tag{3.48}$$

We shall refer to signals satisfying (3.48) as p-norm constrained signals. Note that 1-norm constrained signals are simply class 2 signals. The following theorem provides us with a scaling method for these input signals for each p.

<u>Theorem 3.4</u>: Let f(x) and g(x) be continuous functions on the interval [0,1]. Then

(a) $||f(x)||_\infty = \max_{0 \leq x \leq 1} |f(x)|$

(b) $||f(x)g(x)||_1 \leq ||f(x)||_p ||g(x)||_q$ if $\frac{1}{p} + \frac{1}{q} = 1$, $1 \leq p, q \leq \infty$

(c) $||f(x)||_r \leq ||f(x)||_s$ if $1 \leq r \leq s \leq \infty$

Proof: The proof for (a) and (b) will be omitted here. They can be found in standard works on Lp-spaces. For instance for (a) see [23]. (b) is the well-known Hölder's inequality and can be found in [22] or [23]. We shall prove (c) assuming (b).

From (3.42)

$$||g(x)||_q = \left[\int_0^1 |g(x)|^q dx\right]^{1/q} \qquad 1 \leq q < \infty \qquad (3.49)$$

Taking g(x) = 1, $0 \leq x \leq 1$, we see that

$$||g(x)||_q = 1 \qquad 1 \leq q \leq \infty \qquad (3.50)$$

Therefore from (b)

$$||f(x)||_1 \le ||f(x)||_p \quad 1 \le p \le \infty \tag{3.51}$$

Given $1 \le r \le \infty$ let $u(x) = |f(x)|^r$. If s satisfies $1 \le r \le s \le \infty$, choose p so that $s = pr$. Applying (3.51)

$$||u(x)||_1 \le ||u(x)||_p \tag{3.52}$$

But

$$||u(x)||_1 = \int_0^1 |f(x)|^r dx$$

$$= \left(||f(x)||_r\right)^r \tag{3.53}$$

and

$$||u(x)||_p = \left[\int_0^1 |f(x)|^{rp} dx\right]^{1/p}$$

$$= \left[\int_0^1 |f(x)|^s dx\right]^{r/s}$$

$$= \left(||f(x)||_s\right)^r \tag{3.54}$$

Combining (3.52), (3.53) and (3.54), we have

$$||f(x)||_r \le ||f(x)||_s \quad 1 \le r \le s \le \infty \tag{3.55}$$

Q.E.D.

Using definitions (3.43) and (3.47) part (b) of Theorem 3.4 implies that

$$||F_i(e^{j\omega})X(e^{j\omega})||_1 \leq ||F_i(e^{j\omega})||_p ||X(e^{j\omega})||_q$$

$$\frac{1}{p} + \frac{1}{q} = 1$$

$$1 \leq p,\ q \leq \infty$$

$$1 \leq i \leq Ns \qquad (3.56)$$

But with input $\{x(n)\}$,

$$v_i(n) = \frac{1}{2\pi} \int_0^{2\pi} F_i(e^{j\omega})X(e^{j\omega})e^{j\omega n}d\omega \qquad (3.57)$$

Therefore

$$|v_i(n)| \leq \frac{1}{2\pi} \int_0^{2\pi} |F_i(e^{j\omega})X(e^{j\omega})|d\omega = ||F_i(e^{j\omega})X(e^{j\omega})||_1 \qquad (3.58)$$

Hence

$$|v_i(n)| \leq ||F_i(e^{j\omega})||_p ||X(e^{j\omega})||_q \qquad (3.59)$$

For q-norm constrained signals, i.e. if $||X(e^{j\omega})||_q \leq 1$, (3.59) suggests the following scaling method (p-norm scaling):

$$||F_i(e^{j\omega})||_p = 1 \qquad p = \frac{q}{q-1} \qquad i = 1,\ldots,Ns \qquad (3.60)$$

or stated in terms of $\{S_i\}$,

$$\prod_{j=1}^{i} S_j = \left[||\hat{F}_i(e^{j\omega})||_p \right]^{-1} \qquad i = 1,\ldots,Ns \qquad (3.61)$$

Notice that by virtue of part (a) of Theorem 3.4, ∞-norm scaling is just peak scaling which we have shown to be optimum for class 2, or 1-norm constrained, signals. Furthermore, part (c) of the theorem implies that

$$|x(n)| \leq ||X(e^{j\omega})||_p \leq ||X(e^{j\omega})||_q \qquad 1 \leq p \leq q \leq \infty \qquad (3.62)$$

Therefore we have the hierarchy of classes of signals:

class 1 $\supset$ class 2 $\supset$ p-norm constrained $\supset$ q-norm constrained

if $1 \leq p \leq q \leq \infty$.

In general class 1 and class 2 signals are the most useful to consider. 2-norm constrained signals with 2-norm scaling is useful when all inputs to a filter have finite energy bounded by a known value. For by Parseval's Theorem

$$\sum_n x^2(n) = \frac{1}{2\pi} \int_0^{2\pi} |X(e^{j\omega})|^2 d\omega \qquad (3.63)$$

Hence the energy of $\{x(n)\}$ is simply given by $\left(||X(e^{j\omega})||_2\right)^2$. Thus if the input signals are first scaled so that their maximum energy is 1.0 (or squared dynamic range of filter), then 2-norm scaling is sufficient to ensure no overflow.

However, the p-norm of any $X(e^{j\omega})$ is not defined if $|X(e^{j\omega})|^p$ is not integrable. Therefore all infinite-energy signals, such as infinite-duration periodic signals, are excluded from the 2-norm constrained class. Now any reasonable extension of the definition of p-norms to infinite-energy signals must be consistent with statement (c) of theorem 3.4 and (3.63). Thus we see that if $X(e^{j\omega})$ is the transform of an infinite-energy signal, then $||X(e^{j\omega})||_p$ is necessarily unbounded if $p \geq 2$. Hence infinite-energy signals are excluded from all p-norm constrained classes for $p \geq 2$.

Thus the concept of p-norm constrained signals has little practical usefulness for $p > 1$. Furthermore p-norm scaling for $p < \infty$ is only a sufficient method and no optimality properties can be proved for it. In this thesis we shall be concerned mainly with sum scaling and peak scaling methods for class 1 and class 2 signals respectively.

Clearly, sum scaling and peak scaling can be extended to apply to IIR filters. In fact theorems 3.1 and 3.2 can be readily generalized for IIR filters. However, the input sequence needed in theorem 3.1 to prove necessity in the case of IIR filters is an infinite-duration sequence extending to $-\infty$ with full dynamic range magnitudes, and signs that match those of $\{f_i(k)\}$ for some i. Clearly such an input sequence is highly improbable, hence class 1 signals have been deemed too restrictive a description for ordinary inputs to an IIR filter, resulting in too stringent a scaling method[12].

However, for FIR filters it is not difficult to find an input sequence within dynamic range which will require sum scaling to ensure no overflow, since only a small, finite portion of the sequence need match up with the $\{f_i(k)\}$'s. For example, if $F_1(z)$ has a zero with angle $\frac{\pi}{2} \leq \omega_o < \pi$, then all three samples of $\{f_1(k)\}$ have the same sign, hence an input sequence need only have three consecutive samples of value 1 before $|v_1(n)| = \sum_k |f_1(k)|$ for some n.

Because of the pessimistic nature of sum scaling for IIR filters and also the need to evaluate infinite sums, sum scaling has not been considered in analyses of roundoff noise in cascade IIR filters. However, none of these two reasons are applicable for FIR filters, hence sum scaling will not be neglected in this report. Of course, for any specific filter, the best choice of scaling method depends on the particular application.

4.0 Behavior of Roundoff Noise in Cascade Filters

The previous sections have established the basic properties of linear phase FIR filters, their implementation in cascade form, and methods by which roundoff noise can be analyzed. In the sections following we will make use of the tools developed to investigate the dependence of roundoff noise on sequential ordering of individual sections of a cascade filter and on other filter parameters.

However, let us first establish some facts concerning the characteristics of individual sections of a cascade filter. Let $H_i(z)$ be the transfer function of a filter section which synthesizes a pair of conjugate zeros at $z = re^{\pm j\omega_i}$, $r > 0$, $0 \leq \omega_i \leq \pi$. Then

$$H_i(z) = \beta(1-re^{j\omega_i}z^{-1})(1-re^{-j\omega_i}z^{-1})$$

$$= \beta(1-2r\cos\omega_i z^{-1}+r^2z^{-2}) \qquad (4.1)$$

Referring to (3.1) we have simply $\beta = b_{oi}$, therefore

$$H_i(z) = b_{oi}(1-2r\cos\omega_i z^{-1}+r^2z^{-2}) \qquad (4.2)$$

Now

$$H_i(e^{j\omega}) = b_{oi}(1 - 2r\cos\omega_i\, e^{-j\omega} + r^2e^{-j2\omega})$$

$$= b_{oi}[(r^2 - 2r\cos\omega_i e^{-j\omega} + r^2e^{-j2\omega}) + (1 - r^2)]$$

$$= b_{oi}[(r^2e^{j\omega} - 2r\cos\omega_i + r^2e^{-j\omega})e^{-j\omega} + (1 - r^2)]$$

$$= b_{oi}[2r(r\cos\omega - \cos\omega_i)e^{-j\omega} + (1 - r^2)] \tag{4.3}$$

If r = 1, we have simply

$$H_i(e^{j\omega}) = 2b_{oi}(\cos\omega - \cos\omega_i)e^{-j\omega} \tag{4.4}$$

which has linear phase, as expected.

Clearly, from (4.3)

$$|H_i(e^{j\omega})| \le |b_{oi}|(|2r(r\cos\omega - \cos\omega_i)| + |1 - r^2|)$$

$$\le |b_{oi}|(2r\max_{\omega}|r\cos\omega - \cos\omega_i| + |1 - r^2|) \tag{4.5}$$

But

$$\max_{\omega} |r \cos \omega - \cos \omega_i| = r + |\cos \omega_i| \tag{4.6}$$

Therefore

$$|H_i(e^{j\omega})| \leq |b_{oi}|[2r(r + |\cos \omega_i|) + |1 - r^2|] \tag{4.7}$$

Now from (4.3)

$$H_i(1) = b_{oi}[2r(r - \cos \omega_i) + (1 - r^2)] \tag{4.8}$$

and

$$H_i(-1) = b_{oi}[2r(r + \cos \omega_i) + (1 - r^2)] \tag{4.9}$$

Thus assuming $r \leq 1$, $|H_i(1)|$ is simply the right hand side of (4.7) for $\cos \omega_i \leq 0$ while $|H_i(-1)|$ is the right hand side of (4.7) for $\cos \omega_i \geq 0$. Hence

$$\max_{\omega} |H_i(e^{j\omega})| = \begin{cases} |H_i(-1)| & 0 \leq \omega_i < \frac{\pi}{2} \\ |H_i(1)| & \frac{\pi}{2} \leq \omega_i \leq \pi \end{cases} \tag{4.10}$$

or

$$\max_{\omega} |H_i(e^{j\omega})| = |b_{oi}|(2r(r + |\cos \omega_i|) + (1 - r^2))$$

$$= |b_{oi}|(2r\ |\cos \omega_i| + 1 + r^2) \qquad (4.11)$$

If $r > 1$, let $s = r^{-1}$ and let $H_j(z)$ be a section which produces zeros at $z = se^{\pm j\omega_i}$. Then $H_j(e^{j\omega})$ satisfies (4.10). But then

$$H_i(z) = b_{oi}(1 - 2r \cos \omega_i z^{-1} + r^2 z^{-2})$$

$$= b_{oi}\ r^2 z^{-2}(1 - 2s \cos \omega_i\ z + s^2 z^2)$$

$$= \frac{b_{oi}}{b_{oj}}\ r^2 z^{-2}\ H_j(z^{-1}) \qquad (4.12)$$

Therefore

$$|H_i(e^{j\omega})| = \left| \frac{b_{oi}}{b_{oj}} \right| r^2 \left| H_j(e^{-j\omega}) \right|$$

$$= \left| \frac{b_{oi}}{b_{oj}} \right| r^2 \left| H_j(e^{j\omega}) \right| \qquad (4.13)$$

Hence $|H_i(e^{j\omega})|$ and $|H_j(e^{j\omega})|$ are proportional to each other, thus $|H_i(e^{j\omega})|$ also satisfies (4.10).

Next consider the case when the zeros of $H_i(z)$ are at $z = r, r^{-1}$. Then

$$H_i(z) = b_{oi}(1 - rz^{-1})(1 - r^{-1}z^{-1})$$

$$= b_{oi}(1 - (r + r^{-1})z^{-1} + z^{-2}) \qquad (4.14)$$

and

$$H_i(e^{j\omega}) = b_{oi}(e^{j\omega} - (r + r^{-1}) + e^{-j\omega})e^{-j\omega}$$

$$= b_{oi}(2 \cos \omega - (r + r^{-1}))e^{-j\omega} \qquad (4.15)$$

From (4.15) we see that (4.10) is again satisfied where $\omega_i = 0$ if $r > 0$ and $\omega_i = \pi$ if $r < o$.

Finally, if $H_i(z)$ synthesizes only one zero at $z = -1$, then

$$H_i(e^{j\omega}) = b_{oi}(1 + e^{-j\omega})$$

$$= 2b_{oi} \cos \frac{\omega}{2} e^{-j \frac{\omega}{2}} \qquad (4.16)$$

Thus again, (4.10) is satisfied ($\omega_i = \pi$). Hence (4.10) holds for every section of an FIR filter.

Next, we establish that for all sections

$$\sum_{j=0}^{2} |b_{ji}| = \max_{\omega} |H_i(e^{j\omega})| \qquad (4.17)$$

For zeros at $re^{\pm j\omega_i}$, we have from (4.2)

$$\sum_{j=0}^{2} |b_{ji}| = |b_{oi}|(1 + 2r\,|\cos \omega_i| + r^2) \qquad (4.18)$$

which compared with (4.11) shows that (4.17) is true. If the zeros are at $r^{\pm 1}$, from (4.14)

$$\sum_{j=0}^{2} |b_{ji}| = |b_{oi}|(2 + |r + r^{-1}|) \qquad (4.19)$$

which is seen to be $\max_{\omega} |H_i(e^{j\omega})|$ from (4.15). Finally, a sole zero at $z = -1$ yields

$$\sum_{j=0}^{2} |b_{ji}| = 2|b_{oi}| \qquad (4.20)$$

which by (4.16) again satisfies (4.17).

Thus (4.17) is established. The next few sections will present some of the findings on the behavior of roundoff noise in cascade filters. All filter examples used will be extraripple filters. Figure 4.1 shows a typical extraripple filter and the parameters used to define it. The same symbolic terminology as defined in Figure 4.1 will be used to define all filters in the remainder of this thesis report. Furthermore, all sampling rates will be normalized to unity.

4.1 Dependence of Roundoff Noise on Section Ordering

In section 3.3 it was shown that given a transfer function H(z) to be realized in cascade form and the order in which the factors of H(z) are to be synthesized, there remains N_s degrees of freedom (including gain of filter) in the choice of filter coefficients, where N_s is the number of sections of the filter. Scaling methods were developed to fix these N_s degrees of freedom, and two particular methods, viz. sum scaling and peak scaling, were shown to be optimum for the particular classes of input signals which they assume. These scaling methods will be applied in this section, so that ordering of filter sections will be the sole variable in our investigations.

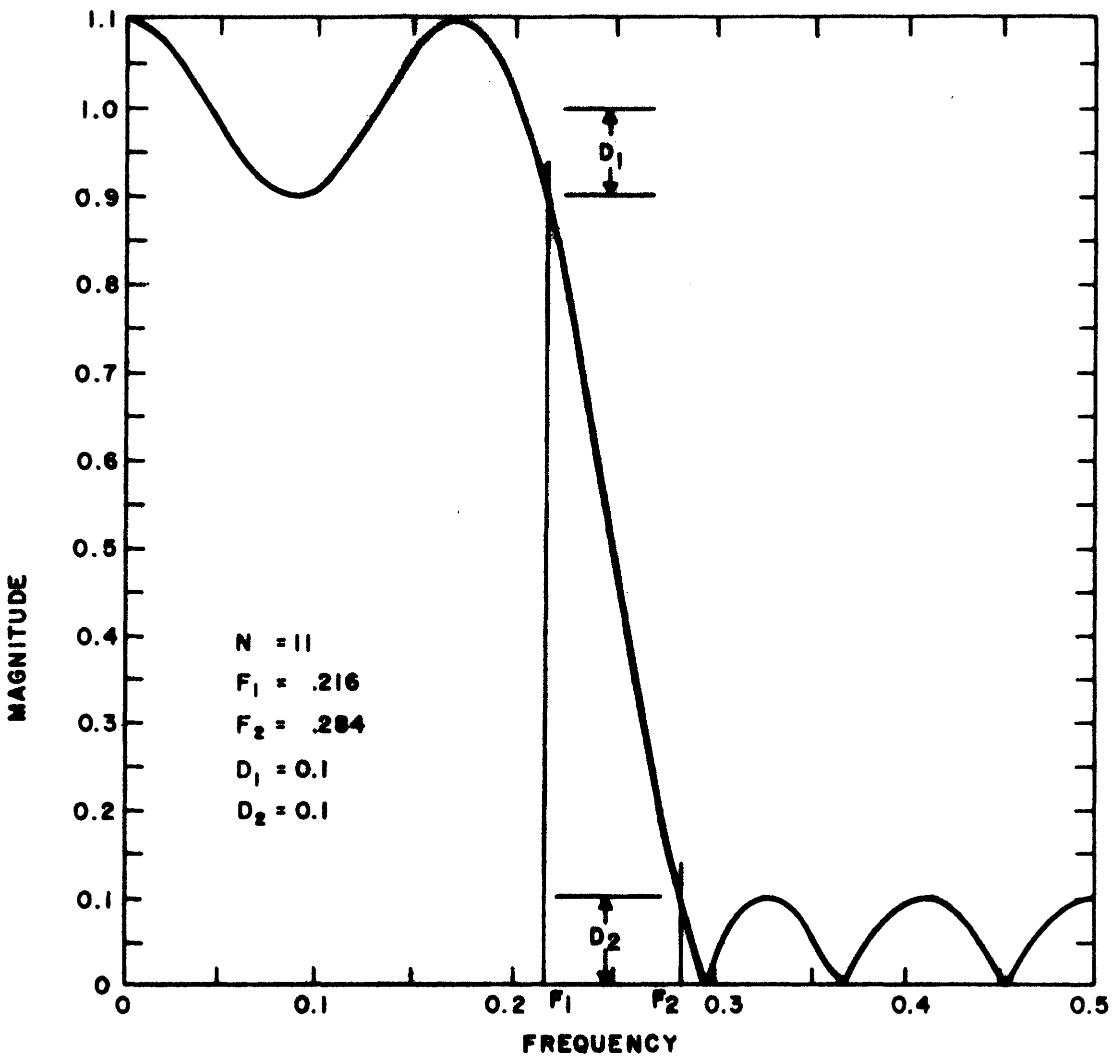


FIG. 4.1 – DEFINITION OF FILTER PARAMETERS

The prime issues in the realization of filters in cascade form are threefold -- scaling, ordering, and section configuration. Because of the simplicity of a 2^{nd} order FIR filter, there is little freedom in the choice of a structure for the sections of a cascade filter. We have assumed thus far the configurations shown in figures 2.2 to 2.4 because they turn out to be the most useful. Other possible configurations will be discussed later on. The major concern of this section is the ordering of sections. Unlike the scaling problem, no workable optimal solution (in terms of feasibility) to the ordering problem has yet been found for cascade filters in general. The dependence of output roundoff noise variance on section ordering given a scaling method is so complex that no simple indicators are provided to assist in any systematic search for an ordering with lowest noise. Any attempt to find the noise variances for all possible orderings of a filter involves on the order of $N_s!$ evaluations, which clearly becomes prohibitive even for moderately large values of N_s. Thus there is little doubt that optimal ordering is by far the most difficult issue to deal with in the design of cascade filters.

Since finding an optimal solution to the ordering problem is very difficult, if not impossible by any feasible means, for all but very low-order filters, it is important to find out how closely a suboptimal solution can approach the optimum and how difficult it would be to find a satisfactory suboptimal solution. Even this concern, however, would be rather unfounded if the roundoff noise level produced by a filter were rather insensitive to ordering. For then the difference in performance between any two orderings may not be sufficient to cause any concern. However, Schüssler has demonstrated that quite the contrary is true[14]. He showed a 33-point FIR filter which, ordered one way, produces $\sigma^2 = 2.4Q^2$ while ordered another way yields $\sigma^2 = 1.5 \times 10^8 Q^2$.* In terms of the noise figure defined in (3.15), this represents a difference of 1.6 bits versus 14.6 bits of noise. The difference is drastic indeed. Hence the problem of finding a proper ordering of sections in the design of a cascade filter cannot be evaded.

An important question to pursue in investigating suboptimal solutions is whether or not there exists some general pattern in which values of noise variances distribute themselves over different orderings. For example, for the 33-point filter mentioned above, are

---

* Assumes all products in each section summed before rounded.

all noise values between the two extremes demonstrated equally likely to occur in terms of occurring in the same number of orderings, or perhaps only a few pathological orderings have noise variances as high as that indicated. On the other hand perhaps only very few orderings have noise variances close to the low value, in which case an optimum solution would be very valuable while a satisfactory suboptimal solution may be just as difficult to obtain as the optimum.

In this section, we will attempt to answer these questions by investigating filters of sufficiently low order so that calculating noise variances of $N_s!$ different orderings is not an unfeasible task. The implications of results obtained will then be generalized. The methods and results will now be presented.

The definitions of sum scaling and peak scaling in section 3.3 indicate that for FIR filters sum scaling is much simpler to perform than peak scaling. To achieve peak scaling, the maxima of the functions $\hat{F}_i(e^{j\omega})$ must be found for all i given an ordering. Even using the FFT this represents considerably more calculations than finding $\sum_{k=0}^{21} |\hat{f}_i(k)|$ for all i. In the 33-point filter mentioned above, Schüssler used peak scaling on both the orderings. We will show in the next section that, given a filter, peak and sum scaling yield noise variances that are not very different (within the same order of magnitude), and, in fact, experimental results indicate that they are essentially in a constant ratio to one another independent of ordering of sections. Hence the general characteristics of the distribution of roundoff noise with respect to orderings should be quite independent of the type of scaling performed. In order to save computation time, sum scaling will be used in our investigations.

Returning to the question of section configuration, for IIR filters Jackson [12] has introduced the concept of transpose configurations to obtain alternate structures for filter sections. However, the application of this concept to Figure 2.2 yields the structure shown in Fig. 4.2, which is seen to have the same noise characteristics as the structure in Fig. 2.2 since by the whiteness assumption on the noise sources, delays have no effect on them. Therefore Fig. 4.2 need not be considered. The only other significant alternate configuration for Fig. 2.2 is shown in Fig. 4.3. The counterpart for Fig. 2.3 is Fig. 4.4, valid when $b_{oi} = b_{2i}$. Both of these new configurations have exactly the same number of multipliers as the original ones. However, one noise source is moved from the output to essentially the input of the section. Thus it is advantageous to use the structures in Fig. 4.3 and 4.4 for the i^{th} section when

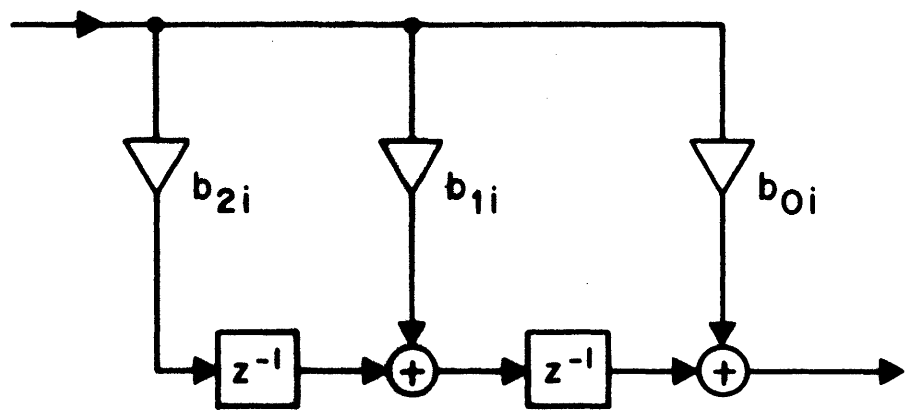


FIG. 4.2 – TRANSPOSE CONFIGURATION OF FIG. 2.2.

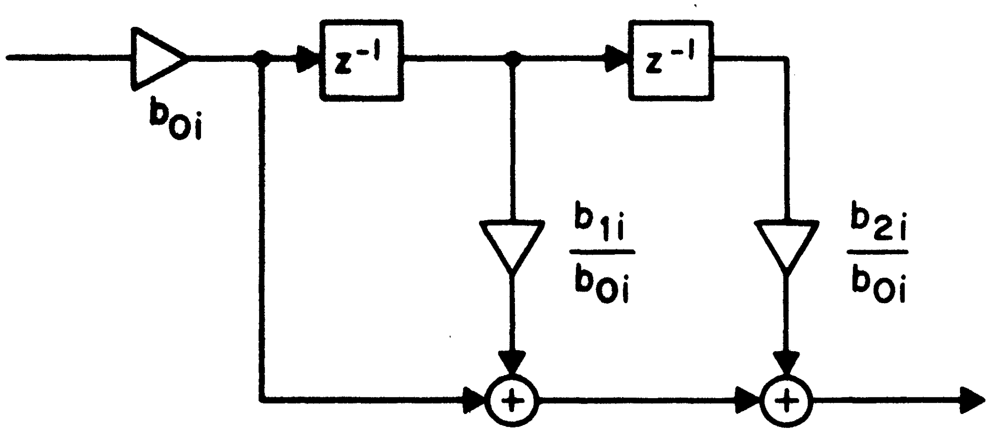


FIG. 4.3 – ALTERNATE CONFIGURATION TO FIG. 2.2.

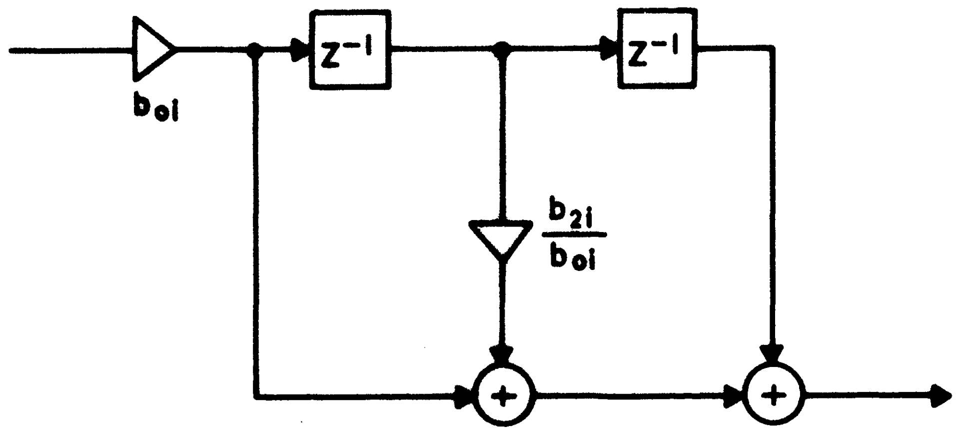


FIG. 4.4 ALTERNATE CONFIGURATION TO FIG. 2.3

$$\frac{1}{b_{oi}^{2}} \sum_k g_{i-1}^2(k) < \sum_k g_i^2(k) \qquad (4.21)$$

where $\{g_i(k)\}$ is as defined in section 3.2. However, in order to have no error-causing internal overflow when the input and output of a section are properly constrained, Fig. 4.3 and 4.4 can be used only when $b_{oi} \leq 1$. If $b_{oi} > 1$, either four multipliers become required or Fig. 4.3 reduces to Fig. 2.2.

In the investigations that follow, for each section of a filter the configuration among Figs. 2.2, 2.3, 4.3 and 4.4 which is applicable and results in the least noise will be employed. It turns out that this flexibility in the choice of configuration has little effect on the noise distribution characteristics of a filter. For low-noise orderings the configurations of Fig. 2.2 and 2.3 are almost always more advantageous. For high-noise orderings the alternate configurations help to reduce the noise variance, but the difference is comparatively small. Thus in actual filter implementations the structures in Fig. 4.3 and 4.4 may be ignored.

Figure 4.5 shows the flow diagram of a computer subroutine which is used to accomplish scaling, choice of configuration, and output noise variance calculation

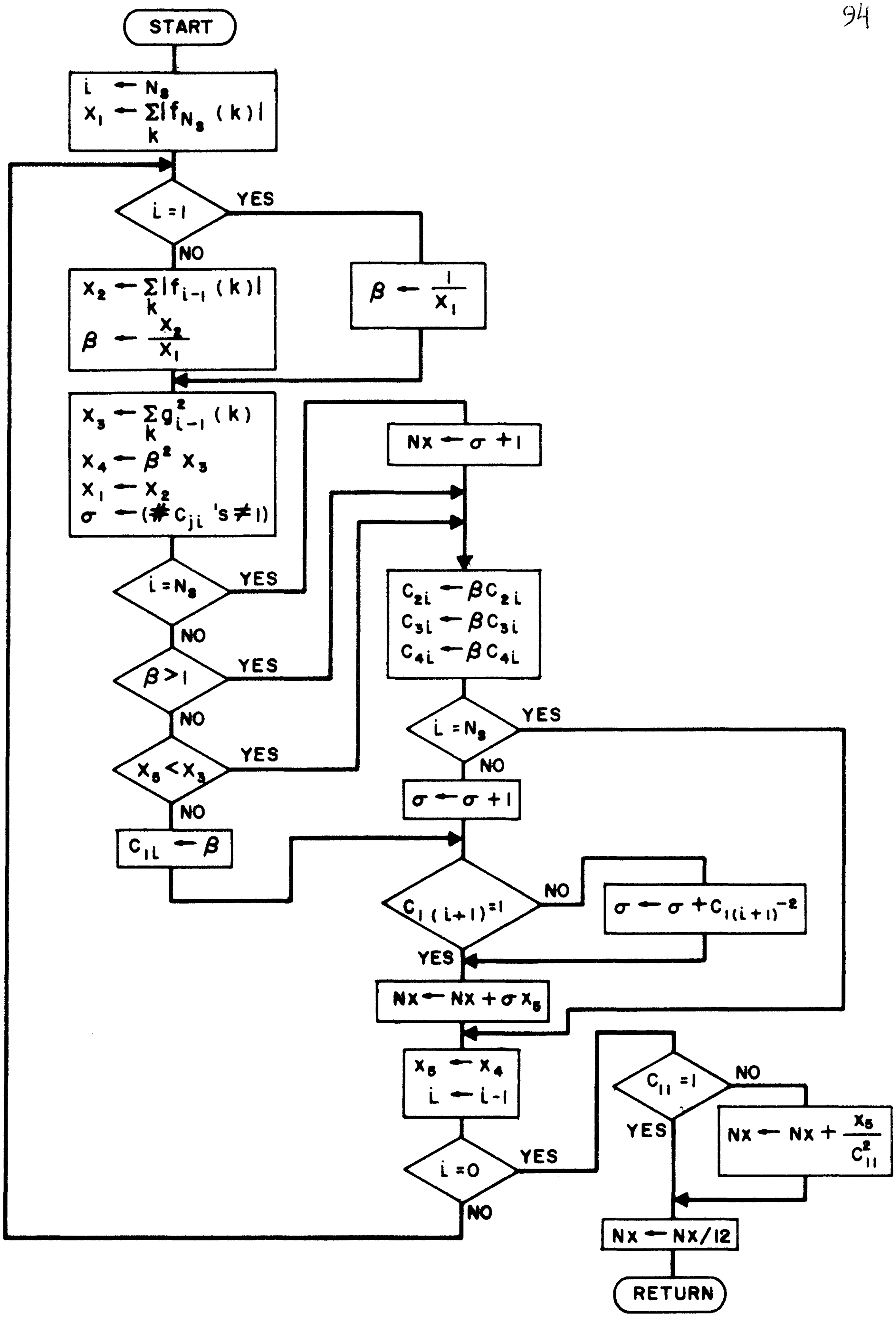


FIG. 4.5 — FLOW CHART OF SCALING AND NOISE CALCULATION SUBROUTINE

given a filter and its ordering. The input to the subroutine consists of Ns (the number of sections) and the sequence $\{C_{ji}, 1 \leq j \leq 4, 1 \leq i \leq N_s\}$, whose elements are unscaled coefficients of the filter, defined by

$$H_i(z) = C_{1i}(C_{2i}+C_{3i}z^{-1}+C_{4i}z^{-2}) \quad 1 \leq i \leq N_s \qquad (4.22)$$

where $H_i(z)$ is, as usual, the i^{th} section in the filter cascade. The sequences $\{f_i(k)\}$ and $\{g_i(k)\}$ in Fig. 4.5 are as previously defined in sections 3.2 and 3.3. The coefficients $\{C_{ji}\}$ on input are assumed to be normalized so that for all i, $C_{1i} = 1$ and at least one of C_{2i} and C_{4i} equals 1. On return $\{C_{ji}\}$ contains the scaled coefficients and NX is the value of output noise variance computed in units of Q^2, where Q is the quantization step size of the filter.

Using this subroutine the noise output of all possible orderings of several FIR filters ranging from $N_s = 3$ to $N_s = 7$ was investigated. Before the results are presented, we shall prove one characteristic of sum scaled filters which, for most filters, reduces the total number of orderings that differ in output noise to at most $N_s!/2$.

Theorem 4.1: Let $\{H_i(z)\}$ and $\{H_i'(z)\}$ be two orderings for H(z), both scaled by sum scaling, thus

$$H(z) = \prod_{i=1}^{N_s} H_i(z) = \prod_{i=1}^{N_s} H_i'(z)$$

Suppose z_i^{-1} is a zero of $H_i'(z)$ whenever z_i is a zero of $H_i(z)$. Then filters ordered according to $\{H_i(z)\}$ and $\{H_i'(z)\}$ produce identical output noise variances.

Proof:

We first establish that if $\{x(n)\}$ is a sequence of length N+1, $\{y(n)\}$ a sequence of length M+1, and

$$p(n) = x(n)*y(n) \qquad (* \text{ denotes convolution})$$

$$q(n) = x(N-n)*y(M-n)$$

then

$$p(n) = q(M+N-n) \tag{4.23}$$

To see this note that

$$q(n) = \sum_{k=0}^{N} x(N-k)y(M-n+k)$$

$$= \sum_{m=0}^{N} x(m)y(M+N-n-m)$$

Hence immediately

$$q(M+N-n) = \sum_{m=0}^{N} x(m)y(n-m) = p(n)$$

Now let $\hat{H}_i(z)$ and $\hat{H}_i'(z)$ be normalized transfer functions as defined in (3.16) so that

$$H_i(z) = S_i \hat{H}_i(z)$$

$$H_i'(z) = S_i' \hat{H}_i'(z)$$

Also let $(\hat{H}_i(z), \{\hat{h}_i(k)\})$, $\left(\prod_{j=1}^{i} \hat{H}_j(z), \{\hat{f}_i(k)\}\right)$, and $\left(\prod_{j=i+1}^{N_s} H_j(z), \{g_i(k)\}\right)$ be transfer function - impulse response pairs, and the same when primes are added. Now for all i such that $H_i(z)$ has 2 zeros (see Theorem 2.1)

$$\frac{\hat{h}_i'(0)}{\hat{h}_i(2)} = \frac{\hat{h}_i'(1)}{\hat{h}_i(1)} = \frac{\hat{h}_i'(2)}{\hat{h}_i(0)}$$

But because of the normalization condition (3.17),

$$\hat{h}_i(k) = \hat{h}_i'(M_i-k) \tag{4.24}$$

where M_i+1 is the length of the sequence $\{\hat{h}_i(k)\}$. In the present case $M_i=2$. If $H_i(z)$ has only 1 zero (viz. at $z=-1$), then

$$\hat{H}_i(z) = \hat{H}_i'(z) = \frac{1+z^{-1}}{2}$$

hence (4.24) is again satisfied.

Now let N_i+1 be the length of sequence $\{\hat{f}_i(k)\}$. We next show by induction that

$$\hat{f}_i(k) = \hat{f}_i'(N_i-k) \qquad 1 \leq i \leq N_s \tag{4.25}$$

For i = 1,

$$\left.\begin{aligned} \hat{f}_1(k) &= \hat{h}_1(k) \\ \hat{f}_1'(k) &= \hat{h}_1'(k) \end{aligned}\right\} \quad k = 0,1,2$$

Hence by (4.24), (4.25) is true.

Suppose (4.25) is true for $i=m$, $m < N_s$. Now

$$\hat{f}_{m+1}(k) = \hat{f}_m(k) * \hat{h}_{m+1}(k)$$

$$\hat{f}_{m+1}'(k) = \hat{f}_m'(k) * \hat{h}_{m+1}'(k)$$

Using (4.24) and the induction hypothesis,

$$\hat{h}_{m+1}(k) = \hat{h}'_{m+1}(M_{m+1}-k)$$

$$\hat{f}_m(k) = \hat{f}'_m(N_m-k)$$

Therefore by (4.23)

$$\hat{f}'_{m+1}(k) = \hat{f}_{m+1}(M_{m+1}+N_m-k)$$

$$= \hat{f}_{m+1}(N_{m+1}-k)$$

Thus (4.25) holds for all i.

Clearly then

$$\sum_{k=0}^{N_i} |\hat{f}_i(k)| = \sum_{k=0}^{N_i} |\hat{f}'_i(k)| \qquad 1 \le i \le Ns$$

Therefore by the definition of sum scaling

$$S_i = S'_i \qquad i = 1,\ldots,N_s \tag{4.26}$$

Using (4.26) we can show in exactly the same way that

$$g_i(k) = g'_i(T_i-k) \qquad i = 0,\ldots,N_s-1$$

where T_i+1 is the length of $\{g_i(k)\}$.

Hence

$$\sum_{k=0}^{T_i} g_i^2(k) = \sum_{k=0}^{T_i} g_i'^2(k) \qquad 0 \le i \le N_s-1$$

Since for all i the i^{th} section of both orderings have the same number of noise sources, we have by (3.5) and (3.7) that their output noise variances are identical.

Q.E.D.

A stronger result than Theorem 4.1 can actually be proved for the case of peak scaling. In particular we can show that if two orderings of a filter differ only in that in one ordering a pair of sections which have reciprocal zeros are interchanged in position, then with peak scaling both these orderings yield the same noise variance. This is easily seen by noting that if $H_i(z)$ and $H_j(z)$ are two sections having reciprocal zeros, by (4.13)

$$|H_i(e^{j\omega})| = \beta |H_j(e^{j\omega})| \tag{4.27}$$

where β is a proportionality constant. Hence the normalized spectra satisfy

$$|\hat{H}_i(e^{j\omega})| = |\hat{H}_j(e^{j\omega})| \tag{4.28}$$

But from (3.10) and (3.37) peak scaling and output noise variance depend only on the magnitude of the individual sections' frequency spectra, hence (4.28) shows that exchanging the positions of $H_i(z)$ and $H_j(z)$ in an ordering does not change the filter's output noise variance under peak scaling.

Since we are concerned only with sum scaling, we shall restrict attention to Theorem 4.1. The result of this theorem can be used in our investigation of all possible noise outputs of a filter by choosing a pair of sections which synthesize reciprocal zeros and then ignoring all orderings in which a particular one of these sections precedes the other in our search over all orderings. To see that this does not change the noise distribution, note that if we divide all orderings into two groups, according to the order in which the pair of sections chosen occurs, then by Theorem 4.1 there exists a one-to-one correspondence in terms of noise output between each ordering of one group and some ordering of the other group. In this way only $N_s!/2$ different orderings need to be scaled and have their output noise variances computed. Of course, the applicability of this procedure depends on the existence of such a pair of sections.

Using the methods and procedures described in this section, the noise distributions of 27 different linear phase, low-pass extraripple filters were investigated. 22 of these filters were 13-point filters, since N=13 represents a good filter length to work with. 13-point filters have six sections each, corresponding to 6! or 720 possible orderings of sections. By reducing redundancy via Theorem 4.1, the number of orderings that are necessary to investigate reduces to 360 for all but 2 of the 22 filters.

The results of the investigations for all 27 filters will eventually be presented. Meanwhile, we focus attention on a typical 13-point filter. Chosen as an example is a filter with 4 ripples in the passband, 3 ripples in the stopband, and passband and stopband tolerances of 0.1 and 0.01 (or -40 dB) respectively. By passband and stopband tolerances it is meant the maximum height of ripples in the respective frequency bands. (For definition of terminology see Fig. 4.1 and page 41). Plots of the impulse response, step response, and magnitude frequency response of the filter are shown in Fig. 4.6. Figure 4.7 shows the positions of the zeros of the filter in the upper half of the z-plane. Each section of the filter is given a number for identification. The zeros that a

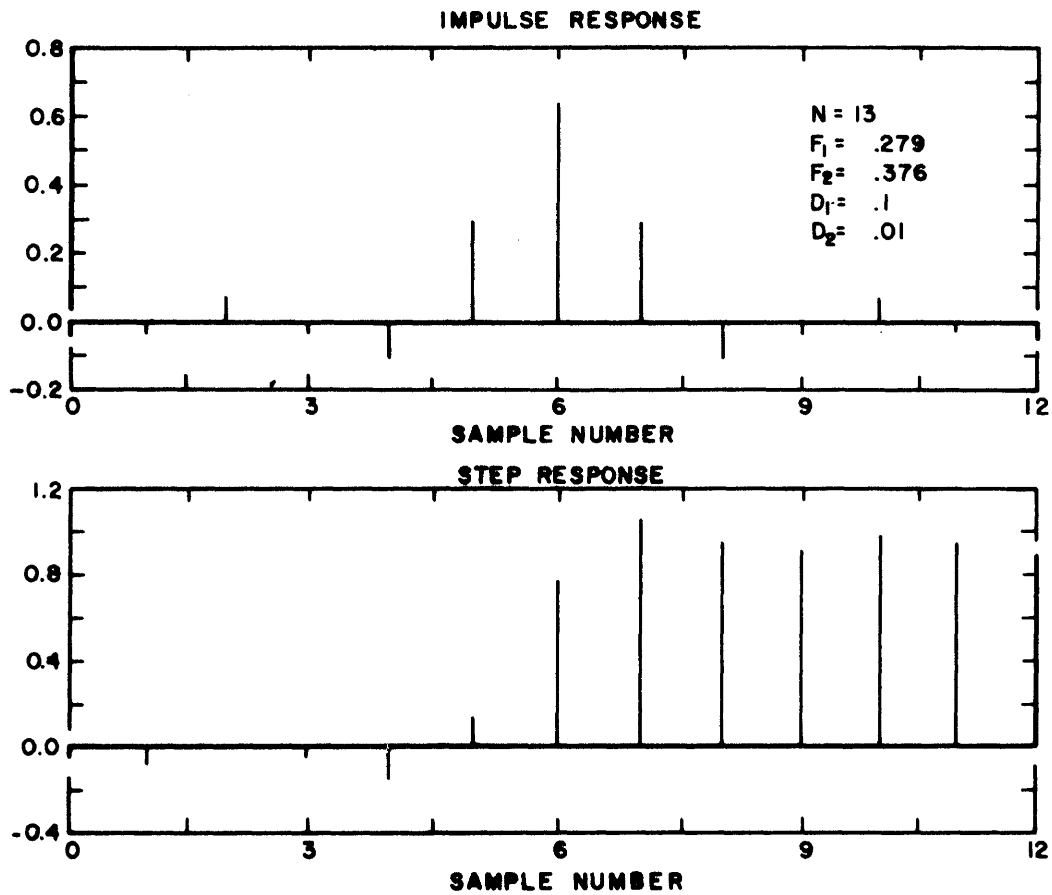


FIG. 4.6 (a) — IMPULSE RESPONSE AND STEP RESPONSE OF TYPICAL 13 POINT FILTER

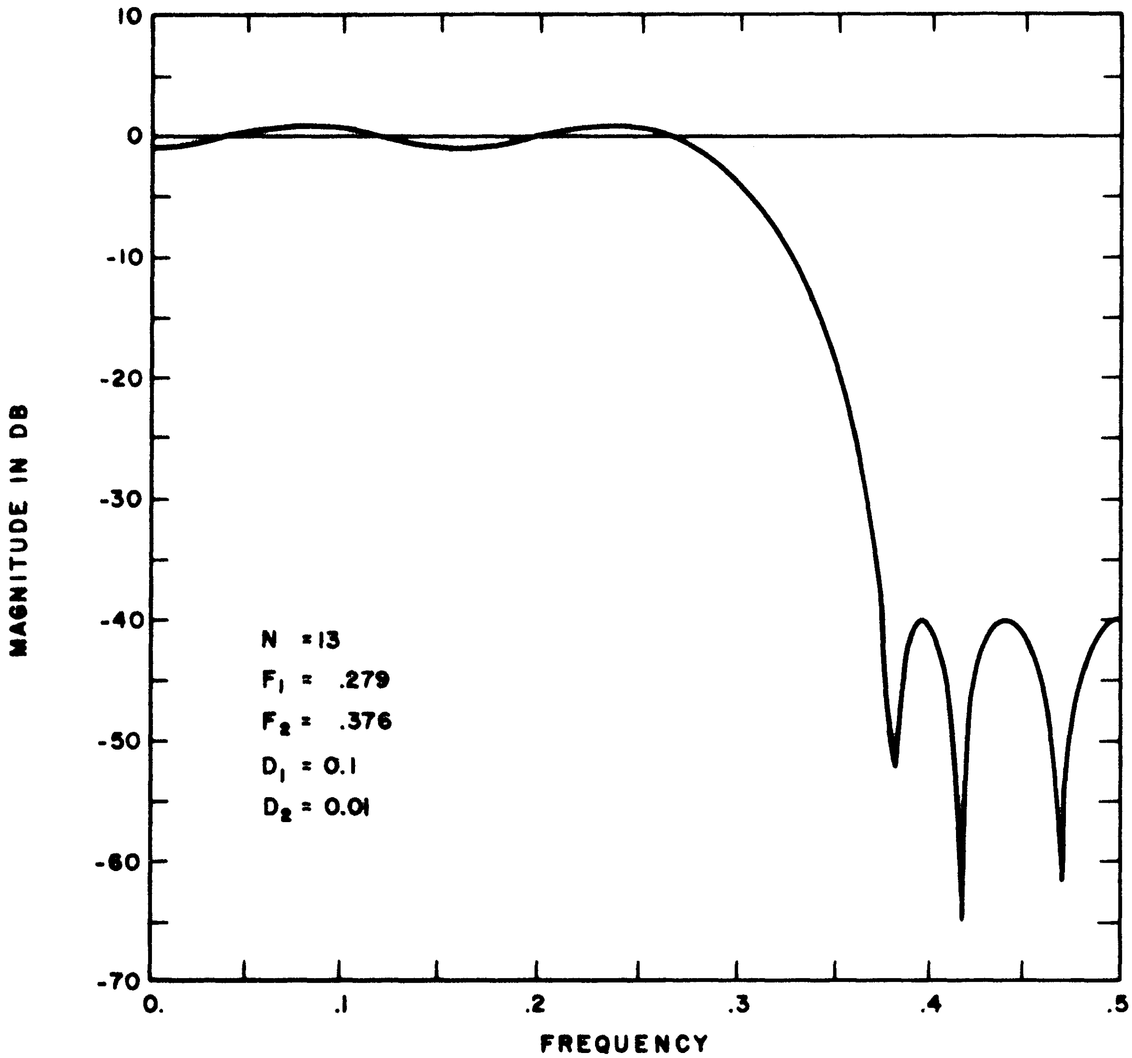


FIG. 4.6(b) — MAGNITUDE FREQUENCY RESPONSE OF TYPICAL 13 POINT FILTER.

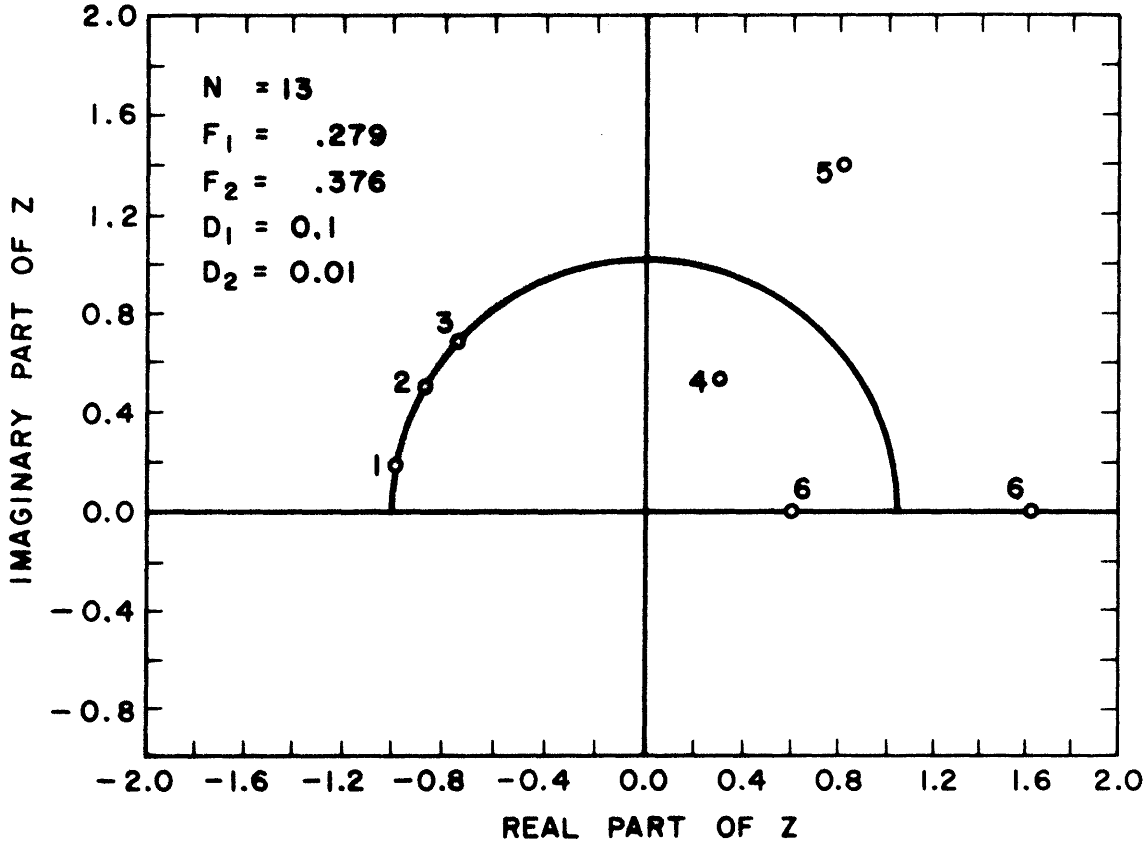


FIG. 4.7 – ZEROS OF FILTER OF FIG. 4.6

section synthesizes are given the same number, and these are shown in Fig. 4.7. Appendix A.1 shows a list in order of increasing noise magnitude of all 360 orderings investigated and their corresponding output noise variances in units of Q^2, computed according to Fig. 4.5. On the Honeywell 6070 machine the total computation time required amounted to approximately 12 seconds. A histogram plot of the noise distribution is shown in Fig. 4.8, and a cumulative distribution plot is shown in Fig. 4.9.

Two characteristics of the histogram shown in Fig. 4.8 are of special importance because they are common to similar plots for all the filters investigated. First of all, most significant is the shape of the distribution. We see that most orderings have very low noise compared to the maximum value possible. In fact, the lowest range of noise variances, in this case between zero and $2Q^2$, is the most probable range in terms of the number of orderings which produce noise variances in this range. The distribution is seen to be highly skewed, with an expected value very close to the low noise end, in this case equal to $19.5Q^2$. In fact, from the cumulative distribution we see that approximately two-thirds of the orderings have noise variances less than

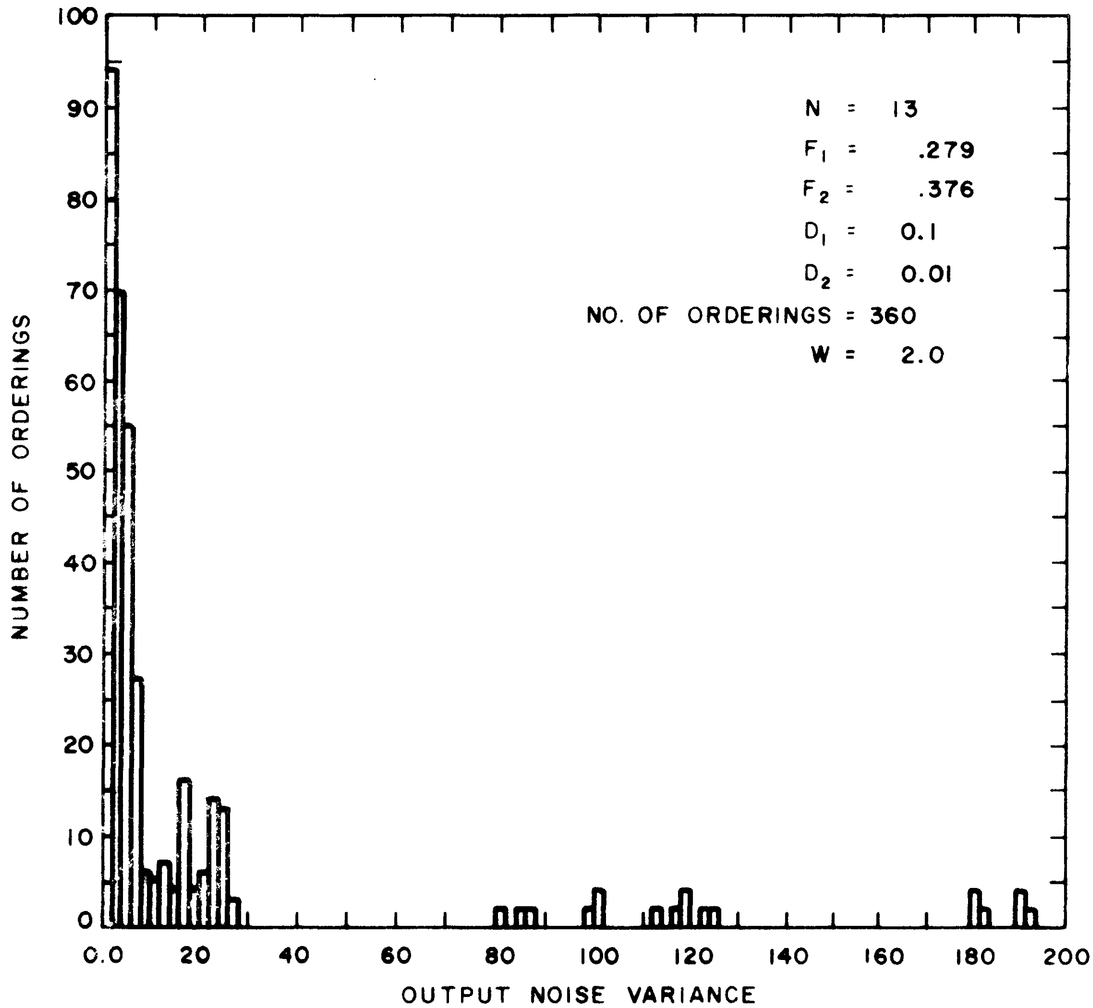


FIG. 4.8 – NOISE DISTRIBUTION HISTOGRAM OF FILTER OF FIG. 4.6

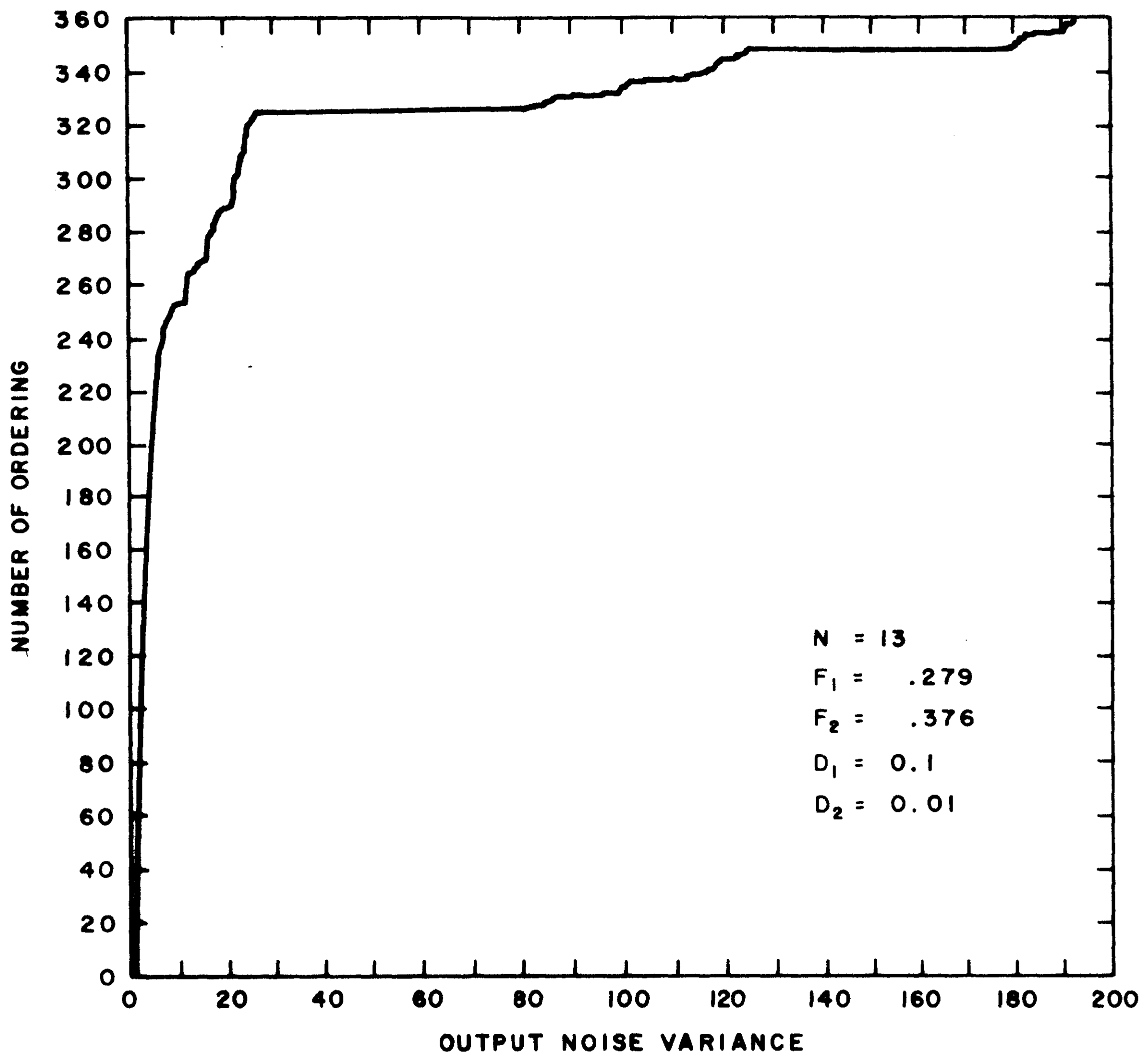


FIGURE 4.9 - CUMULATIVE NOISE DISTRIBUTION OF FILTER OF FIGURE 4.6

4% of the maximum while nine-tenths of them have noise variances less than 14% of the maximum.

The second characteristic is that large gaps occur in the distribution so that noise values within the gaps are not produced by any orderings. While Fig. 4.8 shows this effect only for the higher noise values, a more detailed plot of the distribution in the range from zero to $28Q^2$, as in Fig. 4.10, shows that gaps also occur for lower noise values. Thus noise values tend to occur in several levels of clusters. These observations provide us with the general picture of clusters of noise values which move apart very rapidly as the magnitude of the noise values increases, thus forming a highly skewed noise distribution.

The significance of these results is far-reaching. Given a filter, because of the large abundance of orderings which yield almost the lowest noise variance possible, we conclude that it should not be too difficult to devise a feasible algorithm which will yield an ordering whose noise variance is very close to the minimum. Thus as far as designing practical cascade filters is concerned, it really is not crucial that the optimum ordering be found. In fact, it may be by far more advantageous to use a suboptimal method which can rapidly choose an

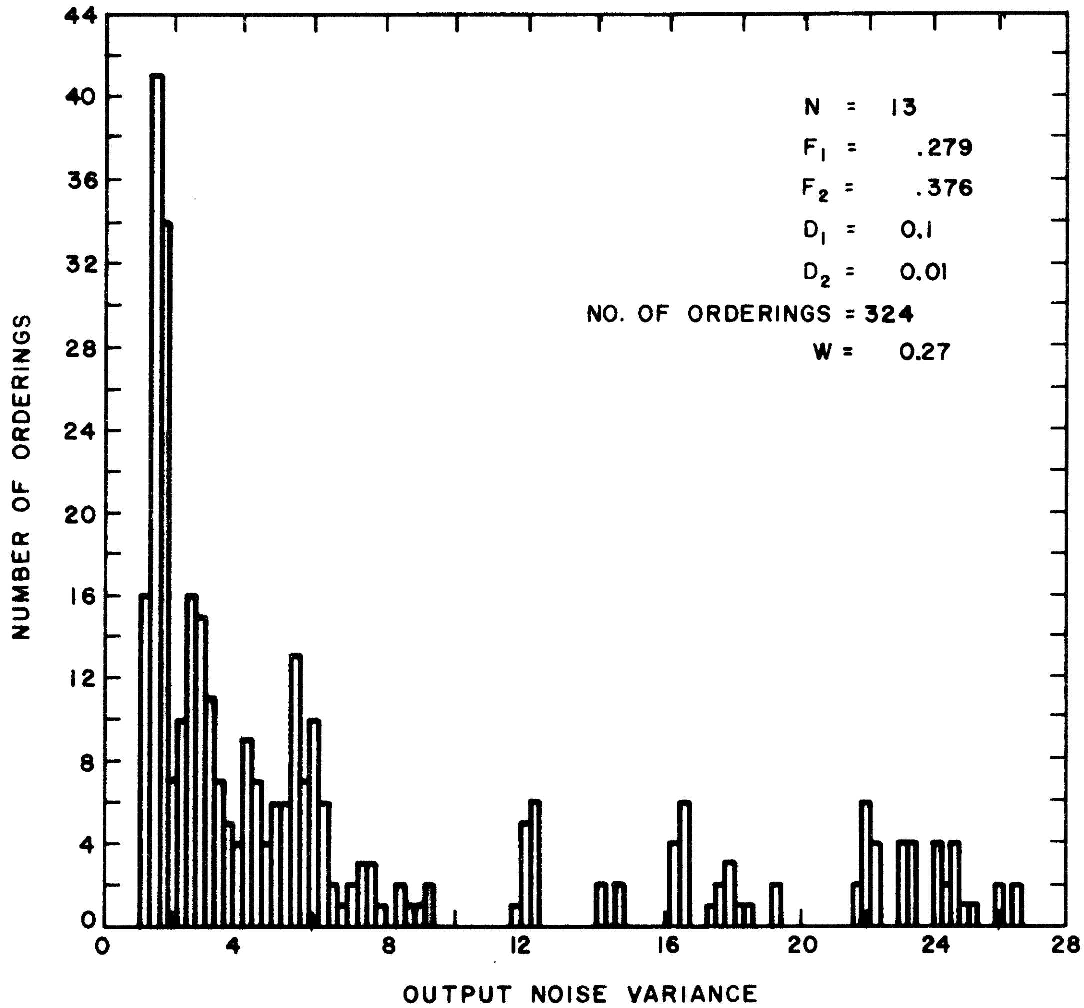


FIG. 4.10 – DETAILED NOISE DISTRIBUTION HISTOGRAM OF FILTER OF FIG. 4.6

ordering that is satisfactory than to try to find the optimum. The amount gained by finding the optimum solution is probably at best not worth the extra effort from the design standpoint. At least up to the present no simple method for finding an optimum ordering has been found.

In section 5.0 we will present a suboptimal method which given a filter yields a low-noise ordering efficiently and has been successfully applied to over 50 filters. But before we do that, we will investigate further the behavior of roundoff noise with respect to scaling and other filter parameters. Also, we will try to understand more on the nature of high noise and low noise orderings, so that they might be more easily recognized.

Before we end this section, we present the noise distribution histograms of an 11-point and two more 13-point filters, in Figs. 4.11 to 4.13. These are seen to exhibit all the characteristics discussed above. The major difference among the noise distributions for the three 13-point filter examples presented lies in the magnitude of the maximum and average noise variances. In fact, the differences are drastic. These differences will be accounted for in section 4.3.

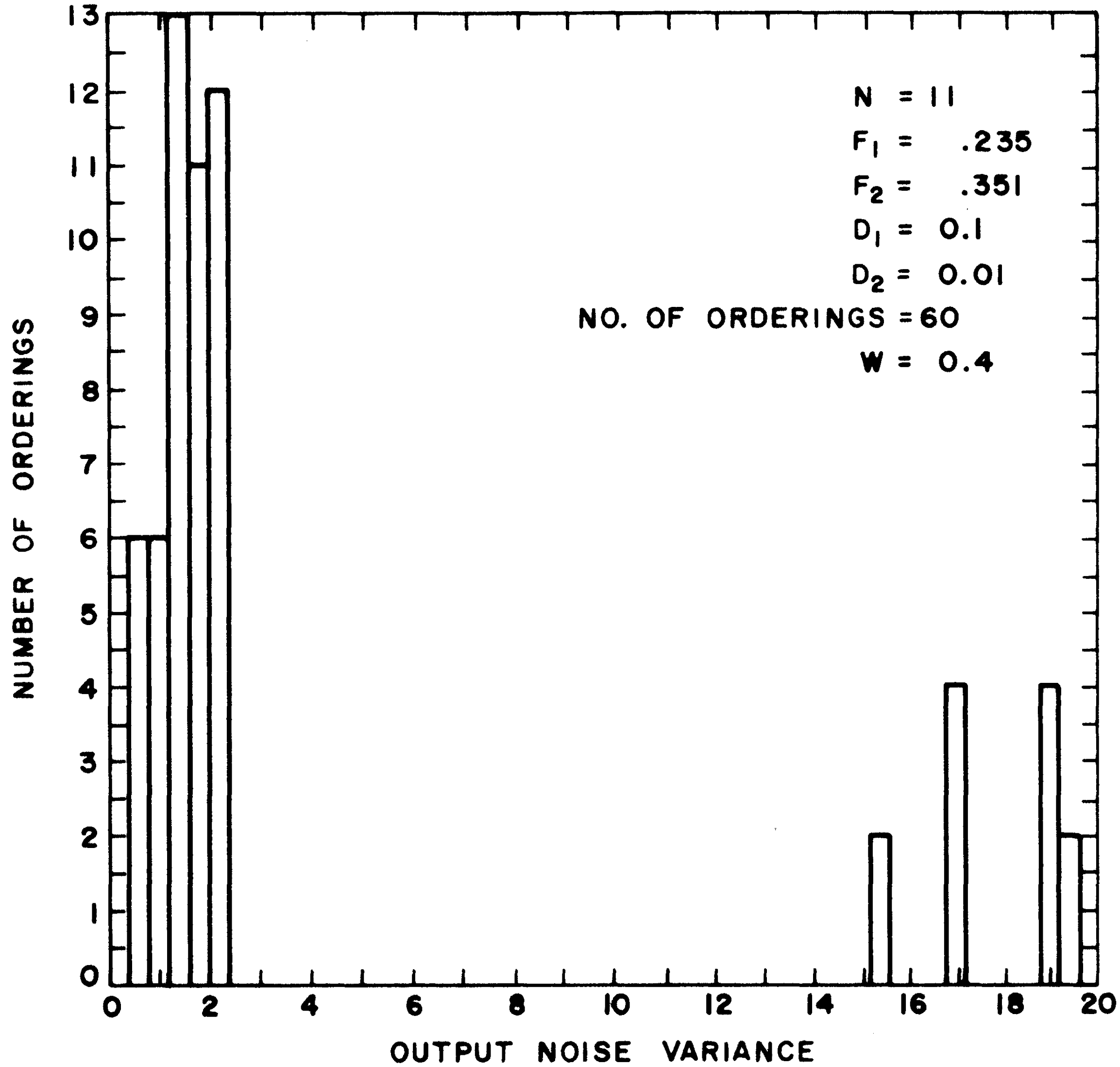


FIG. 4.11 — NOISE DISTRIBUTION HISTOGRAM OF TYPICAL 11 POINT FILTER

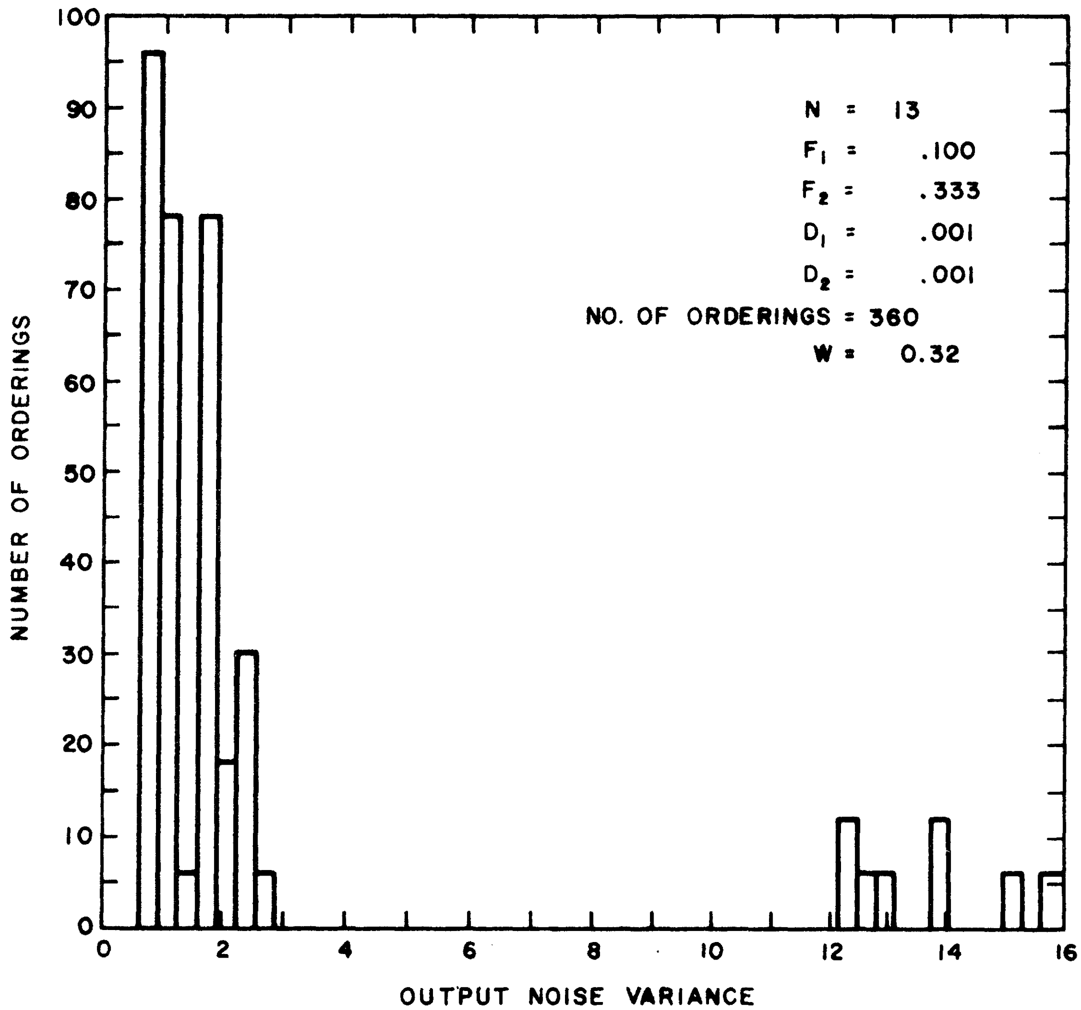


FIG. 4.12 — NOISE DISTRIBUTION HISTOGRAM OF ANOTHER 13 POINT FILTER

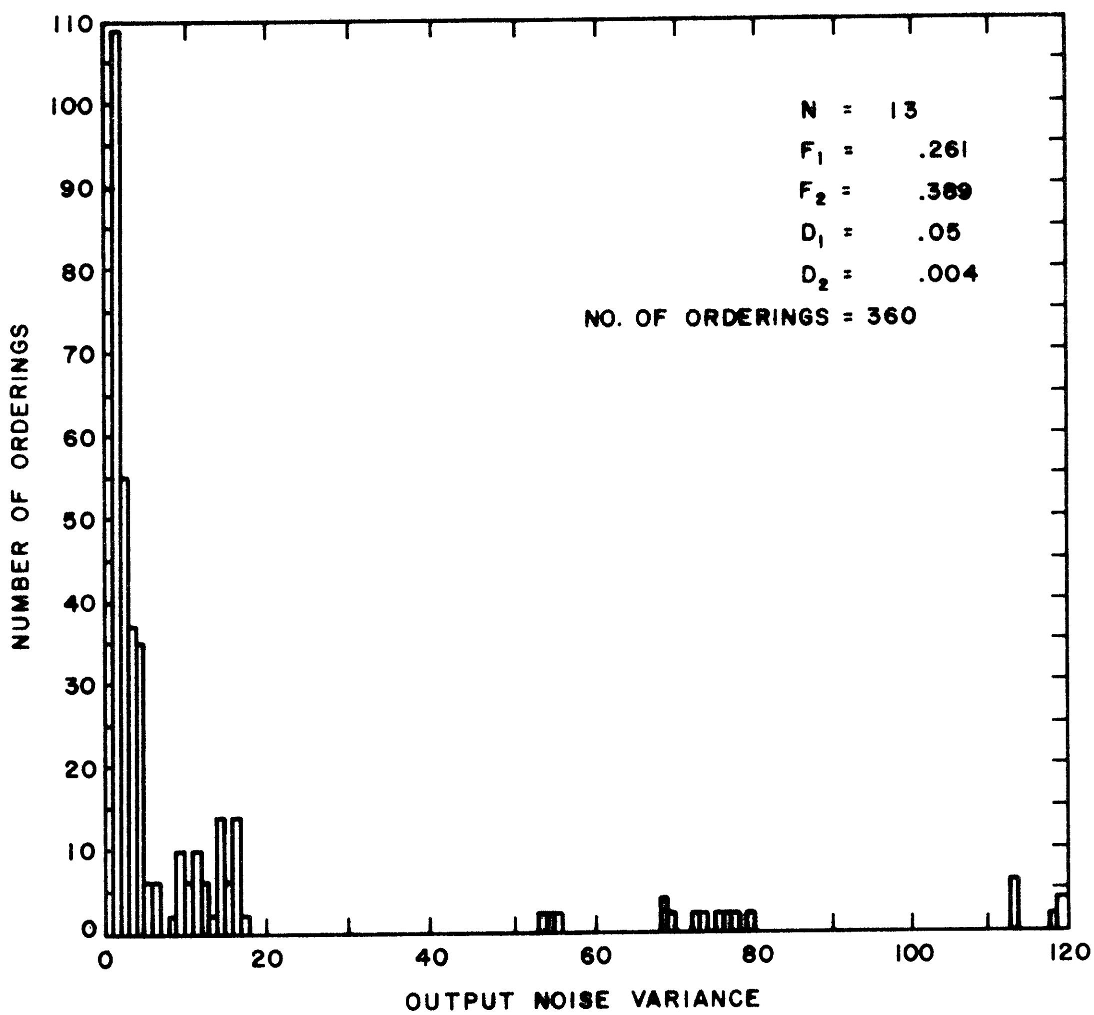


FIG. 4.13 — NOISE DISTRIBUTION HISTOGRAM OF A THIRD 13 POINT FILTER

Finally, the noise distributions for three 15-point, extraripple filters are shown in Figs. 4.14 to 4.16. Each of these involves 2,520 different orderings and requires 118 seconds on the Honeywell 6070 machine. These plots show even stronger emphasis on the distribution characteristics discussed, and together with Fig. 4.11 suggests that the skewed shape and large gaps properties of the noise distribution of a filter must become increasingly pronounced as the order of the filter increases. Thus we expect that our results can be generalized for higher order filters.

4.2 Comparison of Sum Scaling and Peak Scaling

It was mentioned in the previous section that the results obtained on the noise distribution of filters with respect to different orderings ought to be quite independent of whether sum scaling or peak scaling is used. In this section we will show heuristically and support with experimental evidence that this claim is indeed true.

Let $H(z)$ be any transfer function and denote by $\{H_i(z)\}$ an ordering for a filter synthesizing $H(z)$ which is sum scaled and denote by $\{H_i'(z)\}$ the same ordering except peak scaled. Then

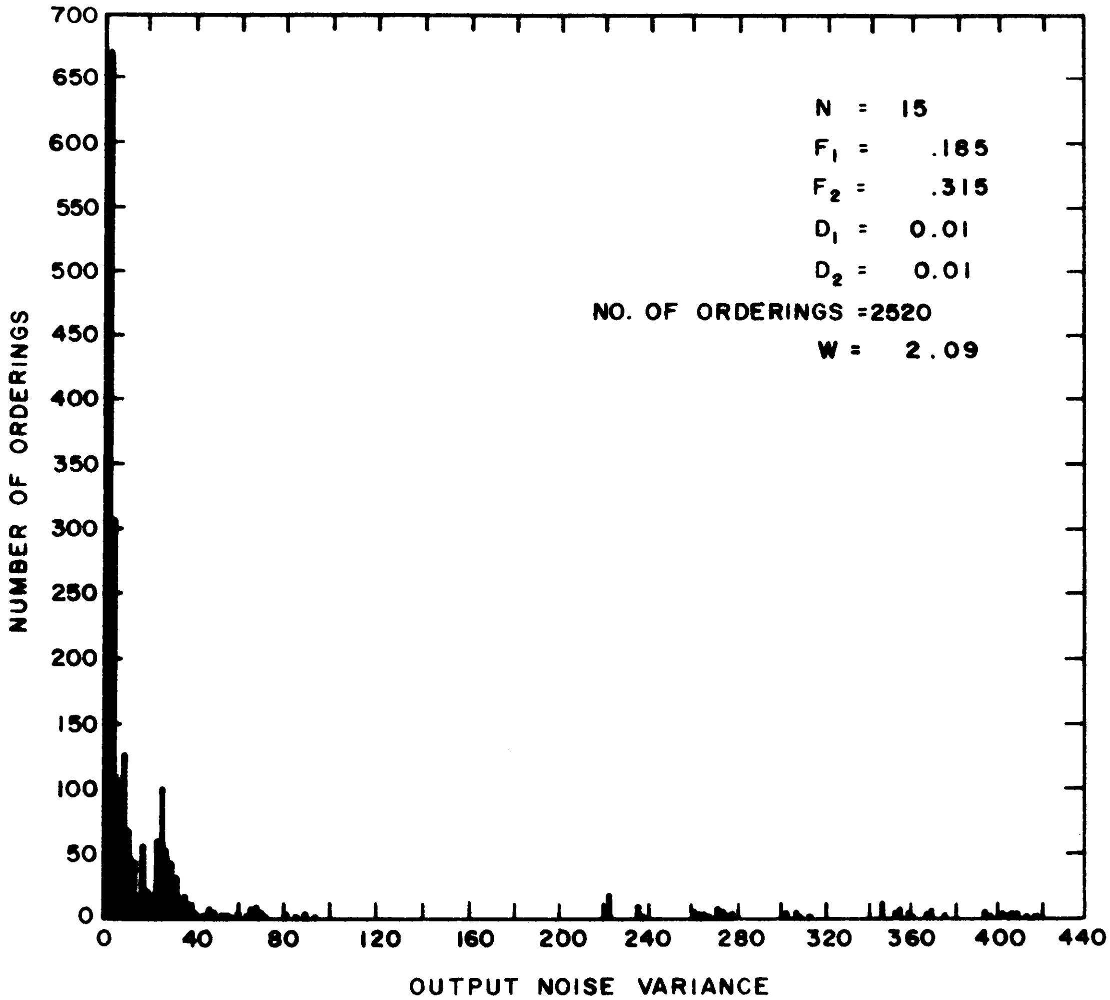


FIG. 4.14 — NOISE DISTRIBUTION HISTOGRAM OF 15 POINT FILTER EXAMPLE 1.

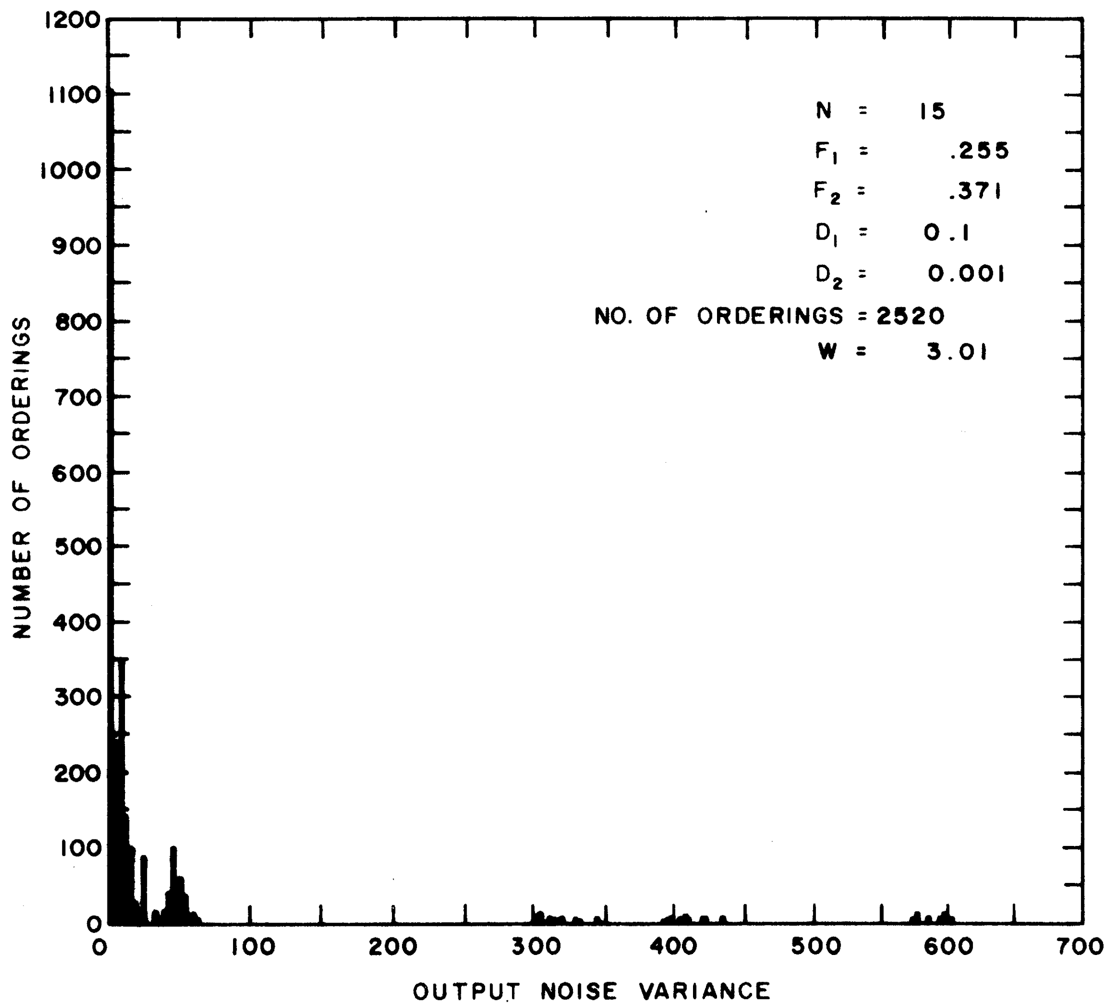


FIG. 4.15 – NOISE DISTRIBUTION HISTOGRAM OF 15 POINT FILTER EXAMPLE 2.

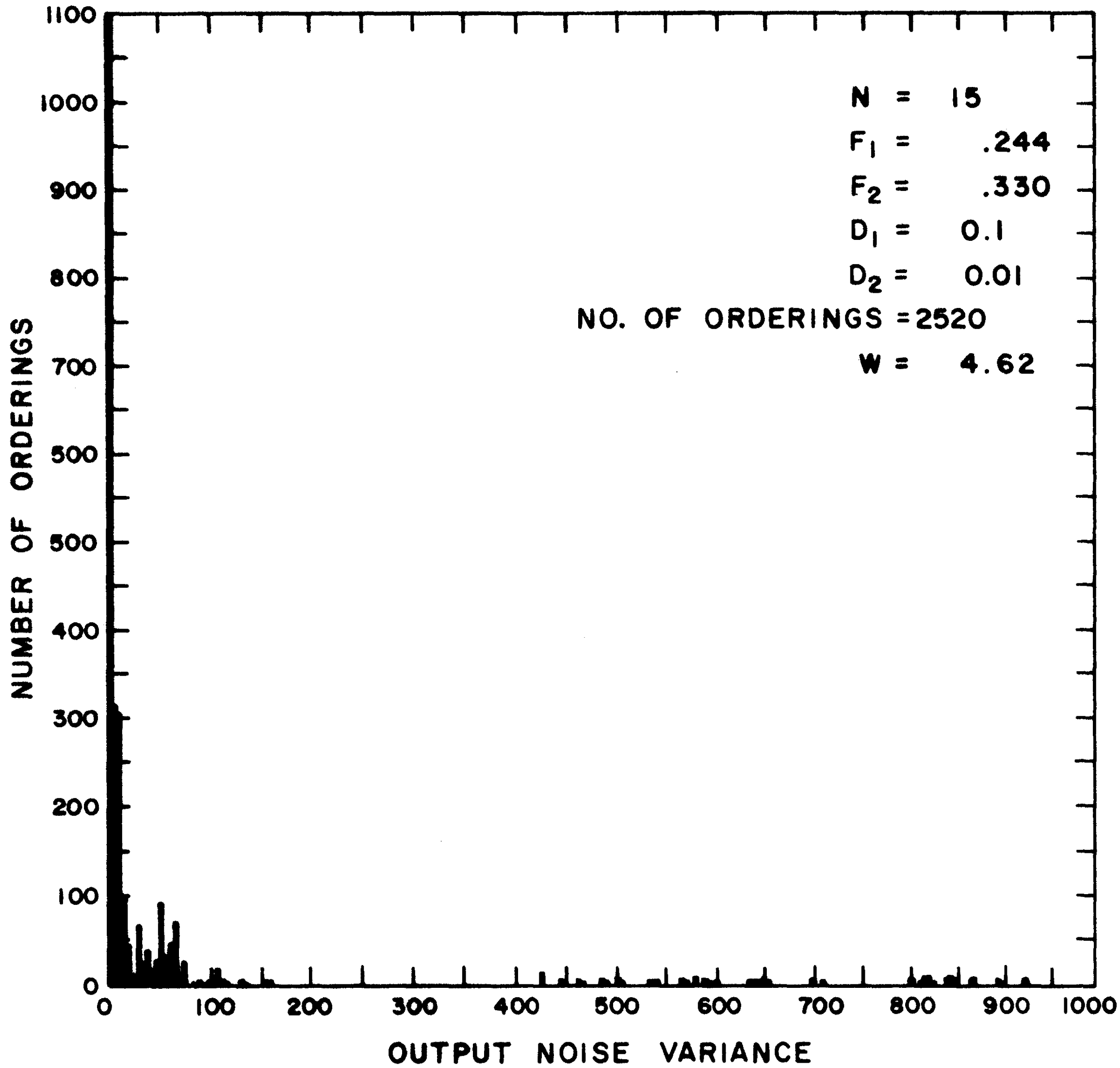


FIG. 4.16 – NOISE DISTRIBUTION HISTOGRAM OF 15 POINT FILTER EXAMPLE 3.

$$H(z) = k_1 \prod_{i=1}^{N_s} H_i(z) = k_2 \prod_{i=1}^{N_s} H_i'(z) \qquad (4.29)$$

where k_1 and k_2 are constants. Let

$$H_i(z) = S_i \hat{H}_i(z), \quad \prod_{j=1}^{N_s} S_j = \beta$$

$$H_i'(z) = S_i' \hat{H}_i(z), \quad \prod_{j=1}^{N_s} S_j' = \beta' \qquad (4.30)$$

where $\hat{H}_i(z)$ is as defined in section 3.3. Furthermore, define as in section 3.3

$$\hat{F}_i(z) = \prod_{j=1}^{i} \hat{H}_j(z) = \sum_k \hat{f}_i(k) z^{-k} \qquad (4.31)$$

Since $\{H_i(z)\}$ is sum scaled, we have from (3.35)

$$\prod_{j=1}^{i} S_j = \left[\sum_k |\hat{f}_i(k)| \right]^{-1} \quad 1 \leq i \leq N_s \qquad (4.32)$$

Recall that condition (4.32) guarantees that no intersection overflow (i.e., error causing) can occur in the filter for all class 1* inputs. Since class 2* is a subset of class 1, the same must also be true for all class 2 inputs. But by Theorem 3.2 and (3.34) this no-overflow condition for class 2 inputs means that the S_i must satisfy (noting from (4.32) that $S_i > 0$ for all i)

---

* See page 62.

$$\prod_{j=1}^{i} S_j \leq \left[\max_{\omega} |\hat{F}_i(e^{j\omega})|\right]^{-1} \quad 1 \leq i \leq N_s \tag{4.33}$$

Turning to $\{H_i'(z)\}$, since it is peak scaled, we have

$$\prod_{j=1}^{i} S_j' = \left[\max_{\omega} |\hat{F}_i(e^{j\omega})|\right]^{-1} \quad 1 \leq i \leq N_s \tag{4.34}$$

Therefore

$$\prod_{j=1}^{i} S_j \leq \prod_{j=1}^{i} S_j' \quad 1 \leq i \leq N_s \tag{4.35}$$

Now the output noise variance due to the i^{th} section for $\{H_i(z)\}$ and $\{H_i'(z)\}$ are respectively

$$\sigma_i^2 = \frac{\beta^2 C_i}{\left(\prod_{j=1}^{i} S_j\right)^2} \quad 1 \leq i \leq N_s \tag{4.36}$$

and

$$\sigma_i'^2 = \frac{\beta'^2 C_i}{\left(\prod_{j=1}^{i} S_j'\right)^2} \quad 1 \leq i \leq N_s \tag{4.37}$$

where

$$C_i = \begin{cases} k_i \dfrac{Q^2}{12} \cdot \dfrac{1}{2\pi} \displaystyle\int_0^{2\pi} \Big| \prod_{j=i+1}^{N_s} \hat{H}_j(e^{j\omega}) \Big|^2 d\omega & 1 \leq i \leq N_s - 1 \\ k_{N_s} \dfrac{Q^2}{12} & i = N_s \end{cases} \tag{4.38}$$

Let

$$\beta = \alpha\beta' \tag{4.39}$$

Then by (4.35) to (4.39)

$$\sigma_i^2 \geq \alpha^2 \sigma_i'^2 \tag{4.40}$$

Therefore

$$\sigma^2 \geq \alpha^2 \sigma'^2 \tag{4.41}$$

where $\sigma^2 = \Sigma\sigma_i^2$ and $\sigma'^2 = \Sigma\sigma_i'^2$.

Now from (4.32) and (4.34) we can write

$$\alpha = \frac{\beta}{\beta'} = \frac{\max_{\omega} |\hat{F}_{N_s}(e^{j\omega})|}{\sum_k |\hat{f}_{N_s}(k)|} \tag{4.42}$$

Clearly, by (4.35) $\beta \leq \beta'$, therefore $\alpha \leq 1$. But

$$\hat{F}_{N_s}(1) = \sum_k \hat{f}_{N_s}(k) \tag{4.43}$$

and

$$\hat{F}_{N_s}(1) = \hat{F}_{N_s}(e^{jo}) \leq \max_{\omega} |\hat{F}_{N_s}(e^{j\omega})| \tag{4.44}$$

Hence

$$\frac{\sum_k \hat{f}_{N_s}(k)}{\sum_k |\hat{f}_{N_s}(k)|} \leq \alpha \leq 1 \tag{4.45}$$

Defining a sequence $\{r(k)\}$ by

$$r(k) = \begin{cases} -\hat{f}_{N_s}(k) & \hat{f}_{N_s}(k) < 0 \\ 0 & \hat{f}_{N_s}(k) \geq 0 \end{cases} \tag{4.46}$$

we can write

$$\frac{\sum_k \hat{f}_{N_s}(k)}{\sum_k |\hat{f}_{N_s}(k)|} = \frac{\sum_k |\hat{f}_{N_s}(k)| - 2 \sum_k r(k)}{\sum_k |\hat{f}_{N_s}(k)|}$$

$$= 1 - 2\varepsilon \tag{4.47}$$

where

$$\varepsilon = \frac{\sum_k r(k)}{\sum_k |\hat{f}_{N_s}(k)|} \tag{4.48}$$

Then

$$1 - 2\varepsilon \leq \alpha \leq 1 \tag{4.49}$$

Loosely speaking, ε is the fraction of the impulse response of $\hat{F}_{N_s}(z)$ which has negative values. Since $\hat{F}_{N_s}(z)$ is simply a constant multiple of H(z), ε is unchanged if $\hat{F}_{N_s}(z)$ is replaced by H(z). For a low pass transfer function H(z), the envelope of its impulse response has the general shape of a truncated $\frac{\sin x}{x}$ curve. Therefore ε is expected to be a small number.

For a low pass filter, we can in addition overbound α more tightly by noting that

$$\sum_k \hat{f}_{N_s}(k) = \hat{F}_{N_s}(1) \geq \max_{\omega} |\hat{F}_{N_s}(e^{j\omega})| - 2\delta \qquad (4.50)$$

where δ is the passband tolerance (i.e. maximum approximation error). Therefore

$$\max_{\omega} |\hat{F}_{N_s}(e^{j\omega})| \leq \sum_k \hat{f}_{N_s}(k) + 2\delta \qquad (4.51)$$

or from (4.42)

$$\alpha \leq \frac{\sum_k \hat{f}_{N_s}(k) + 2\delta}{\sum_k |\hat{f}_{N_s}(k)|} \qquad (4.52)$$

With ε defined as before, we have finally

$$1 - 2\varepsilon \leq \alpha \leq 1 - 2\varepsilon + 2 \frac{\delta}{\sum_k |\hat{f}_{N_s}(k)|} \qquad (4.53)$$

For any reasonably well designed low pass transfer function $\hat{F}_{N_s}(z)$,

$$\delta \ll \max_{\omega} |\hat{F}_{N_s}(e^{j\omega})| \tag{4.54}$$

But from (4.32), (4.34) and (4.35)

$$\max_{\omega} |\hat{F}_{N_s}(e^{j\omega})| \leq \sum_k |\hat{f}_{N_s}(k)| \tag{4.55}$$

Hence

$$\delta \ll \sum_k |\hat{f}_{N_s}(k)| \tag{4.56}$$

Therefore, with little error committed we can write

$$\alpha = 1 - 2\varepsilon \tag{4.57}$$

for a low pass filter.

The relation $\sigma^2 \geq \alpha^2 \sigma'^2$ has the implication that for class 2 input signals peak scaling yields a higher signal-to-noise ratio than sum scaling. To see this note that $\{H_i(z)\}$ has a gain α times as large as that of $\{H_i'(z)\}$. Thus given a class 2 input constrained to the dynamic range, the maximum attainable output signal level is α times as large in $\{H_i(z)\}$ as in $\{H_i'(z)\}$. Therefore if we multiply all outputs of $\{H_i'(z)\}$ by α so that the signal output of both $\{H_i(z)\}$ and $\{H_i'(z)\}$ are identical given identical class 2 inputs, then the ratio of their signal-to-noise ratios will simply be given by the inverse ratio of their noise outputs. But the output noise variance of the modified $\{H_i'(z)\}$ would be given by $\alpha^2\sigma'^2$, hence the signal-to-noise ratios satisfy (S/N for sum scaling and S/N′ for peak scaling):

$$\frac{S/N'}{S/N} = \frac{\sigma}{\alpha\sigma'} \tag{4.58}$$

or since $\sigma \geq \alpha\sigma'$,

$$S/N' \geq S/N \tag{4.59}$$

Of course, this result is to be expected since we have shown in section 3.3 that both sum scaling and peak scaling are optimal within the classes of inputs they assume. Thus since class 2 is a subset of class 1, we would expect the optimal scaling for class 2 signals to yield no worse performance than the optimum for class 1.

In Tables 5.1-5.2, a list of filters and some results of section 5.0 will be presented. Together with these results we have also listed measured values of α for each filter. Observe that for these typical filters α ranges from .5 to 1. Furthermore, for each filter, the last and third last columns of Tables 5.1-5.2 list the noise variances that result from the same ordering using sum scaling and peak scaling respectively. Comparing these, we see that in almost every case

$$\sigma^2 \leq \sigma'^2 \tag{4.60}$$

In particular, we observe that (4.60) holds if α is not too close to 1.0. When α=1, (4.60) can be easily false since (4.36) and (4.37) show that

$$\frac{\sigma_i^{\,2}}{\sigma_i'^{\,2}} = \frac{\left(\prod_{j=1}^{i} S_j'\right)^2}{\left(\prod_{j=1}^{i} S_j\right)^2} \tag{4.61}$$

and it is not difficult to conceive of a filter with α = 1 in which for some i

$$\max_{\omega} |\hat{F}_i(e^{j\omega})| < \sum_k |\hat{f}_i(k)| \tag{4.62}$$

so that $\prod_{j=1}^{i} S_j < \prod_{j=1}^{i} S_j'$. Thus for this i $\sigma_i^2 > \sigma_i'^2$, hence $\sigma^2 > \sigma'^2$, contradicting (4.60). A concrete example is filter no. 40 in Table 5.1.

However, except for the uninteresting cases of filters with all zeros on the unit circle, in general $\alpha < 1$ and (4.60) holds. Thus we may assume for practical arguments that

$$\alpha^2 \leq \frac{\sigma^2}{\sigma'^2} \leq 1 \tag{4.63}$$

From (4.63) we see that the output noise variance for a filter with sum scaling is very comparable, at least in order of magnitude, to that for the same filter ordered the same way with peak scaling applied. In fact, experimental results show that given a filter, the noise variances for sum scaling and peak scaling are in an approximately constant ratio for almost all orderings. An example of this result is shown in Fig. 4.17, where the noise variances for sum scaling and peak scaling of a typical filter are plotted against each other for each ordering. The resulting points are seen to form almost a straight line with slope approximately equal to 2, so that essentially $\sigma'^2 = 2\sigma^2$ for all orderings of this filter.

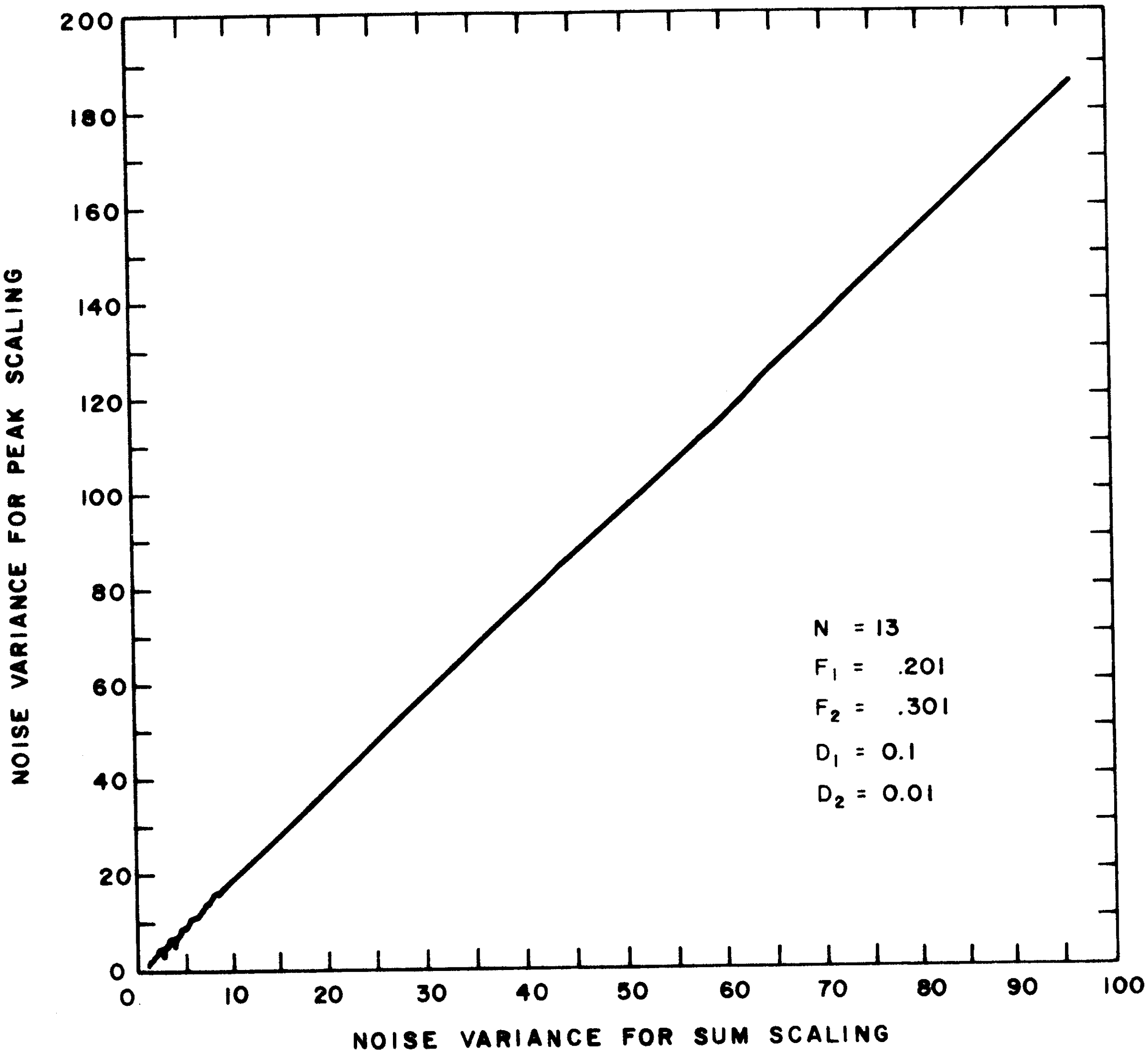


FIG. 4.17 – PEAK SCALING VERSUS SUM SCALING NOISE OUTPUT COMPARISON FOR TYPICAL FILTER.

Thus the noise distribution plots of section 4.1 are essentially unchanged if peak scaling were used instead of sum scaling. As an example we show in Fig. 4.18 and 4.19 the noise distribution plots for sum scaling and peak scaling respectively for the filter of Fig. 4.17. Similar plots for the filter of Fig. 4.6 are shown in Fig. 4.20 and 4.21.

The evaluation of noise variances with peak scaling is done in exactly the same way as that described in Fig. 4.5, except that the statements

$$x_1 \leftarrow \sum_k |f_{N_s}(k)|$$

$$x_2 \leftarrow \sum_k |f_{i-1}(k)|$$

are replaced by

$$x_1 \leftarrow \max_\omega |F_{N_s}(e^{j\omega})|$$

$$x_2 \leftarrow \max_\omega |F_{i-1}(e^{j\omega})|$$

Using a 128-point FFT to evaluate two at a time (exploiting real and imaginary part symmetries) the maxima of the $F_i(e^{j\omega})$, for 360 orderings the computations for peak scaling were found to require four times as much time as that for sum scaling, viz. approximately 48 seconds on the Honeywell 6070 machine.

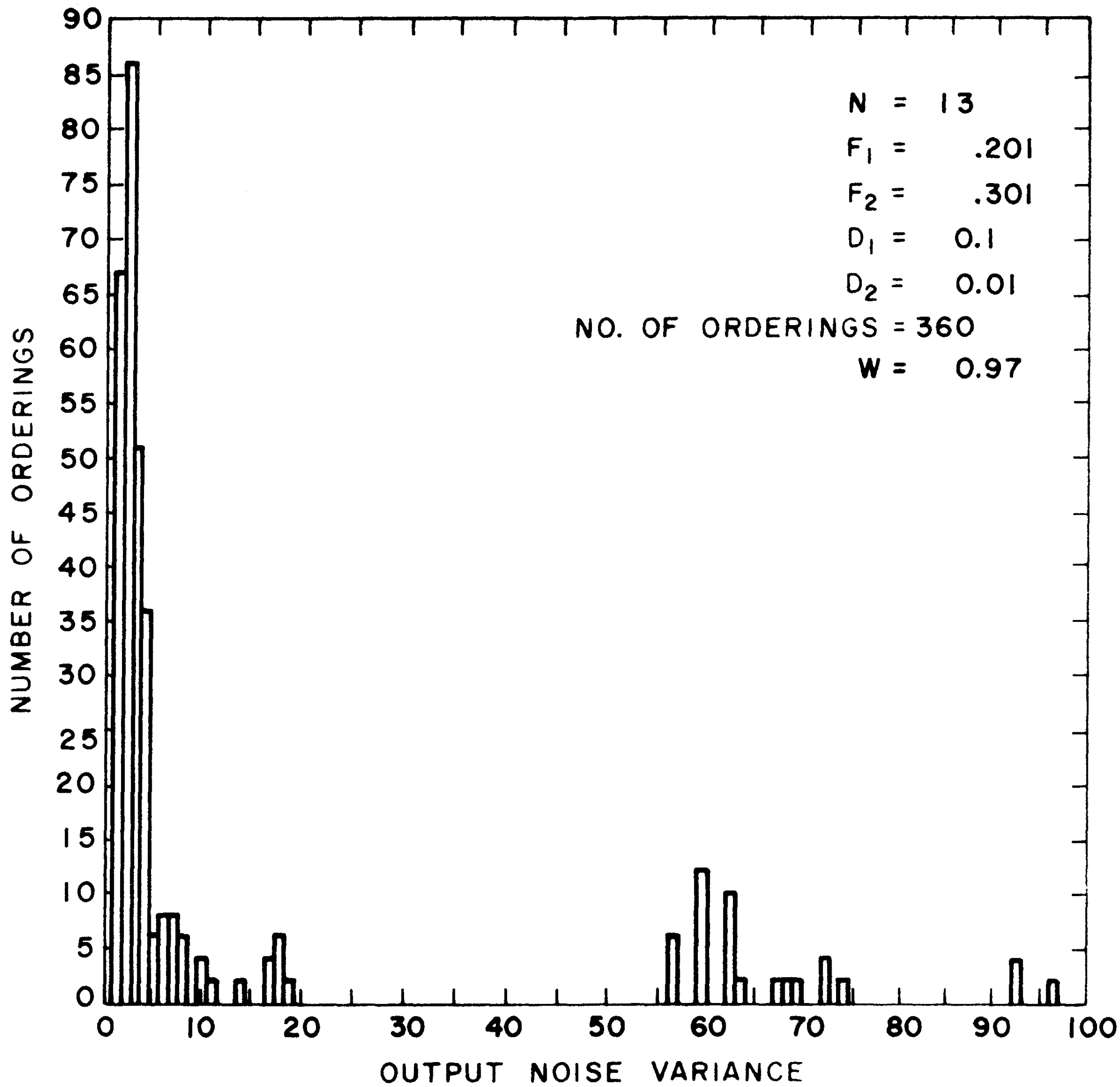


FIG. 4.18 – NOISE DISTRIBUTION HISTOGRAM OF FILTER OF FIG. 4.17 USING SUM SCALING

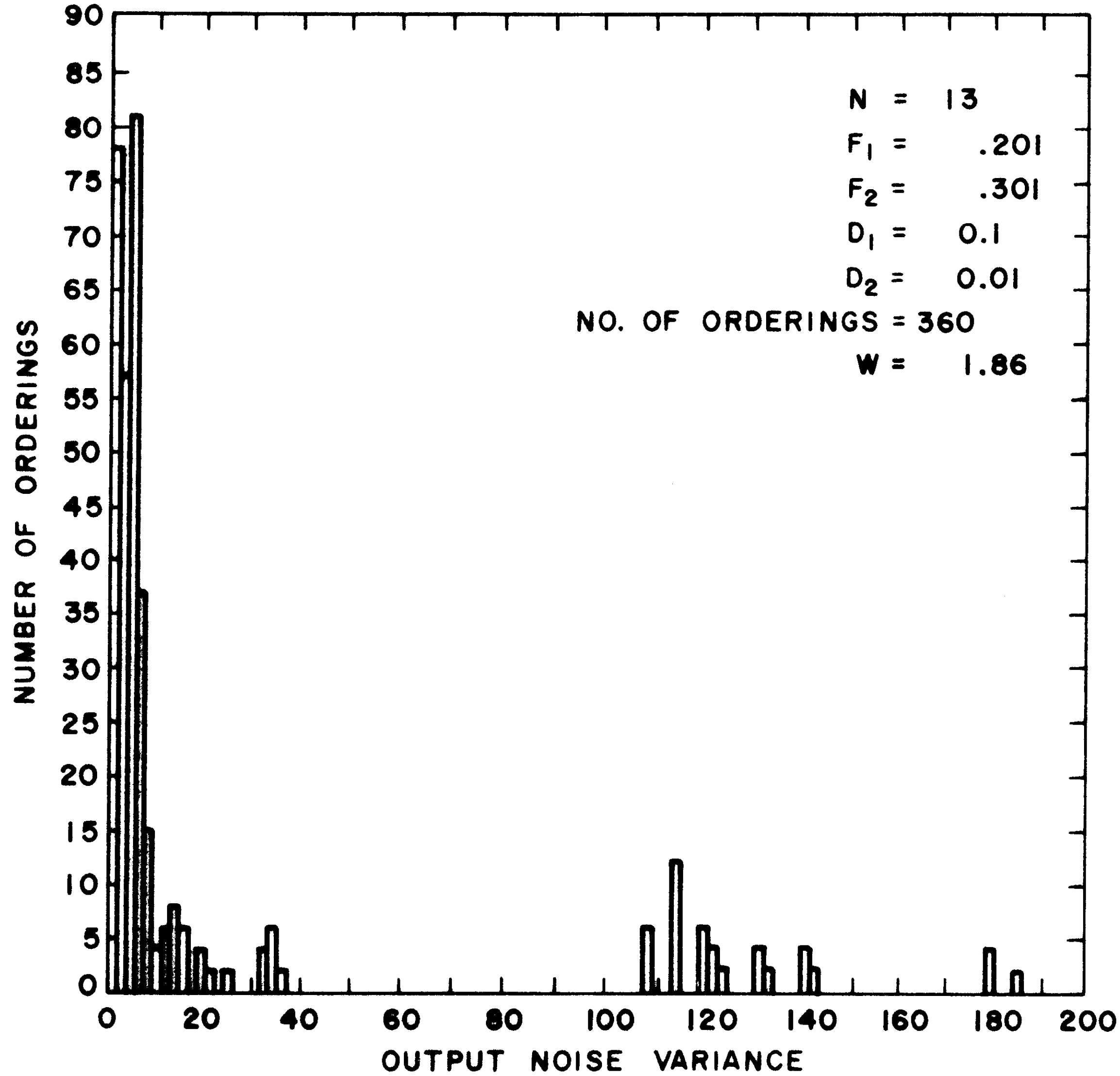


FIG. 4.19 – NOISE DISTRIBUTION HISTOGRAM OF FILTER OF FIG. 4.17 USING PEAK SCALING

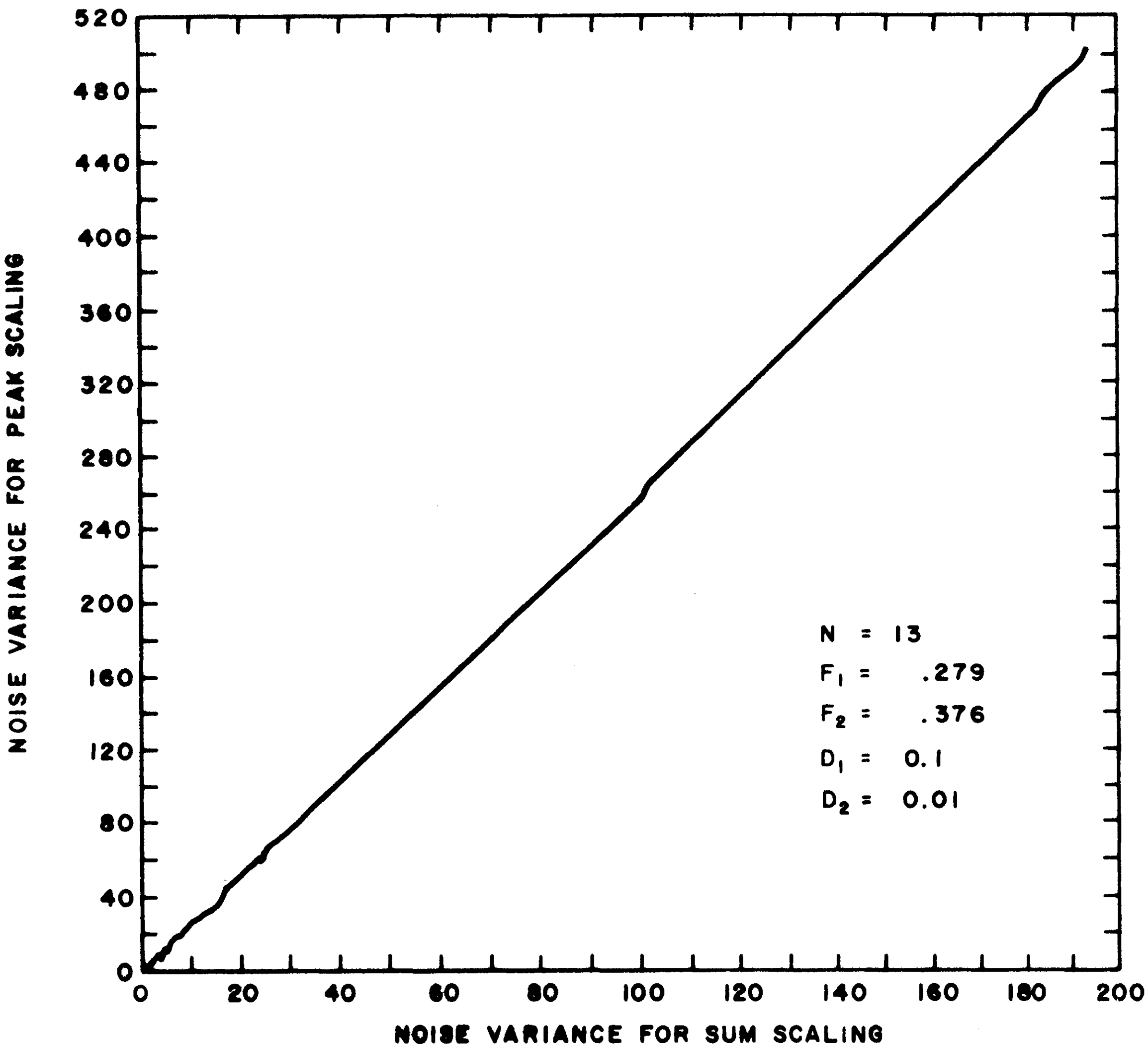


FIGURE 4-20—PEAK SCALING VERSUS SUM SCALING NOISE OUTPUT COMPARISON FOR FILTER OF FIG. 4.6.

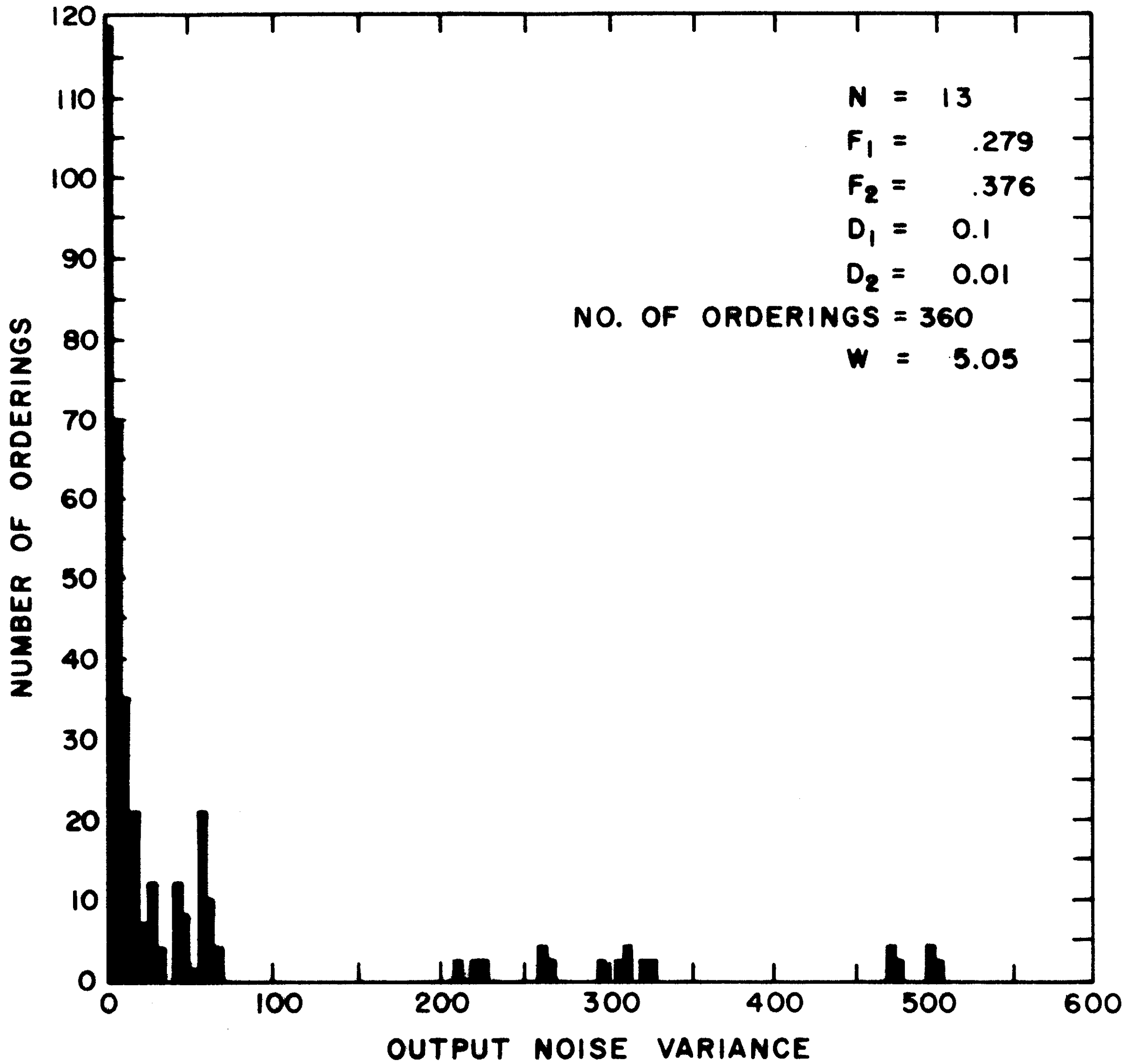


FIG. 4.21 – NOISE DISTRIBUTION HISTOGRAM OF FILTER OF FIG. 4.6 USING PEAK SCALING

4.3 Dependence of Roundoff Noise on Other Filter Parameters

The two preceeding sections have investigated the dependence of roundoff noise output of a cascade filter on scaling and section ordering. It was shown that though different filters may produce very different ranges of output noise variances when ordered in all possible ways, the noise variances for each filter always distribute themselves in essentially the same general pattern. (By different filters we mean filters which realize different transfer functions.) In this section we shall account for the differences in noise variance ranges among different filters by investigating the dependence of noise distributions on parameters which specify the transfer function of a filter.

For simplicity only low-pass filters will be considered. A low-pass transfer function can be specified up to overall gain by four independent parameters. The parameters which we have chosen to be independent variables in our investigations are filter length (N), passband edge (F_1), passband tolerance (D_1), and stopband tolerance (D_2). These four parameters together are sufficient to uniquely specify a transfer function designed via the optimal design technique (see section 2.2).

The noise distributions of several filters with various values of the above parameters were computed using the methods of section 4.1. Sum scaling was employed as the scaling method. Since all these distributions have the same general shape, we can compare them by simply comparing their maximum, average, and minimum values. A list of all the filters whose noise distributions have been computed, including those already discussed in section 4.1, are presented in Table 4.1. These filters are specified by five parameters, namely the four already mentioned plus N_p, the number of ripples in the passband. Since all the filters are extraripple filters, it is more natural to specify N_p than F_1. Of course, N_p and F_1 are not independent. The maximum, average, and minimum values of the noise distributions of each of these filters are listed in Table 4.1. The last column in this table will be discussed in section 5.0.

Filters no. 1 to 5 in Table 4.1 are very similar except for their length in that they all have identical passband and stopband tolerances and approximately the same bandwidth. The maximum, average, and minimum values of their noise distributions are plotted on semi-log coordinates in Fig. 4.22. We see that all these statistics of the distributions have an essentially exponential dependence on filter length. The less regular behavior of the minimum values is believed to be caused by differences in bandwidth (F_1) among the filters.

Table 4.1

List of Filters and Their Noise Distribution Statistics

						Noise Variance			
#	N	N_p	F_1	D_1	D_2	Max	Avg	Min	Alg. 1
1	7	2	.212	.1	.01	1.24	.84	.37	.37
2	9	3	.281	.1	.01	6.26	2.54	.73	.73
3	11	3	.235	.1	.01	19.41	4.79	.68	.68
4	13	4	.279	.1	.01	192.86	19.55	1.10	1.10
5	15	4	.244	.1	.01	923.63	54.45	1.02	1.16
6	13	3	.100	.001	.001	15.84	3.01	.65	.69
7	13	4	.261	.05	.004	119.48	12.91	.96	1.02
8	13	1	.012	.01	.01	9.91	1.61	.32	.35
9	13	2	.067	.01	.01	16.30	2.94	.44	.47
10	13	3	.138	.01	.01	42.63	5.94	.71	.73
11	13	4	.213	.01	.01	69.76	8.52	.82	.91
12	13	5	.288	.01	.01	76.43	11.01	1.44	1.52
13	13	6	.364	.01	.01	52.54	10.33	1.92	2.43
14	13	3	.201	.1	.01	96.25	12.09	.81	
15	13	3	.179	.05	.01	69.26	9.02	.76	
16	13	3	.154	.02	.01	50.63	6.87	.72	
17	13	3	.123	.005	.01	37.36	5.33	.70	
18	13	3	.106	.002	.01	32.83	4.80	.69	
19	13	3	.095	.001	.01	30.53	4.53	.69	
20	13	3	.124	.01	.1	132.57	17.56	1.02	
21	13	3	.129	.01	.05	85.84	11.45	.83	
22	13	3	.135	.01	.02	54.94	7.47	.75	
23	13	3	.141	.01	.005	35.59	5.07	.68	
24	13	3	.144	.01	.002	26.44	4.37	.68	
25	13	3	.146	.01	.001	22.52	4.07	.70	
26	15	4	.185	.01	.01	417.08	27.38	1.00	
27	15	4	.255	.1	.001	601.83	35.15	1.02	

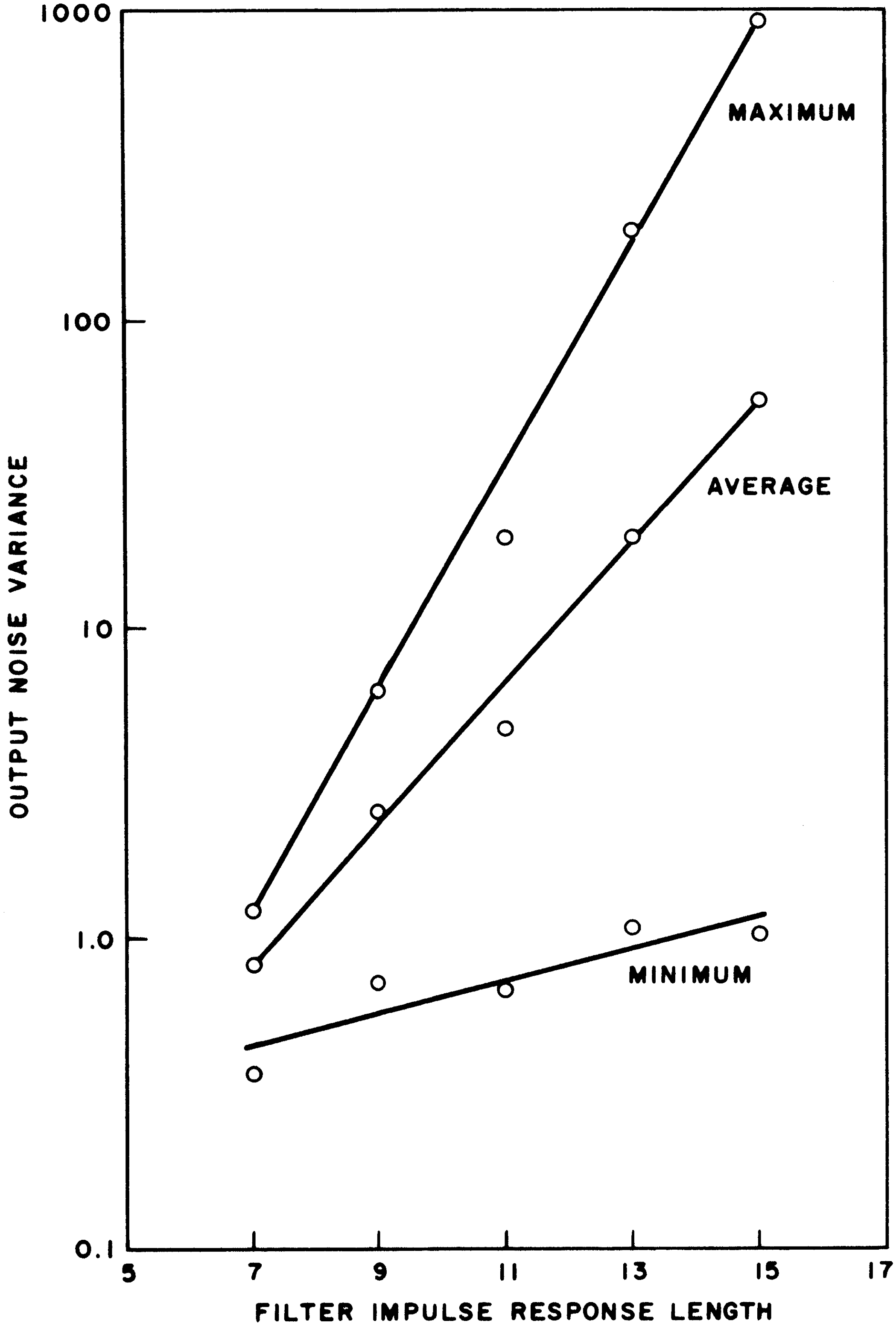


FIGURE 4.22 - OUTPUT NOISE VARIANCE AS A FUNCTION OF FILTER LENGTH

Figure 4.23 shows a similar plot of the same distribution statistics for filters no. 8 to 13 as a function of F_1. These filters have identical values of $N_,$ D_1, and D_2, and represent all six possible extraripple filters that have these parameter specifications. From Fig. 4.23 we see that with those parameters mentioned held fixed the noise output of a cascade filter tends to increase with increasing bandwidth.

Filters no. 14 to 25 all have fixed values of N, N_p, and either D_1 or D_2. Plots of the distribution statistics of these filters as functions of D_1 and D_2 are shown respectively in Figs. 4.24 and 4.25. These plots indicate that as the transfer function approximation error for a filter decreases, so does its noise output. Though the plots were made holding N_p rather than F_1 fixed, we see that at least for the filters used in Fig. 4.25, bandwidth increases with decreasing approximation error. Since the noise output of a filter is found to increase with bandwidth, we expect that noise would still decrease with stopband tolerance D_2 if F_1 were fixed instead of N_p. In any event the variation of F_1 among these filters is small.

Figures 4.23 to 4.25 are all plots of statistics for 13-point filters. Notice how the maximum, average, and minimum curves all tend to move together. In particular the

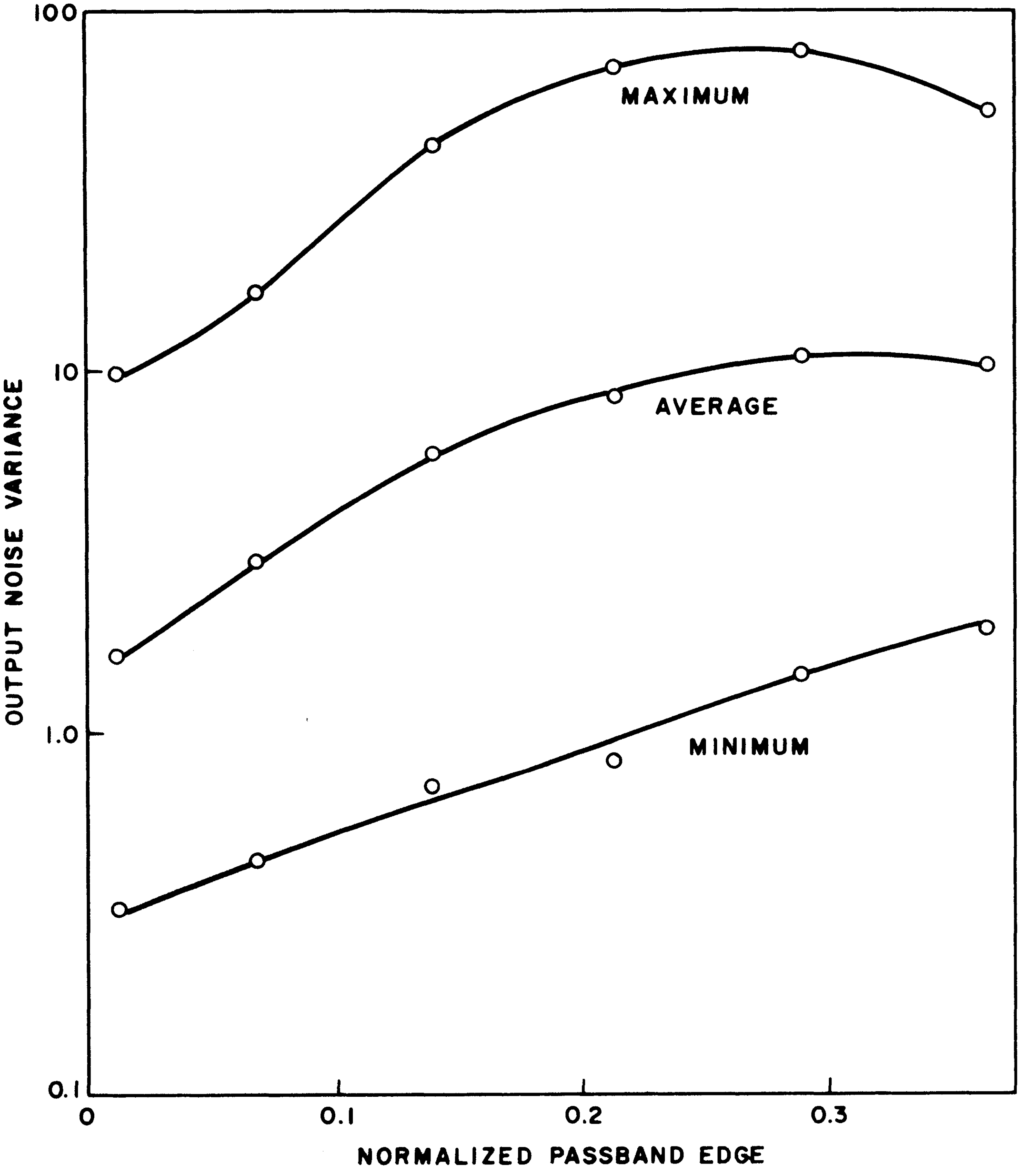


FIGURE 4.23 - OUTPUT NOISE VARIANCE AS A FUNCTION OF BANDWIDTH.

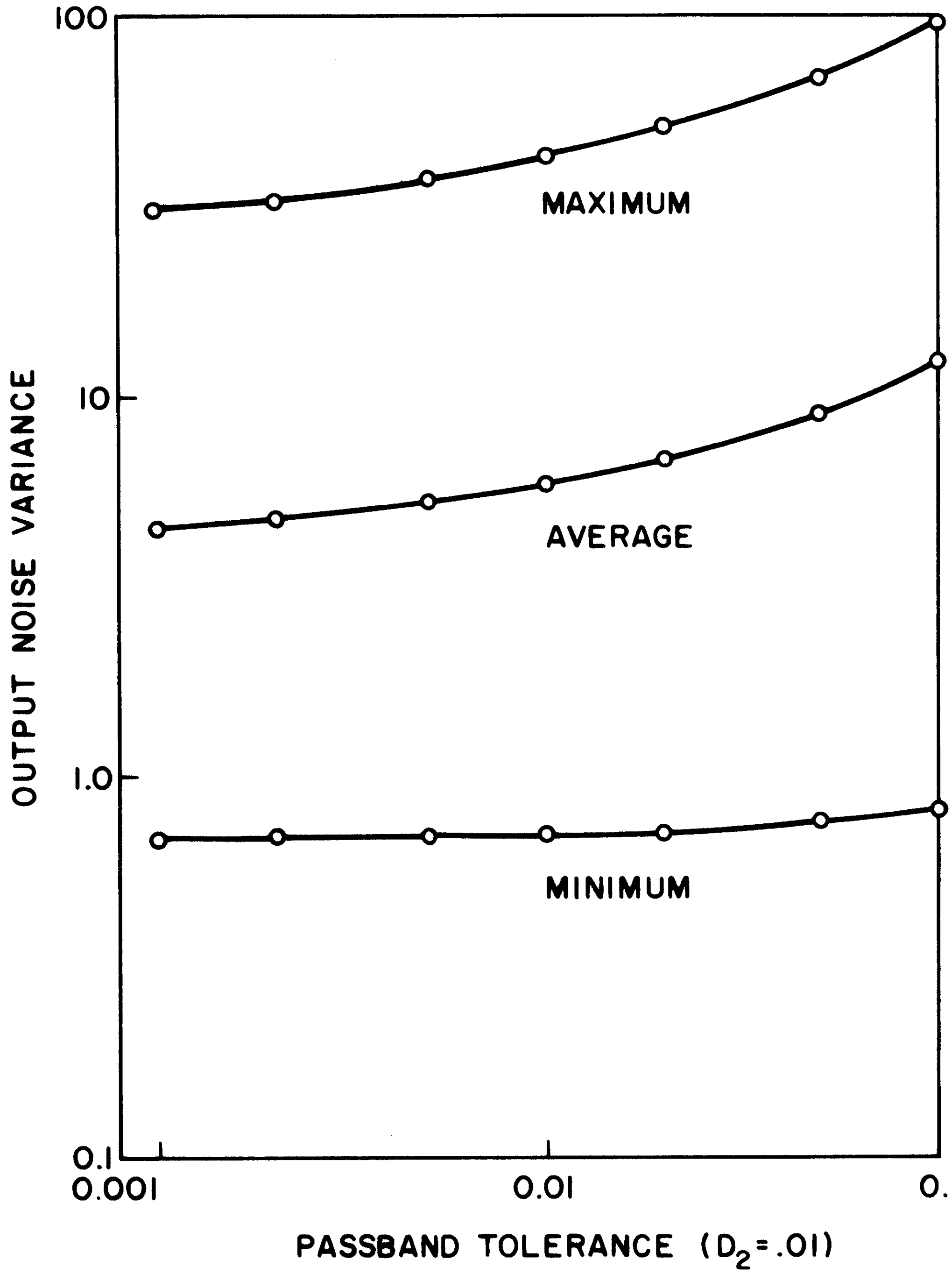


FIG. 4.24 – OUTPUT NOISE VARIANCE AS A FUNCTION OF PASSBAND APPROXIMATION ERROR.

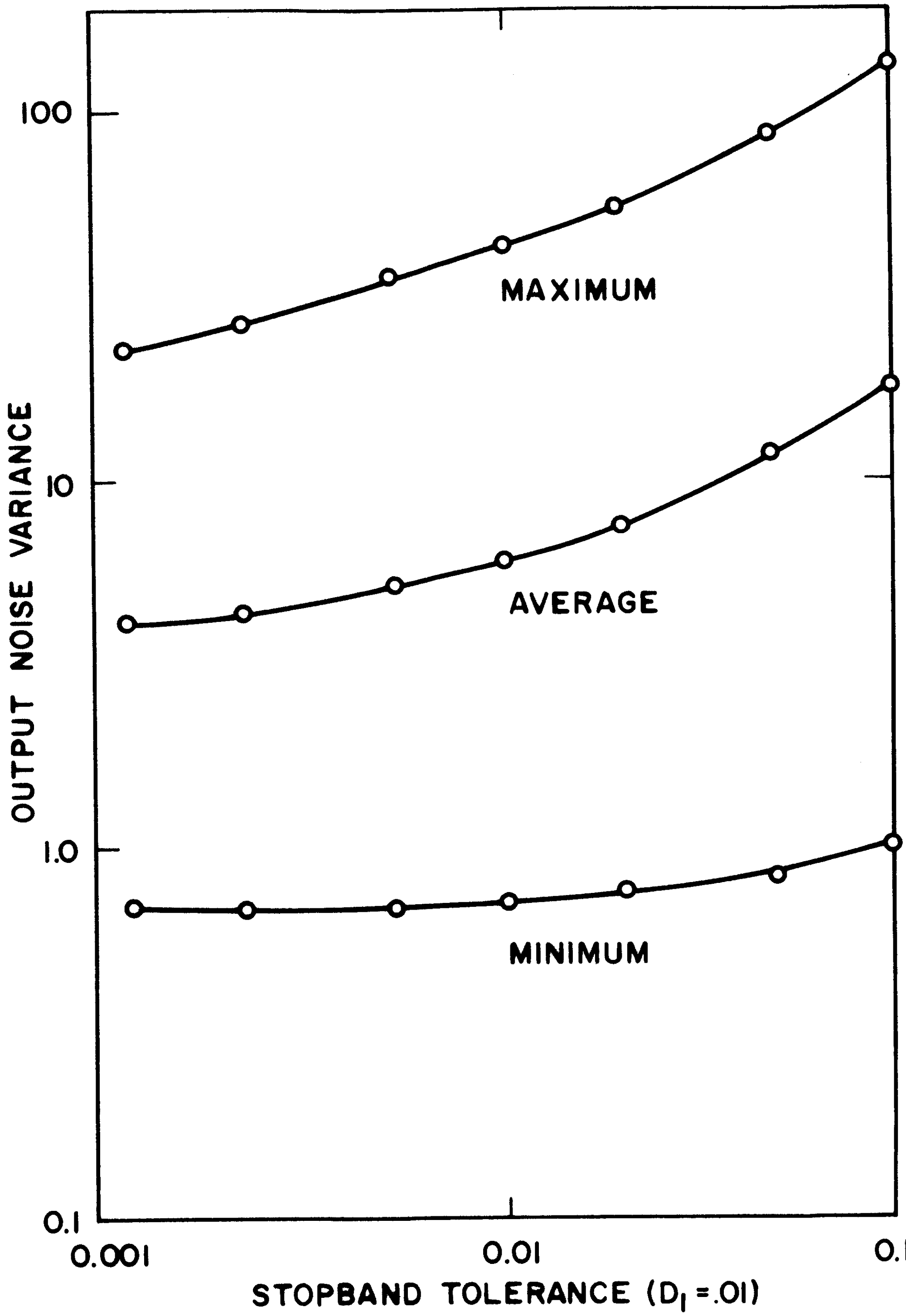


FIG. 4.25 OUTPUT NOISE VARIANCE AS A FUNCTION OF STOPBAND APPROXIMATION ERROR.

average curve almost always stays approximately halfway on the logarithmic scale between the maximum and minimum curves. This phenomenon is, of course, simply a manifestation of the empirical finding that noise distributions of different filters have essentially the same shape independent of differences in transfer characteristics.

To summarize, we have found experimentally that with other parameters fixed, the roundoff noise output of a filter tends to increase with all four independent parameters N, F_1, D_1, and D_2 which specify its transfer function. In particular, noise output tends to grow exponentially with N. We did not show that the noise output of a filter with a fixed ordering and scaled a given way always varies in the way indicated when its transfer function parameters are perturbed. What we have shown is perhaps a more useful result from the design viewpoint. Our findings imply that, other things being equal, a transfer function with, for instance, a higher value of D_2, is likely when realized in a cascade form to result in a higher noise output than a transfer function with a smaller value of D_2 realized by the same method. Though these results were found using only low-order filters, we expect them to generalize for higher order filters as well. Section 5 will present experimental evidence to confirm this expectation.

4.4 Spectral Peaking and Roundoff Noise

Given a filter, we have seen that its output noise variance can be very different depending on how its sections are ordered. This difference arises from complicated reasons which involve differing spectral shapes of different combinations of individual filter sections and the interactive scaling of signal levels within a filter necessitated by dynamic range limitations. As such, these reasons are too complex to provide a good feel for judging by inspection whether a given ordering ought to have a relatively high noise or a reasonably low noise output without performing involved calculations. In this section, we shall attempt to devise more intuitive means for judging the relative noise level of a filter by bringing out one characteristic of an ordered filter which is associated with the noise level of the filter.

While specific points in the following arguments cannot be proved to be valid in general, they are very reasonable to assume and predictions based on them are supported by experimental evidence. Thus the arguments are useful to advance. More importantly, they have provided valuable insights to the author on the behavior of roundoff noise in cascade filters. Taking advantage of the similarity between sum scaled and peak scaled filters in terms of noise output we shall for simplicity restrict our discussions to peak scaled filters.

Let H(z) be the transfer function of a peak scaled filter with section transfer functions $\{H_i(z)\}$. Then

$$H(z) = \prod_{j=1}^{N_s} H_j(z) = \beta \prod_{j=1}^{N_s} \hat{H}_j(z) \qquad (4.64)$$

where the $\hat{H}_j(z)$'s are as defined in section 3.3 and β is a constant. Let $\beta = C^{N_s}$, $C > 0$, and define new transfer functions

$$\overline{H}_i(z) = C\hat{H}_i(z) \quad 1 \leq i \leq N_s \qquad (4.65)$$

Then

$$H(z) = \prod_{j=1}^{N_s} \overline{H}_j(z) \qquad (4.66)$$

From (4.17) and the definition of $\hat{H}_i(z)$ we have clearly

$$\max_{\omega} |\hat{H}_i(e^{j\omega})| = 1 \quad 1 \leq i \leq N_s \qquad (4.67)$$

Hence

$$\max_{\omega} |\overline{H}_i(e^{j\omega})| = C \quad 1 \leq i \leq N_s \qquad (4.68)$$

Now define a sequence $\{r_i\}$ by

$$H_i(z) = r_i \overline{H}_i(z) \quad 1 \le i \le N_s \tag{4.69}$$

Then

$$\prod_{j=1}^{N_s} r_j = 1 \tag{4.70}$$

and peak scaling implies for all i

$$\prod_{j=1}^{i} r_j = \left[\max_{\omega} \left| \prod_{j=1}^{i} \overline{H}_j(e^{j\omega}) \right| \right]^{-1} \tag{4.71}$$

Clearly r_i is simply related to S_i by

$$S_i = Cr_i \tag{4.72}$$

Stated simply, the $\overline{H}_i(e^{j\omega})$'s are factors of $H(e^{j\omega})$ which have maxima normalized to the same value such that their product has a maximum equal to unity.

Now define a number Pk by

$$Pk = \max(p_1, p_2) \tag{4.73}$$

where

$$p_1 = \max_{1 \le i \le N_s - 1} \left(\max_{\omega} \left| \prod_{j=1}^{i} \overline{H}_j(e^{j\omega}) \right| \right)$$

$$p_2 = \max_{1 \le i \le N_s - 1} \left(\max_{\omega} \left| \prod_{j=i+1}^{N_s} \overline{H}_j(e^{j\omega}) \right| \right) \tag{4.74}$$

We will argue that given an ordering a large value of Pk indicates a high noise output while a low value of Pk indicates a low noise output.

To see this, define

$$\overline{G}_i(z) = \left(\max_{\omega} \left| \prod_{j=i+1}^{N_s} \overline{H}_j(e^{j\omega}) \right| \right)^{-1} \prod_{j=i+1}^{N_s} \overline{H}_j(z) \qquad (4.75)$$

Then

$$\max_{\omega} |\overline{G}_i(e^{j\omega})| = 1 \qquad (4.76)$$

Now the noise output due to the i^{th} section is given by

$$\sigma_i^2 = \left(\prod_{j=i+1}^{N_s} r_j \right)^2 k_i \frac{Q^2}{12} \cdot \frac{1}{2\pi} \int_0^{2\pi} \left| \prod_{j=i+1}^{N_s} \overline{H}_j(e^{j\omega}) \right|^2 d\omega$$

$$1 \le i \le N_s - 1 \qquad (4.77)$$

(We will not consider $i=N_s$ since $\sigma_{N_s}^2$ is independent of ordering.) But

$$\prod_{j=i+1}^{N_s} r_j = \left(\prod_{j=1}^{i} r_j \right)^{-1} = \max_{\omega} \left| \prod_{j=1}^{i} \overline{H}_j(e^{j\omega}) \right| \qquad (4.78)$$

and

$$\int_0^{2\pi} \left| \prod_{j=i+1}^{N_s} \overline{H}_j(e^{j\omega}) \right|^2 d\omega = \left(\max_{\omega} \left| \prod_{j=i+1}^{N_s} \overline{H}_j(e^{j\omega}) \right| \right)^2 \int_0^{2\pi} \left| \overline{G}_i(e^{j\omega}) \right|^2 d\omega \tag{4.79}$$

Therefore defining

$$A_i = \max_{\omega} \left| \prod_{j=1}^{i} \overline{H}_j(e^{j\omega}) \right|$$

$$B_i = \max_{\omega} \left| \prod_{j=i+1}^{N_s} \overline{H}_j(e^{j\omega}) \right|$$

$$C_i = k_i \frac{Q^2}{12} \cdot \frac{1}{2\pi} \int_0^{2\pi} \left| \overline{G}_i(e^{j\omega}) \right|^2 d\omega \tag{4.80}$$

we have

$$\sigma_i^2 = A_i^2 \, B_i^2 \, C_i \qquad 1 \leq i \leq N_s - 1 \tag{4.81}$$

For the moment assume that C_i is a constant factor independent of ordering. Then σ_i^2 is proportional to $(A_i B_i)^2$. Note that for any i, A_i and B_i are the maxima of two functions whose product is $H(e^{j\omega})$. Furthermore, for some i either $A_i = Pk$ or $B_i = Pk$. Now suppose $Pk >> C$. Without loss of generality we may assume $A_i = Pk$. We then argue that $A_i B_i >> C$.

Clearly $C \geq 1$, since we must have

$$\max_{\omega} |\overline{H}_i(e^{j\omega})| \geq \left(\max_{\omega} |\prod_{j=1}^{N_s} \overline{H}_j(e^{j\omega})|\right)^{\frac{1}{N_s}} = 1 \qquad (4.82)$$

In fact, C is found to be an increasing function of N_s given other parameters fixed. A plot of measured values of C for typical filters is shown in Fig. 4.26. These filters are listed in Table 4.2. We see that typically $C > 2$.

For simplicity of notation let

$$A(z) = \prod_{j=1}^{i} \overline{H}_j(z)$$

$$B(z) = \prod_{j=i+1}^{N_s} \overline{H}_j(z) \qquad (4.83)$$

so that $A_i = \max_{\omega}|A(e^{j\omega})|$ and $B_i = \max_{\omega}| B(e^{j\omega})|$. Clearly $A_i = |A(e^{j\omega_o})|$ for some ω_o. Now $A(z)B(z) = H(z)$, and $H(z)$ is a function with zeros only in the z-plane other than the origin. Also, at least in the case of well-designed band-select filters, the zeros of $H(z)$ are well

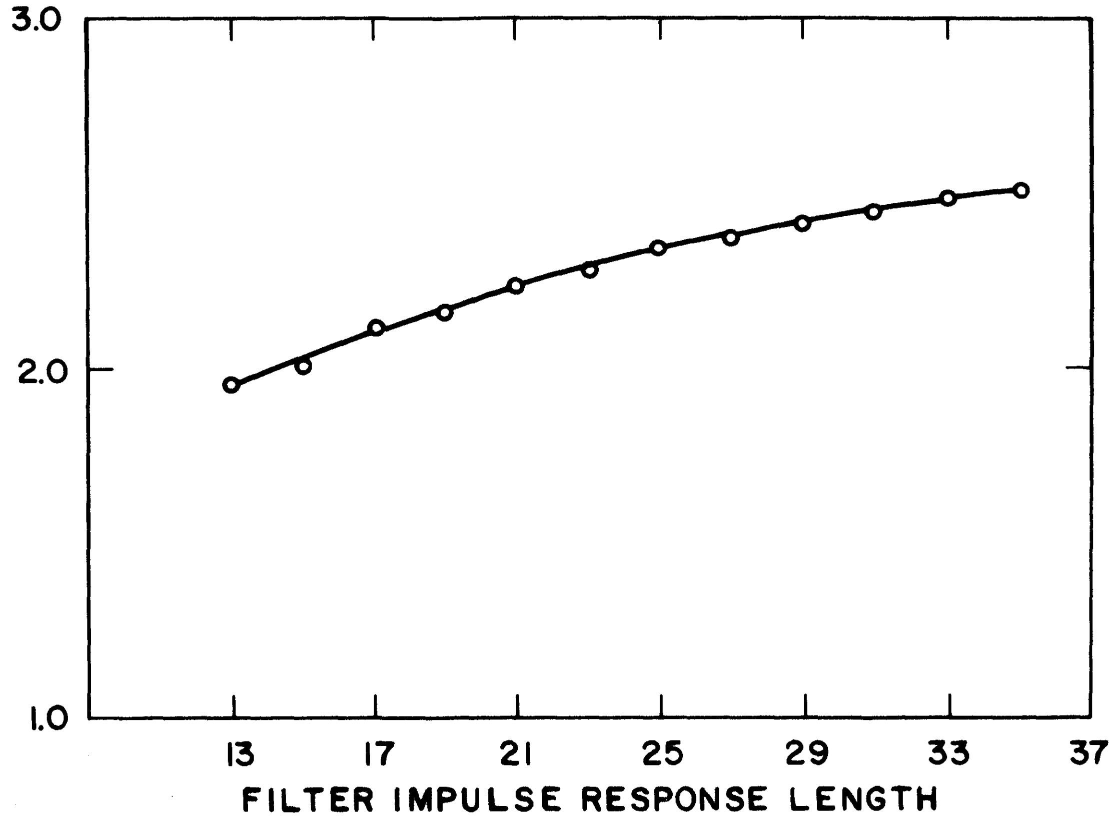


FIG. 4.26 VALUES OF PARAMETER C FOR DIFFERENT FILTER LENGTHS

Table 4.2

Tabulation of Filters Used for Fig. 4.26

D1 = .01
D2 = .01

N	N_p	F_1	C
13	4	.213	1.97
15	4	.185	2.02
17	5	.222	2.12
19	5	.198	2.16
21	6	.227	2.24
23	6	.207	2.28
25	7	.231	2.34
27	7	.214	2.37
29	8	.234	2.42
31	8	.218	2.45
33	9	.236	2.49
35	9	.222	2.51

spaced and spread out around the unit circle. Plots of the zeros of typical filters are shown in Fig. 4.7 and 4.27. Furthermore, $|H(e^{j\omega})| \leq 1$. Thus in order for $A(e^{j\omega})$ to have a large peak at ω_o, several zeros of $H(z)$ which occur in the vicinity of $z = e^{j\omega_o}$ must be missing from $A(z)$, while most of the remaining zeros must be in $A(z)$. This means that $B(z)$ has a concentration of zeros around $e^{j\omega_o}$. Recalling from (4.3) and (4.10) the shape of the frequency spectra associated with individual zeros, we see that most factors of $B(e^{j\omega})$ must have maxima which occur at exactly the same ω. Since each maximum typically has value $C > 2$, $B(e^{j\omega})$ is very likely to have a peak which is at least 1, or $B_i \geq 1$. Thus $A_iB_i >> C$. By the same token if $B_i = Pk$ and $Pk >> C$ then $A_iB_i >> C$.

Hence if $Pk >> C$, then for at least one i $\sigma_i^2 = (A_iB_i)^2C_i$ where $A_iB_i >> C$. Compared with a nominal value of say $A_iB_i = C$ the resulting difference in output noise variance can be great. When Pk takes on its lowest possible value, viz. $Pk = C$, the σ_i^2's are comparatively small for all i, hence we may expect that the resulting σ^2 is among the lowest values possible. Thus we have established a correlation between high values of Pk and high noise, and low values of Pk and low noise.

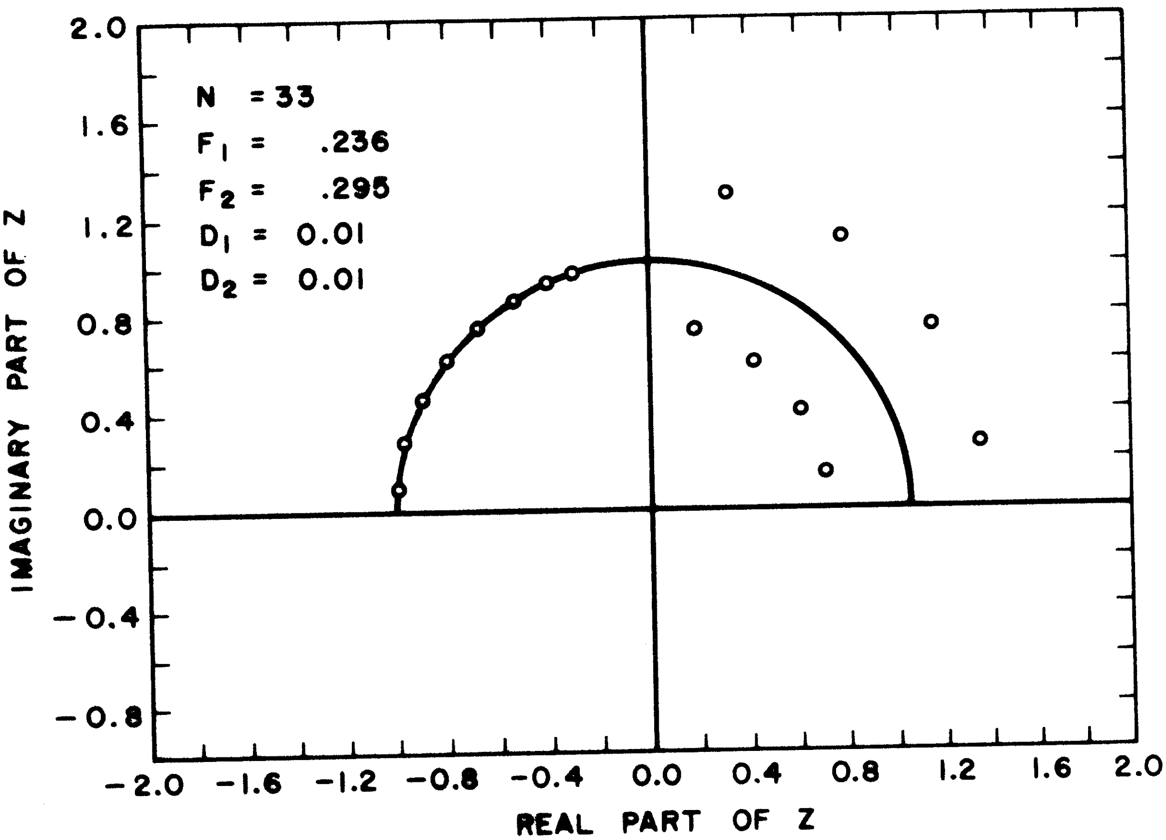


FIG. 4.27 (a) — ZEROS OF TYPICAL 33 POINT FILTER

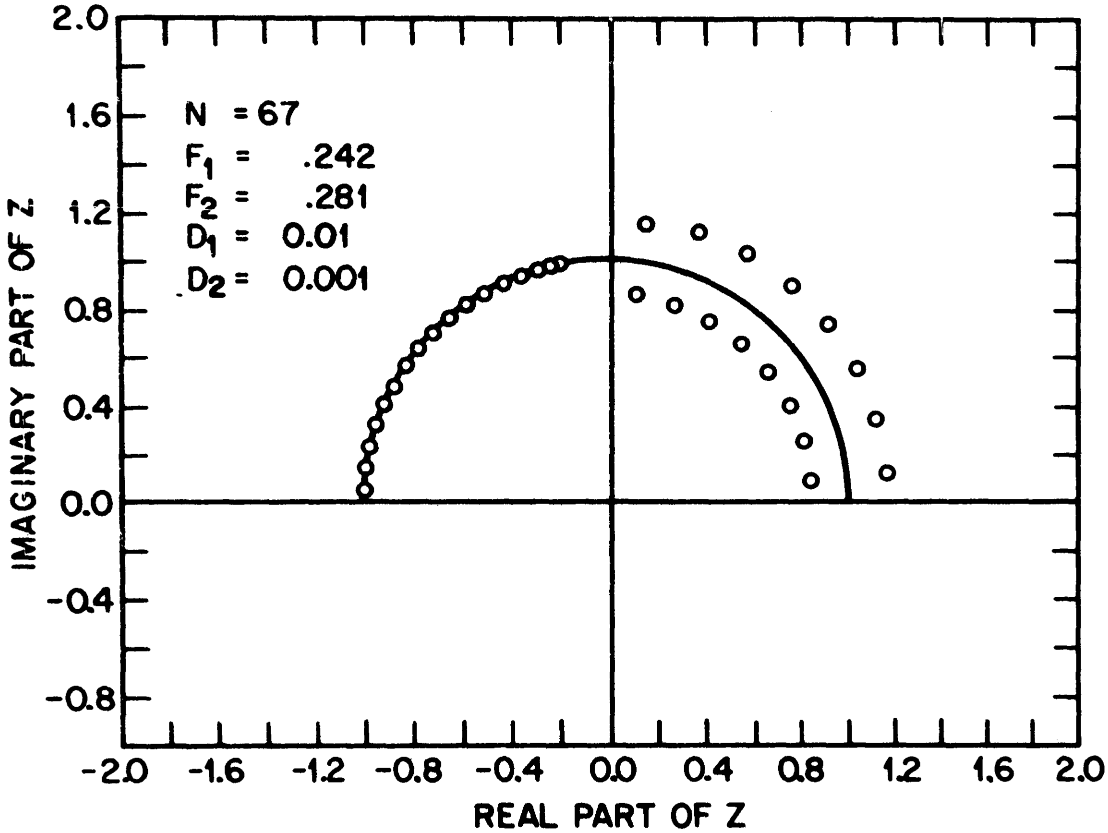


FIG. 4.27 (b) - ZEROS OF TYPICAL 67 POINT FILTER

Concerning the assumption that C_i is constant independent of ordering, it is reasonable as long as only order of magnitude estimates are of interest. Since by definition $\max_{\omega} |\overline{G}_i(e^{j\omega})| = 1$ independent of ordering and i, we can expect that variations in C_i with ordering is much less than variations in $(A_i B_i)^2$.

What all these arguments lead to is the result that the output noise variance from a cascade filter can be minimized by choosing an ordering which yields a minimal value of Pk. The usefulness of this result lies in the fact that Pk is a parameter whose magnitude can be judged by inspection much more easily than the magnitude of σ^2. Thus a means is provided for judging whether an ordering is likely to have high noise or low noise without having to calculate σ^2. For example, we can conclude that an ordering which groups together either at the beginning or at the end of a filter several zeros all from either the left half or the right half of the z-plane is likely to yield very high noise. This observation is based on the fact that zeros from the same half of the z-plane produce frequency spectra whose maxima occur at exactly the same ω (namely 0 or π). Hence several zeros from the same half of the z-plane can build up a large peak in $A(e^{j\omega})$ or $B(e^{j\omega})$ for several i. On the other hand, a scheme which orders sections so that the angle of the zeros synthesized by each section lies closest to the ω at which the maximum of the spectrum of the combination of the preceeding sections occurs is likely to yield a low noise filter.

The above observations are found to be true for all the filters whose noise distributions were investigated. For example, from the list of all orderings and noise variances for the filter of Fig. 4.6, shown in Appendix A.1, we see that those orderings which group together all three sections 4, 5, and 6 of this filter (see Fig. 4.7) either at the beginning or at the end of the filter are precisely those which have the highest noise, viz. with $\sigma^2 > 81Q^2$. Furthermore, the next highest in noise output are those orderings in which sections 4 and 6 or 5 and 6 occur side by side at the beginning or the end of the filter. Similarly, a different 13-point filter, one whose highest output noise variance is only $15.8\ Q^2$ and whose noise distribution was shown in Fig. 4.12, also confirms our remarks. Its list of orderings and noise values in Appendix A.2 shows that all noise variances above 2.8 Q^2 are produced by orderings which group together sections 5 and 6 at the beginning or the end of the filter. From the plot of its zeros in Fig. 4.28 we see that sections 5 and 6 are those sections which synthesize the passband zeros on the right half of the z-plane.

Using the results on the noise distribution of a filter and the results of this section, we can say that the comparatively few orderings of a filter which have unusually high noise can be avoided simply by judiciously choosing zeros for each section so that no large peaking in the spectrum either as seen from the input to each section or from each section to the output is allowed to occur. In particular this can be done by ensuring that from the input to each section the zeros synthesized well represent all values of ω, i.e., the variation in the density over ω of zeros chosen should be minimal.

Although the correlation between high peaking (large Pk) and high noise was not rigorously shown, we can overbound the possible values of output noise variances in terms of Pk. For since $A_i B_i \leq Pk^2$ and by (4.76) and (4.80)

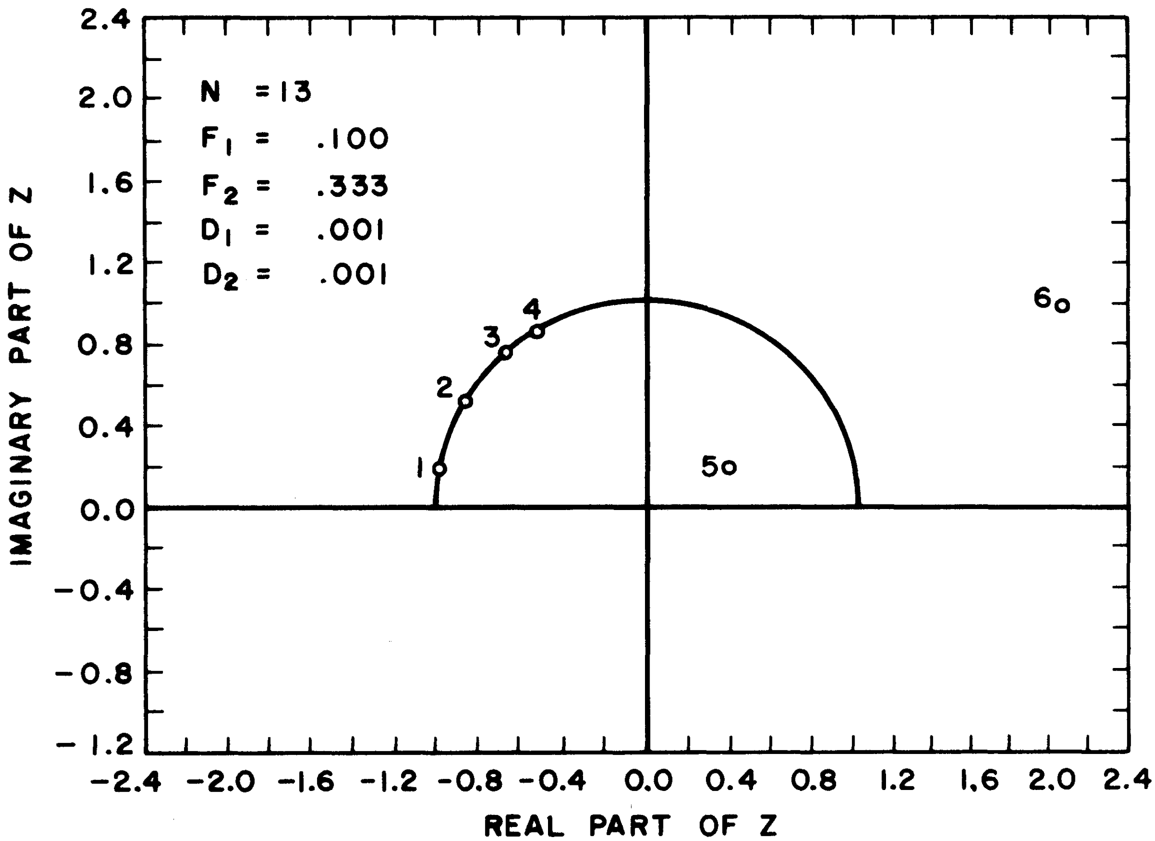


FIG. 4.28 — ZEROS OF FILTER OF FIG. 4.12

$$C_i < k_i \frac{Q^2}{12} \leq \frac{Q^2}{4} \qquad (4.84)$$

we have

$$\sigma_i^2 < \frac{Pk^4}{4} Q^2 \qquad 1 \leq i \leq N_s-1 \qquad (4.85)$$

Hence

$$\sigma^2 \leq (N_s-1)\, \sigma_i^2 + \frac{Q^2}{4}$$

$$< \left[\frac{(N_s-1)\, Pk^4 + 1}{4} \right] Q^2 \qquad (4.86)$$

Considering the ordering in which all zeros of a filter are sequenced according to increasing angle, we see that Pk can increase as $C^{N_s/2}$. Hence the high noise values of a filter can increase exponentially with N_s (recall that C also increases with N_s). This was shown experimentally to be true for small N_s in the previous section. On the other hand, (4.86) shows that if Pk can always be chosen small (eg. Pk = C), then noise variances bounded by an approximately linear increase with N_s can be guaranteed. Thus for large N_s the difference can be very important. In general we cannot guarantee that

Pk = C can be attained. However, in several of the filters experimentally investigated (N_s = 6) Pk = C was indeed attained for several orderings. In any event by virtue of (4.86) the minimization of Pk is certainly a working means for minimizing roundoff noise.

The values of Pk for all orderings of several filters have been measured. The results for a typical 13-point filter, namely that listed as no. 14 in Table 4.1, are listed in Appendix A.3. The designation of orderings refers to the numbering scheme in Fig. 4.29 and all noise variances are calculated using peak scaling on the filters. The list is arranged in order of increasing noise value. We see that the orderings with the highest value of Pk are indeed those with the highest noise, viz. having $\sigma^2 > 109\ Q^2$, while where Pk has the lowest value, $\sigma^2 < 1.7\ Q^2$. Thus our arguments are supported. Those orderings with noise values between these extremes are less well behaved in terms of correlation between σ^2 and Pk. For this filter there are only 4 possible values for Pk. However, for higher order filters we would expect a much larger spectrum of values for Pk.

If Pk and σ^2 were in fact well correlated, we would expect each ordering of a filter to have a noise

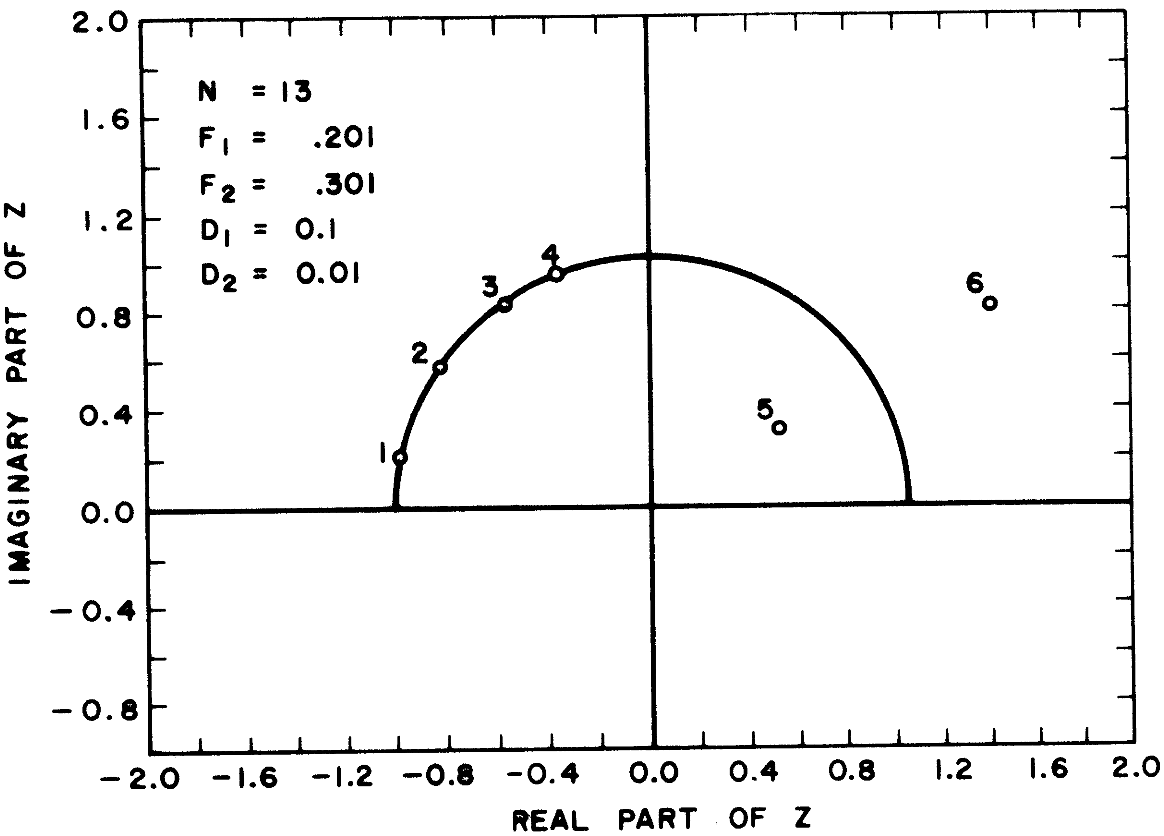


FIG. 4.29 – ZEROS OF FILTER #14 OF TABLE 4.1

value comparable to that of its reverse, since both have the same value of Pk. Appendix A.4 lists the σ^2 and Pk values of each ordering of filter no. 14 and those of its reverse on the same line. (Note that sections 5 and 6 synthesize reciprocal zeros, hence their relative order does not matter.) The overall list is ordered according to increasing noise values on the left side. Comparing the right and left sides of this list we see that indeed reversed orderings have comparable noise variances. This experimental result is a further confirmation that Pk and σ^2 are well correlated.

As a final illustration, plots of the spectra $\{|F_i(e^{j\omega})|\}$ and $\{|G_i(e^{j\omega})|\}$ (as defined in sections 3.2 and 3.3) for a high noise and a low noise ordering of filter no. 14 (peak scaled) are shown in Appendices B.1 and B.2 respectively. From Appendix A.3 we see that the former ordering, namely 213456, has $\sigma^2 = 186\ Q^2$ while the latter, 351462, has $\sigma^2 = 1.1\ Q^2$. Note that as expected, for the high noise ordering the spectra $|G_i(e^{j\omega})|$ have large maxima for at least one i, reaching a value of 60, while for the low noise ordering $|G_i(e^{j\omega})| < 2.2$ for all i. Since, in reference to (4.80), $A_i = 1$ and $B_i = \max_{\omega} |G_i(e^{j\omega})|$ for all i in both orderings, we see that indeed the high noise ordering has large values of A_iB_i. Along the same lines, we see also that C_i, the integral of $|G_i(e^{j\omega})|^2$ with its maximum normalized to unity, does not vary too much between the two orderings.

Finally, we point out that the high noise ordering has its zeros sequenced almost exactly in the order they occur around the unit circle, while the sequencing of zeros in the low noise ordering obeys the rule of good representation of all values of ω, thus resulting in a sequence which "jumps around" a great deal around the unit circle. Furthermore notice how the spectrum of each section in the low noise ordering tends to suppress the peak in the spectrum of the combination of previous sections. Our deductions are thus further supported.

In the next section we shall see how the results of this section give basis to an algorithm which can be used successfully to find low noise orderings for cascade filters.

5.0 An Algorithm for Obtaining a Low Noise Ordering for a Cascade Filter

An extensive analysis of roundoff noise in cascade form FIR filters has been presented in the previous sections. However, an investigation of roundoff noise would not be complete without studying the practical question which in the first place had motivated all the analyses and experimentation. The question is, given an FIR transfer function desired to be realized in cascade form, how does one systematically choose an ordering for the filter sections so that roundoff noise can be kept to a minimum?

A partial answer to this question has already been given in the previous section. However, no completely systematic method has yet been devised for selecting an ordering for a filter guaranteed to have low noise. Ultimately, one wishes to find an algorithm which, when implemented on a computer, can automatically choose a proper ordering in a feasible length of time.

Avenhaus has studied an analogous problem for cascade IIR filters and has presented an algorithm for finding a "favorable" ordering of filter sections[16]. His algorithm consists of two major steps; a "preliminary determination" and a "final determination." In this section we shall describe an algorithm for ordering FIR filters which is based upon the procedure used in the "preliminary determination" step of Avenhaus' algorithm. We have found that a procedure appended to our algorithm similar to Avenhaus' final determination step adds little that is really worth the extra computation time to the already very good solution obtainable by the first step. Hence such a procedure is not included in our algorithm.

No statement was made by Avenhaus as to what range of noise values can be expected of filters ordered by his algorithm, nor did he claim that his algorithm always yields a low noise ordering (relatively speaking, of course). However, based on the results of sections 4.1 to 4.4, we shall be able to argue heuristically that our algorithm always yields filters which have output noise variances among the lowest possible. Together with extensive experimental confirmation, these arguments enable us to be confident that our algorithm produces solutions that are very close to the optimum.

Application of Avenhaus' procedure to FIR filters also enables us to introduce modifications which reduce significantly the amount of computation time required. Finally, while IIR filters of higher order than the classic 22nd order bandstop filter quoted by Avenhaus are of very little practical usefulness because of high coefficient sensitivity problems, practical FIR filters can well have orders over 100. Though the same basic algorithm should still work for high orders, care must be exercised in performing details to avoid large roundoff errors in the computations. Through proper initialization our algorithm has been successfully tested for filters of order up to at least 128.

With these remarks as introduction we now describe the basic procedure or algorithm proposed by Avenhaus. The procedure is simply the following. To order a filter of Ns sections, begin with i=Ns and permanently build into position i in the cascade the filter section which, together with all the sections already built in, results in the smallest possible variance for the output noise component due to noise sources in the i^{th} section of the cascade. Because in an FIR cascade filter noise is injected only into the output of each section, for FIR filters we need to modify the procedure and consider the output noise due to the section in position i-1 rather than i when choosing a section for position i. But the i^{th} section is determined before the $(i-1)^{th}$ section, hence the number of noise sources at the output of the $(i-1)^{th}$ section is unknown at the time that a section for position i is to be chosen. This problem is overcome by assuming all sections to have the same number of noise sources. Then σ_i^2 is simply proportional to $\sum_k g_i^2(k)$ independent of what the i^{th} section is (see section 3.2 for notational definitions).

Hence the revised basic algorithm for ordering FIR cascade filters is: <u>beginning with i=Ns, permanently build into position i the section which, together with the sections already built in, causes the smallest possible value for $\sum_k g_{i-1}^2(k)$.</u> Once this basic algorithm is determined, we need only decide on a scaling method and a computational algorithm for accomplishing the desired scaling and noise evaluation before an ordering algorithm is completed. Prior to discussing these issues, let us consider why the basic algorithm described above is always able to find a low noise ordering.

The reason why the algorithm might not be able to find a low noise ordering is that rather than minimizing $\sum \sigma_i^2$ directly, it minimizes each σ_i^2 individually where for σ_j^2, $1 \leq j \leq Ns-1$, the search is essentially conducted over only (j+1)! out of the total of Ns! possible orderings. Now this set of (j+1)! orderings depends on which sections were chosen for positions j+2 to Ns in the cascade if j < Ns-1. Hence in choosing a section for position j, previous choices might prevent attainment of a sufficiently small value for σ_{j-1}^2.

The basis for our arguments is the results of section 4.4. Let H(z) be an appropriately scaled filter. Given j, $1 \leq j \leq Ns-1$, suppose σ_i^2 is small for all $i \geq j$. Then the zeros of $\prod_{i=j+1}^{N_s} H_i(z)$ must be well spread around the unit circle in the z-plane since a clustering would cause large peaking in $\prod_{i=k+1}^{N_s} \overline{H}_i(e^{j\omega})$ for some $k \geq j$, hence a large value of σ_k^2. But this means that the remaining zeros of H(z), namely those in $\prod_{i=1}^{j} H_i(z)$, must also be well spread around the unit circle, since the zeros of H(z) are distributed almost uniformly around the unit circle. Hence it ought certainly to be possible to find some pair of zeros in $\prod_{i=1}^{j} H_i(z)$ which, when assigned to position j, causes little peaking in $\prod_{i=1}^{j-1} \overline{H}_i(z)$ or $\prod_{i=j}^{N_s} \overline{H}_i(z)$, and thus results in a small value for σ_{j-1}^2. By induction, then, σ_i^2 can be chosen small for all i.

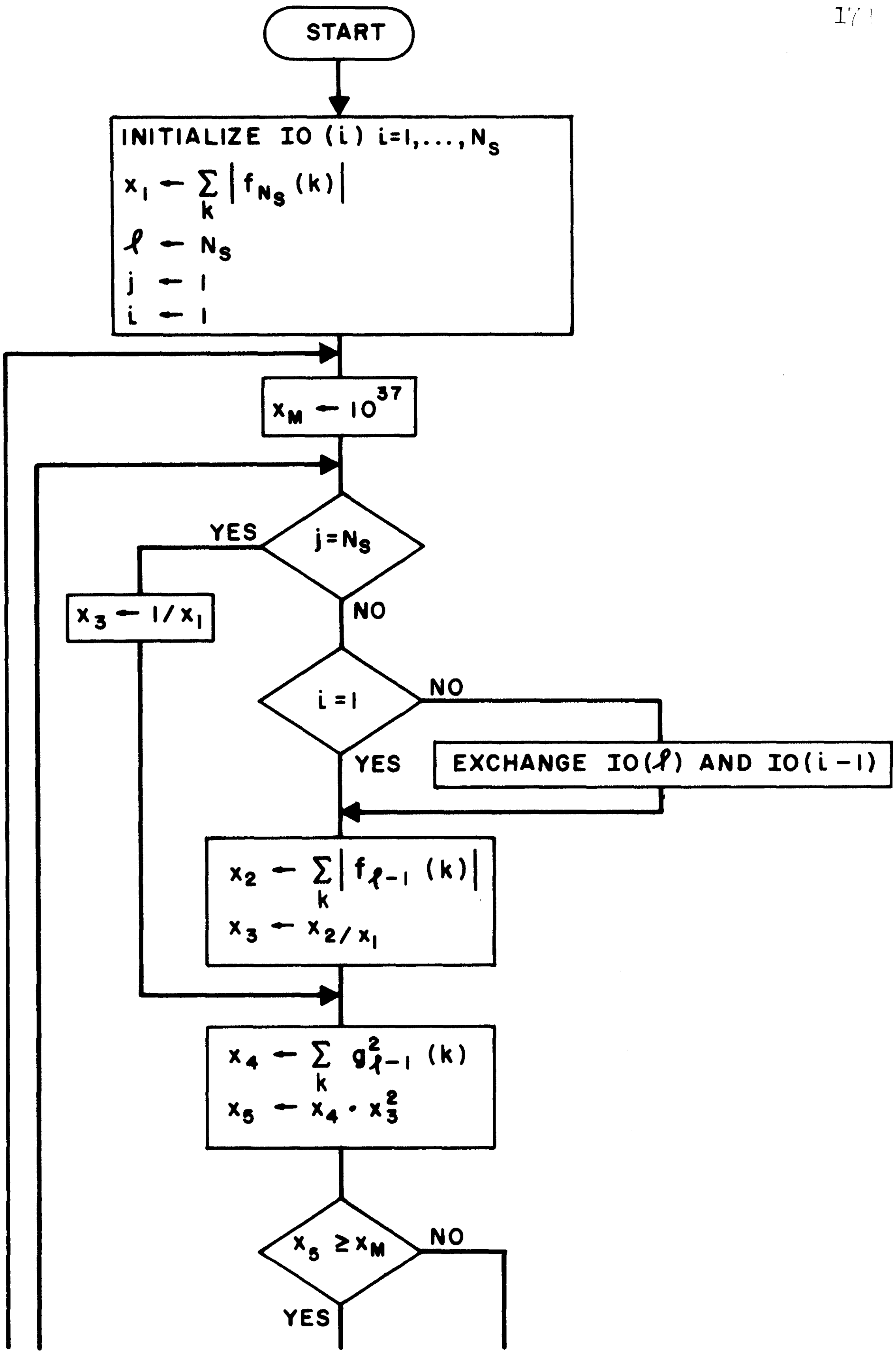
For small j it is true that there are very few zeros left as candidates for position j, but in these positions little peaking in the spectra can occur since the overall spectrum $\prod_{i=1}^{N_s} H_i(e^{j\omega})$ must be a well behaved filter characteristic. Typically in a high noise ordering σ_j^2 reaches a peak for j somewhere in the middle between 1 and Ns, while σ_j^2 for small j has little contribution to σ^2. Hence the choice of sections for small j is not too crucial. Of course, the eligible candidates are still well-spaced zeros as for larger j, so that peaking should not be a problem.

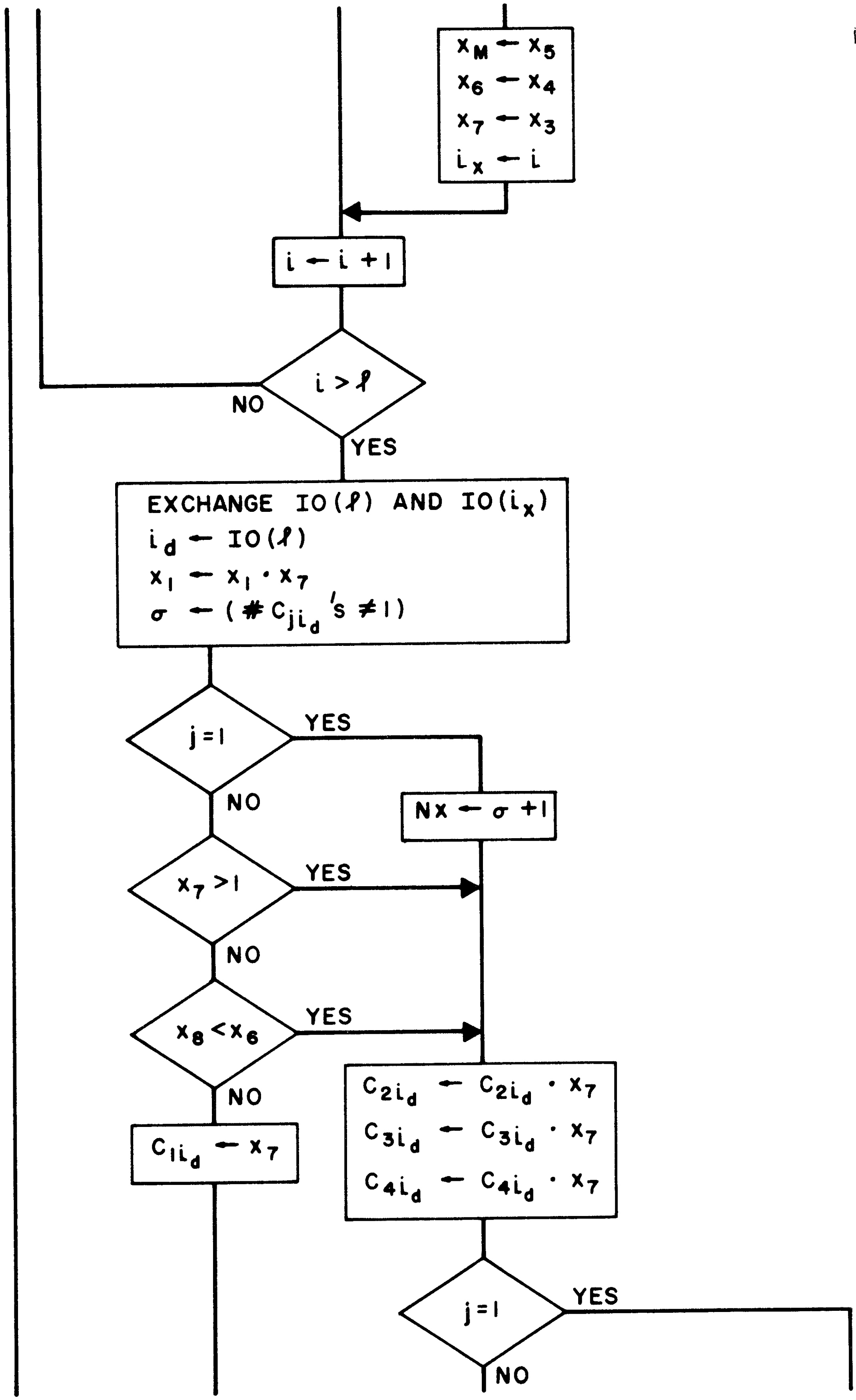
Note that the reason the algorithm works so well is tied in with the result of section 4.1 that most orderings of a filter have comparatively low noise. That most orderings of a filter have low noise is principally because there are far more ways to sequence zeros around a circle so that they are well "interlaced" and do not cluster than if they are to form clusters. Because it is not difficult to find low noise arrangements of zeros, we are able to minimize $\sum \sigma_i^2$ by minimizing each σ_i^2 independently, searching over a much smaller domain. If we were not able to segment the sum $\sum \sigma_i^2$, searching for a minimum would be essentially an impossible task because of time limitations.

Having discussed why the basic algorithm works, we now turn to the practical problem of implementing it. First of all, we have the choice of scaling method to use in computing the $\sum_k g_i^2(k)$. As in the calculation of noise distributions in section 4.1, sum scaling is to be preferred since it can be carried out the fastest. Figure 5.1 shows a flow chart of the ordering algorithm in which sum scaling is employed. Calculation of σ^2 (NX in the flow chart) is done exactly the same way as in the algorithm of Fig. 4.5.

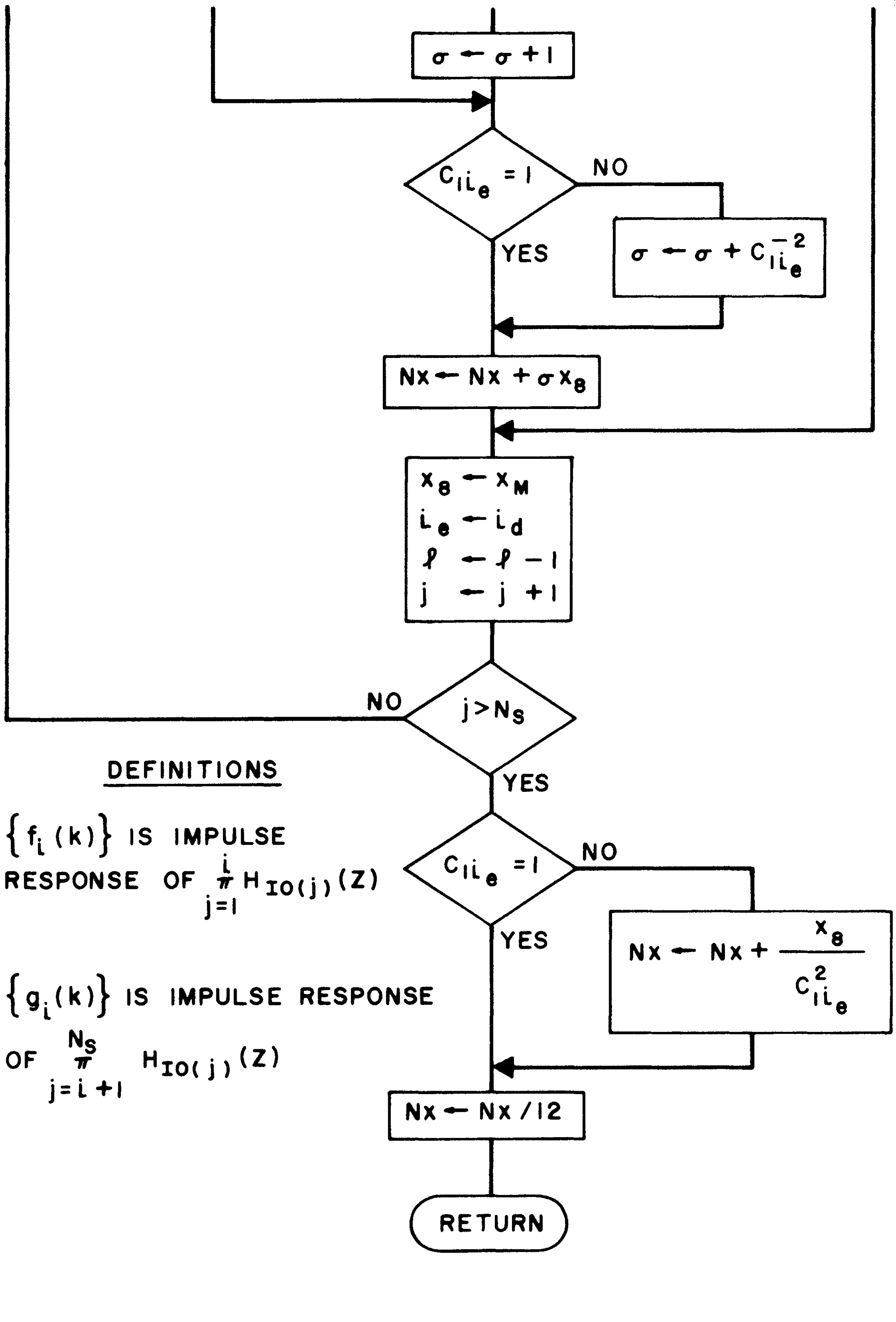
Using this ordering algorithm, over 50 filters have been ordered and the noise variances in units of Q^2 (Q = quantization step size) of the resulting filters are shown in the last columns of Tables 4.1, 5.1 and 5.2. Note that these noise variances are computed with sum scaling applied to the filters. The corresponding noise variance values for peak scaling have also been computed for the filters of Table 5.1. These are shown in the third to the last column of that table. The comparability of these noise values to those for sum scaling has already been pointed out in section 4.2.

For an alternative implementation of the basic ordering algorithm, peak scaling can be used. To distinguish between the two different resulting algorithms,


START
INITIALIZE IO (i) i=1,...,N_S
x_1 ← Σ_k |f_{N_S}(k)|
ℓ ← N_S
j ← 1
i ← 1
x_M ← 10^37
j=N_S
YES
NO
x_3 ← 1/x_1
i = 1
NO
YES
EXCHANGE IO(ℓ) AND IO(i−1)
x_2 ← Σ_k |f_{ℓ−1}(k)|
x_3 ← x_2/x_1
x_4 ← Σ_k g²_{ℓ−1}(k)
x_5 ← x_4 · x_3²
x_5 ≥ x_M
NO
YES



$x_M \leftarrow x_5$
$x_6 \leftarrow x_4$
$x_7 \leftarrow x_3$
$i_x \leftarrow i$
$i \leftarrow i+1$
$i > \ell$
NO
YES
EXCHANGE IO(ℓ) AND IO(i_x)
$i_d \leftarrow$ IO(ℓ)
$x_1 \leftarrow x_1 \cdot x_7$
$\sigma \leftarrow (\# C_{ji_d}$'s $\neq 1)$
$j=1$
YES
NO
$Nx \leftarrow \sigma + 1$
$x_7 > 1$
YES
NO
$x_8 < x_6$
YES
NO
$C_{1i_d} \leftarrow x_7$
$C_{2i_d} \leftarrow C_{2i_d} \cdot x_7$
$C_{3i_d} \leftarrow C_{3i_d} \cdot x_7$
$C_{4i_d} \leftarrow C_{4i_d} \cdot x_7$
$j=1$
YES
NO



DEFINITIONS

$\{f_i(k)\}$ IS IMPULSE RESPONSE OF $\prod_{j=1}^{i} H_{IO(j)}(Z)$

$\{g_i(k)\}$ IS IMPULSE RESPONSE OF $\prod_{j=i+1}^{N_S} H_{IO(j)}(Z)$

FIG. 5.1 – FLOW CHART OF ORDERING ALGORITHM

Table 5.1

List of Filters and the Results of Ordering Algorithms

D1 = .01
D2 = .001

					Noise Variance		
					Peak Scaling		Sum Scaling
#	N	N_p	F_1	α	Alg. 1	Alg. 2	Alg. 1
28	13	4	.219	.65	1.25	1.26	0.90
29	15	4	.193	.68	1.23	1.22	1.02
30	17	5	.230	.61	1.99	2.49	1.37
31	19	5	.207	.64	1.93	1.92	1.47
32	21	6	.236	.59	2.50	2.61	1.58
33	23	6	.216	.61	2.57	2.91	1.77
34	25	7	.240	.57	3.75	3.62	2.35
35	27	7	.223	.59	3.95	4.11	2.45
36	29	8	.243	.55	4.54	5.04	2.67
37	31	8	.227	.57	5.27	5.88	2.74
38	33	9	.244	.54	7.81	6.67	4.59
39	35	9	.231	.55	6.01	6.43	3.72
40	33	1	.005	1.0	0.47	0.48	0.53
41	33	2	.029	.82	0.60	0.67	0.60
42	33	3	.059	.73	0.89	1.00	0.80
43	33	4	.090	.68	1.43	1.36	1.16
44	33	5	.121	.63	2.29	1.84	1.71
45	33	6	.152	.60	2.48	2.70	1.61
46	33	7	.183	.58	3.47	3.37	2.30
47	33	8	.214	.61	4.72	5.23	3.38
48	33	10	.275	.52	10.04	8.16	4.83
49	33	11	.305	.52	15.68	11.35	8.30
50	33	12	.334	.50	13.43	14.88	6.27
51	33	13	.363	.50	21.35	17.62	9.14
52	33	14	.392	.50	41.64	31.41	15.40
53	33	15	.419	.51	55.20	41.13	22.12
54	33	16	.448	.53	89.52	65.66	38.23

Table 5.2

List of Filters and the Results of Ordering Algorithms

$N = 33$
$N_p = 8$

					Noise Variance		
					Peak Scaling		Sum Scaling
#	F_1	D_1	D_2	α	Alg. 1	Alg. 2	Alg. 1
55	.211	.01	.002	.59	5.63	5.33	3.34
56	.208	.01	.005	.57	5.13	5.69	3.18
57	.205	.01	.01	.56	5.05	5.27	3.34
58	.202	.01	.02	.55	7.63	8.31	4.01
59	.197	.01	.05	.53	11.34	12.53	6.92
60	.193	.01	.1	.51	46.33	22.99	16.88
61	.238	.1	.01	.58	9.90	9.01	5.61
62	.227	.05	.01	.58	8.91	7.35	5.52
63	.214	.02	.01	.56	8.87	5.75	4.32
64	.196	.005	.01	.56	5.47	4.69	3.68
65	.185	.002	.01	.56	5.95	4.08	3.41
66	.178	.001	.01	.57	4.10	4.11	2.85

we shall refer to the former (sum scaling) as alg. 1 and the latter as alg. 2. The only changes to Fig. 5.1 needed to realize alg. 2 rather than alg. 1 is to replace $\sum_k |f_i(k)|$ by $\max_\omega |F_i(e^{j\omega})|$ for given i whenever it appears. Results of using alg. 2 on the filters of Tables 5.1 and 5.2 are shown in the second last column of those tables. Observe that though the two algorithms in general yield different orderings for a given filter, the resulting noise variances are very comparable. Thus with both alg. 1 and alg. 2 we can obtain two separate low noise orderings for a given filter.

At this point let us digress for a moment to examine more closely the results presented in Tables 4.1, 5.1 and 5.2. Note from Table 4.1 how close to the minimum, if not the very minimum, a noise variance alg. 1 is able to result in. From this observation and the results of section 4.3 on the dependence of the minimum noise variance for a filter on different parameters, we are quite confident that the noise variances shown in Tables 5.1-5.2 are also very close to the minimum possible. The filters of Tables 5.1-5.2 were chosen intentionally to show the dependence of the results of the ordering algorithms on various transfer function parameters. We see that the noise variances indeed behave in the way that we would expect from the results of section 4.3. In particular, σ^2 is seen to be essentially an increasing function of N, F_1, D_1, as well as D_2. The results of Tables 5.1-5.2 are then a confirmation of the expectation that the results of section 4.3 on the general dependence of noise on transfer function parameters can be generalized to higher order filters.

We now return to the description of the algorithms. Even with a scaling method decided upon, the questions still remain of how $\sum_k g_i^2(k)$ and $\sum_k |f_i(k)|$ or $\max_\omega |F_i(e^{j\omega})|$ are to be computed and how the sequence $\{IO(i)\}$ is to be initialized. In obtaining the results of Tables 5.1-5.2 we have simply done the following. $\sum_k g_i^2(k)$ and $\sum_k |f_i(k)|$ were computed by evaluating $\{g_i(k)\}$ or $\{f_i(k)\}$ through simulation in the time domain (i.e. convolution). $\max_\omega |F_i(e^{j\omega})|$ was determined by transforming $\{f_i(k)\}$ via an FFT and then maximizing. Finally, $\{IO(i)\}$ was initialized to $IO(i) = i$, $i = 1,\ldots,N_s$. We shall see later that these procedures must be modified for higher order filters. But meanwhile let us consider what these procedures imply in terms of dependence of computation time on filter length.

Clearly, in algorithmically computing the impulse response of an N-point filter via convolution, the number of multiplies and adds required to calculate each point varies as N, hence the time required to evaluate the entire impulse response must vary approximately as N^2. Now in the basic algorithm there are two nested loops, where the number of times the operations within the inner loop are performed is given by

$$\sum_{i=1}^{N_s} i = \frac{N_s(N_s+1)}{2}$$

$$\cong \frac{N^2}{8}$$

Clearly for alg. 1 the evaluation of $\sum_k |f_{\ell-1}(k)|$ and $\sum_k g^2_{\ell-1}(k)$ dominates all operations within the inner loop in terms of time required. Since the total number of points required to evaluate for computing $\{f_{\ell-1}(k)\}$ and $\{g_{\ell-1}(k)\}$ together turns out to be a constant independent of ℓ, the combined operations must have approximately an N^2 time dependence. Hence we would predict that the computation time required for alg. 1 must be approximately proportional to N^4. This prediction is verified in Fig. 5.2

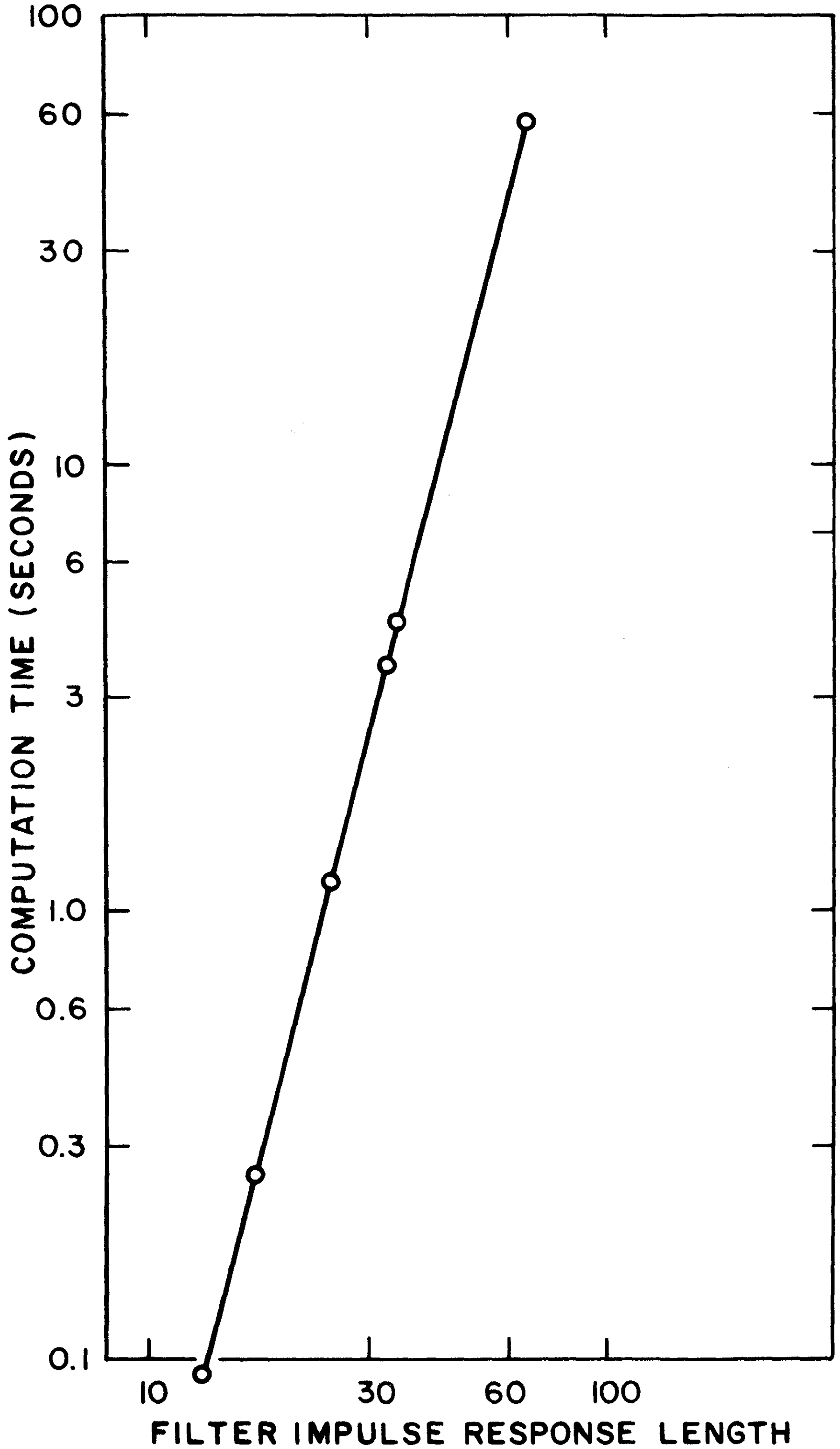


FIG. 5.2 COMPUTATION TIME VERSUS FILTER LENGTH FOR ORDERING ALGORITHM

where computation time for alg. 1 on the Honeywell 6070 computer is plotted against N on log-log coordinates for various values of N. As expected, these points lie on a straight line with a slope very nearly equal to 4.

For alg. 2 exactly the same procedures as in alg. 1 are carried out except that after each evaluation of $\{f_i(k)\}$ an FFT is performed. Thus for a given N alg. 2 always requires more time than alg. 1, with the exact difference depending on the number of points employed in the FFT.

For filters of length greater than approximately 41, it is found that accuracy in the evaluation of impulse response samples by the methods described rapidly breaks down. This phenomenon is chiefly due to the fact that the initial ordering used is a very bad one. In particular, we have seen that this ordering (i.e., IO(i) = i) has a noise variance which is among the highest possible and which increases exponentially with N. Thus all attempts at evaluating the impulse response of the filter by simulation in the time domain is marred by roundoff noise.

A natural possibility for resolving this problem is to perform calculations in the frequency domain. This we have tried as a modification to alg. 2. In particular, rather than computing $F_i(e^{j\omega})$ from $\{f_i(k)\}$, we evaluate it as a product of $H_j(e^{j\omega})$, $j = 1,\ldots,i$, where each $H_j(e^{j\omega})$ is computed from the coefficients of section j via an FFT. In this way the accuracy problem was solved, but computation time increased significantly. As an example the 67-point filter listed in Table 5.3 was ordered using this method. The resulting noise variance was a reasonable 26.6 Q^2, but even with a 256-point FFT the computation time required amounted to 7.2 minutes, more than 7 times that required for alg. 1 to order the same filter.

A far better solution is as follows. Recall from section 4.1 that most orderings of a filter have relatively low noise. Thus if we were to choose an ordering at random, we ought to end up with an ordering which has relatively low noise. The strategy is then to use a random ordering as an initial ordering for alg. 1. A given ordering of a sequence of numbers $\{IO(i),\ i = 1,\ldots,N_s\}$ can be easily randomized using the following shuffling algorithm[25]:

step 1: Set $j \leftarrow N_s$.

step 2: Generate a random number U, uniformly distributed between zero and one.

step 3: Set $k \leftarrow \lfloor jU \rfloor + 1$. (Now k is a random integer between 1 and j.) Exchange $IO(k) \leftrightarrow IO(j)$.

step 4: Decrease j by 1. If $j > 1$, return to step 2.

By adding a step to randomize the initial ordering IO(i) = i in alg. 1, the inaccuracy problem was eliminated. The interesting question now arises that since most orderings of a filter have relatively low noise, can we not obtain a good ordering simply by choosing one at random? The answer is yes, but as we shall shortly see, a random ordering is by far not as good as one which can be obtained using the ordering algorithm.

The extra step of randomizing the initial ordering for alg. 1 requires negligible additional computation time, and a filter with impulse response length as high as 129 has been successfully ordered in this way. The time required to order this filter was approximately 13.5 minutes. Except for time limitations, there is no reason why even higher order filters cannot be similarly ordered. The results of using the modified alg. 1 (denoted alg. 1') on this filter as well as a few other filters are shown in Table 5.3. Also shown in this table are the noise variances of these filters when they are in the sequential ordering IO(i) = i (where computable within the numerical range of the computer) as well as when they are in a random ordering (obtained by randomizing {IO(i)} where IO(i) = i, as described above). Because of the potentially very large roundoff noise encounterable in these orderings, the noise variances were computed using frequency domain techniques. In particular, each $H_j(e^{j\omega})$ is evaluated via an FFT; peak scaling is then performed; and finally σ_i^2 is computed via $\frac{1}{2\pi}\int_0^{2\pi} |G_i(e^{j\omega})|^2 d\omega$ rather than $\sum_k g_i^2(k)$.

From Table 5.3 we see that though the noise variances of the random orderings are certainly a great deal lower than those of the corresponding sequential orderings, they are far from being as low as those obtained by alg. 1'. Thus it is certainly advantageous to use alg. 1' to find proper orderings for cascade filters. In practice cascade FIR filters of orders over approximately 50 are of little interest since there exist more efficient

Table 5.3

List of Filters and the Results of Alg. 1′

D_1 = .01

					Noise Variance			
					Ordering		Alg. 1′	
#	N	N_p	F_1	D_2	Sequential	Random	Sum sc.	Peak sc.
38	33	9	.244	.001	1.0×10^{11}	6.2×10^{3}	4.59	7.81
67	47	12	.237	.001	4.3×10^{17}	2.2×10^{6}	6.47	12.07
68	67	17	.242	.001	3.3×10^{27}	1.5×10^{6}	16.77	30.03
69	101	25	.241	.001	$>10^{38}$	1.4×10^{5}	41.93	73.55
70	129	20	.153	.0001	-	5.5×10^{11}	17.98	37.54

ways than the cascade form to implement filters of higher orders. For filters of at most 50^{th} order the computation time required for alg. 1 is less than 20 seconds on the Honeywell 6070 computer. Thus alg. 1 (or 1') is also a very efficient means for ordering cascade filters.

In all the examples given, one can do little better in trying to find orderings with lower noise. With the possible exception of the uninteresting wide band filter, #54 in Table 5.1, all filters have less than 4 bits of noise (as defined in (3.15)) after ordering by alg. 1, while the great majority have less than 3 bits. Thus we do not expect that these noise figures can be further reduced by much more than a bit or so.

In summary, an algorithm has been described which enables a filter designer with access to a general-purpose computer to determine efficiently for a cascade FIR filter an ordering which has very low noise. The noise figures obtained are in general sufficiently close to the optimum so that little further improvement can be made. Thus for all practical design purposes, it is believed that the ordering problem for cascade FIR filters has been solved.

6.0 Conclusions

In this thesis a comprehensive investigation of the problem of roundoff noise in cascade FIR filters has been presented. We have considered the central issues of scaling and ordering for cascade filters. In particular, several methods of scaling to meet dynamic range constraints have been discussed, with emphasis on two types of methods which were named sum scaling and peak scaling. The effects of these two types of scaling methods were compared and found to be closely related. With regard to ordering, a specific algorithm has been presented to automatically choose a proper ordering for any given FIR filter.

In addition to these central issues, the dependence of roundoff noise on various filter transfer function parameters has also been determined. This knowledge enables a designer to predict, based on known results, the level of noise to expect in new situations. Finally, an explanation of why some orderings of a filter have low noise while others have high noise has been developed in terms of characteristics of a filter which provide good intuitive "feel." Based on the notions developed we are able to characterize and recognize high noise orderings and to explain one of the results of our research that given a filter some orderings have relatively very high noise but most orderings have relatively low noise.

While a fairly complete study of roundoff noise in cascade FIR filters has been presented in this thesis, none of the issues involved in other types of quantization effects has been touched upon. In particular, the question still remains as to what is the best way to obtain transfer functions whose coefficients are quantized. Furthermore, in addition to the cascade form, many other structures exist in which FIR transfer functions can be realized[14]. Some of these structures may prove to be particularly advantageous under certain circumstances. In order that these structures may be intelligently compared, a great deal more must be understood concerning quantization effects in them. Ultimately it would be desirable to know for any given filter and application just what is the best structure to use. These are some of the problems open to further research.

BIBLIOGRAPHY

1. B. Gold and C. M. Radar, Digital Processing of Signals, McGraw-Hill, 1969.
2. T. W. Parks and J. H. McClellan, "Chebyshev Approximation for Non-Recursive Digital Filters with Linear Phase," IEEE Transactions on Circuit Theory, Vol. CT-19, March 1972.
3. E. Hofstetter, A. V. Oppenheim and J. Siegel, "A New Technique for the Design of Non-Recursive Digital Filters," Proc. Fifth Annual Princeton Conference on Information Sciences and Systems, 1971.
4. L. R. Rabiner, B. Gold and C. A. McGonegal, "An Approach to the Approximation Problem for Non-Recursive Digital Filters," IEEE Transactions on Audio and Electro-acoustics, Vol. AU 18, No. 2, June, 1970.
5. L. R. Rabiner, "Techniques for Designing Finite-Duration Impulse-Response Digital Filters," IEEE Transactions on Communication Technology, Vol. COM-19, April, 1971.
6. O. Herrmann, and W. Schüssler, "On the Accuracy Problem in the Design of Non-Recursive Digital Filters," Archiv Der Elektrischen Übertragung, Band 24, 1970.
7. E. Avenhaus, "On the Design of Digital Filters With Coefficients of Limited Word Length," IEEE Transactions on Audio and Electroacoustics, Vol. AU-20, No. 3, August 1972.

8. J. B. Knowles and E. M. Olcayto, "Coefficient Accuracy and Digital Filter Response," IEEE Transactions on Circuit Theory, Vol. CT-15, March 1968.

9. C. M. Rader and B. Gold, "Effects of Parameter Quantization on the Poles of a Digital Filter," Proceedings of the IEEE, May 1967.

10. C. J. Weinstein, "Quantization Effects in Digital Filters," Technical Report 468, Lincoln Laboratory, Lexington, Mass., November 1969.

11. L. B. Jackson, "Roundoff-Noise Analysis for Fixed-Point Digital Filters Realized in Cascade or Parallel Form," IEEE Transactions on Audio and Electroacoustics, Vol. AU-18, June 1970.

12. L. B. Jackson, "On the Interaction of Roundoff Noise and Dynamic Range in Digital Filters," The Bell System Technical Journal, Vol. 49, February 1970.

13. B. Gold and C. M. Rader, "Effects of Quantization Noise in Digital Filters," Presented at 1966 Spring Joint Computer Conference, AFIPS Proc., 28, 1966.

14. W. Schüssler, "On Structures for Nonrecursive Digital Filters," Archiv Für Elektronik Und Übertragungstechnik, Band 26, 1972.

15. L. R. Rabiner and R. W. Schafer, "Recursive and Nonrecursive Realizations of Digital Filters Designed by Frequency Sampling Techniques," IEEE Transactions on Audio and Electroacoustics, Vol. AU-19, September 1971.

16. E. Avenhaus, "Realizations of Digital Filters with a Good Signal-to-Noise Ratio," Nachrichtentechnische Zeitscrift, May 1970.

17. L. R. Rabiner and K. Steiglitz, "The Design of Wide-Band Recursive and Nonrecursive Digital Differentiators", IEEE Transactions on Audio and Electroacoustics, Vol. AU-18, no. 2, June 1970.

18. T. W. Parks and L. R. Rabiner, "On the Transition Width of Finite Impulse Response Digital Filters," submitted to IEEE Transactions on Audio and Electroacoustics.

19. O. Herrmann, "Design of Nonrecursive Digital Filters with Linear Phase," Electronics Letters, Vol. 6, no. 11, 1970.

20. L. R. Rabiner, "The Design of Finite Impulse Response Digital Filters Using Linear Programming Techniques," The Bell System Technical Journal, Vol. 51, No. 6, July-August, 1972.

21. Y. Chu, Digital Computer Design Fundamentals, New York, McGraw-Hill, 1962.

22. W. H. Fleming, Functions of Several Variables, Addison-Wesley, 1965, pp. 200-204.

23. J. R. Rice, The Approximation of Functions, Addison-Wesley, 1964, pp. 4-10.

24. J. F. Kaiser, "Digital Filters," ch. 7 in Systems Analysis by Digital Computer, F. F. Kuo and J. F. Kaiser, Eds., New York, Wiley, 1966.

25. D. E. Knuth, Seminumerical Algorithms - vol. 2 of The Art of Computer Programming, Addison-Wesley, 1969, pp. 125.

26. R. W. Hankins, "Design Procedure for Equiripple Non-recursive Digital Filters," S. M. Thesis, Dept. of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Mass., 1971.

27. W. R. Bennett, "Spectra of Quantized Signals," The Bell System Technical Journal, Vol. 27, July 1948.

28. J. B. Knowles and R. Edwards, "Effect of a Finite-Word-Length Computer in a Sampled-Data Feedback System," Proc. IEE, Vol. 112, No. 6, June 1965.

29. T. Kaneko and B. Liu, "Round-off Error of Floating-Point Digital Filters," Proc. Sixth Annual Allerton Conf. on Circuit and System Theory, October 1968, pp. 219-227.

30. C. J. Weinstein and A. V. Oppenheim, "A comparison of Roundoff Noise in Floating Point and Fixed Point Digital Filter Realizations," Proc. IEEE (letters), Vol. 57, 1969, pp. 1181-1183.

APPENDIX A

The following pages are computer listings of the results of searches over all orderings of different filters. Each ordering is identified by a sequence of numbers, where each number identifies a section of the filter. The left-most number in each sequence corresponds to the input section. A filter section is identified by the same number as that which labels the zeros it synthesizes in the plot of its zeros (see figures). All noise variances listed are in units of Q^2, where Q is the quantization step size.

App. A.1 - List of Orderings and Noise Variances of Filter of Fig. 4.6

ORDER	NOISE	ORDER	NOISE	ORDER	NOISE
263451	1.0983	145263	1.1104	145362	1.1131
163452	1.1382	245163	1.1601	245361	1.1605
362451	1.1834	246351	1.2305	162453	1.2456
361452	1.2561	261453	1.2783	143652	1.2841
146352	1.3245	415263	1.3298	415362	1.3325
243651	1.3356	345261	1.3546	345162	1.3568
246153	1.3652	346251	1.3660	341652	1.3666
425163	1.3687	425361	1.3692	146253	1.3763
163425	1.3797	342651	1.4009	263415	1.4151
142653	1.4160	241653	1.4332	426351	1.4392
346152	1.4489	162435	1.4582	261435	1.4909
361425	1.4976	143562	1.4977	362415	1.5002
413652	1.5034	435261	1.5227	435162	1.5249
143625	1.5256	436251	1.5341	431652	1.5347
416352	1.5439	423651	1.5442	264351	1.5470
246315	1.5474	142563	1.5642	146325	1.5660
432651	1.5690	426153	1.5739	246135	1.5778
341562	1.5803	241563	1.5814	146235	1.5889
416253	1.5957	341625	1.6081	436152	1.6170
142635	1.6286	412653	1.6354	421653	1.6418
241635	1.6458	243615	1.6524	243561	1.6539
164352	1.6583	264153	1.6817	346215	1.6829
346125	1.6904	364251	1.7040	164253	1.7101
413562	1.7171	342615	1.7177	342561	1.7192
413625	1.7449	431562	1.7483	426315	1.7560
431625	1.7762	412563	1.7835	416325	1.7854
426135	1.7865	364152	1.7869	421563	1.7900
416235	1.8083	412635	1.8480	436215	1.8509
421635	1.8544	436125	1.8585	423615	1.8610
423561	1.8626	264315	1.8638	432615	1.8858
432561	1.8873	264135	1.8943	164325	1.8998
164235	1.9227	364215	2.0208	364125	2.0284
136452	2.0700	316452	2.1489	236451	2.2117
623451	2.2534	632451	2.2904	613452	2.3028
136425	2.3115	631452	2.3632	316425	2.3904
326451	2.3999	245631	2.4004	612453	2.4102
621453	2.4335	245613	2.4699	134652	2.4854
236415	2.5285	613425	2.5443	314652	2.5643
623415	2.5703	631425	2.6047	632415	2.6073
425631	2.6090	612435	2.6228	621435	2.6461
425613	2.6785	145632	2.6801	134562	2.6991
145623	2.6993	624351	2.7021	326415	2.7167
134625	2.7269	126453	2.7639	314562	2.7779
234651	2.7927	314625	2.8058	634251	2.8110
614352	2.8229	624153	2.8368	614253	2.8747
634152	2.8939	415632	2.8994	216453	2.9085
415623	2.9187	126435	2.9765	324651	2.9809
624315	3.0190	624135	3.0495	614325	3.0644
614235	3.0873	234615	3.1095	234561	3.1110
216435	3.1211	634215	3.1278	634125	3.1354
124653	3.2439	324615	3.2977	324561	3.2992
214653	3.3885	124563	3.3920	124635	3.4565
246531	3.4977	214563	3.5367	246513	3.5672
214635	3.6011	426531	3.7063	426513	3.7758
264531	3.8141	264513	3.8836	163245	4.0265
146532	4.0376	146523	4.0569	162345	4.0697
261345	4.1025	361245	4.1445	345621	4.2234
416532	4.2570	345612	4.2737	416523	4.2763
263145	4.2914	164532	4.3715	362145	4.3766
164523	4.3907	435621	4.3915	435612	4.4417

451263	4.5778	451362	4.5806	452163	4.6348
452361	4.6353	453261	4.7361	453162	4.7383
136245	4.9584	624531	4.9693	145236	4.9857
245136	5.0354	316245	5.0372	624513	5.0388
613245	5.1911	415236	5.2051	612345	5.2343
425136	5.2440	631245	5.2515	621345	5.2577
236145	5.4048	145326	5.4322	142536	5.4395
623145	5.4466	241536	5.4567	632145	5.4836
614532	5.5361	614523	5.5553	126345	5.5880
326145	5.5930	415326	5.6515	412536	5.6589
421536	5.6653	345126	5.6759	216345	5.7326
143526	5.8168	245316	5.8434	435126	5.8440
452631	5.8751	341526	5.8993	452613	5.9446
413526	6.0362	345216	6.0375	346521	6.0426
425316	6.0520	431526	6.0674	346512	6.0928
451632	6.1475	451623	6.1668	435216	6.2055
436521	6.2106	436512	6.2609	243516	6.3368
364521	6.3805	342516	6.4021	364512	6.4307
423516	6.5454	432516	6.5701	134526	7.0181
314526	7.0970	124536	7.2674	214536	7.4120
634521	7.4875	634512	7.5377	453621	7.6049
453612	7.6551	234516	7.7939	324516	7.9820
451236	8.4532	452136	8.5101	451326	8.8996
453126	9.0573	452316	9.3181	453216	9.4189
462351	11.8333	463251	11.8896	461352	11.9658
462153	11.9680	463152	11.9725	461253	12.0176
462315	12.1501	462135	12.1806	463215	12.2064
461325	12.2073	463125	12.2140	461235	12.2302
462531	14.1004	462513	14.1699	461532	14.6789
461523	14.6982	143265	16.3035	341265	16.3860
142365	16.4329	241365	16.4501	413265	16.5228
431265	16.5541	463521	16.5661	463512	16.6163
412365	16.6522	421365	16.6587	134265	17.5048
314265	17.5837	243165	17.7595	342165	17.8249
423165	17.9682	432165	17.9929	124365	18.2607
214365	18.4053	234165	19.2166	324165	19.4048
642351	21.7670	643251	21.8232	641352	21.8995
642153	21.9017	643152	21.9061	641253	21.9513
642315	22.0838	642135	22.1143	643215	22.1401
641325	22.1410	643125	22.1476	641235	22.1639
143256	23.0109	341256	23.0934	142356	23.1403
241356	23.1575	413256	23.2302	431256	23.2615
412356	23.3596	421356	23.3661	642531	24.0341
642513	24.1036	134256	24.2122	314256	24.2911
243156	24.4670	342156	24.5323	641532	24.6126
641523	24.6319	423156	24.6756	432156	24.7003
124356	24.9681	214356	25.1128	234156	25.9240
324156	26.1122	643521	26.4998	643512	26.5500
132645	81.4953	312645	81.5742	231645	85.2733
321645	85.4615	123645	87.0142	213645	87.1589
456231	99.7815	456213	99.8510	456132	100.2220
456123	100.2410	456321	101.7430	456312	101.7940
132465	112.8460	312465	112.9250	231465	116.6240
321465	116.8120	123465	118.3650	213465	118.5090
132456	119.5530	312456	119.6320	231456	123.3310
321456	123.5190	123456	125.0720	213456	125.2170
465231	180.9150	465213	180.9850	465132	181.3550
465123	181.3750	465321	182.8770	465312	182.9270
645231	190.8490	645213	190.9180	645132	191.2890
645123	191.3080	645321	192.8110	645312	192.8610

App. A.2 - List of Orderings and Noise Variances of Filter of Fig. 4.12

ORDER	NOISE	ORDER	NOISE	ORDER	NOISE
135462	0.6493	315462	0.6545	125463	0.6576
215463	0.6592	251463	0.6595	351462	0.6635
254163	0.6670	145362	0.6688	354162	0.6699
145263	0.6704	235461	0.6741	152463	0.6763
153462	0.6773	325461	0.6777	415362	0.6785
253461	0.6789	245163	0.6795	415263	0.6800
135264	0.6818	352461	0.6819	154362	0.6821
254361	0.6832	154263	0.6837	315264	0.6870
425163	0.6876	354612	0.6876	354261	0.6876
125364	0.6886	235164	0.6889	215364	0.6902
251364	0.6904	354621	0.6908	325164	0.6925
253164	0.6937	245361	0.6957	351264	0.6960
352164	0.6967	425361	0.7037	345162	0.7056
451362	0.7060	152364	0.7072	451263	0.7076
453162	0.7076	452163	0.7077	254613	0.7082
254631	0.7090	153264	0.7098	435162	0.7100
235641	0.7154	325641	0.7190	253641	0.7202
235614	0.7207	245613	0.7207	245631	0.7215
352641	0.7232	345612	0.7233	345261	0.7233
452361	0.7239	325614	0.7243	453612	0.7253
453261	0.7253	253614	0.7255	135642	0.7260
345621	0.7265	435612	0.7277	435261	0.7278
453621	0.7285	352614	0.7285	425613	0.7288
425631	0.7295	435621	0.7309	315642	0.7312
135624	0.7344	315624	0.7396	145632	0.7399
351642	0.7402	145623	0.7423	125643	0.7424
215643	0.7439	251643	0.7442	125634	0.7484
351624	0.7486	452613	0.7489	415632	0.7495
452631	0.7497	215634	0.7500	251634	0.7502
415623	0.7520	154632	0.7532	153642	0.7540
154623	0.7556	152643	0.7610	153624	0.7624
152634	0.7670	451632	0.7771	451623	0.7795
134562	1.0838	314562	1.0890	143562	1.0960
413562	1.1057	256413	1.1153	256431	1.1161
256341	1.1264	341562	1.1304	256314	1.1317
431562	1.1349	256143	1.1443	256134	1.1504
521463	1.1698	531462	1.1704	534162	1.1769
524163	1.1774	156432	1.1806	512463	1.1821
156423	1.1830	513462	1.1831	156342	1.1843
514362	1.1879	532461	1.1889	523461	1.1892
514263	1.1895	156243	1.1915	156324	1.1928
524361	1.1935	534612	1.1945	534261	1.1946
124563	1.1973	541362	1.1974	156234	1.1975
534621	1.1977	214563	1.1989	142563	1.1989
541263	1.1990	543162	1.1990	542163	1.1991
521364	1.2007	531264	1.2029	532164	1.2036
523164	1.2040	241563	1.2078	412563	1.2086
512364	1.2130	542361	1.2153	513264	1.2156
421563	1.2159	543612	1.2167	543261	1.2167
524613	1.2186	524631	1.2193	543621	1.2199
234561	1.2232	324561	1.2267	532641	1.2302
523641	1.2305	532614	1.2354	523614	1.2358
243561	1.2383	542613	1.2403	542631	1.2411
423561	1.2463	531642	1.2471	521643	1.2545
531624	1.2556	514632	1.2590	513642	1.2598
521634	1.2605	514623	1.2614	342561	1.2638
512643	1.2668	513624	1.2682	432561	1.2683
541632	1.2685	541623	1.2709	512634	1.2728
356412	1.3532	356421	1.3564	356241	1.3818
356214	1.3871	356142	1.3926	356124	1.4010

526413	1.6257	526431	1.6264	526341	1.6368
526314	1.6420	526143	1.6547	526134	1.6607
132564	1.6755	312564	1.6807	231564	1.6816
321564	1.6851	123564	1.6862	516432	1.6863
213564	1.6878	516423	1.6888	516342	1.6901
516243	1.6973	516324	1.6986	516234	1.7033
135246	1.7039	135426	1.7069	125436	1.7071
215436	1.7087	251436	1.7089	315246	1.7091
125346	1.7107	235146	1.7110	315426	1.7121
215346	1.7123	251346	1.7125	325146	1.7146
253146	1.7158	254136	1.7165	351246	1.7181
352146	1.7188	145236	1.7199	351426	1.7211
152436	1.7258	145326	1.7264	235416	1.7269
354126	1.7275	245136	1.7290	152346	1.7293
415236	1.7295	325416	1.7305	253416	1.7317
153246	1.7319	154236	1.7332	352416	1.7347
153426	1.7349	254316	1.7360	415326	1.7361
425136	1.7371	154326	1.7398	354216	1.7405
245316	1.7485	425316	1.7566	451236	1.7571
452136	1.7572	345126	1.7632	451326	1.7636
453126	1.7652	435126	1.7676	345216	1.7761
452316	1.7767	453216	1.7781	435216	1.7806
536412	1.8602	536421	1.8634	456312	1.8735
456321	1.8767	536241	1.8888	456213	1.8910
456231	1.8918	536214	1.8940	536142	1.8995
536124	1.9080	456132	1.9091	456123	1.9115
134526	2.1414	314526	2.1466	143526	2.1536
413526	2.1633	341526	2.1880	431526	2.1925
521436	2.2193	521346	2.2229	531246	2.2250
532146	2.2258	523146	2.2261	524136	2.2269
531426	2.2280	512436	2.2316	534126	2.2345
512346	2.2351	513246	2.2377	514236	2.2390
513426	2.2407	532416	2.2417	523416	2.2421
514326	2.2455	524316	2.2464	124536	2.2468
534216	2.2474	214536	2.2484	142536	2.2484
541236	2.2484	542136	2.2486	541326	2.2550
543126	2.2566	241536	2.2573	412536	2.2580
421536	2.2653	542316	2.2681	543216	2.2695
234516	2.2760	324516	2.2796	243516	2.2911
423516	2.2992	342516	2.3166	432516	2.3211
546312	2.3649	546321	2.3681	546213	2.3824
546231	2.3832	546132	2.4005	546123	2.4029
132546	2.6977	312546	2.7028	231546	2.7037
321546	2.7073	123546	2.7083	213546	2.7099
561432	12.2206	561423	12.2230	561342	12.2243
561243	12.2315	561324	12.2328	561234	12.2375
562413	12.3073	562431	12.3081	562341	12.3184
562314	12.3237	562143	12.3363	562134	12.3424
563412	12.5146	563421	12.5178	563241	12.5432
563214	12.5485	563142	12.5540	563124	12.5624
564312	12.9438	564321	12.9470	564213	12.9614
564231	12.9621	564132	12.9794	564123	12.9819
134256	13.8218	314256	13.8270	143256	13.8341
413256	13.8437	124356	13.8653	214356	13.8669
142356	13.8669	341256	13.8685	431256	13.8729
241356	13.8758	412356	13.8766	421356	13.8839
234156	15.2546	324156	15.2581	243156	15.2697
423156	15.2777	342156	15.2952	432156	15.2997
132456	15.8294	312456	15.8346	231456	15.8354
321456	15.8390	123456	15.8400	213456	15.8416

App. A.3 - List of Orderings, Noise Variances, and Parameter Pk for Filter no. 14

ORDER	NOISE	PEAK	ORDER	NOISE	PEAK
153462	0.9940	1.9732	154362	0.9949	1.9732
253461	0.9958	1.9732	254361	1.0031	1.9732
254163	1.0290	1.9732	251463	1.0558	1.9732
154263	1.0605	1.9732	154632	1.0702	3.8934
154623	1.0832	3.8934	152463	1.0912	1.9732
351462	1.1021	1.9732	253641	1.1040	3.8934
354162	1.1313	1.9732	235461	1.1328	3.8934
352461	1.1393	1.9732	325461	1.1474	3.8934
135462	1.1527	3.8934	253614	1.1556	3.8934
145362	1.1574	3.8934	315462	1.1647	3.8934
354261	1.1709	1.9732	153642	1.1958	3.8934
415362	1.1986	3.8934	254631	1.2114	3.8934
245361	1.2118	3.8934	145263	1.2229	3.8934
254613	1.2270	3.8934	145632	1.2327	3.8934
245163	1.2377	3.8934	235641	1.2411	3.8934
153624	1.2448	3.8934	145623	1.2456	3.8934
352641	1.2475	3.8934	425361	1.2555	3.8934
325641	1.2556	3.8934	415263	1.2642	3.8934
415632	1.2739	3.8934	425163	1.2815	3.8934
415623	1.2869	3.8934	235614	1.2926	3.8934
352614	1.2991	3.8934	351642	1.3039	3.8934
325614	1.3071	3.8934	253164	1.3348	1.9732
351624	1.3529	3.8934	135642	1.3545	3.8934
315642	1.3665	3.8934	251364	1.3688	1.9732
153264	1.3725	1.9732	251643	1.3807	3.8934
345162	1.3976	3.8934	135624	1.4035	3.8934
152364	1.4042	1.9732	315624	1.4155	3.8934
152643	1.4161	3.8934	251634	1.4167	3.8934
245631	1.4201	3.8934	435162	1.4268	3.8934
245613	1.4356	3.8934	345261	1.4372	3.8934
152634	1.4521	3.8934	425631	1.4639	3.8934
435261	1.4664	3.8934	235164	1.4718	3.8934
352164	1.4783	1.9732	425613	1.4794	3.8934
351264	1.4807	1.9732	325164	1.4863	3.8934
135264	1.5313	3.8934	315264	1.5433	3.8934
451362	1.6191	1.9732	453162	1.6473	1.9732
452361	1.6626	1.9732	451263	1.6846	1.9732
453261	1.6870	1.9732	452163	1.6886	1.9732
451632	1.6944	3.8934	451623	1.7073	3.8934
452631	1.8710	3.8934	452613	1.8865	3.8934
125463	2.3505	3.8934	215463	2.4160	3.8934
125364	2.6636	3.8934	125643	2.6754	3.8934
125634	2.7115	3.8934	215364	2.7291	3.8934
215643	2.7409	3.8934	256341	2.7615	7.6822
215634	2.7770	3.8934	256314	2.8131	7.6822
256431	2.8283	7.6822	256413	2.8438	7.6822
256143	2.9285	7.6822	256134	2.9645	7.6822
513462	3.1508	3.8934	514362	3.1518	3.8934
523461	3.1662	3.8934	524361	3.1734	3.8934
524163	3.1994	3.8934	531462	3.2149	3.8934
514263	3.2173	3.8934	521463	3.2262	3.8934
514632	3.2270	3.8934	514623	3.2400	3.8934
534162	3.2441	3.8934	512463	3.2481	3.8934
532461	3.2521	3.8934	523641	3.2744	3.8934
534261	3.2837	3.8934	523614	3.3259	3.8934
513642	3.3526	3.8934	532641	3.3604	3.8934
524631	3.3818	3.8934	524613	3.3973	3.8934
513624	3.4016	3.8934	532614	3.4119	3.8934
531642	3.4167	3.8934	156432	3.4641	7.6822
531624	3.4657	3.8934	156423	3.4771	7.6822

523164	3.5051	3.8934	513264	3.5294	3.8934
521364	3.5392	3.8934	156342	3.5485	7.6822
521643	3.5511	3.8934	512364	3.5611	3.8934
512643	3.5730	3.8934	521634	3.5871	3.8934
532164	3.5911	3.8934	531264	3.5935	3.8934
156324	3.5975	7.6822	512634	3.6090	3.8934
156243	3.6786	7.6822	156234	3.7146	7.6822
541362	3.7157	3.8934	543162	3.7440	3.8934
542361	3.7593	3.8934	541263	3.7813	3.8934
543261	3.7836	3.8934	542163	3.7852	3.8934
541632	3.7910	3.8934	541623	3.8040	3.8934
542631	3.9676	3.8934	542613	3.9831	3.8934
134562	4.0256	7.6822	314562	4.0376	7.6822
354621	4.0545	3.8934	354612	4.0570	3.8934
254136	4.1207	3.8934	251436	4.1475	3.8934
154236	4.1522	3.8934	152436	4.1829	3.8934
143562	4.2290	7.6822	234561	4.2579	7.6822
413562	4.2702	7.6822	324561	4.2724	7.6822
153426	4.2729	3.8934	154326	4.2739	3.8934
145236	4.3146	3.8934	345621	4.3207	3.8934
345612	4.3233	3.8934	245136	4.3294	3.8934
435621	4.3500	3.8934	435612	4.3525	3.8934
415236	4.3558	3.8934	425136	4.3732	3.8934
351426	4.3811	3.8934	354126	4.4103	3.8934
341562	4.4268	7.6822	135426	4.4317	3.8934
145326	4.4364	3.8934	315426	4.4437	3.8934
431562	4.4560	7.6822	415326	4.4776	3.8934
453621	4.5705	3.8934	453612	4.5730	3.8934
253416	4.6673	3.8934	254316	4.6746	3.8934
345126	4.6765	3.8934	435126	4.7058	3.8934
451236	4.7763	3.8934	452136	4.7802	3.8934
235416	4.8043	3.8934	352416	4.8108	3.8934
325416	4.8188	3.8934	354216	4.8424	3.8934
245316	4.8832	3.8934	451326	4.8981	3.8934
453126	4.9263	3.8934	425316	4.9270	3.8934
253146	4.9289	3.8934	526341	4.9319	7.6822
251346	4.9630	3.8934	153246	4.9667	3.8934
526314	4.9834	7.6822	152346	4.9984	3.8934
526431	4.9986	7.6822	526413	5.0142	7.6822
243561	5.0437	7.6822	235146	5.0659	3.8934
352146	5.0724	3.8934	351246	5.0748	3.8934
325146	5.0805	3.8934	423561	5.0874	7.6822
526143	5.0988	7.6822	345216	5.1087	3.8934
135246	5.1254	3.8934	526134	5.1348	7.6822
315246	5.1374	3.8934	435216	5.1379	3.8934
342561	5.3292	7.6822	452316	5.3341	3.8934
432561	5.3584	7.6822	453216	5.3584	3.8934
125436	5.4422	3.8934	215436	5.5077	3.8934
516432	5.6210	7.6822	516423	5.6340	7.6822
516342	5.7053	7.6822	516324	5.7543	7.6822
516243	5.8354	7.6822	516234	5.8715	7.6822
534621	6.1673	3.8934	534612	6.1698	3.8934
125346	6.2577	3.8934	524136	6.2911	3.8934
514236	6.3090	3.8934	521436	6.3178	3.8934
215346	6.3232	3.8934	512436	6.3397	3.8934
513426	6.4298	3.8934	514326	6.4308	3.8934
531426	6.4939	3.8934	534126	6.5231	3.8934
543621	6.6671	3.8934	543612	6.6697	3.8934
523416	6.8377	3.8934	524316	6.8449	3.8934
541236	6.8729	3.8934	542136	6.8769	3.8934
532416	6.9236	3.8934	534216	6.9552	3.8934

541326	6.9947	3.8934	543126	7.0230	3.8934
523146	7.0993	3.8934	513246	7.1235	3.8934
521346	7.1333	3.8934	512346	7.1552	3.8934
532146	7.1852	3.8934	531246	7.1876	3.8934
134526	7.3046	7.6822	314526	7.3166	7.6822
542316	7.4307	3.8934	543216	7.4551	3.8934
143526	7.5080	7.6822	413526	7.5492	7.6822
142563	7.6270	7.6822	412563	7.6682	7.6822
341526	7.7057	7.6822	431526	7.7350	7.6822
234516	7.9294	7.6822	324516	7.9439	7.6822
241563	8.1837	7.6822	421563	8.2275	7.6822
243516	8.7151	7.6822	423516	8.7589	7.6822
342516	9.0006	7.6822	432516	9.0299	7.6822
124563	10.1927	7.6822	214563	10.2582	7.6822
142536	10.7186	7.6822	412536	10.7599	7.6822
356241	11.2025	7.6822	356142	11.2393	7.6822
356214	11.2540	7.6822	241536	11.2754	7.6822
356124	11.2883	7.6822	421536	11.3192	7.6822
124536	13.2844	7.6822	536241	13.3153	7.6822
214536	13.3498	7.6822	536142	13.3521	7.6822
536214	13.3668	7.6822	536124	13.4011	7.6822
356421	13.8100	7.6822	356412	13.8125	7.6822
231564	15.7456	7.6822	321564	15.7602	7.6822
132564	15.7646	7.6822	312564	15.7766	7.6822
536421	15.9228	7.6822	536412	15.9254	7.6822
231546	19.3398	7.6822	321546	19.3543	7.6822
132546	19.3587	7.6822	312546	19.3707	7.6822
123564	21.8119	7.6822	213564	21.8774	7.6822
123546	25.4061	7.6822	213546	25.4716	7.6822
456132	32.3851	7.6822	456123	32.3981	7.6822
456231	32.4994	7.6822	456213	32.5149	7.6822
546132	34.4817	7.6822	546123	34.4947	7.6822
546231	34.5960	7.6822	546213	34.6115	7.6822
456321	35.0402	7.6822	456312	35.0427	7.6822
546321	37.1368	7.6822	546312	37.1393	7.6822
134256	109.2160	15.1583	314256	109.2280	15.1583
143256	109.4200	15.1583	413256	109.4610	15.1583
341256	109.6170	15.1583	431256	109.6470	15.1583
562341	113.5220	15.1583	562314	113.5730	15.1583
562431	113.5880	15.1583	562413	113.6040	15.1583
562143	113.6890	15.1583	562134	113.7250	15.1583
561432	114.1440	15.1583	561423	114.1570	15.1583
561342	114.2280	15.1583	561324	114.2770	15.1583
561243	114.3590	15.1583	561234	114.3950	15.1583
234156	119.7630	15.1583	324156	119.7770	15.1583
243156	120.5480	15.1583	423156	120.5920	15.1583
342156	120.8340	15.1583	432156	120.8630	15.1583
563241	121.2620	15.1583	563142	121.2990	15.1583
563214	121.3140	15.1583	563124	121.3480	15.1583
563421	123.8700	15.1583	563412	123.8720	15.1583
142356	130.6920	15.1583	412356	130.7330	15.1583
241356	131.2490	15.1583	421356	131.2920	15.1583
124356	133.2570	15.1583	214356	133.3230	15.1583
564132	140.4500	15.1583	564123	140.4630	15.1583
564231	140.5650	15.1583	564213	140.5800	15.1583
564321	143.1050	15.1583	564312	143.1080	15.1583
231456	179.8560	15.1583	321456	179.8710	15.1583
132456	179.8750	15.1583	312456	179.8870	15.1583
123456	185.9220	15.1583	213456	185.9880	15.1583

App. A.4 - Comparison of Orderings and their Reverse for Filter no. 14

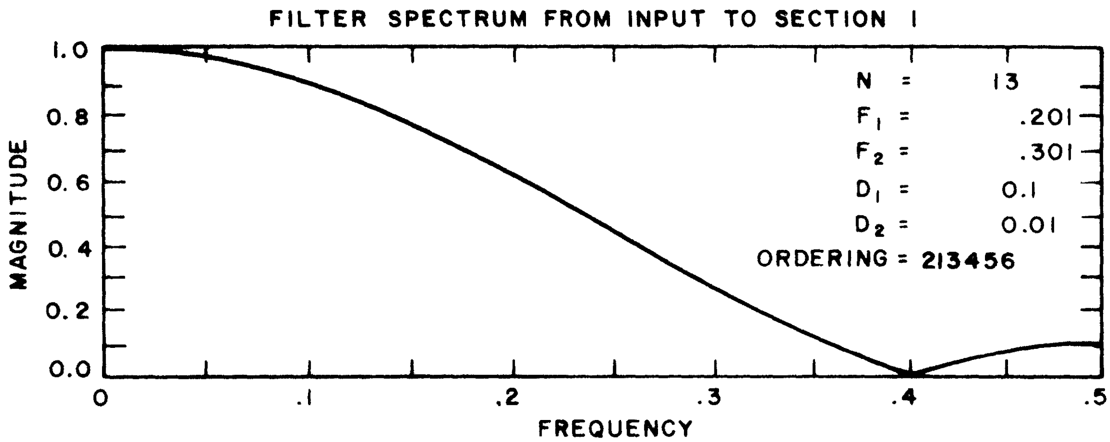
ORDER	NOISE	PEAK	ORDER	NOISE	PEAK
153462	0.9940	1.9732	254361	1.0031	1.9732
154362	0.9949	1.9732	253461	0.9958	1.9732
254163	1.0290	1.9732	351462	1.1021	1.9732
251463	1.0558	1.9732	354162	1.1313	1.9732
154263	1.0605	1.9732	352461	1.1393	1.9732
154632	1.0702	3.8934	235461	1.1328	3.8934
154623	1.0832	3.8934	325461	1.1474	3.8934
152463	1.0912	1.9732	354261	1.1709	1.9732
253641	1.1040	3.8934	145362	1.1574	3.8934
135462	1.1527	3.8934	254631	1.2114	3.8934
253614	1.1556	3.8934	415362	1.1986	3.8934
315462	1.1647	3.8934	254613	1.2270	3.8934
153642	1.1958	3.8934	245361	1.2118	3.8934
145263	1.2229	3.8934	352641	1.2475	3.8934
145632	1.2327	3.8934	235641	1.2411	3.8934
245163	1.2377	3.8934	351642	1.3039	3.8934
153624	1.2448	3.8934	425361	1.2555	3.8934
145623	1.2456	3.8934	325641	1.2556	3.8934
415263	1.2642	3.8934	352614	1.2991	3.8934
415632	1.2739	3.8934	235614	1.2926	3.8934
425163	1.2815	3.8934	351624	1.3529	3.8934
415623	1.2859	3.8934	325614	1.3071	3.8934
253164	1.3348	1.9732	451362	1.6191	1.9732
135642	1.3545	3.8934	245631	1.4201	3.8934
315642	1.3665	3.8934	245613	1.4356	3.8934
251364	1.3688	1.9732	453162	1.6473	1.9732
153264	1.3725	1.9732	452361	1.6626	1.9732
251643	1.3807	3.8934	345162	1.3976	3.8934
135624	1.4035	3.8934	425631	1.4639	3.8934
152364	1.4042	1.9732	453261	1.6870	1.9732
315624	1.4155	3.8934	425613	1.4794	3.8934
152643	1.4161	3.8934	345261	1.4372	3.8934
251634	1.4167	3.8934	435162	1.4268	3.8934
152634	1.4521	3.8934	435261	1.4664	3.8934
235164	1.4718	3.8934	451632	1.6944	3.8934
352164	1.4783	1.9732	451263	1.6846	1.9732
351264	1.4807	1.9732	452163	1.6886	1.9732
325164	1.4863	3.8934	451623	1.7073	3.8934
135264	1.5313	3.8934	452631	1.8710	3.8934
315264	1.5433	3.8934	452613	1.8865	3.8934
125463	2.3505	3.8934	354621	4.0545	3.8934
215463	2.4160	3.8934	354612	4.0570	3.8934
125364	2.6636	3.8934	453621	4.5705	3.8934
125643	2.6754	3.8934	345621	4.3207	3.8934
125634	2.7115	3.8934	435621	4.3500	3.8934
215364	2.7291	3.8934	453612	4.5730	3.8934
215643	2.7409	3.8934	345612	4.3233	3.8934
256341	2.7615	7.6822	143562	4.2290	7.6822
215634	2.7770	3.8934	435612	4.3525	3.8934
256314	2.8131	7.6822	413562	4.2702	7.6822
256431	2.8283	7.6822	134562	4.0256	7.6822
256413	2.8438	7.6822	314562	4.0376	7.6822
256143	2.9285	7.6822	341562	4.4268	7.6822
256134	2.9645	7.6822	431562	4.4560	7.6822
513462	3.1508	3.8934	254316	4.6746	3.8934
514362	3.1518	3.8934	253416	4.6673	3.8934
523461	3.1652	3.8934	154326	4.2739	3.8934
524361	3.1734	3.8934	153426	4.2729	3.8934
524163	3.1994	3.8934	351426	4.3811	3.8934
531462	3.2149	3.8934	254136	4.1207	3.8934

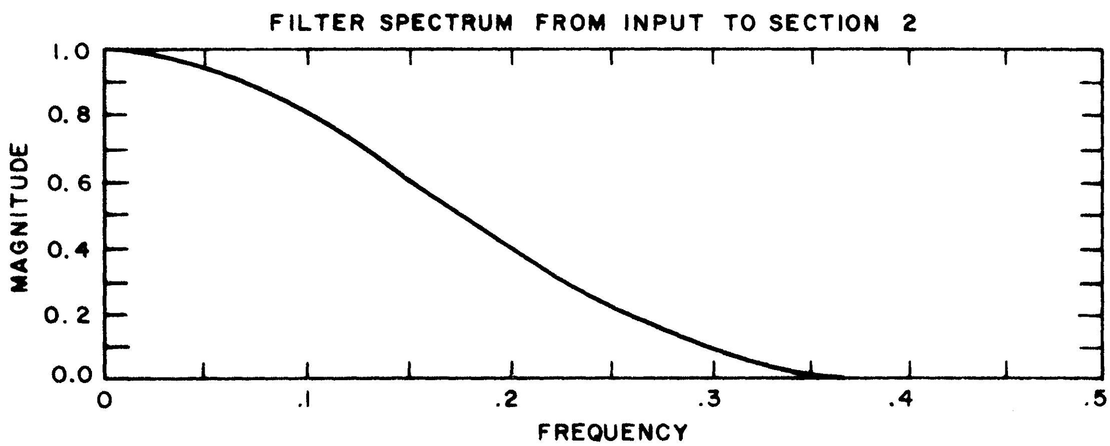
514263	3.2173	3.8934	352416	4.8108	3.8934
521463	3.2262	3.8934	354126	4.4103	3.8934
514632	3.2270	3.8934	235416	4.8043	3.8934
514623	3.2400	3.8934	325416	4.8188	3.8934
534162	3.2441	3.8934	251436	4.1475	3.8934
512463	3.2481	3.8934	354216	4.8424	3.8934
532461	3.2521	3.8934	154236	4.1522	3.8934
523641	3.2744	3.8934	145326	4.4364	3.8934
534261	3.2837	3.8934	152436	4.1829	3.8934
523614	3.3259	3.8934	415326	4.4776	3.8934
513642	3.3526	3.8934	245316	4.8832	3.8934
532641	3.3604	3.8934	145236	4.3146	3.8934
524631	3.3818	3.8934	135426	4.4317	3.8934
524613	3.3973	3.8934	315426	4.4437	3.8934
513624	3.4016	3.8934	425316	4.9270	3.8934
532614	3.4119	3.8934	415236	4.3558	3.8934
531642	3.4167	3.8934	245136	4.3294	3.8934
156432	3.4641	7.6822	234561	4.2579	7.6822
531624	3.4657	3.8934	425136	4.3732	3.8934
156423	3.4771	7.6822	324561	4.2724	7.6822
523164	3.5051	3.8934	451326	4.8981	3.8934
513264	3.5294	3.8934	452316	5.3341	3.8934
521364	3.5392	3.8934	453126	4.9263	3.8934
156342	3.5485	7.6822	243561	5.0437	7.6822
521643	3.5511	3.8934	345126	4.6765	3.8934
512364	3.5611	3.8934	453216	5.3584	3.8934
512643	3.5730	3.8934	345216	5.1087	3.8934
521634	3.5871	3.8934	435126	4.7058	3.8934
532164	3.5911	3.8934	451236	4.7763	3.8934
531264	3.5935	3.8934	452136	4.7802	3.8934
156324	3.5975	7.6822	423561	5.0874	7.6822
512634	3.6090	3.8934	435216	5.1379	3.8934
156243	3.6786	7.6822	342561	5.3292	7.6822
156234	3.7146	7.6822	432561	5.3584	7.6822
541362	3.7157	3.8934	253146	4.9289	3.8934
543162	3.7440	3.8934	251346	4.9630	3.8934
542361	3.7593	3.8934	153246	4.9667	3.8934
541263	3.7813	3.8934	352146	5.0724	3.8934
543261	3.7836	3.8934	152346	4.9984	3.8934
542163	3.7852	3.8934	351246	5.0748	3.8934
541632	3.7910	3.8934	235146	5.0659	3.8934
541623	3.8040	3.8934	325146	5.0805	3.8934
542631	3.9676	3.8934	135246	5.1254	3.8934
542613	3.9831	3.8934	315246	5.1374	3.8934
526341	4.9319	7.6822	143526	7.5080	7.6822
526314	4.9834	7.6822	413526	7.5492	7.6822
526431	4.9986	7.6822	134526	7.3046	7.6822
526413	5.0142	7.6822	314526	7.3166	7.6822
526143	5.0988	7.6822	341526	7.7057	7.6822
526134	5.1348	7.6822	431526	7.7350	7.6822
125436	5.4422	3.8934	534621	6.1673	3.8934
215436	5.5077	3.8934	534612	6.1698	3.8934
516432	5.6210	7.6822	234516	7.9294	7.6822
516423	5.6340	7.6822	324516	7.9439	7.6822
516342	5.7053	7.6822	243516	8.7151	7.6822
516324	5.7543	7.6822	423516	8.7589	7.6822
516243	5.8354	7.6822	342516	9.0006	7.6822
516234	5.8715	7.6822	432516	9.0299	7.6822
125346	6.2577	3.8934	543621	6.6671	3.8934
524136	6.2911	3.8934	531426	6.4939	3.8934
514236	6.3090	3.8934	532416	6.9236	3.8934

521436	6.3178	3.8934	534126	6.5231	3.8934
215346	6.3232	3.8934	543612	6.6697	3.8934
512436	6.3397	3.8934	534216	6.9552	3.8934
513426	6.4298	3.8934	524316	6.8449	3.8934
514326	6.4308	3.8934	523416	6.8377	3.8934
541236	6.8729	3.8934	532146	7.1852	3.8934
542136	6.8769	3.8934	531246	7.1876	3.8934
541326	6.9947	3.8934	523146	7.0993	3.8934
543126	7.0230	3.8934	521346	7.1333	3.8934
513246	7.1235	3.8934	542316	7.4307	3.8934
512346	7.1552	3.8934	543216	7.4551	3.8934
142563	7.6270	7.6822	356241	11.2025	7.6822
412563	7.6682	7.6822	356214	11.2540	7.6822
241563	8.1837	7.6822	356142	11.2393	7.6822
421563	8.2275	7.6822	356124	11.2883	7.6822
124563	10.1927	7.6822	356421	13.8100	7.6822
214563	10.2582	7.6822	356412	13.8125	7.6822
142536	10.7186	7.6822	536241	13.3153	7.6822
412536	10.7599	7.6822	536214	13.3668	7.6822
241536	11.2754	7.6822	536142	13.3521	7.6822
421536	11.3192	7.6822	536124	13.4011	7.6822
124536	13.2844	7.6822	536421	15.9228	7.6822
214536	13.3498	7.6822	536412	15.9254	7.6822
231564	15.7456	7.6822	456132	32.3851	7.6822
321564	15.7602	7.6822	456123	32.3981	7.6822
132564	15.7646	7.6822	456231	32.4994	7.6822
312564	15.7766	7.6822	456213	32.5149	7.6822
231546	19.3398	7.6822	546132	34.4817	7.6822
321546	19.3543	7.6822	546123	34.4947	7.6822
132546	19.3587	7.6822	546231	34.5960	7.6822
312546	19.3707	7.6822	546213	34.6115	7.6822
123564	21.8119	7.6822	456321	35.0402	7.6822
213564	21.8774	7.6822	456312	35.0427	7.6822
123546	25.4061	7.6822	546321	37.1368	7.6822
213546	25.4716	7.6822	546312	37.1393	7.6822
134256	109.2160	15.1583	562431	113.5880	15.1583
314256	109.2280	15.1583	562413	113.6040	15.1583
143256	109.4200	15.1583	562341	113.5220	15.1583
413256	109.4610	15.1583	562314	113.5730	15.1583
341256	109.6170	15.1583	562143	113.6890	15.1583
431256	109.6470	15.1583	562134	113.7250	15.1583
561432	114.1440	15.1583	234156	119.7630	15.1583
561423	114.1570	15.1583	324156	119.7770	15.1583
561342	114.2280	15.1583	243156	120.5480	15.1583
561324	114.2770	15.1583	423156	120.5920	15.1583
561243	114.3590	15.1583	342156	120.8340	15.1583
561234	114.3950	15.1583	432156	120.8630	15.1583
563241	121.2620	15.1583	142356	130.6920	15.1583
563142	121.2990	15.1583	241356	131.2490	15.1583
563214	121.3140	15.1583	412356	130.7330	15.1583
563124	121.3480	15.1583	421356	131.2920	15.1583
563421	123.8700	15.1583	124356	133.2570	15.1583
563412	123.8720	15.1583	214356	133.3230	15.1583
564132	140.4500	15.1583	231456	179.8560	15.1583
564123	140.4630	15.1583	321456	179.8710	15.1583
564231	140.5650	15.1583	132456	179.8750	15.1583
564213	140.5800	15.1583	312456	179.8870	15.1583
564321	143.1050	15.1583	123456	185.9220	15.1583
564312	143.1080	15.1583	213456	185.9880	15.1583

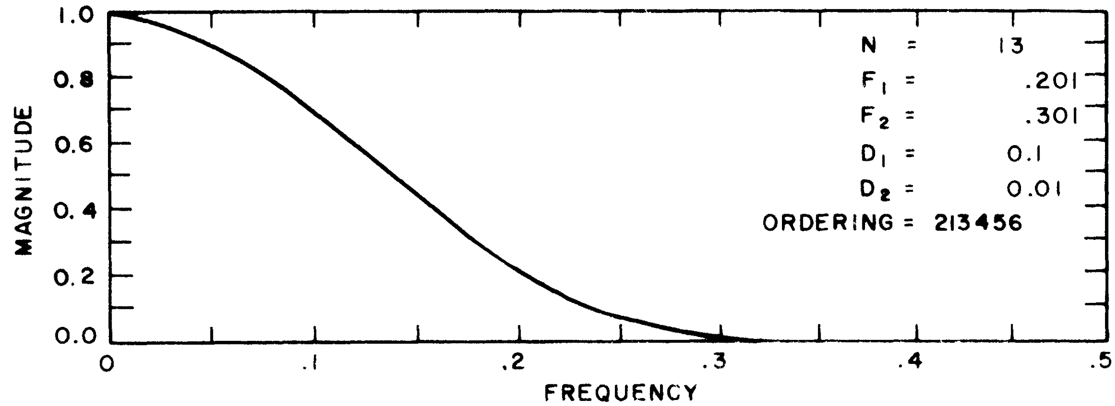
Appendix B.1

Plots of Subfilter Spectra for a High Noise Ordering of Filter no. 14

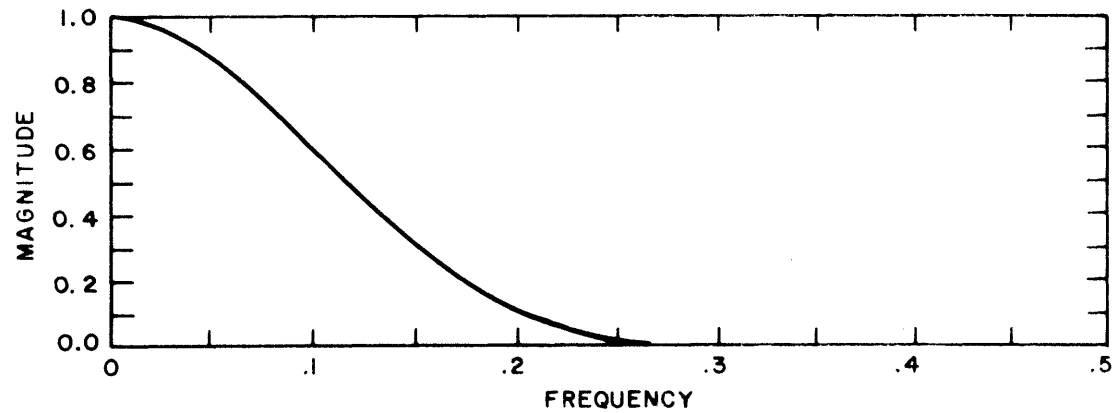

FILTER SPECTRUM FROM INPUT TO SECTION 1
MAGNITUDE
1.0
0.8
0.6
0.4
0.2
0.0
0
.1
.2
.3
.4
.5
FREQUENCY
N = 13
F1 = .201
F2 = .301
D1 = 0.1
D2 = 0.01
ORDERING = 213456

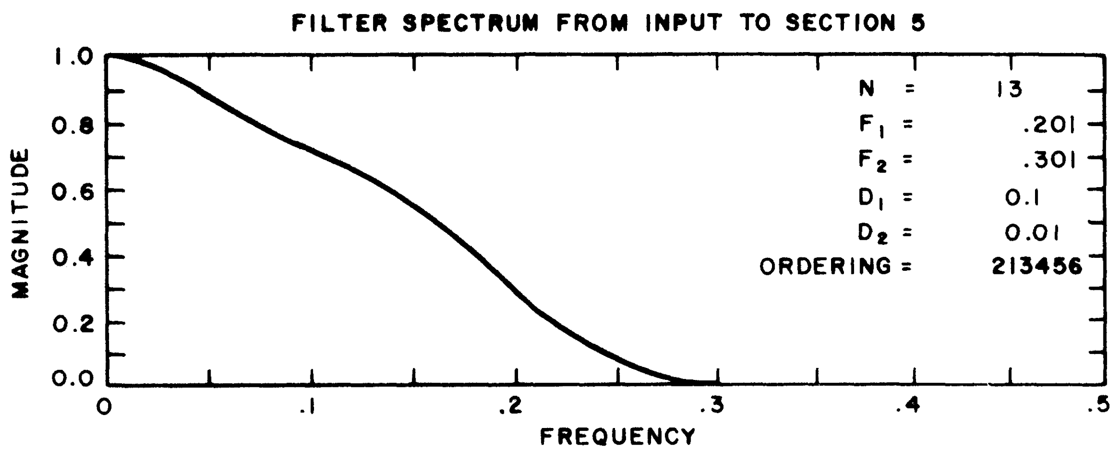

FILTER SPECTRUM FROM INPUT TO SECTION 2
MAGNITUDE
1.0
0.8
0.6
0.4
0.2
0.0
0
.1
.2
.3
.4
.5
FREQUENCY

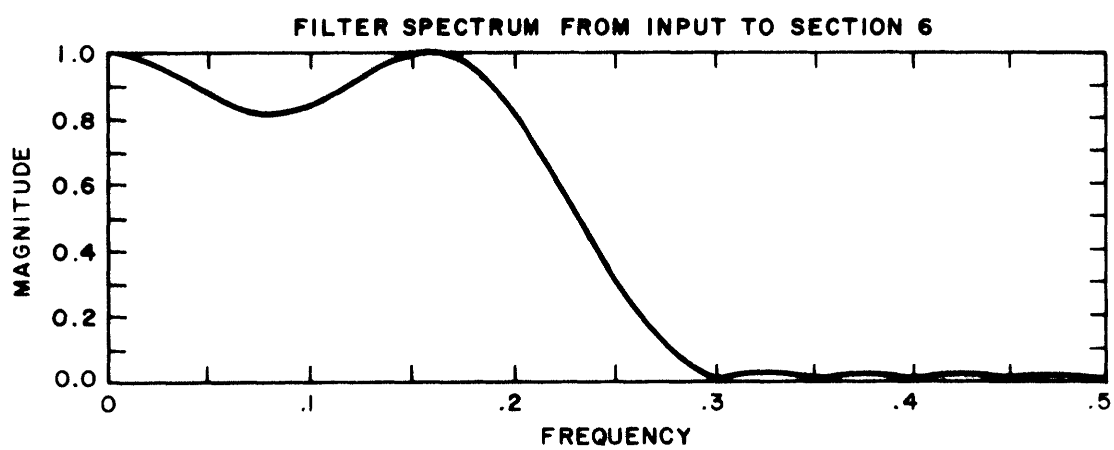
FILTER SPECTRUM FROM INPUT TO SECTION 3

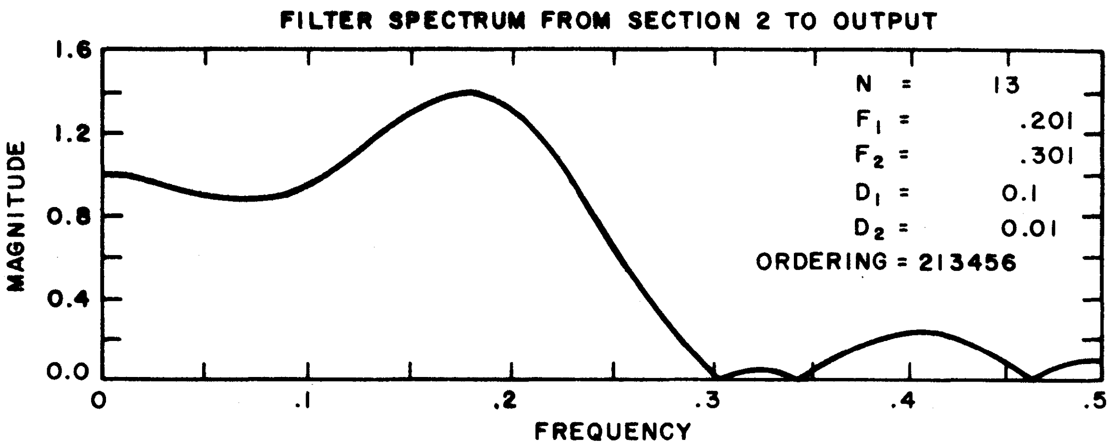


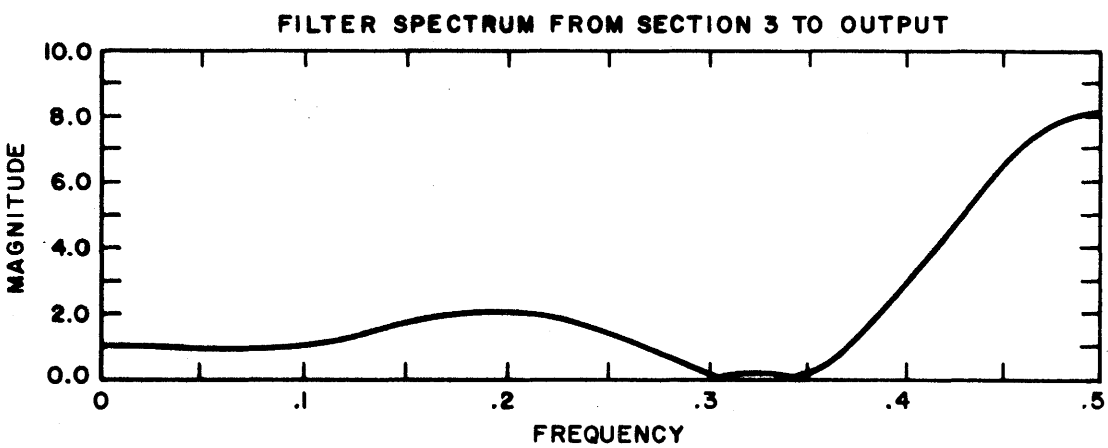
FILTER SPECTRUM FROM INPUT TO SECTION 4

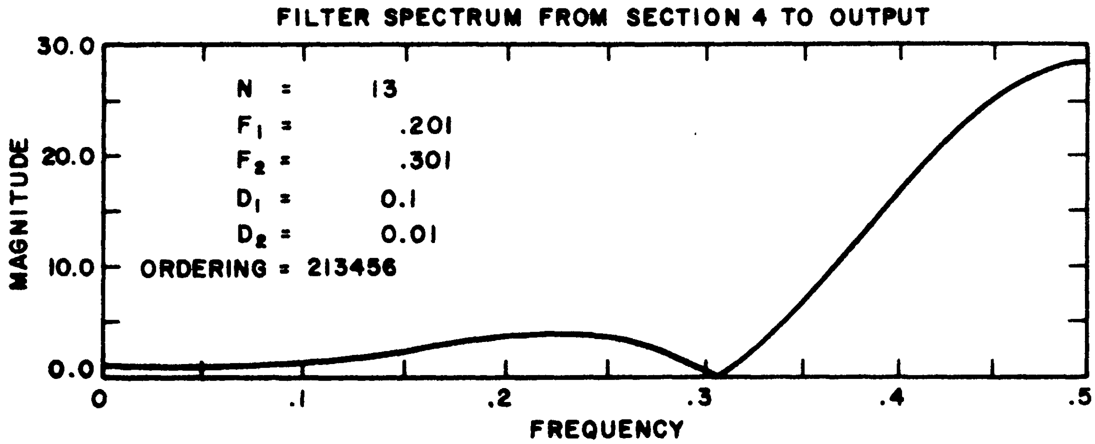



FILTER SPECTRUM FROM INPUT TO SECTION 5
MAGNITUDE
1.0
0.8
0.6
0.4
0.2
0.0
0
.1
.2
.3
.4
.5
FREQUENCY
N = 13
F1 = .201
F2 = .301
D1 = 0.1
D2 = 0.01
ORDERING = 213456

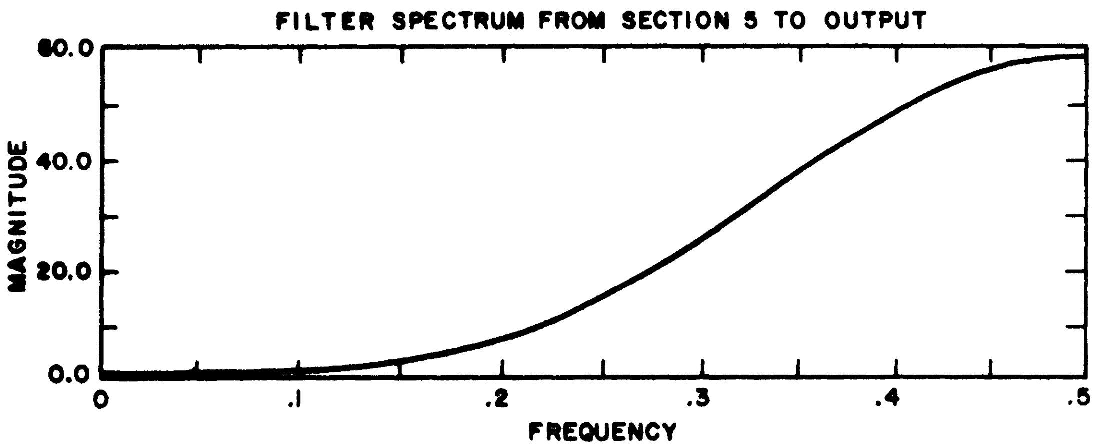

FILTER SPECTRUM FROM INPUT TO SECTION 6
MAGNITUDE
1.0
0.8
0.6
0.4
0.2
0.0
0
.1
.2
.3
.4
.5
FREQUENCY


FILTER SPECTRUM FROM SECTION 2 TO OUTPUT
MAGNITUDE
FREQUENCY
N = 13
F1 = .201
F2 = .301
D1 = 0.1
D2 = 0.01
ORDERING = 213456

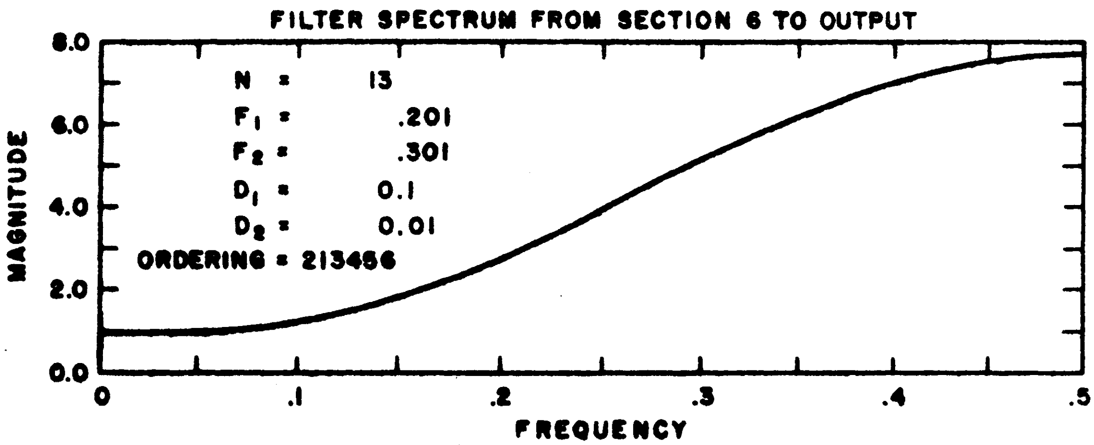

FILTER SPECTRUM FROM SECTION 3 TO OUTPUT
MAGNITUDE
FREQUENCY



FILTER SPECTRUM FROM SECTION 4 TO OUTPUT
30.0
20.0
10.0
0.0
MAGNITUDE
N = 13
F1 = .201
F2 = .301
D1 = 0.1
D2 = 0.01
ORDERING = 213456
0
.1
.2
.3
.4
.5
FREQUENCY



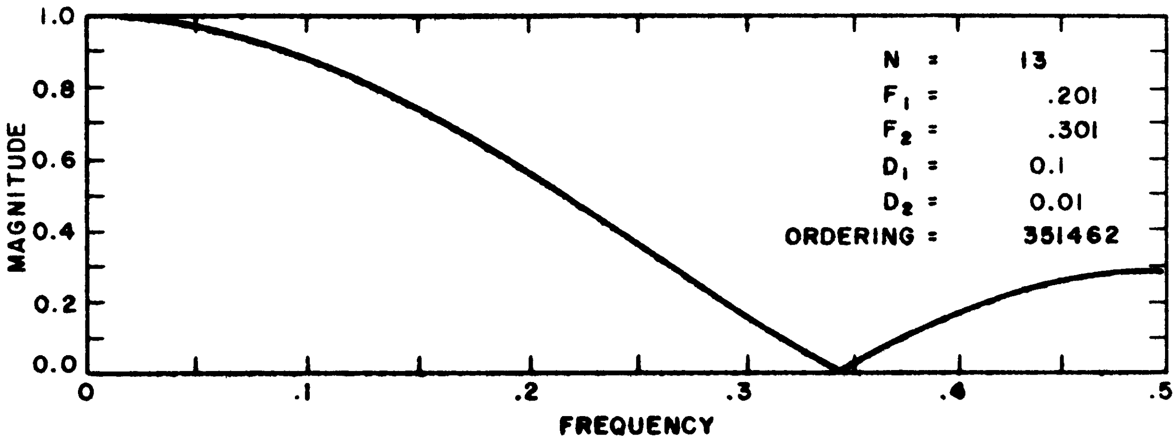
FILTER SPECTRUM FROM SECTION 5 TO OUTPUT
60.0
40.0
20.0
0.0
MAGNITUDE
0
.1
.2
.3
.4
.5
FREQUENCY


FILTER SPECTRUM FROM SECTION 6 TO OUTPUT
N = 13
F1 = .201
F2 = .301
D1 = 0.1
D2 = 0.01
ORDERING = 213456
MAGNITUDE
8.0
6.0
4.0
2.0
0.0
0
.1
.2
.3
.4
.5
FREQUENCY

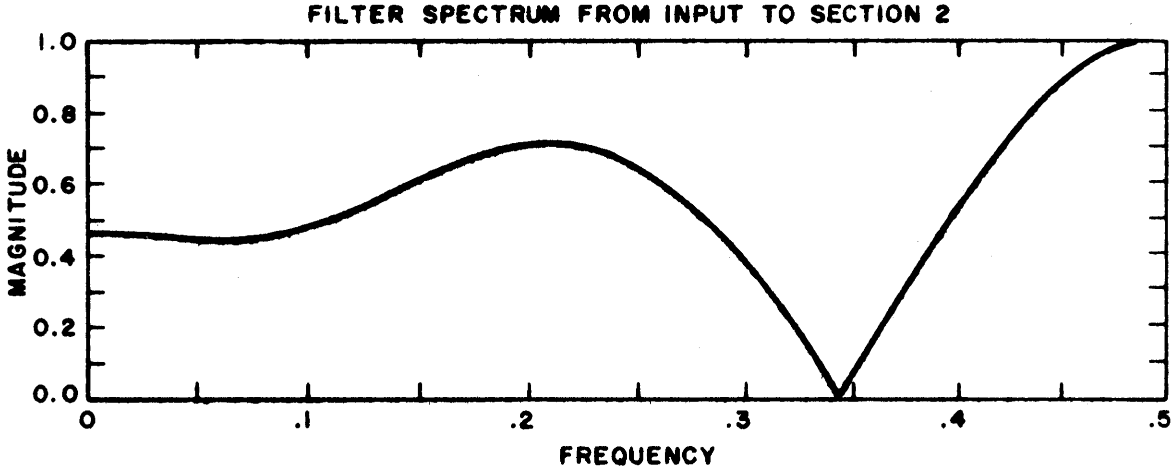
Appendix B.2

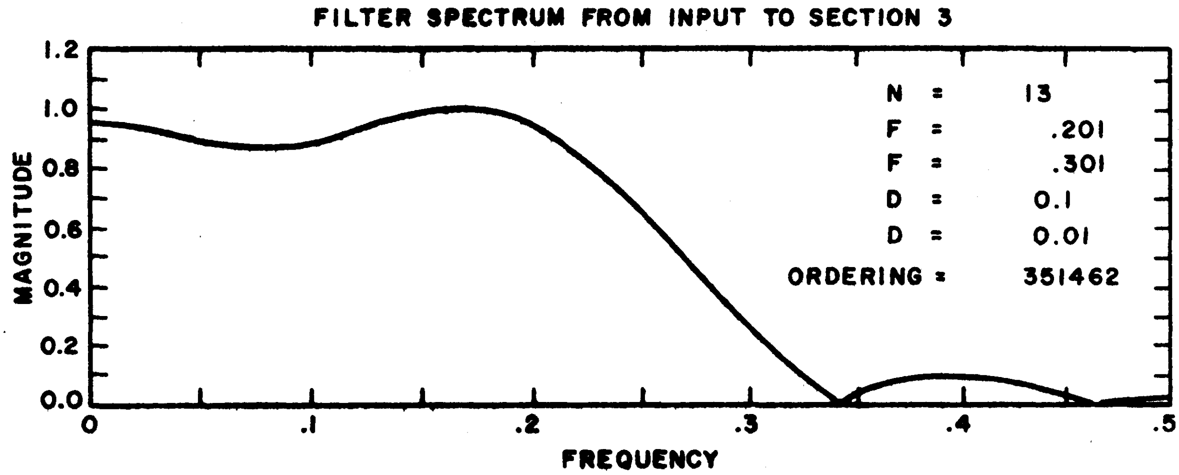
Plots of Subfilter Spectra for a Low Noise Ordering of Filter no. 14

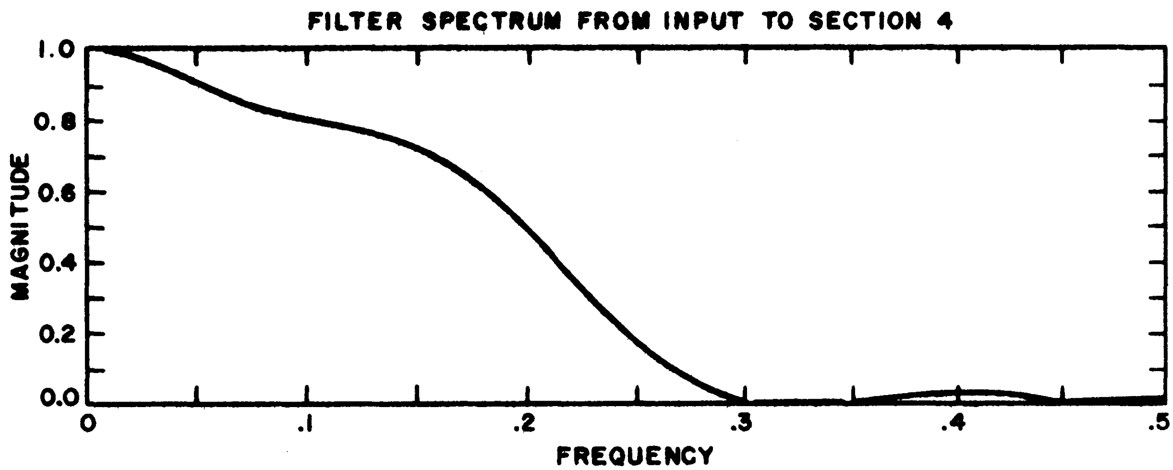
FILTER SPECTRUM FROM INPUT TO SECTION 1



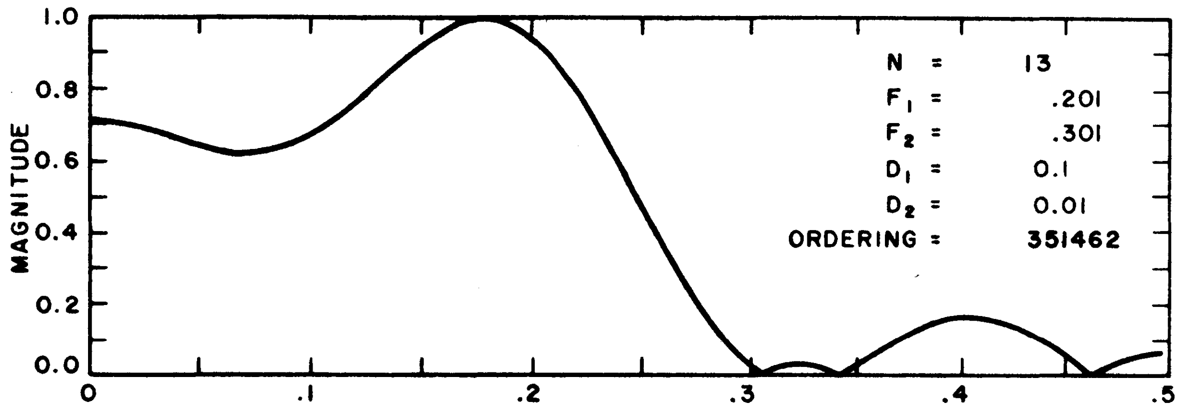
FILTER SPECTRUM FROM INPUT TO SECTION 2




FILTER SPECTRUM FROM INPUT TO SECTION 3
MAGNITUDE
1.2
1.0
0.8
0.6
0.4
0.2
0.0
N = 13
F = .201
F = .301
D = 0.1
D = 0.01
ORDERING = 351462
0
.1
.2
.3
.4
.5
FREQUENCY

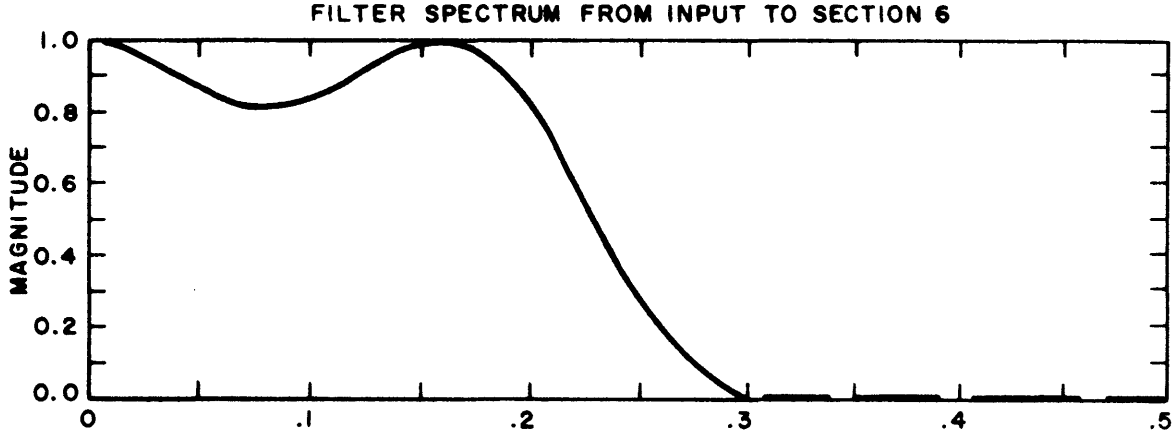

FILTER SPECTRUM FROM INPUT TO SECTION 4
MAGNITUDE
1.0
0.8
0.6
0.4
0.2
0.0
0
.1
.2
.3
.4
.5
FREQUENCY

FILTER SPECTRUM FROM INPUT TO SECTION 5



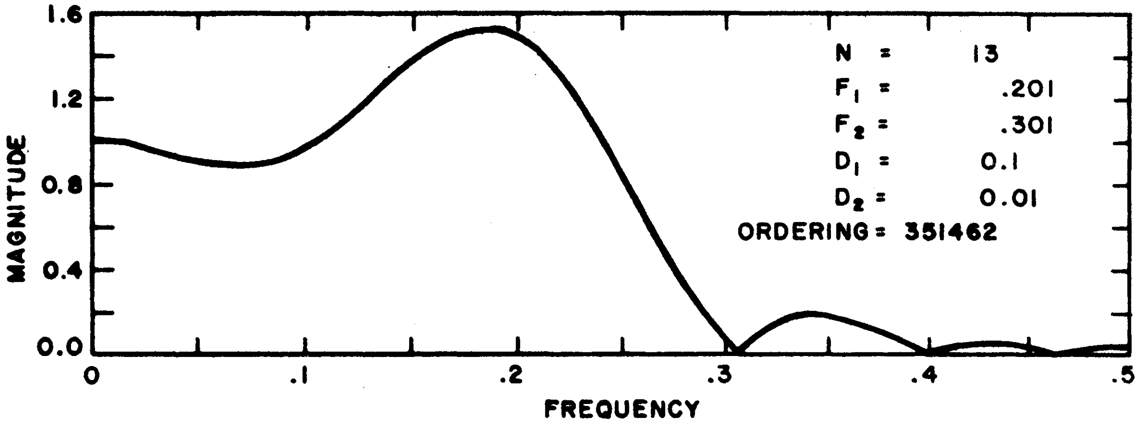
FREQUENCY

FILTER SPECTRUM FROM INPUT TO SECTION 6

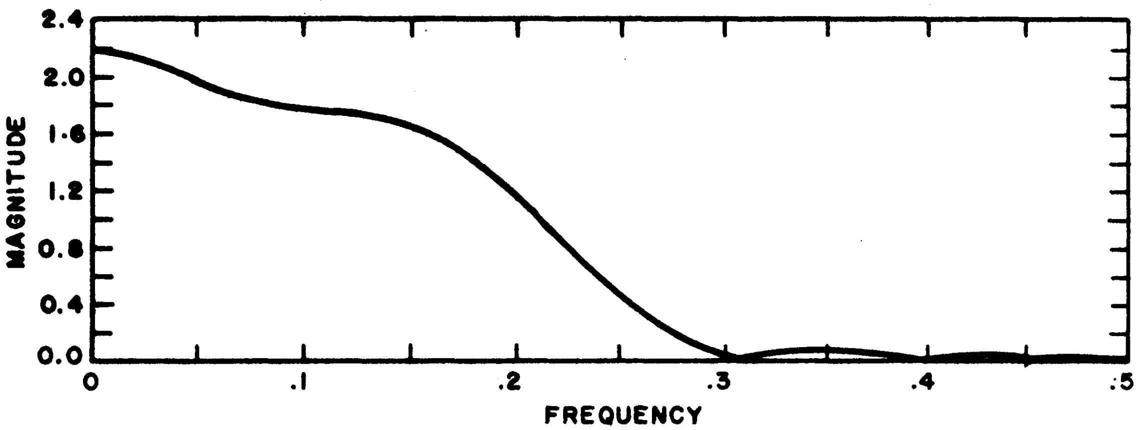


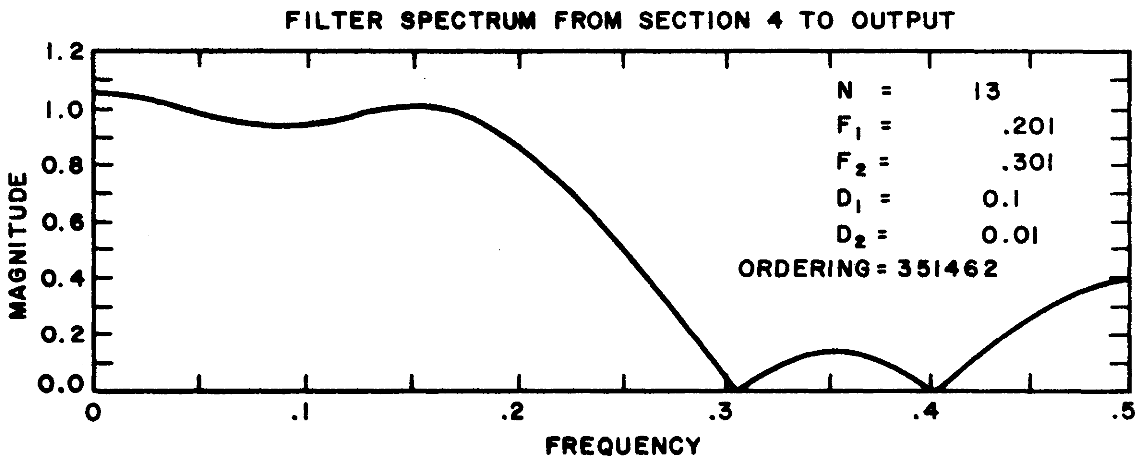
FREQUENCY

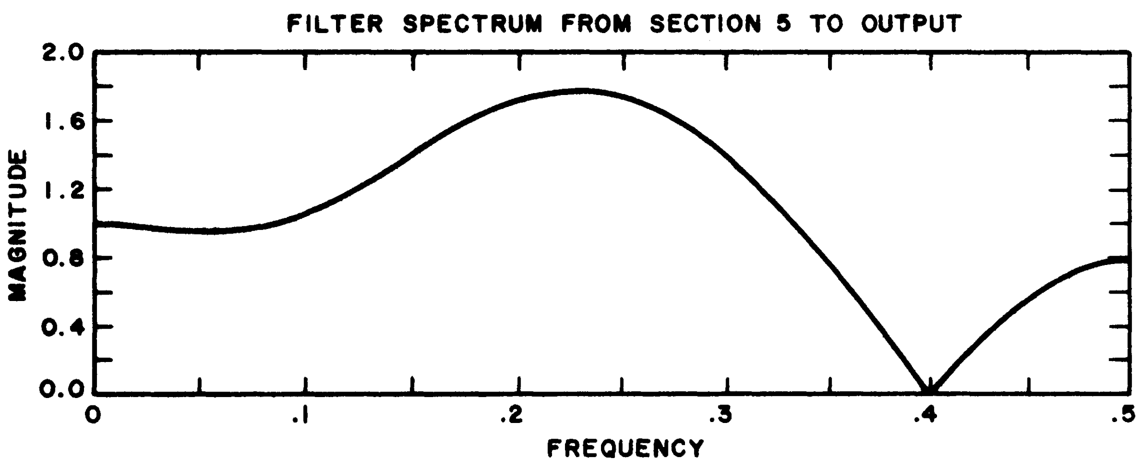
FILTER SPECTRUM FROM SECTION 2 TO OUTPUT

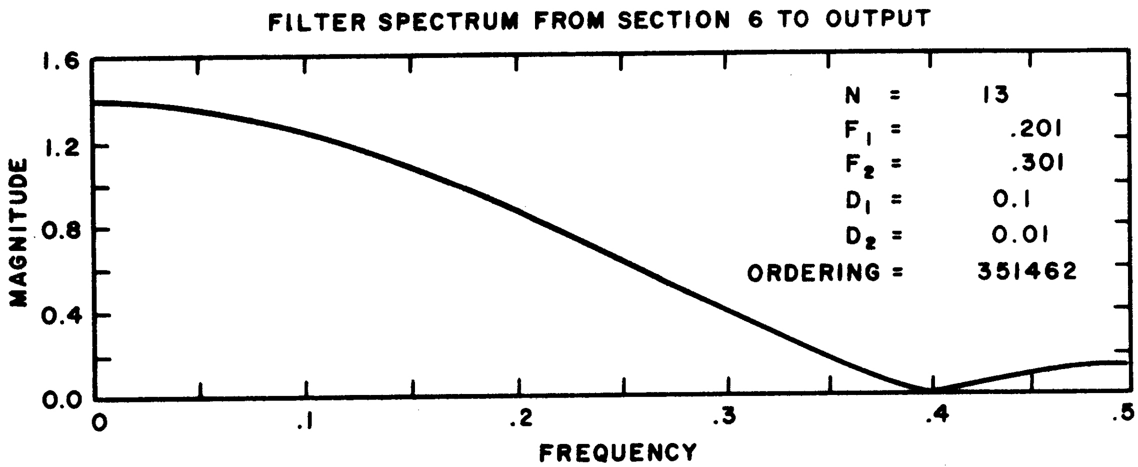


FILTER SPECTRUM FROM SECTION 3 TO OUTPUT




FILTER SPECTRUM FROM SECTION 4 TO OUTPUT
MAGNITUDE
FREQUENCY
N = 13
F1 = .201
F2 = .301
D1 = 0.1
D2 = 0.01
ORDERING = 351462


FILTER SPECTRUM FROM SECTION 5 TO OUTPUT
MAGNITUDE
FREQUENCY


FILTER SPECTRUM FROM SECTION 6 TO OUTPUT
MAGNITUDE
1.6
1.2
0.8
0.4
0.0
0
.1
.2
.3
.4
.5
FREQUENCY
N = 13
F1 = .201
F2 = .301
D1 = 0.1
D2 = 0.01
ORDERING = 351462