Copyright © 1973 American Telephone and Telegraph Company
THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 3, March, 1973
Printed in U.S.A.

An Algorithm for Minimizing Roundoff Noise in Cascade Realizations of Finite Impulse Response Digital Filters

By D. S. K. CHAN and L. R. RABINER

(Manuscript received September 14, 1972)

Experimental results on roundoff noise in cascade realizations of Finite Impulse Response (FIR) digital filters are presented in this paper. The entire roundoff noise distribution (i.e., over all possible orderings) is given for several low-order filters using both sum and peak scaling. Based on observations about this distribution, as well as intuitive arguments about the effects of ordering on roundoff noise, an algorithm for minimizing roundoff noise is presented. Experimental verification of this algorithm for a wide range of filters is given.*

I. INTRODUCTION

As discussed in previous works,[1,2] the implementation of FIR filters using finite precision arithmetic has become an important issue in recent years. For cascade realizations of FIR filters, roundoff noise is a crucial problem. In Refs. 1 and 2, some of the theoretical bases for the analysis of roundoff noise in the FIR cascade form have been considered. This paper presents a large body of experimental results which depict the dependence of roundoff noise on several of the important parameters of a cascade FIR low-pass filter. Most importantly, these results point to an algorithm which can find efficiently, for a cascade filter, an ordering which has a noise variance very close to the minimum possible. Experimental verification of this algorithm for a wide range of filters is presented.

Low-pass, extraripple[3] filters are used throughout these investigations as being representative of FIR filters. It will be seen that most results

* This paper is based on a thesis[1] submitted in partial fulfillment of the requirements for the degrees of Bachelor of Science and Master of Science in the Department of Electrical Engineering at the Massachusetts Institute of Technology in September 1972.

should not depend on the type of filter used. Figure 1 shows the magnitude response of a typical extraripple filter (which by definition has linear phase) and the parameters which define it. Of the four parameters F_1, F_2, D_1, and D_2, only three can be independently specified. The parameters N (impulse response length), F_1, D_1, and D_2 will be chosen as independent variables in these investigations. The studied ranges of variation of these parameters are as follows: $7 \leqq N \leqq 129$, $0.1 \geqq D_1 \geqq 0.001$, $0.1 \geqq D_2 \geqq 0.001$, $0 < F_1 < 0.5$, and $0 < F_2 < 0.5$ (sampling frequency $= 1$). These ranges comprise a large range of the significant values that these parameters take on. In the present state of the art in real-time digital filter hardware, 128th-order ($N = 129$) is the highest order that can be implemented in cascade form at a sampling rate of 10 kHz (e.g., typical for speech processing).[4,5] Furthermore, the stated ranges for D_1 and D_2 are those significant for many speech processing systems.[6]

While a great deal of experimental data has been collected, only representative examples will be presented here. For more examples see Ref. 1.

II. PRELIMINARY REMARKS

In Ref. 2 it is 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 remain 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 are developed to fix these N_s degrees of freedom, and two particular methods, viz., sum scaling and peak scaling,* are shown to be optimum for the particular classes of input signals which they assume. These scaling methods will be applied in this paper.

The prime issues in the realization of filters in cascade form are threefold–scaling, ordering, and section configuration. Because of the simplicity of a 2nd-order FIR filter, there is little freedom in the choice of a structure for the sections of a cascade filter. In Ref. 2, the configuration shown in Fig. 2a is assumed because it turns out to be the most useful in a practical situation. Another configuration, Fig. 2b, is also mentioned in Ref. 2. Because, as seen from Fig. 2c, these two configurations can be readily accommodated in a more general sub-

* Sum (peak) scaling is defined to be a method of scaling where the scale factors for the cascade sections are chosen so that the sum of the impulse response magnitudes (peak of the magnitude of the frequency response) up to that section does not exceed one.

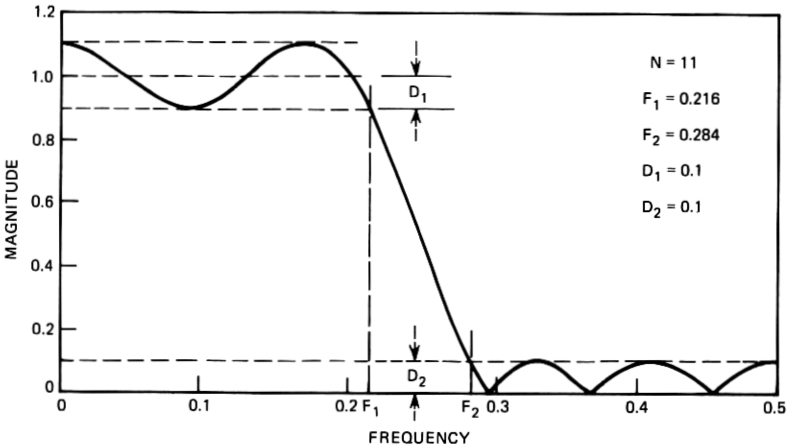


Fig. 1—Definition of filter parameters D_1, D_2, F_1, and F_2.

filter structure, it is here assumed that the configurations of Figs. 2a and b are both used in a cascade structure, depending on whether $b_{0i} \neq b_{2i}$ or $b_{0i} = b_{2i}$ respectively. The configuration of Fig. 2b has the advantage of having lower noise than the configuration of Fig. 2a. The option of summing all products in an increased length register before rounding is also possible for all configurations. However, the gain in signal-to-noise ratio does not seem to be worth the required sacrifice in speed (assuming serial arithmetic).[7] In any case, all resulting noise variances would simply be scaled down by a factor of from 2 to 3 if this strategy were used instead of rounding after each multiplication, as assumed here.

Other possible section configurations will be discussed later on. Since scaling is treated in depth in Ref. 2, the major concern here 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 structures 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 with time constraint is by far the most difficult issue to deal with in the design of filters in cascade form.

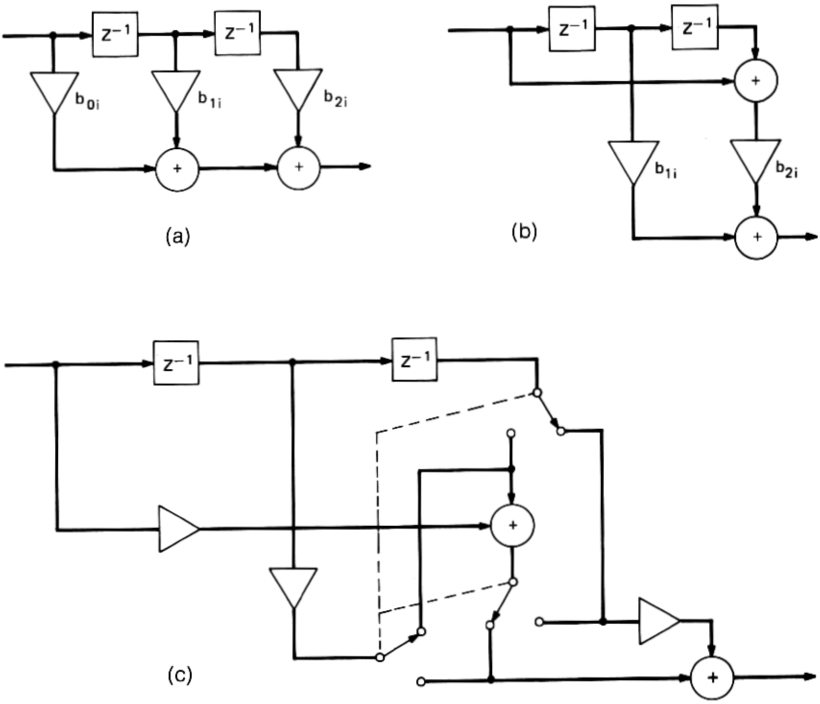


Fig. 2—(a) Cascade form filter section. (b) Alternate cascade filter section. (c) General cascade filter section.

Since finding an optimal solution to the ordering problem through exhaustive searching is very time consuming (if not impossible in any feasible amount of time) 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 unfounded if the roundoff noise level produced by a filter were rather insensitive to ordering. 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.[7,8] He showed a 33-point FIR filter which, ordered one way, produces $\sigma^2 = 2.4\ Q^2$ (where Q is the quantization step size of the filter), while ordered another way yields $\sigma^2 = 1.5 \times 10^8\ Q^2$ (assuming all products in each section are summed before rounding). This represents a difference of 1.6 bits versus 14.6 bits of noise. Clearly, the difference is large enough so that 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 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 variance 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 the next section, these questions will be answered 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.

III. CALCULATION OF NOISE DISTRIBUTIONS

3.1 *Methods*

The definitions of sum scaling and peak scaling in Ref. 2 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 $\sum_{k=0}^{2i} |\hat{f}_i(k)|$ for all i as is necessary for sum scaling. In the 33-point filter mentioned above, Schüssler used peak scaling on both the orderings. It will be shown in Section IV 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 these investigations.

Returning to the question of section configuration, for Infinite Impulse Response (IIR) filters Jackson[9] has introduced the concept of transpose configurations to obtain alternate structures for filter sec-

* $\hat{F}_i(e^{j\omega})$ and $\{\hat{f}_i(k)\}$ are defined in Ref. 2 and are the frequency response and impulse response of the cascade of sections 1 to i.

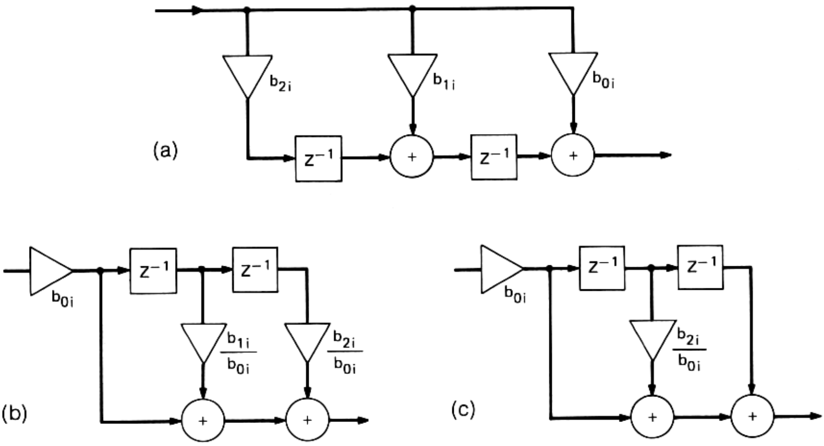


Fig. 3—(a) Transpose configuration of Fig. 2a. (b) Alternate configuration of Fig. 2a. (c) Alternate configuration of Fig. 2b.

tions. However, the application of this concept to Fig. 2a yields the structure shown in Fig. 3a, which is seen to have the same noise characteristics as the structure in Fig. 2a since, by the whiteness assumption on the noise sources, delays have no effect on them. Therefore, the structure of Fig. 3a has no advantages over other structures as far as roundoff noise is concerned. The only other significant alternate configuration for Fig. 2a is shown in Fig. 3b. The counterpart for Fig. 2b is Fig. 3c and is valid when $b_{0i} = b_{2i}$. 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 Figs. 3b and c for the ith section when

$$\frac{1}{b_{0i}^2} \sum_k g_{i-1}^2(k) < \sum_k g_i^2(k) \tag{1}$$

where $\{g_i(k)\}$ is, as defined in Ref. 2, the impulse response of the equivalent filter seen by the ith noise source. However, in order to have no error-causing internal overflow when the input and output of a section are properly constrained, the structures of Figs. 3b and c can be used only when $b_{0i} \leqq 1$. If $b_{0i} > 1$, either four multipliers are required, or Fig. 3b reduces to Fig. 2a.

In the investigations that follow, for each section of a filter, the configuration among Figs. 2a, 2b, 3b, and 3c which is applicable and

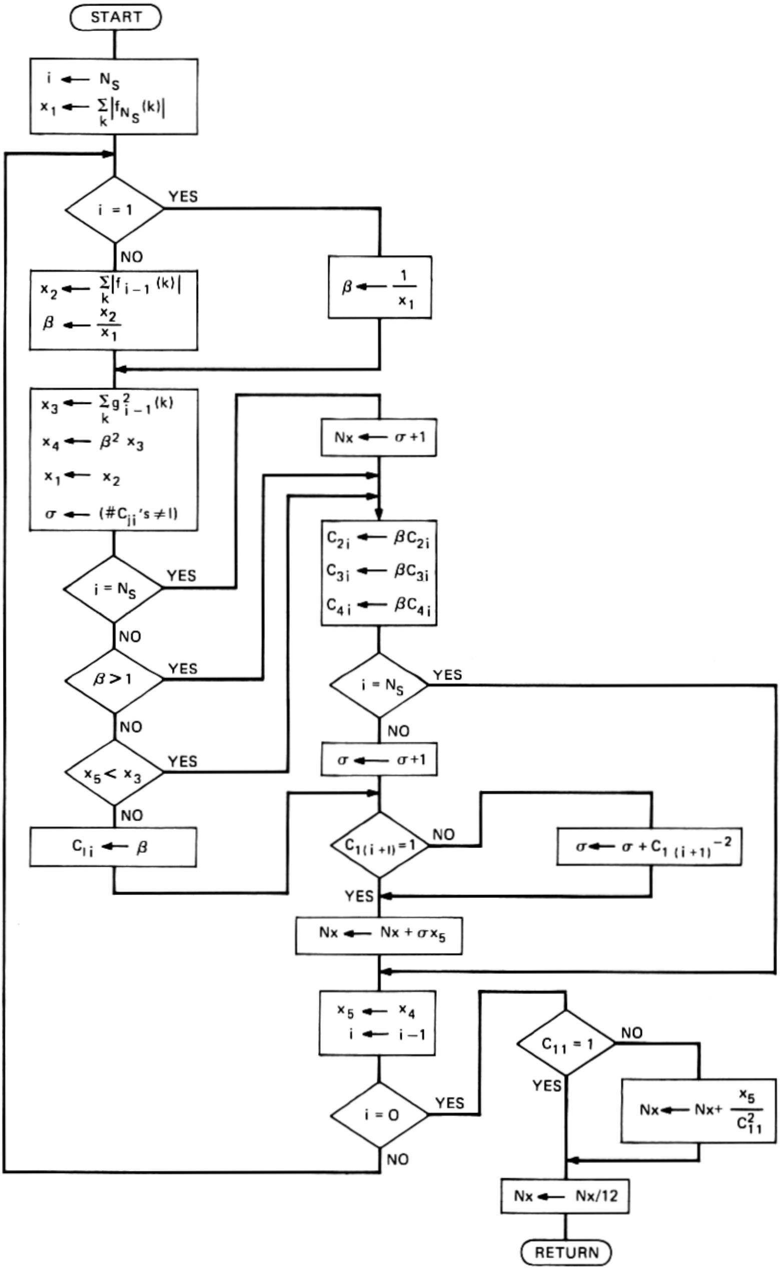

START
$i \leftarrow N_S$
$x_1 \leftarrow \sum_k |f_{N_S}(k)|$
$i = 1$
YES
NO
$\beta \leftarrow \frac{1}{x_1}$
$x_2 \leftarrow \sum_k |f_{i-1}(k)|$
$\beta \leftarrow \frac{x_2}{x_1}$
$x_3 \leftarrow \sum_k g_{i-1}^2(k)$
$x_4 \leftarrow \beta^2 x_3$
$x_1 \leftarrow x_2$
$\sigma \leftarrow (\#C_{ji}\text{'s} \neq 1)$
$Nx \leftarrow \sigma + 1$
$i = N_S$
YES
NO
$\beta > 1$
YES
NO
$x_5 < x_3$
YES
NO
$C_{1i} \leftarrow \beta$
$C_{2i} \leftarrow \beta C_{2i}$
$C_{3i} \leftarrow \beta C_{3i}$
$C_{4i} \leftarrow \beta C_{4i}$
$i = N_S$
YES
NO
$\sigma \leftarrow \sigma + 1$
$C_{1(i+1)} = 1$
NO
$\sigma \leftarrow \sigma + C_{1(i+1)}^{-2}$
YES
$Nx \leftarrow Nx + \sigma x_5$
$x_5 \leftarrow x_4$
$i \leftarrow i - 1$
$i = 0$
YES
NO
$C_{11} = 1$
NO
YES
$Nx \leftarrow Nx + \frac{x_5}{C_{11}^2}$
$Nx \leftarrow Nx/12$
RETURN

Fig. 4—Flow chart of scaling and noise calculation subroutine.

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 Figs. 2a and b are almost always more advantageous than the other configurations. 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 Figs. 3b and c may be ignored.

Figure 4 is the flow diagram of a computer subroutine which is used to accomplish scaling, choice of configuration, and output noise variance calculation given a filter and its ordering. The input to the subroutine consists of N_s (the number of sections) and the sequence $\{C_{ji},\ 1 \leqq j \leqq 4,\ 1 \leqq i \leqq N_s\}$, the elements of which are unscaled coefficients of the filter, defined by

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

where $H_i(z)$ is the ith section in the filter cascade. The sequences $\{f_i(k)\}$ and $\{g_i(k)\}$ in Fig. 4 are the impulse responses of the cascade of the first i sections and the last $(N_s - i)$ sections respectively. 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. By Theorem 7(*ii*) in Ref. 2, for any filter with at least one set of two complex conjugate pairs of reciprocal zeros, there are at most $N_s!/2$ orderings that differ in output noise variance. This is true since if all orderings are divided into two groups, according to the order in which the reciprocal zeros are synthesized in the cascade, then Theorem 7(*ii*) establishes a one-to-one correspondence between each ordering of one group and some ordering of the other group. Thus, in the investigation of all possible noise outputs of a filter, where possible, a pair of sections which synthesize reciprocal zeros is chosen and all orderings in which a particular one of these sections precedes the other are ignored.

3.2 *Discussion of Results*

Using the methods and procedures described, the noise distributions of 27 different linear phase, low-pass extraripple filters were investi-

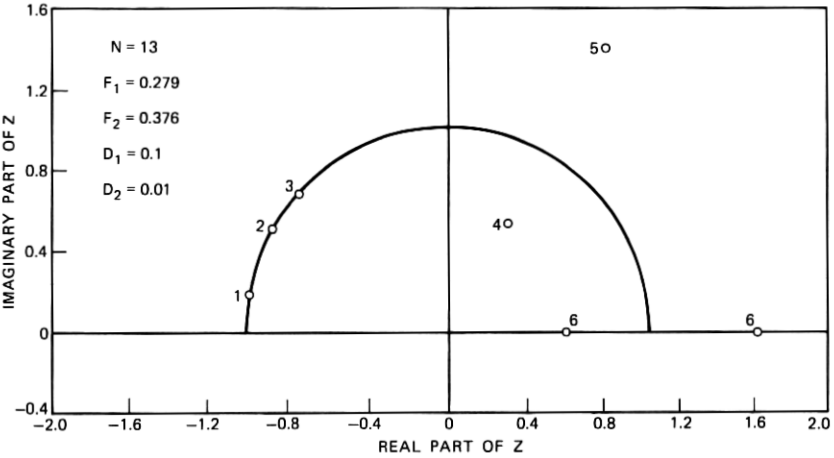


Fig. 5—Positions of the zeros of one filter.

gated. Twenty-two of these filters are 13-point filters, since $N = 13$ represents a reasonable filter length to work with. Thirteen-point filters have six sections each, corresponding to 6! or 720 possible orderings of sections. By reducing redundancy via Theorem 7 of Ref. 2, the number of orderings it is 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, attention is focused on a typical 13-point filter. As an example, a filter with 4 ripples in the passband, 3 ripples in the stopband, and passband and stopband tolerances of 0.1 and 0.01, respectively, is used. By passband and stopband tolerances is meant the maximum height of ripples in the respective frequency bands. Figure 5 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 section synthesizes are given the same number, and these are shown in Fig. 5. Table I 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. A histogram plot of the noise distribution is shown in Fig. 6a, and a cumulative distribution plot is shown in Fig. 6b.

Two characteristics of the histogram shown in Fig. 6a 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. It is seen that most orderings have very low noise compared to

TABLE I—NOISE VARIANCE OF ALL 360 ORDERINGS OF A 13-POINT FILTER

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

Table I—*Continued.*

Order	Noise	Order	Noise	Order	Noise
264531	3.8141	341526	5.8993	413265	16.5228
264513	3.8836	452613	5.9446	431265	16.5541
163245	4.0265	413526	6.0362	463521	16.5661
146532	4.0376	345216	6.0375	463512	16.6163
146523	4.0569	346521	6.0426	412365	16.6522
162345	4.0697	425316	6.0520	421365	16.6587
261345	4.1025	431526	6.0674	134265	17.5048
361245	4.1445	346512	6.0928	314265	17.5837
345621	4.2234	451632	6.1475	243165	17.7595
416532	4.2570	451623	6.1668	342165	17.8249
345612	4.2737	435216	6.2055	423165	17.9682
416523	4.2763	436521	6.2106	432165	17.9929
263145	4.2914	436512	6.2609	124365	18.2607
164532	4.3715	243516	6.3368	214365	18.4053
362145	4.3766	364521	6.3805	234165	19.2166
164523	4.3907	342516	6.4021	324165	19.4048
435621	4.3915	364512	6.4307	642351	21.7670
435612	4.4417	423516	6.5454	643251	21.8232
451263	4.5778	432516	6.5701	641352	21.8995
451362	4.5806	134526	7.0181	642153	21.9017
452163	4.6348	314526	7.0970	643152	21.9061
452361	4.6353	124536	7.2674	641253	21.9513
453261	4.7361	214536	7.4120	642315	22.0838
453162	4.7383	634521	7.4875	642135	22.1143
136245	4.9584	634512	7.5377	643215	22.1401
624531	4.9693	453621	7.6049	641325	22.1410
145236	4.9857	453612	7.6551	643125	22.1476
245136	5.0354	234516	7.7939	641235	22.1639
316245	5.0372	324516	7.9820	143256	23.0109
624513	5.0388	451236	8.4532	341256	23.0934
613245	5.1911	452136	8.5101	142356	23.1403
415236	5.2051	451326	8.8996	241356	23.1575
612345	5.2343	453126	9.0573	413256	23.2302
425136	5.2440	452316	9.3181	431256	23.2615
631245	5.2515	453216	9.4189	412356	23.3596
621345	5.2577	462351	11.8333	421356	23.3661
236145	5.4048	463251	11.8896	642531	24.0341
145326	5.4322	461352	11.9658	642513	24.1036
142536	5.4395	462153	11.9680	134256	24.2122
623145	5.4466	463152	11.9725	314256	24.2911
241536	5.4567	461253	12.0176	243156	24.4670
632145	5.4836	462315	12.1501	342156	24.5323
614532	5.5361	462135	12.1806	641532	24.6126
614523	5.5553	463215	12.2064	641523	24.6319
126345	5.5880	461325	12.2073	423156	24.6756
326145	5.5930	463125	12.2140	432156	24.7003
415326	5.6515	461235	12.2302	124356	24.9681
412536	5.6589	462531	14.1004	214356	25.1128
421536	5.6653	462513	14.1699	234156	25.9240
345126	5.6759	461532	14.6789	324156	26.1122
216345	5.7326	461523	14.6982	643521	26.4998
143526	5.8168	143265	16.3035	643512	26.5500
245316	5.8434	341265	16.3860	132645	81.4953
435126	5.8440	142365	16.4329	312645	81.5742
452631	5.8751	241365	16.4501	231645	85.2733

TABLE I—*Continued.*

Order	Noise	Order	Noise	Order	Noise
321645	85.4615	231465	116.6240	465213	180.9850
123645	87.0142	321465	116.8120	465132	181.3550
213645	87.1589	123465	118.3650	465123	181.3750
456231	99.7815	213465	118.5090	465321	182.8770
456213	99.8510	132456	119.5530	465312	182.9270
456132	100.2220	312456	119.6320	645231	190.8490
456123	100.2410	231456	123.3310	645213	190.9180
456321	101.7430	321456	123.5190	645132	191.2890
456312	101.7940	123456	125.0720	645123	191.3080
132465	112.8460	213456	125.2170	645321	192.8110
312465	112.9250	465231	180.9150	645312	192.8610

the maximum value possible. In fact, the lowest range of noise variance, in this case between zero and 2 Q^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 19.5 Q^2. In fact, from the cumulative distribution it is seen that approximately two-thirds of the orderings have noise variances less than 4 percent of the maximum, while nine-tenths of them have noise variances less than 14 percent 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. 6a shows this effect only for the higher noise values, a more detailed plot of the distribution in the range from zero to 28 Q^2, as in Fig. 6c, shows that gaps also occur for lower noise values. Thus noise values tend to occur in several levels of clusters. These observations provide the general picture of clusters of noise values, the separation of which increases rapidly as a function of the magnitude of the noise values, thus forming a highly skewed noise distribution.

The significance of these results is far-reaching. Given a specific filter, because of the abundance of orderings which yield almost the lowest noise variance possible, it is concluded that it should not be too difficult to devise a feasible algorithm which will yield an ordering, the noise variance of which 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 far more advantageous to use a suboptimal method, which can rapidly choose an ordering that is satisfactory, than to try to find the optimum. The reduction in roundoff noise gained by finding the optimum solu-

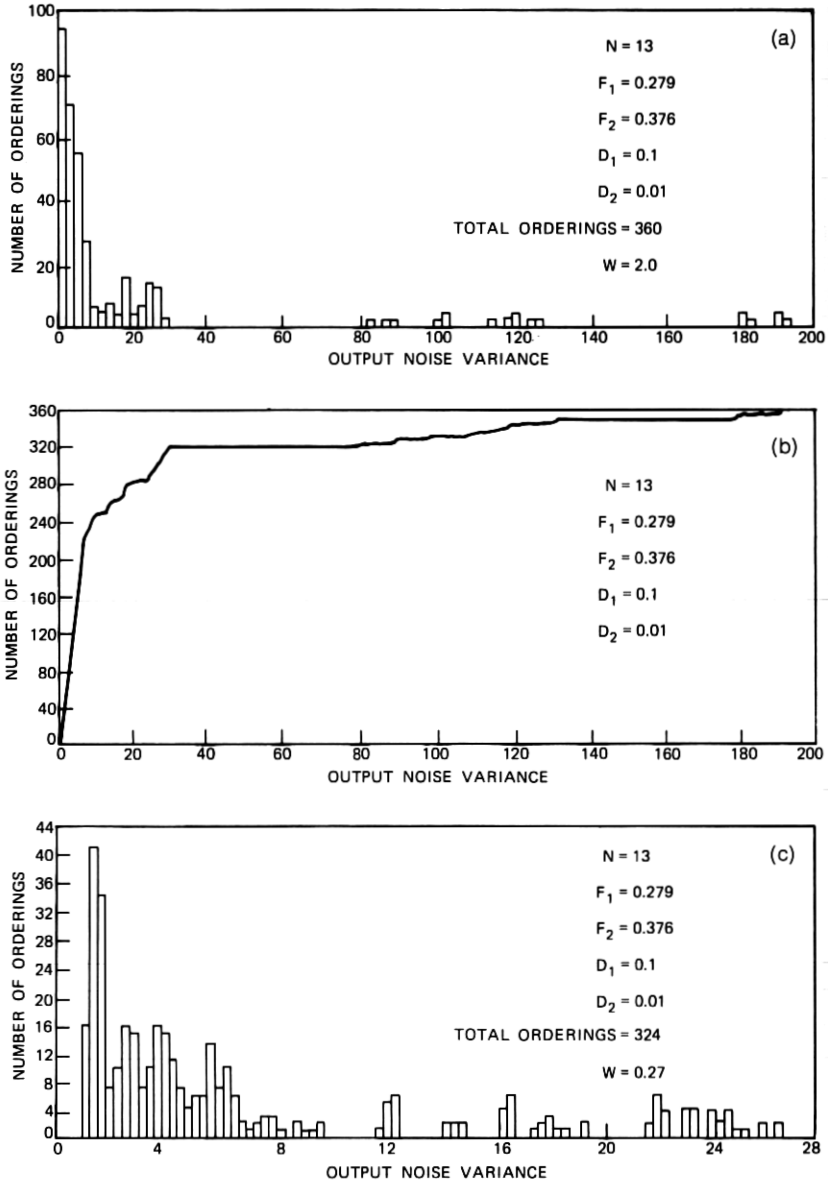


Fig. 6—(a) Noise distribution histogram of filter of Fig. 5. (b) Cumulative noise distribution of filter of Fig. 5. (c) Detailed noise distribution histogram of filter of Fig. 5.

tion is probably, at best, not worth the extra effort from the design standpoint. At least up to the present, no efficient method for finding an optimum ordering has been found.

In Section VI, a suboptimal method is presented which, given a filter, yields a low-noise ordering efficiently and has been successfully applied to a wide range of filters. Before presenting the algorithm, the behavior of roundoff noise with respect to scaling and other filter parameters will be further investigated. Also, the nature of high-noise and low-noise orderings will be discussed, so that they can be more easily recognized.

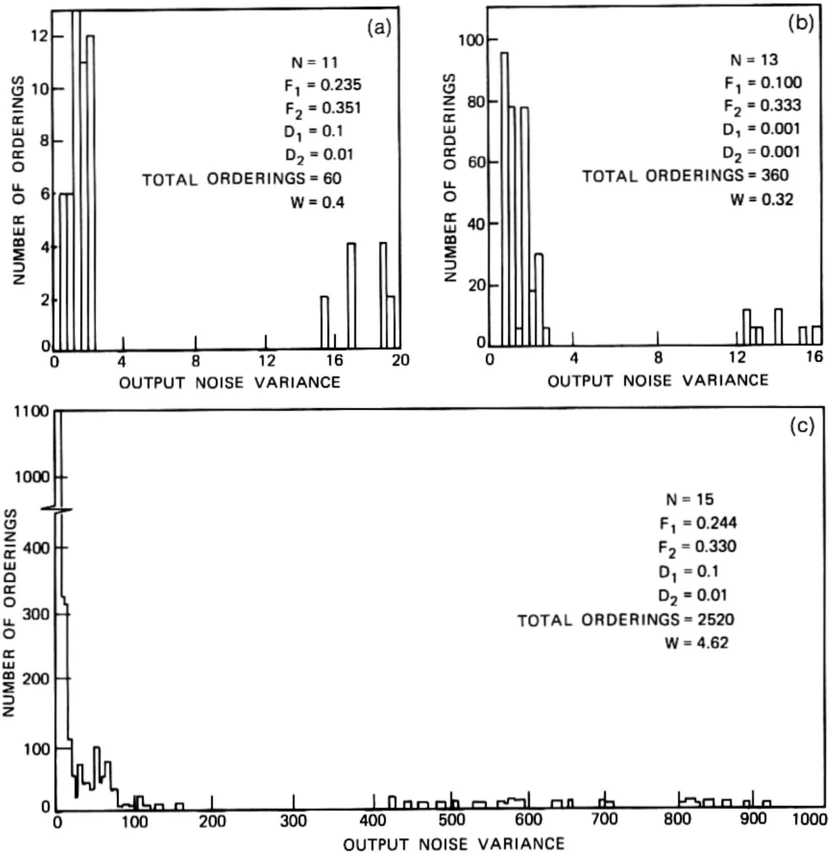


Fig. 7—(a) Noise distribution histogram of typical 11-point filter. (b) Noise distribution histogram of another 13-point filter. (c) Noise distribution histogram of a 15-point filter.

Before ending this section, the noise distribution histograms of an 11-point and one more 13-point filter are presented in Figs. 7a and b. These are seen to exhibit all the characteristics discussed above. The major difference between the noise distributions of the two 13-point filter examples presented lies in the magnitude of the maximum and average noise variances. This difference will be accounted for presently.

Also presented is the noise distribution for a 15-point filter, in Fig. 7c. The calculation of this distribution involves 2520 different orderings. This histogram shows even stronger emphasis on the distribution characteristics discussed and, together with Figs. 6a and 7a, suggests that the skewed shape and large-gap properties of the noise distribution of a filter become increasingly pronounced as the order of the filter increases. Thus it is expected that the results presented can be generalized for higher-order filters.

3.3 *Dependence of Distributions on Transfer Function Parameters*

From all the calculated noise distributions of the previous section, the interesting fact is observed 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. The differences in noise variance ranges among different filters is accounted for by investigating the dependence of noise distributions on parameters which specify the transfer function of a filter.

The noise distributions of several low-pass extraripple filters with various values of the parameters, N, F_1, D_1, and D_2 were computed using the methods described. Since all these distributions have the same general shape, they can be compared by simply examining their maximum, average, and minimum values. A list of all the filters, the noise distributions of which have been computed, including those already discussed, is presented in Table II. 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 II. The last column in this table will be discussed in Section VI.

Filters numbered 1 to 5 in Table II are very similar except for their order in that they all have identical passband and stopband tolerances and approximately the same low-pass bandwidth. The maximum, average, and minimum values of their noise distributions are plotted

Table II—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	0.212	0.1	0.01	1.24	0.84	0.37	0.37
2	9	3	0.281	0.1	0.01	6.26	2.54	0.73	0.73
3	11	3	0.235	0.1	0.01	19.41	4.79	0.68	0.68
4	13	4	0.279	0.1	0.01	192.86	19.55	1.10	1.10
5	15	4	0.244	0.1	0.01	923.63	54.45	1.02	1.16
6	13	3	0.100	0.001	0.001	15.84	3.01	0.65	0.69
7	13	4	0.261	0.05	0.004	119.48	12.91	0.96	1.02
8	13	1	0.012	0.01	0.01	9.91	1.61	0.32	0.35
9	13	2	0.067	0.01	0.01	16.30	2.94	0.44	0.47
10	13	3	0.138	0.01	0.01	42.63	5.94	0.71	0.73
11	13	4	0.213	0.01	0.01	69.76	8.52	0.82	0.91
12	13	5	0.288	0.01	0.01	76.43	11.01	1.44	1.52
13	13	6	0.364	0.01	0.01	52.54	10.33	1.92	2.43
14	13	3	0.201	0.1	0.01	96.25	12.09	0.81	
15	13	3	0.179	0.05	0.01	69.26	9.02	0.76	
16	13	3	0.154	0.02	0.01	50.63	6.87	0.72	
17	13	3	0.123	0.005	0.01	37.36	5.33	0.70	
18	13	3	0.106	0.002	0.01	32.83	4.80	0.69	
19	13	3	0.095	0.001	0.01	30.53	4.53	0.69	
20	13	3	0.124	0.01	0.1	132.57	17.56	1.02	
21	13	3	0.129	0.01	0.05	85.84	11.45	0.83	
22	13	3	0.135	0.01	0.02	54.94	7.47	0.75	
23	13	3	0.141	0.01	0.005	35.59	5.07	0.68	
24	13	3	0.144	0.01	0.002	26.44	4.37	0.68	
25	13	3	0.146	0.01	0.001	22.52	4.07	0.70	
26	15	4	0.185	0.01	0.01	417.08	27.38	1.00	
27	15	4	0.255	0.1	0.001	601.83	35.15	1.02	

on semilog coordinates in Fig. 8a. It is seen 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.

Figure 8b shows a similar plot of the same distribution statistics for filters numbered 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. 8b it is seen that with those parameters mentioned above held fixed, the noise output of a cascade filter tends to increase with increasing bandwidth.

Filters numbered 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. 8c and 8d. These plots indicate that, as the transfer function approximation error for a filter decreases, so does its noise output. Though the plots are made holding N_p rather than F_1 fixed, it is seen that, at least for the filters used in Fig. 8d, bandwidth increases with decreasing approximation error. Since the noise output of a filter is found to increase with band-

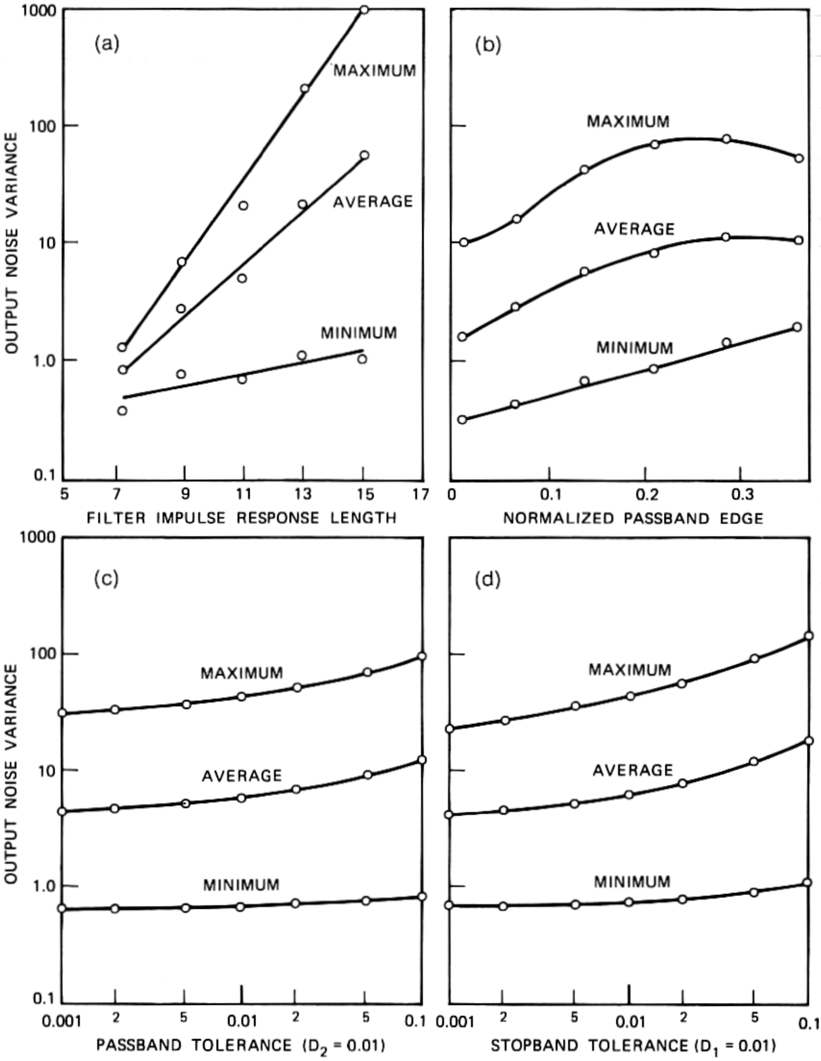


Fig. 8—(a) Output noise variance as a function of filter length. (b) Output noise variance as a function of bandwidth. (c) Output noise variance as a function of passband approximation error. (d) Output noise variance as a function of stopband approximation error.

width, it is expected 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 8b to 8d 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 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, it has been found experimentally that with other parameters fixed, the roundoff noise output of a filter tends to increase with increases in 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. It was not shown 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 has been shown is perhaps a more useful result from the design viewpoint. These 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, it is expected they could be generalized for higher-order filters as well. Section VI will present experimental evidence to confirm this expectation.

IV. COMPARISON OF SUM SCALING AND PEAK SCALING

The claim was made earlier that the results obtained on the noise distribution of filters ought to be quite independent of whether sum scaling or peak scaling is used. This claim will now be sustained heuristically and experimentally.

Let α denote the ratio of the maximum gain (over all frequencies) of a low-pass extraripple filter scaled by sum scaling to the maximum gain of the same filter peak scaled. Then it must be true that $\alpha \leqq 1$, since by definition the maximum gain for peak scaling is exactly one, while for sum scaling it must be no more than one if class 2 signals (which are a subset of class 1) are to be properly constrained (by Theorem 3, Ref. 2). Furthermore, it can be easily shown[1] that

$\alpha \geqq 1 - 2\epsilon$ where ϵ is defined as

$$\epsilon = \frac{\sum_k r(k)}{\sum_k |h(k)|} \tag{3}$$

with $\{h(k)\}$ being the impulse response of the filter and $\{r(k)\}$ being the magnitude of the negative portion of $\{h(k)\}$, i.e.,

$$r(k) = \begin{cases} -h(k) & h(k) < 0 \\ 0 & h(k) \geqq 0 \end{cases}. \tag{4}$$

For a low-pass filter, the envelope of $\{h(k)\}$ has the general shape of a truncated $(\sin x)/x$ curve, hence ϵ is expected to be a small number.

Table III—List of Filters and the Results of Ordering Algorithms

$D1 = 0.01$ $D2 = 0.001$					Noise Variance		
					Peak Scaling		Sum Scaling
#	N	N_p	F_1	α	Alg 1	Alg 2	Alg 1
28	13	4	0.219	0.65	1.25	1.26	0.90
29	15	4	0.193	0.68	1.23	1.22	1.02
30	17	5	0.230	0.61	1.99	2.49	1.37
31	19	5	0.207	0.64	1.93	1.92	1.47
32	21	6	0.236	0.59	2.50	2.61	1.58
33	23	6	0.216	0.61	2.57	2.91	1.77
34	25	7	0.240	0.57	3.75	3.62	2.35
35	27	7	0.223	0.59	3.95	4.11	2.45
36	29	8	0.243	0.55	4.54	5.04	2.67
37	31	8	0.227	0.57	5.27	5.88	2.74
38	33	9	0.244	0.54	7.81	6.67	4.59
39	35	9	0.231	0.55	6.01	6.43	3.72
40	33	1	0.005	1.0	0.47	0.48	0.53
41	33	2	0.029	0.82	0.60	0.67	0.60
42	33	3	0.059	0.73	0.89	1.00	0.80
43	33	4	0.090	0.68	1.43	1.36	1.16
44	33	5	0.121	0.63	2.29	1.84	1.71
45	33	6	0.152	0.60	2.48	2.70	1.61
46	33	7	0.183	0.58	3.47	3.37	2.30
47	33	8	0.214	0.61	4.72	5.23	3.38
48	33	10	0.275	0.52	10.04	8.16	4.83
49	33	11	0.305	0.52	15.68	11.35	8.30
50	33	12	0.334	0.50	13.43	14.88	6.27
51	33	13	0.363	0.50	21.35	17.62	9.14
52	33	14	0.392	0.50	41.64	31.41	15.40
53	33	15	0.419	0.51	55.20	41.13	22.12
54	33	16	0.448	0.53	89.52	65.66	38.23

Table IV—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	0.211	0.01	0.002	0.59	5.63	5.33	3.34
56	0.208	0.01	0.005	0.57	5.13	5.69	3.18
57	0.205	0.01	0.01	0.56	5.05	5.27	3.34
58	0.202	0.01	0.02	0.55	7.63	8.31	4.01
59	0.197	0.01	0.05	0.53	11.34	12.53	6.92
60	0.193	0.01	0.1	0.51	46.33	22.99	16.88
61	0.238	0.1	0.01	0.58	9.90	9.01	5.61
62	0.227	0.05	0.01	0.58	8.91	7.35	5.52
63	0.214	0.02	0.01	0.56	8.87	5.75	4.32
64	0.196	0.005	0.01	0.56	5.47	4.69	3.68
65	0.185	0.002	0.01	0.56	5.95	4.08	3.41
66	0.178	0.001	0.01	0.57	4.10	4.11	2.85

In fact, it is easily shown[1] that, for a low-pass filter, if the passband tolerance D_1 is much less than the maximum passband gain, as is usually the case, then to an excellent approximation $\alpha = 1 - 2\epsilon$. It is now shown that if σ^2 is the output noise variance of a filter with sum scaling and σ'^2 is the output noise variance of the same filter except with peak scaling, then $\sigma^2 \geqq \alpha^2(\sigma'^2)$. To show this, the optimality properties for sum scaling and peak scaling, proved in Ref. 2, Theorem 4, are invoked. Since the gain of the sum-scaled filter is α times that of the peak-scaled filter, the ratio of their signal-to-noise ratios for class 2 inputs must equal α times the inverse ratio of their rms noise values (square root of variance), i.e., with S/N for sum scaling and S/N' for peak scaling,

$$\frac{S/N}{S/N'} = \alpha \cdot \frac{\sigma'}{\sigma}. \tag{5}$$

But since peak scaling is optimum for class 2 inputs and class 2 is a subset of class 1, then $S/N' \geqq S/N$. Thus $\sigma^2 \geqq \alpha^2 \sigma'^2$. For an alternative derivation see Ref. 1.

In Tables III and IV, a list of filters and some results of Section VI are presented. Together with these results are listed measured values of α for each filter. Observe that, for these typical filters, α ranges from 0.5 to 1. Furthermore, for each filter, the last and third to last columns

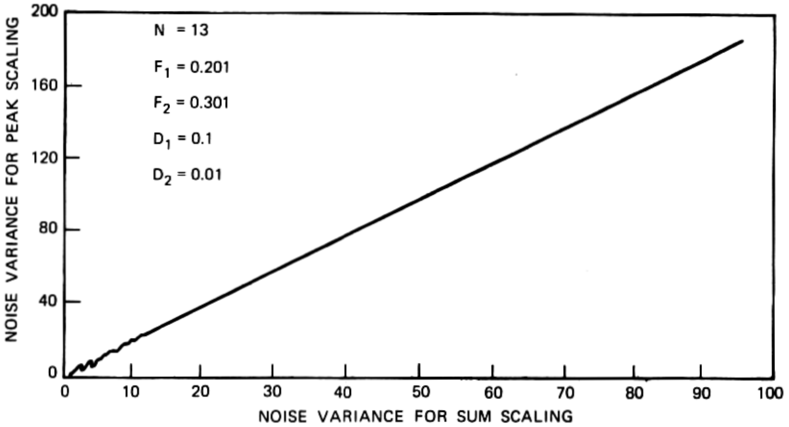


Fig. 9—Peak scaling versus sum scaling noise output comparison for a typical filter.

of Tables III and IV list the noise variances that result from the same ordering using sum scaling and peak scaling respectively. Comparing these, it is seen that, in almost every case, $\sigma^2 \leqq (\sigma'^2)$.[2] In particular, this is true if α is not too close to 1.0. The case where $\sigma^2 > (\sigma'^2)$ and $\alpha = 1$ is filter number 40 in Table III. However, except for the uninteresting cases of filters with all zeros on the unit circle, in general, $\alpha < 1$, and it is expected that $\sigma^2 \leqq (\sigma'^2)$. Thus for all practical purposes it can be assumed that

$$\alpha^2 \leqq \frac{\sigma^2}{\sigma'^2} \leqq 1. \tag{6}$$

From eq. (6), it is seen that the output noise variance for a filter with sum scaling is 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. 9, 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.

Thus the noise distributions of the previous section are essentially unchanged if peak scaling is used instead of sum scaling. To illustrate this, Fig. 10 shows the noise distribution for the filter of Fig. 6 with peak scaling used instead of sum scaling.

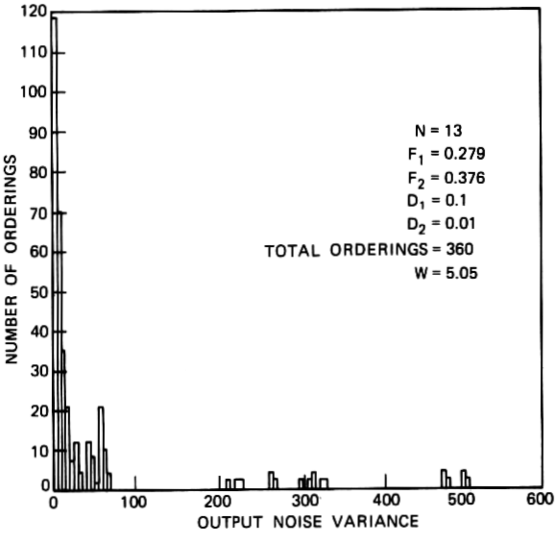


Fig. 10—Noise distribution histogram of filter of Fig. 5 using peak scaling.

The evaluation of noise variances with peak scaling is done in essentially the same way as that described in Fig. 4. Using a 128-point FFT to evaluate two transforms at a time (exploiting real and imaginary part symmetries) to give 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.

V. AN INTUITIVE EXPLANATION OF ROUNDOFF NOISE DEPENDENCE ON ORDERING IN TERMS OF SPECTRAL PEAKING

That roundoff noise is distributed in the way shown with respect to orderings for a filter is an intriguing fact which is by no means obvious. The dependence of roundoff noise on ordering involves complicated matters like 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, this dependence is much too complicated to visualize intuitively. It is proposed that the relative level of roundoff noise in a filter is adequately determined in order of magnitude by the amount of peaking in certain subfilter spectra. Thus it will be shown that, since the dependence on ordering of the amount of peaking of these spectra is not too difficult to visualize, by judging the relative amount of peaking of these spectra, the relative merit of an ordering in terms of high-noise or low-noise output can be determined by inspection. These findings explain the general shape of noise distributions.

Given a linear phase filter with z-transform $H(z)$, define transfer functions $\bar{H}_i(z)$, $i = 1, \cdots, N_s$, to have the property that $\bar{H}_i(z)$ is proportional to the transfer function for the ith section of the filter, and each $|\bar{H}_i(e^{j\omega})|$ for all i has a maximum over ω equal to C, where C is chosen so that the overall filter frequency response $H(e^{j\omega})$ has a maximum in magnitude equal to one. Clearly $C \geqq 1$. Define transfer functions

$$\begin{aligned} \bar{A}_i(z) &= \prod_{j=1}^{i} \bar{H}_j(z) \\ \bar{B}_i(z) &= \prod_{j=i+1}^{N_s} \bar{H}_j(z) \end{aligned} \tag{7}$$

and define a number Pk to be the largest value of $\max_\omega |\bar{A}_i(e^{j\omega})|$ or $\max_\omega |\bar{B}_i(e^{j\omega})|$ for all i. It will be argued 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 $\bar{G}_i(e^{j\omega})$ to be $\bar{B}_i(e^{j\omega})$ with its maximum in magnitude over ω normalized to unity. Then it can be easily shown[1] that if

$$\begin{aligned} A_i &= \max_\omega |\bar{A}_i(e^{j\omega})| \\ B_i &= \max_\omega |\bar{B}_i(e^{j\omega})| \\ C_i &= k_i \frac{Q^2}{12} \cdot \frac{1}{2\pi} \int_0^{2\pi} |\bar{G}_i(e^{j\omega})|^2 d\omega \end{aligned} \tag{8}$$

(where k_i is the number of noise sources in the ith section), then the output noise variance due to the ith section is given by

$$\sigma_i^2 = A_i^2 B_i^2 C_i \qquad 1 \leqq i \leqq N_s - 1. \tag{9}$$

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 the product of which is $H(e^{j\omega})$. Furthermore, for some i either $A_i = Pk$ or $B_i = Pk$. Now suppose $Pk \gg C$. Without loss of generality, it may be assumed $A_i = Pk$. Then argue that $A_i B_i \gg C$.

Clearly $A_i = |\bar{A}_i(e^{j\omega_0})|$ for some ω_0. Now $\bar{A}_i(z)\bar{B}_i(z) = H(z)$, 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 spaced and spread out around the unit circle. The zeros of a typical filter are shown in Fig. 11a. Furthermore, $|H(e^{j\omega})| \leqq 1$. Thus in order for $|\bar{A}_i(e^{j\omega})|$ to have a large peak at ω_0, several zeros of $H(z)$ which occur in the vicinity of $z = e^{\pm j\omega_0}$ must be missing from $\bar{A}_i(z)$, while most of the remaining zeros must be in $\bar{A}_i(z)$. This means that $\bar{B}_i(z)$ has a concentration of zeros around $e^{\pm j\omega_0}$. By the result of Theorem 6(*i*) in Ref. 2, which says that the maximum of the magnitude frequency response of a filter section occurs at either $\omega = 0$ or $\omega = \pi$ depending on whether the zeros synthesized are in the left half or the right half of the z-plane, it is seen that most factors of $\bar{B}_i(e^{j\omega})$ must have maxima in magnitude which occur at exactly the same ω.

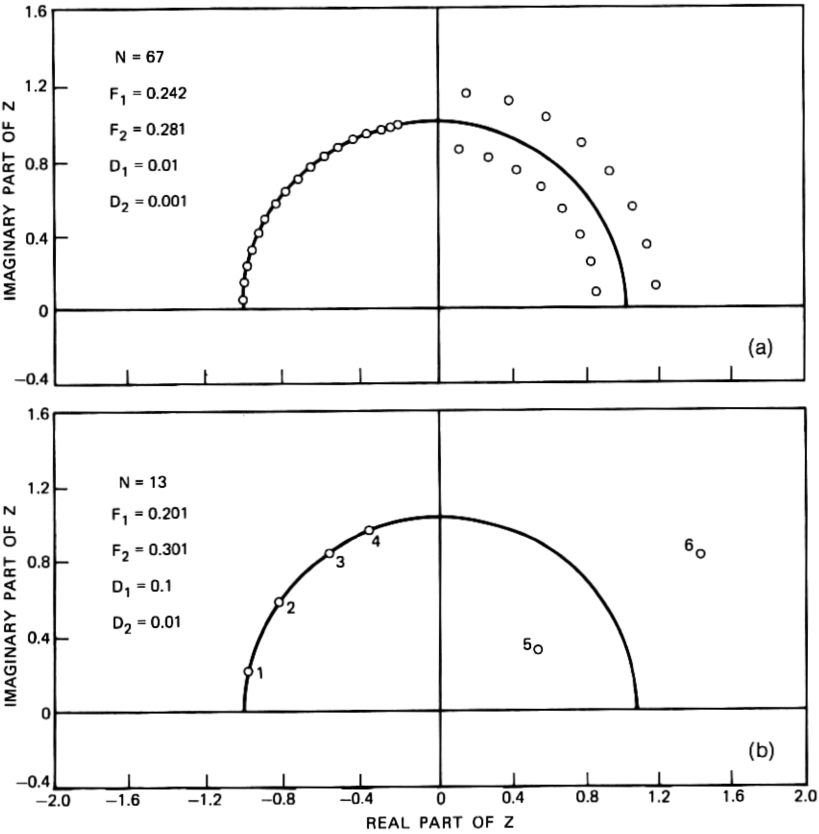


Fig. 11—(a) Zero positions of a typical 67-point filter. (b) Zero positions of filter number 14 of Table II.

Now C is found to be an increasing function of N_s^1, where typically $C > 2$. Hence $|\bar{B}_i(e^{j\omega})|$ is very likely to have a peak which is at least 1, or $B_i \geqq 1$. Thus $A_iB_i \gg C$. By the same token, if $B_i = Pk$ and $Pk \gg C$, then $A_iB_i \gg C$.

Hence if $Pk \gg C$, then for at least one i, $\sigma_i^2 = (A_iB_i)^2C_i$ where $A_iB_i \gg C$. 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 it is expected that the resulting σ^2 is among the lowest values possible. Thus there exists a correlation between high values of Pk and high noise, and low values of Pk and low noise.

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 $\max_\omega|\bar{G}_i(e^{j\omega})| = 1$ independent of ordering and i, it can be expected that variations in C_i with ordering are much less than variations in $(A_iB_i)^2$.

Based on these results, it can be concluded 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 since zeros from the same half of the z-plane produce frequency spectra the maxima of which occur at exactly the same ω, several zeros from the same half of the z-plane can build up a large peak in the product of their spectra $\bar{H}_i(e^{j\omega})$. 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 preceding sections occurs is likely to yield a low-noise filter.

The above observations are found to be true for all the filters the noise distributions of which were investigated. For example, from Table I, it is seen that those orderings which group together all three sections 4, 5, and 6 of the filter of Fig. 5 either at the beginning or at the end of the filter are precisely those which have the highest noise. Namely, they account for all noise variance values above 26.6 Q^2. Using the results on the noise distribution of a filter and the results of this section, it can thus be said 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.

These observations have the implication that low-noise orderings for a filter are those which choose zeros in such an order that they "jump around" about the unit circle and are well "interlaced," whereas high-noise orderings are those which allow clusters of adjacent zeros to be sequenced adjacently in the filter cascade. Since there are certainly far more ways to sequence zeros so that they satisfy the former property than the latter, it is clear why most orderings of a filter have low noise.

The values of Pk for all orderings of several filters were measured,[1] and they show good correlation with σ^2, thus supporting these arguments. For reasons of space, the results are not tabulated here. However, Figs. 12 and 13 show plots of the spectra from each section to the output of a high-noise and a low-noise ordering for a typical 13-point filter, the zeros of which are shown in Fig. 11b. Both orderings are peak scaled, so that the spectra from the input to each section of the filter have maxima equal to one. Thus, in reference to eq. (9), A_iB_i is equal to the maximum of the spectrum from section $(i + 1)$ to the output. The ordering of Fig. 12 has $\sigma^2 = 186\ Q^2$, while that of Fig. 13 has $\sigma^2 = 1.1\ Q^2$. It is seen that, as expected, A_iB_i has a large value for at least one i in the high-noise ordering, reaching a value of 60, while for the low-noise ordering $A_iB_i < 2.2$ for all i. Furthermore C_i, which is proportional to the integral of the square of the spectrum from section $(i + 1)$ to output with its maximum normalized to unity, does not vary too much between the two orderings. Finally, note that the spectrum of each section in the low-noise ordering does indeed tend to suppress the peak in the spectrum of the combination of previous sections. Thus the arguments of this section are supported.

VI. AN ALGORITHM FOR OBTAINING A LOW-NOISE ORDERING FOR THE CASCADE FORM

An extensive analysis of roundoff noise in cascade form FIR filters has been presented in this paper and in Ref. 2. 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

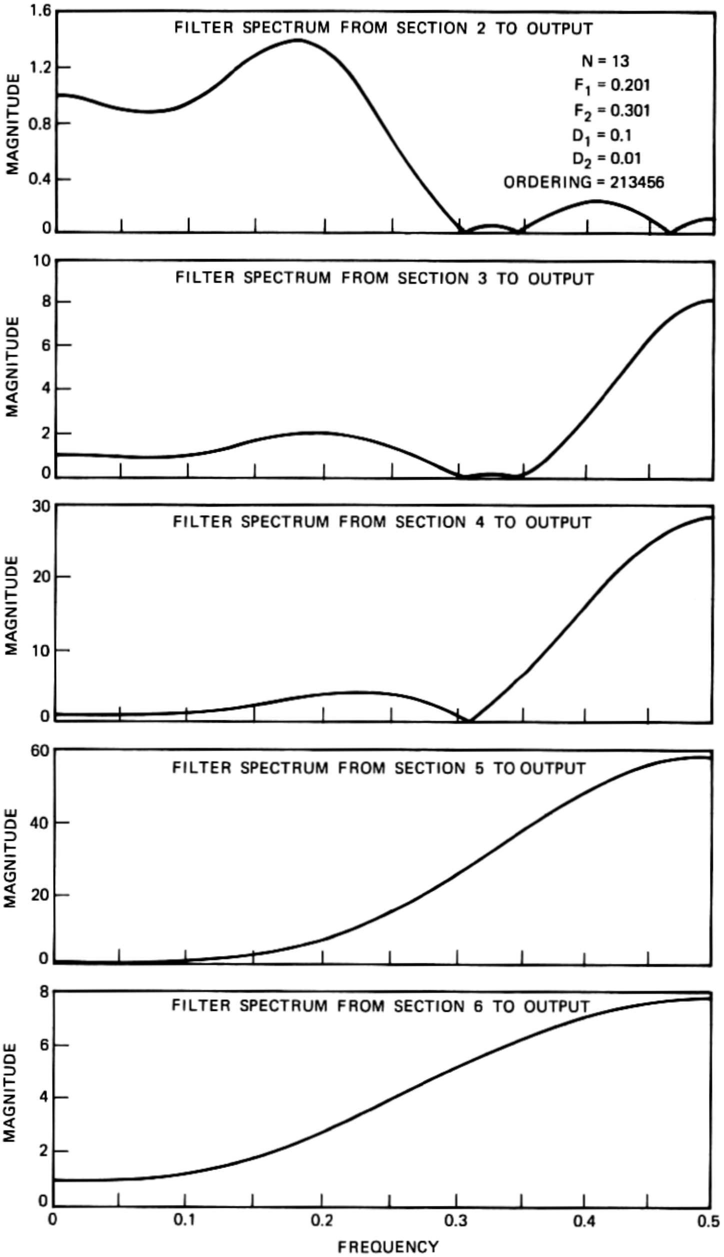


Fig. 12—Series of spectra from section i to the output where i = 2, 3, 4, 5, 6, for a high-noise ordering.

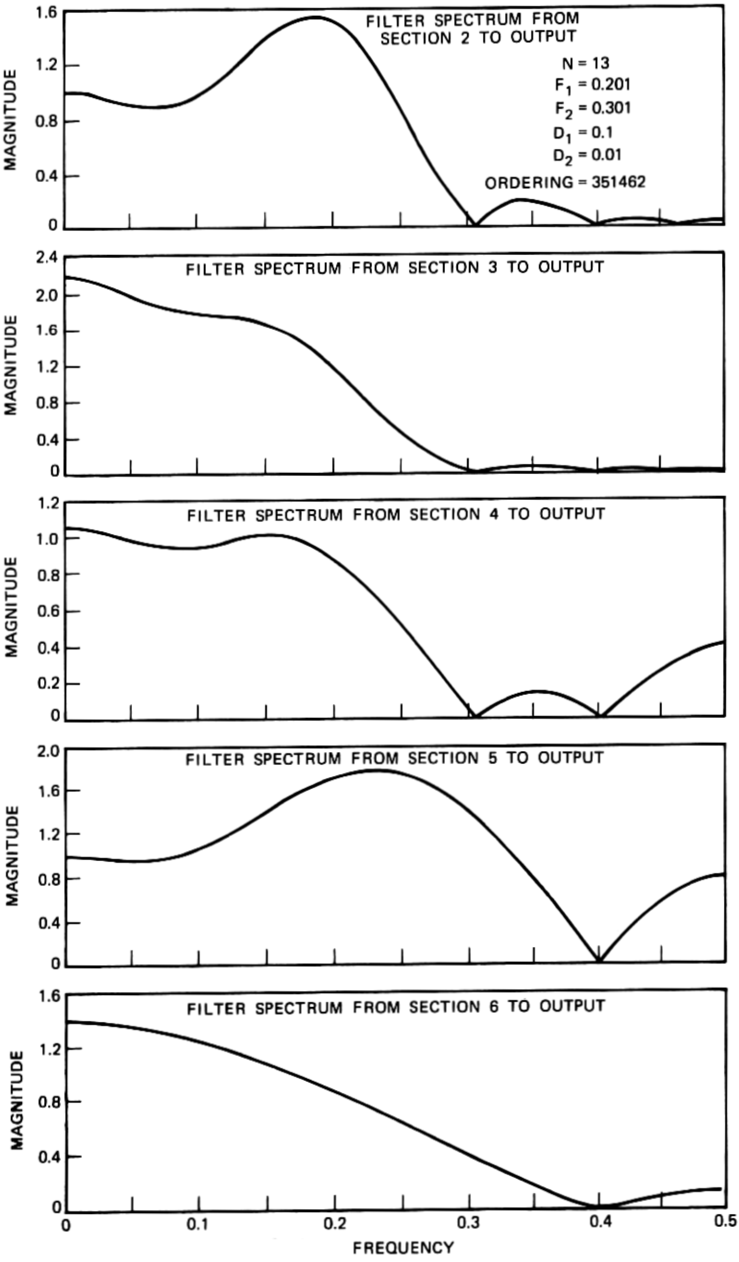


Fig. 13—Series of spectra from section i to the output where i = 2, 3, 4, 5, 6, for a low-noise ordering.

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.[10] His algorithm consists of two major steps; a "preliminary determination" and a "final determination." In this section an algorithm is described for ordering FIR filters which is based upon the procedure used in the "preliminary determination" step of Avenhaus' algorithm. It has been found that a procedure appended to the proposed 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 this 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 III through V, it will be argued heuristically that the proposed algorithm always yields filters which have output noise variances among the lowest possible. Together with extensive experimental confirmation, these arguments provide confidence that the proposed algorithm produces solutions that are very close to the optimum.

Application of Avenhaus' procedure to FIR filters allows the introduction of modifications which reduce significantly the amount of computation time required. Also, while IIR filters seldom require a higher order than the classic 22nd-order bandstop filter quoted by Avenhaus, practical FIR filters can easily require 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, the proposed algorithm has been successfully tested for filters of order up to 128.

6.1 *Description and Discussion of Basic Algorithm*

The basic procedure or algorithm proposed by Avenhaus is simply the following. To order a filter of N_s sections, begin with $i = N_s$ 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 ith section of the cascade. Because in the FIR cascade form noise is injected only into the output of each section, for FIR filters the procedure needs to be modified by considering the output noise due to the section in position $i - 1$ rather than i when choosing a section for position i. But the ith 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 ith section is, where $\{g_i(k)\}$ is the impulse response of the part of the filter from section $(i + 1)$ to the output.

Hence the revised basic algorithm for ordering FIR cascade filters is: *beginning with $i = N_s$, permanently build into position i the section which, together with the sections already built in, causes the smallest possible value for $\sum_k g_{i-1}^2(k)$*. Once this basic algorithm is determined, it is necessary only to 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, $1 \leqq j \leqq N_s - 1$, the search is essentially conducted over only $(j + 1)!$ out of the total of $N_s!$ possible orderings. Now this set of $(j + 1)!$ orderings depends on which sections were chosen for positions $j + 2$ to N_s in the cascade if $j < N_s - 1$. 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 the following arguments is presented in Section V. Let $H(z)$ be an appropriately scaled filter. Given l, $1 \leqq l \leqq N_s - 1$, suppose σ_i^2 is small for all $i \geqq l$. Then the zeros of $\prod_{i=l+1}^{N_s} H_i(z)$ must be well spread around the unit circle in the z-plane since a clustering would cause large peaking in $\prod_{i=k+1}^{N_s} \bar{H}_i(e^{j\omega})$ for some $k \geqq l$, hence a large value of σ_k^2. But this means that the remaining zeros of $H(z)$, namely those in $\prod_{i=1}^{l} H_i(z)$, 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 $\prod_{i=1}^{l} H_i(z)$ which, when assigned to position l, causes little peaking in $\prod_{i=1}^{l-1} \bar{H}_i(e^{j\omega})$ or $\prod_{i=l}^{N_s} \bar{H}_i(e^{j\omega})$, and thus results in a small value for σ_{l-1}^2. By induction, then, σ_i^2 can be chosen small for all i.

For small l it is true that there are very few zeros left as candidates for position l, 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_l^2 reaches a peak for l somewhere in the middle between 1 and N_s, while σ_l^2 for small l has little contribution to σ^2, the total output noise variance. Hence the choice of sections for small l is not too crucial. Of course, the eligible candidates are still well-spaced zeros as for larger l, 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 III that most orderings of a filter have comparatively low noise. Because it is not difficult to find low-noise arrangements of zeros, $\sum \sigma_i^2$ can be minimized approximately by minimizing each σ_i^2 independently, searching over a much smaller domain. If the sum $\sum \sigma_i^2$ could not be segmented, searching for a minimum would be essentially an impossible task because of time limitations.

6.2 *Two Versions of the Complete Algorithm*

Having discussed why the basic algorithm works, the practical problem of implementing it is now discussed. First of all, there is 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 III, sum scaling is to be preferred since it can be carried out the fastest. Figure 14 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.

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 II through IV. 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 Tables III and IV. These are shown in the third to the last column of the tables. The comparability of these noise values to those for sum scaling is a confirmation of the results of Section IV.

For an alternative implementation of the basic ordering algorithm, peak scaling can be used. To distinguish between the two different resulting algorithms, the former (sum scaling) will be referred to as alg. 1 and the latter as alg. 2. The only change needed in Fig. 14 to realize alg. 2 rather than alg. 1 is to replace $\sum_k |f_i(k)|$ by $\max_\omega |F_i(e^{j\omega})|$

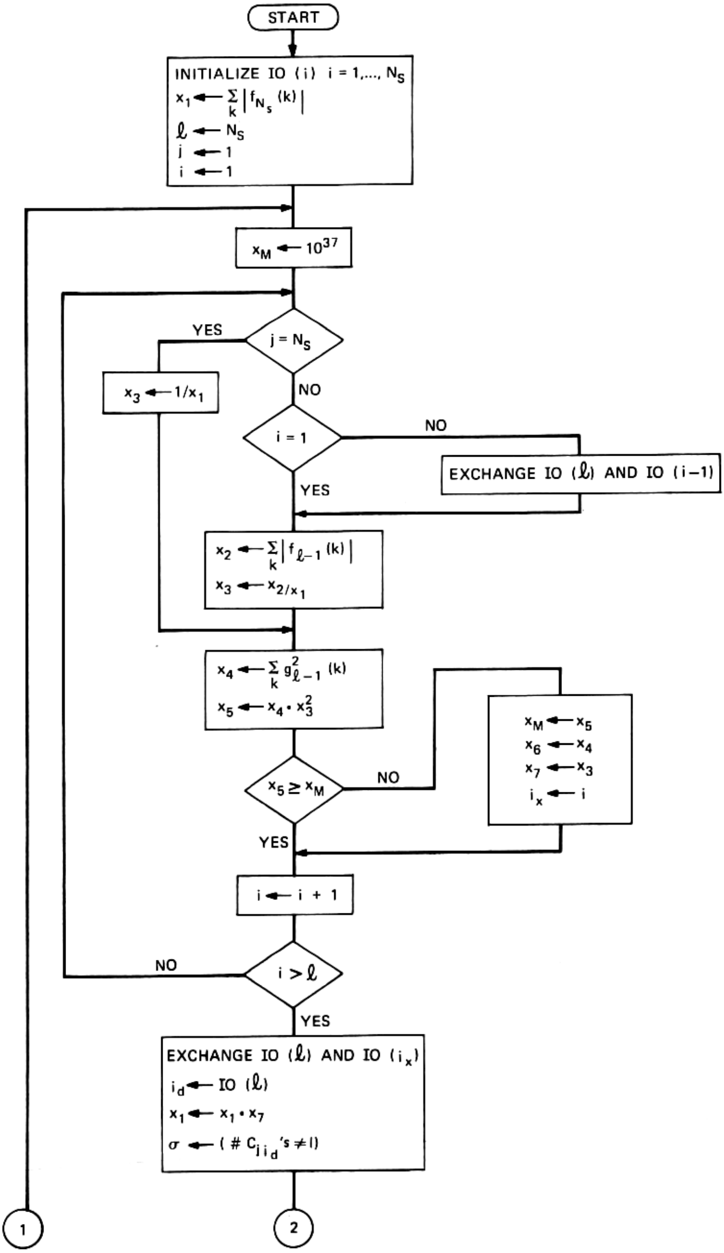
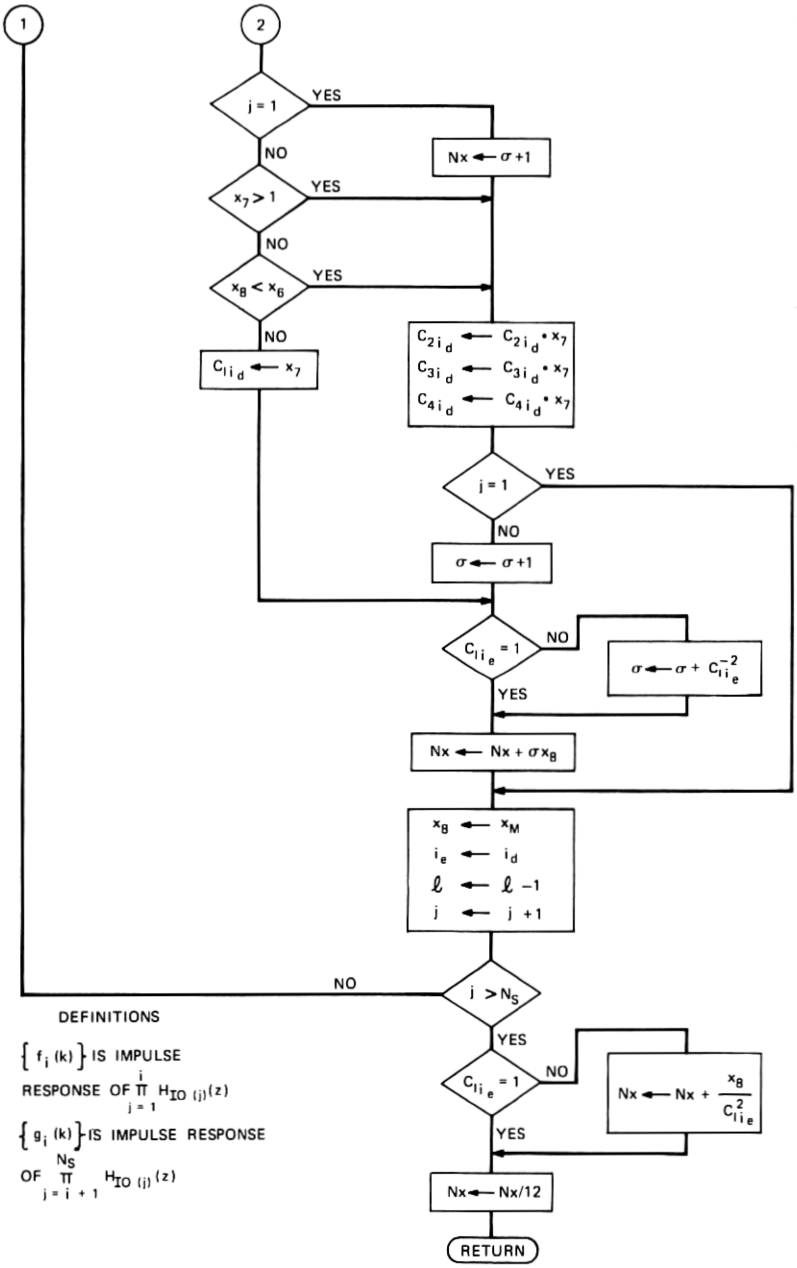

START
INITIALIZE IO (i) $i = 1, \ldots, N_S$
$x_1 \leftarrow \sum_k |f_{N_s}(k)|$
$\ell \leftarrow N_S$
$j \leftarrow 1$
$i \leftarrow 1$
$x_M \leftarrow 10^{37}$
$j = N_S$
YES
NO
$x_3 \leftarrow 1/x_1$
$i = 1$
NO
YES
EXCHANGE IO (ℓ) AND IO (i−1)
$x_2 \leftarrow \sum_k |f_{\ell-1}(k)|$
$x_3 \leftarrow x_{2/x_1}$
$x_4 \leftarrow \sum_k g^2_{\ell-1}(k)$
$x_5 \leftarrow x_4 \cdot x_3^2$
$x_5 \geq x_M$
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)
$x_1 \leftarrow x_1 \cdot x_7$
$\sigma \leftarrow$ (# $C_{j i_d}$'s $\neq$ I)
1
2

Fig. 14—Flow chart of ordering algorithm.


1
2
$j = 1$
YES
$Nx \leftarrow \sigma + 1$
NO
$x_7 > 1$
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
$\sigma \leftarrow \sigma + 1$
$C_{1i_e} = 1$
NO
$\sigma \leftarrow \sigma + C_{1i_e}^{-2}$
YES
$Nx \leftarrow Nx + \sigma x_8$
$x_8 \leftarrow x_M$
$i_e \leftarrow i_d$
$\ell \leftarrow \ell - 1$
$j \leftarrow j + 1$
NO
$j > N_S$
YES
$C_{1i_e} = 1$
NO
$Nx \leftarrow Nx + \frac{x_8}{C_{1i_e}^2}$
YES
$Nx \leftarrow Nx/12$
RETURN

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. 14 (continued).

for given i whenever it appears. Results of using alg. 2 on the filters of Tables III and IV are shown in the next to 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 comparable. Thus, using both alg. 1 and alg. 2, two separate low-noise orderings for a given filter can be obtained. Further discussion of the results will be given shortly.

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 $\{IO(i)\}$ which describes how the filter sections are ordered (see Fig. 14) is to be initialized. In obtaining the results of Tables III and IV, the following has been done: $\sum_k g_i^2(k)$ 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) and $\max_\omega |F_i(e^{j\omega})|$ was determined by transforming $\{f_i(k)\}$ via an FFT and then finding the maximum value. Finally, $\{IO(i)\}$ was initialized to $IO(i) = i$, $i = 1, \cdots, N_s$. It will be seen that these procedures must be modified for higher-order filters. But meanwhile, the implications of these procedures in terms of dependence of computation time on filter length is considered.

Clearly, in algorithmically computing the impulse response of an N-point filter via convolution, the number of multiplies and adds

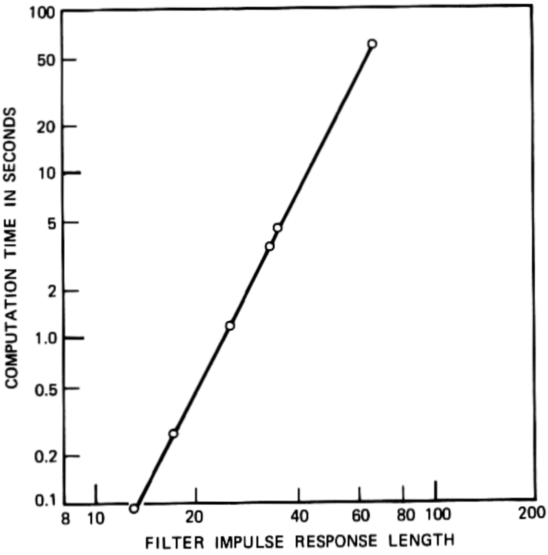


Fig. 15—Computation time versus filter length for ordering algorithm.

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 $\sum_k |f_{l-1}(k)|$ and $\sum_k g_{l-1}^2(k)$ dominates all operations within the inner loop in terms of time required. Since the total number of points that must be evaluated in order to compute $\{f_{l-1}(k)\}$ and $\{g_{l-1}(k)\}$ together turns out to be a constant independent of l, the combined operations must have approximately an N^2 time dependence. Hence it is predicted that the computation time required for alg. 1 must be approximately proportional to N^4. This prediction is verified in Fig. 15, 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.

6.3 *Modification of Algorithm for Higher Filter Orders*

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, it is not difficult to see that this ordering (i.e., $\mathrm{IO}(i) = i$) has a noise variance which is among the highest possible and which increases at least exponentially with N. Thus all attempts at evaluating the impulse response of the filter by simulation in the time domain are marred by roundoff noise.

A natural possibility for resolving this problem is to perform calculations in the frequency domain. This has been tried as a modification to alg. 2. In particular, rather than computing $F_i(e^{j\omega})$ from

$\{f_i(k)\}$, it is evaluated as a product of $H_l(e^{j\omega})$, $l = 1, \cdots, i$, where $H_l(e^{j\omega})$ is computed from the coefficients of section l 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 V 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. The conclusion of Section III–that most orderings of a filter have relatively low noise–means that, if an ordering were chosen at random, it ought to have 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 $\{\mathrm{IO}(i), i = 1, \cdots, N_s\}$ can be easily randomized using the following shuffling algorithm:[11]

```
Step 1: Set $j \leftarrow N_s$.
Step 2: Generate a random number $U$, uniformly distributed between zero and one.
Step 3: Set $k \leftarrow [jU] + 1$. (Now $k$ is a random integer between 1 and $j$.) Exchange $\mathrm{IO}(k) \leftrightarrow \mathrm{IO}(j)$.
Step 4: Decrease $j$ by 1. If $j > 1$, return to Step 2.
```

By adding a step to randomize the initial ordering $\mathrm{IO}(i) = i$ 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 in a relative sense, but, as will shortly be seen, a random ordering is far from being 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. Further results are described in the next section.

6.4 *Discussion of Results*

Note in Table II that alg. 1 can result in a noise variance which is very close to the minimum, if not the minimum. From this observation and the conclusions of Section III on the dependence of the minimum

TABLE V—LIST OF FILTERS AND THE RESULTS OF ALG 1′

					Noise Variance			
$D_1 = 0.01$					Ordering		Alg 1′	
#	N	N_p	F_1	D_2	Sequential	Random	Sum sc	Peak sc
38	33	9	0.244	0.001	1.0×10^{11}	6.2×10^{3}	4.59	7.81
67	47	12	0.237	0.001	4.3×10^{17}	2.2×10^{6}	6.47	12.07
68	67	17	0.242	0.001	3.3×10^{27}	1.5×10^{6}	16.77	30.03
69	101	25	0.241	0.001	$> 10^{38}$	1.4×10^{5}	41.93	73.55
70	129	20	0.153	0.0001	—	5.5×10^{11}	17.98	37.54

noise variance for a filter on different parameters, we can be quite confident that the noise variances shown in Tables III and IV are also very close to the minimum possible. The filters of Tables III and IV were chosen intentionally to show the dependence of the results of the ordering algorithms on various transfer function parameters. It is seen that the noise variances indeed behave in the way that would be expected from the results of Section III. 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 III and IV are then a confirmation of the expectation that the conclusions of Section III on the general dependence of noise on transfer function parameters can be generalized to higher-order filters.

The results of using the modified alg. 1 (denoted alg. 1′) on a few filters are shown in Table V. Also shown in this table, for comparison, are the noise variances of these filters when they are in the sequential ordering $\mathrm{IO}(i) = i$ (where computable within the numerical range of the computer) as well as when they are in a random ordering (obtained by randomizing $\{\mathrm{IO}(i)\}$ where $\mathrm{IO}(i) = i$, 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 $H_l(e^{j\omega})$ is evaluated via an FFT, peak scaling is then performed, and finally σ_i^2 is computed via $1/(2\pi) \int_0^{2\pi} |G_i(e^{j\omega})|^2 d\omega$ rather than $\sum_k g_i^2(k)$.

From Table V it is seen 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

(or 1′) to find proper orderings for filters in cascade form. In all the examples given in Tables II through V, one can do little better in trying to find orderings with lower noise. With the possible exception of the uninteresting wideband filter, number 54 in Table III, all filters have less than 4 bits of noise after ordering by alg. 1, while the great majority have less than 3 bits. Thus it is not expected that these noise figures can be further reduced by much more than a bit or so.

Finally, in practice, cascade FIR filters of orders over approximately 50 are generally of little interest since there exist more efficient ways than the cascade form to implement filters of higher orders. For filters of, at most, 50th order, the computation time required for alg. 1 is less than 20 seconds on the Honeywell 6070 computer. Thus alg. 1 (or 1′) is a very efficient means for ordering cascade filters.

VII. SUMMARY

In this paper, experimental results have been presented which show that most orderings for an FIR filter in cascade form have very low noise relatively, and that the shape of the distribution of noise with respect to ordering is essentially independent of transfer function parameters as well as method of scaling (sum or peak). An explanation of these properties has been proposed, based on a characterization of high-noise and low-noise orderings. Furthermore, the dependence of noise on transfer function parameters and scaling has been investigated. These results point to an algorithm for ordering cascade FIR filters which has been implemented and tested for filters with a wide range of values of transfer function parameters. In every case, the algorithm gave results within expectation which were deduced to be very close to the optimum. Justification for the success of the algorithm has also been given.

REFERENCES

1. Chan, D. S. K., "Roundoff Noise in Cascade Realization of Finite Impulse Response Digital Filters," S. B. and S. M. Thesis, Dept. of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, September, 1972.
2. Chan, D. S. K., and Rabiner, L. R., "Theory of Roundoff Noise in Cascade Realizations of Finite Impulse Response Digital Filters," B.S.T.J., this issue, pp. 329–345.
3. Parks, T. W., and McClellan, J. H., "Chebyshev Approximation for Non-Recursive Digital Filters with Linear Phase," IEEE Trans. Circuit Theory, *CT-19* (March 1972), pp. 189–194.
4. Jackson, L. B., Kaiser, J. F., and McDonald, H. S., "An Approach to the Implementation of Digital Filters," IEEE Trans. Audio Electroacoustics, *AU-16*, No 3 (September 1968), pp. 413–421.

5. Schafer, R. W., "A Survey of Digital Speech Processing Techniques," IEEE Trans. Audio Electroacoustics, *AU-20*, No. 4 (March 1972), pp. 28–35.
6. Schafer, R. W., and Rabiner, L. R., "Design and Simulation of a Speech Analysis Synthesis System Based on Short-Time Fourier Analysis," IEEE Trans. Audio Electroacoustics, June 1973, in preparation.
7. Schüssler, W., private communication.
8. Schüssler, W., "On Structures for Nonrecursive Digital Filters," Archiv Für Electronik Und Übertragungstechnik, *26*, 1972.
9. Jackson, L. B., "On the Interaction of Roundoff Noise and Dynamic Range in Digital Filters," B.S.T.J., *49*, No. 2 (February 1970), pp. 159–184.
10. Avenhaus, E., "Realizations of Digital Filters with a Good Signal-to-Noise Ratio," Nachrichtentechnische Zeitscrift, May 1970.
11. Knuth, D. E., *Seminumerical Algorithms*, Vol. 2 of *The Art of Computer Programming*, Reading, Massachusetts: Addison-Wesley, 1969, p. 125.