

# Design of Finite Impulse Response Digital Filters with Nonlinear Phase Response

ELI GOLDBERG, ROBERT KURSHAN, MEMBER, IEEE, AND DAVID MALAH, MEMBER, IEEE

**Abstract**—Most of the existing literature on FIR digital filters is concerned with linear-phase (LP) filters. However, several papers have appeared on the subject of nonlinear-phase (NLP) filters, mainly proposing methods for designing minimum-phase filters, or approximating a desired phase response. In this paper, an investigation is made of one such method, based upon a selection of zeros from a prototype LP filter. It is shown that with respect to *minimizing the order* of a filter subject to given gain response specifications, this is the most efficient method for designing FIR filters. Coefficient quantization error is analyzed for filters generated by this method. A practical comparison is given between the resulting filter and the corresponding minimal order LP filter. It is shown that while most LP filters can be implemented more efficiently than NLP filters by taking into account the symmetry of their coefficients, for filters with very wide passband and for certain special purpose filters such as CCD and those used for filtering a delta-modulated or ADPCM signal, an NLP implementation is usually more efficient. In addition, an alternate design algorithm is proposed for NLP filters which decreases ripple magnitude. The resulting filters, while not of minimal order, can be efficiently implemented by decomposing the filter into LP stopband and NLP passband sections, which is especially attractive for narrow passbands.

## I. INTRODUCTION

FINITE impulse response (FIR) digital filters have been extensively studied in the past years for several reasons. Among these are the properties of stability, high-speed implementation using the FFT or other number-theoretic transforms, and optional exact linear-phase. Most of the work done concentrated on linear-phase (LP) filters, mainly because the optimizing methods for designing FIR filters [1], [2] are based on the assumption that the filter coefficients are symmetrical (or antisymmetrical) and this implies that the filter is LP. However, some research has been done on nonlinear-phase (NLP) digital filters. Holt *et al.* [3] demonstrated how to use the Remez exchange algorithm in order to approximate both amplitude and phase response. The amplitude response performance of the resulting filter is somewhat worse [4] than that of optimal (i.e., minimal order) LP filters. Herrman *et al.* [5] showed how to design a minimum-phase FIR filter, using a prototype LP filter and selecting half of its zeros. Another approach for designing a minimum-phase filter was suggested by Burriss [6], using the relation between gain and phase re-

sponse of minimum-phase filters. It appeared that minimum-phase filters, designed using the last approach, needed fewer coefficients than optimal LP filters with the same gain response specifications. However, this was not proved.

In this paper, some aspects of a design method similar to the one proposed by Herrman *et al.* [5] are discussed. First, the prototype LP filter is designed using more efficient design methods than the one used in [5]. Then, a discussion is given on the problem of defining the specifications on the prototype LP filter. Next, the performance of the resulting NLP filter is theoretically and practically compared to that of the optimal (minimal order) LP filter. Finally, some practical limitations in using the proposed design method are brought up, and an alternative design method is proposed in order to avoid these limitations.

## II. DESIGN OF NONLINEAR-PHASE FIR DIGITAL FILTERS

Let  $H(z)$  be the transfer function of an NLP FIR digital filter of order  $m$  (or length  $m + 1$ ):

$$H(z) = \sum_{n=0}^m h(n) z^{-n}. \quad (1)$$

Define an FIR filter whose transfer function is

$$\hat{H}(z) \triangleq H(z)H(z^{-1})z^{-m} = \sum_{n=0}^{2m} \hat{h}(n) z^{-n} \quad (2)$$

where

$$\hat{h}(k) = \hat{h}(2m - k) = \sum_{j=0}^k h(j)h(j + m - k), \quad k = 0, 1, \dots, m. \quad (3)$$

Equation (3) implies that  $\hat{H}(z)$  is a linear-phase filter since its coefficients are symmetrical. The frequency response of  $\hat{H}(z)$  can be described as

$$\begin{aligned} \hat{H}(e^{j\omega}) &= H(e^{j\omega}) \cdot H(e^{-j\omega}) \cdot e^{-j\omega m} \\ &= |H(e^{j\omega})|^2 \cdot e^{-j\omega m} = \hat{P}(e^{j\omega}) \cdot e^{-j\omega m} \end{aligned} \quad (4)$$

where  $\hat{P}(e^{j\omega})$  is a real, nonnegative function.

Let us examine the connection between  $H(z)$  and  $\hat{H}(z)$  from the  $z$ -plane point of view. From (2) it is clear that if  $z_i$  is a zero of  $H(z)$ , then both  $z_i$  and  $z_i^{-1}$  are zeros of  $\hat{H}(z)$ . An example of a possible zero positioning for  $H(z)$  and  $\hat{H}(z)$ , which corresponds to a low-pass filter, is shown in Fig. 1 where a circled dot represents a double zero. It is easy to

Manuscript received June 30, 1980; revised March 11, 1981.

E. Goldberg is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa, Israel.

R. Kurshan is with the Mathematics and Statistics Research Center, Bell Laboratories, Murray Hill, NJ 07974.

D. Malah is with the Department of Acoustics Research, Bell Laboratories, Murray Hill, NJ 07974, on leave from the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa, Israel.

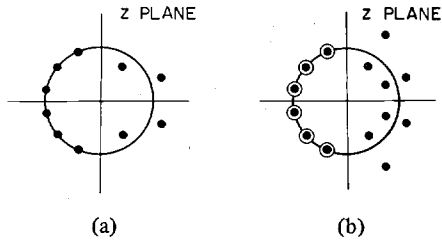


Fig. 1. Positioning of zeros for (a)  $H(z)$ , and (b)  $\hat{H}(z)$ .

see that if  $z_i$  is a zero of  $\hat{H}(z)$ , so are  $z_i^{-1}$ ,  $z_i^*$ , and  $(z_i^*)^{-1}$  ( $z^*$  denotes the complex conjugate of  $z$ ), consistent with the fact that  $\hat{H}(z)$  is a linear-phase filter. Notice that all of the zeros of  $\hat{H}(z)$  on the unit circle appear in a multiplicity of two. This is necessitated by the fact that  $\hat{P}(e^{j\omega})$  cannot be negative, and therefore cannot intersect the  $\omega$ -axis but only be tangent to it.

From the previous discussion, the following algorithm can be used for designing an NLP FIR digital filter  $H(z)$  of order  $m$  (i.e., whose length is  $m+1$ ) and whose gain response  $|H(e^{j\omega})|$  approximates  $P(e^{j\omega})$  for some given  $P$ .

a) Design an LP FIR digital filter  $\hat{H}(z)$  of order  $2m$  whose gain response  $\hat{P}(e^{j\omega})$  approximates  $P^2(e^{j\omega})$  (see below).

b) Find the zeros of  $\hat{H}(z)$  [since the zeros are in reciprocal and conjugate pairs, once a zero on (off) the unit circle is known, the order of the problem may be reduced by 2 (4)].

c) Pick one zero from each pair of reciprocal zeros of  $\hat{H}(z)$  and replace each double zero on the unit circle with a single zero, thereby defining  $H(z)$ . (If minimum phase filter is needed, the zeros inside and on the unit circle are selected.)

d) Determine the unit sample response  $\{h(n)\}_{n=0}^m$  satisfying  $H(z) = \sum_{n=0}^m h(n) z^{-n}$ .

An alternative to steps b), c), and d) is to solve the set of  $m+1$  quadratic equations (3). Unfortunately, no analytic solution has been found to this problem.

### III. SELECTION OF DESIGN SPECIFICATIONS

When executing step a) in the proposed algorithm, it is necessary to properly define the specifications of the prototype LP filter, so that the desired specifications of the NLP filter are met. Generally, the way of doing it depends on the particular method used for designing the prototype LP filter. In this section, two of the known optimal design methods are considered, the Remez type exchange algorithm [7] and the use of extremal polynomials [2]. The first is preferable over the method used in [5], as it permits the designer to specify the band edges. Moreover, currently it can be implemented more efficiently. The second is equivalent to the first, but permits control of behavior in transition bands [2].

Our discussion is restricted to multipassband/stopband filters. However, similar considerations can be applied for different shapes of gain response.

#### A. The Remez Type Exchange Algorithm

Consider the bandpass filter shown in Fig. 2. Fig. 2(a) shows the gain response function  $P(e^{j2\pi f})$  of an equiripple LP filter whose desired value in the  $i$ th band is  $D_i$  and whose deviation from the desired value in the  $i$ th band is  $\delta_i$ . Fig. 2(b) shows

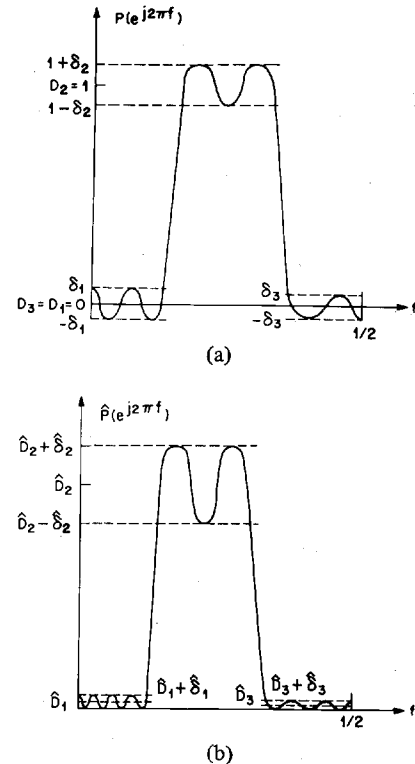


Fig. 2. Gain response function of (a) ordinary and (b) prototype optimal linear-phase bandpass filter.

the gain response function  $\hat{P}(e^{j2\pi f})$  of the prototype LP filter which has to be designed, so that the specifications on the resulting NLP filter are identical to those of the filter of Fig. 2(a). The sets  $\{\hat{D}_i\}$  and  $\{\hat{\delta}_i\}$  can easily be derived from  $\{D_i\}$  and  $\{\delta_i\}$ . For example, if the  $i$ th band is a passband, then we normalize  $D_i = 1$  and therefore

$$\begin{aligned} (1 + \delta_i)^2 &= \hat{D}_i + \hat{\delta}_i \\ (1 - \delta_i)^2 &= \hat{D}_i - \hat{\delta}_i \end{aligned} \quad (5)$$

and hence

$$\begin{aligned} \hat{\delta}_i &= 2\delta_i \\ \hat{D}_i &= 1 + \delta_i^2 \approx 1. \end{aligned} \quad (6)$$

On the other hand, if the  $i$ th band is a stopband, then

$$\hat{D}_i = \frac{1}{2} \delta_i^2 = \hat{\delta}_i \quad (7)$$

so that the gain response of the resulting NLP filter in the  $i$ th band is bounded between zero and

$$\delta_i = (2\hat{\delta}_i)^{1/2}.$$

When using the Remez exchange algorithm directly, the set of deviations  $\{\delta_i\}$  is the "output" of the design procedure, while the "input" design parameter is a weighting function  $W(e^{j2\pi f})$  (or, in the case of multiband filters, a set of weighting values  $\{W_i\}$ ). However, it is not difficult to use "trial-and-error" methods [8] in order to have more flexibility in selecting the input design parameters. On the other hand, when designing NLP filters using the proposed methods, the set of deviations  $\{\delta_i\}$  must be included among the input design parameters. For example, suppose we want

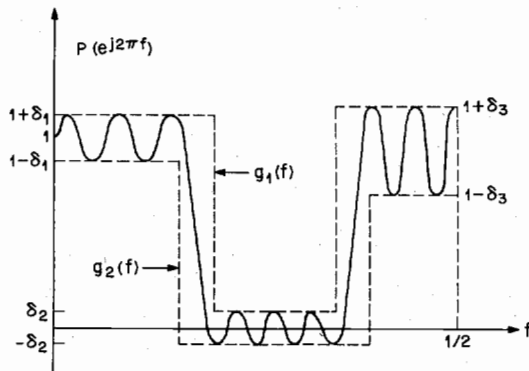


Fig. 3. Design of FIR digital filter using extremal polynomials.

to design an NLP filter whose weighting values are  $W_1$  and  $W_2$  (so  $\delta_1/\delta_2 = W_2/W_1$ ), with the weighting values of the prototype LP filter denoted as  $\hat{W}_1$  and  $\hat{W}_2$ . Without loss of generality assume that  $\hat{W}_1 = 1$ . Then  $\hat{W}_2$  can be derived as follows.

$$\hat{W}_2 = \frac{\delta_1}{\delta_2} = \frac{2\delta_1}{\frac{1}{2}\delta_2^2} = 4 \frac{\delta_1}{\delta_2} \cdot \frac{1}{\delta_2} = 4 \cdot \frac{W_2}{W_1} \cdot \frac{1}{\sqrt{2\delta_2}} = f(\delta_2). \quad (8)$$

If  $\delta_2$  is not known before the design of  $\hat{P}(e^{j2\pi f})$  is carried out, then  $\hat{W}_2$  cannot be derived and  $\hat{P}(e^{j2\pi f})$  cannot be designed.

### B. The Use of Extremal Polynomials

When using this method for designing an LP filter, one has to define two functions,  $g_1(f)$  and  $g_2(f) < g_1(f)$  (see Fig. 3). The gain response function  $P(e^{j2\pi f})$  oscillates between these two functions. When an NLP filter has to be designed, it is clear that the function  $\hat{P}(e^{j2\pi f})$  has to be bounded by the functions  $\hat{g}_1(f)$  and  $\hat{g}_2(f)$ , where

$$\hat{g}_1(f) = g_1^2(f)$$

$$\hat{g}_2(f) = \begin{cases} 0, & g_2(f) \leq 0 \\ g_2^2(f), & g_2(f) > 0. \end{cases} \quad (9)$$

## IV. ON THE EFFICIENCY OF NLP DIGITAL FILTERS

An interesting question which may be raised concerning the proposed design algorithm is how the order of the resulting NLP filter is related to the order of other kinds of FIR filters (LP or NLP), having the same gain response specification. In this section, this question is considered. A weighting function  $W$  is used in the next lemma to allow for different error amplitudes in different bands.

**Lemma 1:** Let  $H_1(z)$  and  $H_2(z)$  be the transfer functions of two LP filters defined on the same frequency band  $B$ . Associated with each filter is a weighting function  $W(e^{j2\pi f})$ . Suppose the order  $N_i$  of  $H_i$  is even and denote the gain response of  $H_i$  by  $P_i(e^{j2\pi f})$  ( $i = 1, 2$ ). Define the (Chebyshev) error in the  $i$ th filter with respect to a desired gain response  $D$  as

$$E_i = \max_{f \in B} \{ |W(e^{j2\pi f})| |P_i(e^{j2\pi f}) - D(e^{j2\pi f})| \}. \quad (10)$$

If  $E_1$  is minimal for the given  $N_1$  then

$$N_1 > N_2 \Rightarrow E_1 < E_2.$$

*Proof:* Express  $P_1$  and  $P_2$  as cosine sums:

$$P_1(e^{j2\pi f}) = \sum_{k=0}^{N_1/2} a(k) \cos 2\pi f k$$

$$P_2(e^{j2\pi f}) = \sum_{k=0}^{N_2/2} b(k) \cos 2\pi f k. \quad (11)$$

Assuming  $N_1 > N_2$ ,  $P_2$  can also be described as

$$P_2(e^{j2\pi f}) = \sum_{k=0}^{N_1/2} b(k) \cos 2\pi f k \quad (12)$$

where  $b(k) = 0$  for  $N_2/2 < k \leq N_1/2$ .

Since the set  $\{a(k)\}_{k=0}^{N_1/2}$  minimizes the Chebyshev error  $E_1$ , this set uniquely satisfies (10) [7]. As  $a(N_1/2) \neq 0 = b(N_1/2)$ , it must follow that  $E_1 < E_2$ .

**Theorem 1:** Among all FIR filters defined on the same region of frequencies  $B$  and having the same gain response specifications, the NLP filter designed using the algorithm proposed in Section II has the smallest possible order.

*Proof:* Let  $H_1(z)$  and  $H_2(z)$  be the transfer functions of two FIR filters having the orders  $N_1$  and  $N_2$ , respectively, and suppose that  $H_1(z)$  is designed using the proposed algorithm. The following filters are next defined.

$$\hat{H}_1(z) = z^{-N_1} H_1(z) \cdot H_1(z^{-1})$$

$$\hat{H}_2(z) = z^{-N_2} H_2(z) \cdot H_2(z^{-1}). \quad (13)$$

As shown before,  $\hat{H}_1(z)$  and  $\hat{H}_2(z)$  are both LP filters having the orders  $\hat{N}_1 = 2N_1$  and  $\hat{N}_2 = 2N_2$ , respectively. If  $H_1(z)$  and  $H_2(z)$  have the same gain specifications, so have  $\hat{H}_1(z)$  and  $\hat{H}_2(z)$ . Suppose now that  $N_2 < N_1$  (and thus  $\hat{N}_2 < \hat{N}_1$ ), and design an optimal LP filter  $\tilde{H}_2(z)$  whose length is  $\hat{N}_2$ , assuming the same desired specifications and weighting functions as for  $\hat{H}_1$  and  $\hat{H}_2$ . The resulting Chebyshev error of  $\tilde{H}_2(z)$  will be denoted as  $\tilde{E}$ . Since both  $\hat{H}_1(z)$  and  $\tilde{H}_2(z)$  are optimal filters, and since  $\hat{N}_2 < \hat{N}_1$ , it follows from Lemma 1 that

$$\tilde{E} > \hat{E} \quad (14)$$

where  $\hat{E}$  is defined as the Chebyshev error of  $\hat{H}_1$  and  $\hat{H}_2$ .

$\hat{H}_2(z)$  and  $\tilde{H}_2(z)$  are both linear-phase filters of the same order, having the same desired and weighting functions. Since  $\tilde{H}_2(z)$  is optimal,

$$\tilde{E} \leq \hat{E}. \quad (15)$$

Clearly, there is a contradiction between (15) and (14). We must, therefore, assume that

$$N_2 \geq N_1.$$

Two important conclusions are deduced from Theorem 1. The first one is that as far as the filter's order is concerned, the proposed method produces the most efficient NLP filters. The resulting filter will, therefore, be referred to as an "optimal" NLP filter. The second conclusion is that optimal NLP filters need strictly fewer coefficients than optimal LP

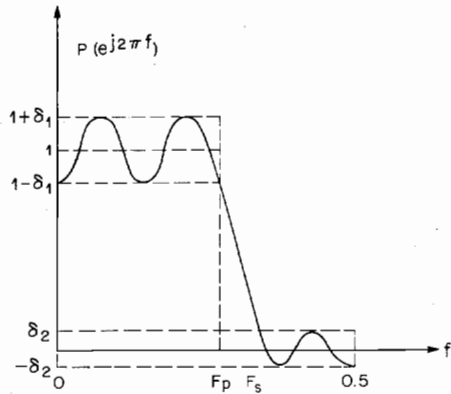


Fig. 4. Gain response of optimal linear-phase low-pass filter.

filters with the same gain response specifications (for order  $N > 2$ ). This conclusion follows from Theorem 1 and the algorithm and is intuitively understood since NLP filters are free of the constraint of having symmetrical coefficients. The question is how much we can reduce the filter's order by giving up the phase linearity.

In order to answer this question, the following measure is defined with respect to given gain specifications.

$$E_R \triangleq \frac{N_{NLP}}{N_{LP}} \quad (16)$$

where  $N_{NLP}$  and  $N_{LP}$  are the respective lengths of optimal NLP and LP filters having the same gain specifications. Since the relations between optimal LP filter parameters are generally unknown,  $E_R$  is computed only for the case of low-pass filter, for which good approximated relationships are known [9] between the filter's parameters (see Fig. 4). For a given set of specifications,  $E_R$  is given by the following equation.

$$E_R = \frac{1}{2} [N(2\delta_1, \frac{1}{2}\delta_2^2, \Delta F) + 1] / N(\delta_1, \delta_2, \Delta F) \quad (17)$$

where  $\Delta F = F_s - F_p$  is the transition width and  $N(\delta_1, \delta_2, \Delta F)$  is the length of the optimal LP low-pass filter whose parameters are  $\delta_1$ ,  $\delta_2$ , and  $\Delta F$ . Equation (17) is derived directly from (16) using the results of the previous section.

The filter length  $N(\delta_1, \delta_2, \Delta F)$  was found in [9] empirically, for  $\delta_1 \geq \delta_2$ . While the region of interest for determining the minimum value of  $E_R$  corresponds to the case when  $\delta_1 \geq \delta_2$ , the error caused by using the equation for  $N(\delta_1, \delta_2, \Delta F)$  given in [9] with  $\delta_1 < \delta_2$  is quite small (within a few percent) [14]. The curves in Fig. 5 show the computed values in both regions.

Three types of curves are derived and depicted in Fig. 5. Fig. 5(a), parts 1 and 2 show  $E_R$  plotted against stopband rejection in dB and transition width  $\Delta F$ , respectively, for the case of narrow-band filters ( $F_p \rightarrow 0$ ), Fig. 5(c), parts 1 and 2 for the case of wide-band filters ( $F_s \rightarrow 0.5$ ), and Fig. 5(b), parts 1 and 2 for "ordinary" filters (where  $F_p$  and  $F_s$  are not near the edges). This partition results from the fact that different relations between the filter's parameters exist for each of these three types of filters [9]. Part 1 of (a), (b), and (c) show  $E_R$  as a function of the stopband rejection  $DL_2$  for varying passband rejections  $DL_1$ , where the passband and stopband rejections are defined as

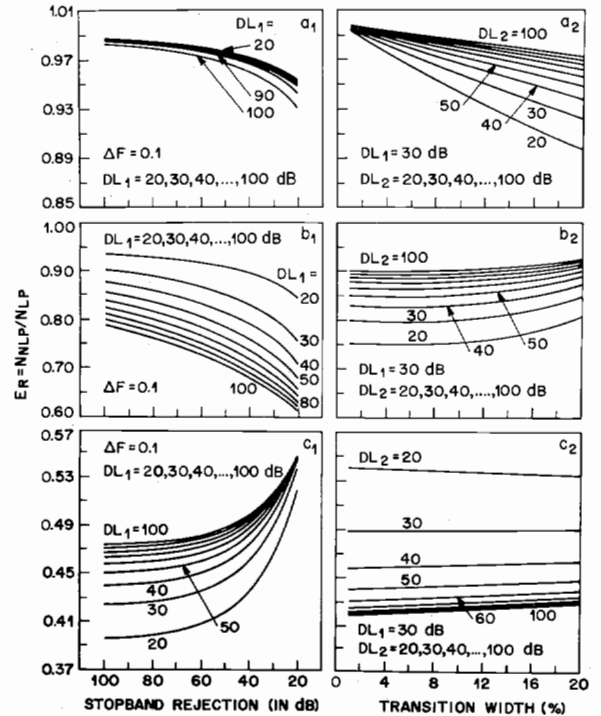


Fig. 5. Dependence of  $E_R$  on the filter parameters for a low-pass filter.

$$DL_i = -20 \log_{10}(\delta_i), \quad i = 1, 2. \quad (18)$$

In part 2 of (a), (b), and (c),  $E_R$  is depicted as a function of the transition width  $\Delta F$ , for varying values of  $DL_2$ . An examination of Fig. 5 reveals the following observations.

1) For a wide range of parameters ( $0.01 \leq \Delta F \leq 0.2$ ,  $0.00001 \leq \delta_2 \leq \delta_1 \leq 0.1$ ),  $E_R$  is bounded by

$$0.4 < E_R < 1. \quad (19)$$

2) As  $F_p$  increases (from 0 to 0.5),  $E_R$  decreases from 1 to about 0.4.

3) The dependence of  $E_R$  on the transition width  $\Delta F$  is very weak.

4) Except for the case of wide-band filters,  $E_R$  tends to decrease with decreasing stopband rejection and increasing passband rejection. The behavior for wide-band filters is reversed.

The fact that  $E_R < 1$  is predicted by Theorem 1. Let us explain the behavior of  $E_R$  at its low end. Let  $H$  and  $H_1$  be the transfer functions of an optimal NLP and an optimal LP low-pass filter, respectively, each having the same gain response specifications  $P$ ,  $\delta_1$ ,  $\delta_2$  and of respective orders  $N$  and  $N_1$ . Then  $E_R = (N + 1)/(N_1 + 1)$ . The LP filter with transfer function

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

and order  $\hat{N} = 2N$  has gain response  $P$  with deviations  $2\delta_1$  and  $\delta_2^2/2$  with respect to which  $\hat{N}$  is minimal for LP filters, as has been demonstrated. Since filter order decreases monotonically with increasing deviation, in comparing the LP  $H_1$  with the LP  $\hat{H}$ , the effect of relaxing  $\delta_1$  to  $2\delta_1$  would tend to cause  $N_1 > \hat{N}$  while the effect of contracting  $\delta_2$  to  $\delta_2^2/2$  (for  $\delta_2 \ll 1$ ) would tend to cause  $N_1 < \hat{N}$ . As relaxation in the passband proceeds linearly while contraction in the stopband proceeds

geometrically, one would expect the latter effect to dominate, with the result that  $N_1 \leq \hat{N}$  or  $E_R = (N+1)/(N_1+1) = (\frac{1}{2}) \cdot (\hat{N}+2)/(N_1+1) \geq \frac{1}{2}$ . However, the effect on the order of changing the deviation bound  $\delta_i$  is also proportional to the bandwidth of the  $i$ th band [9]. Thus, for a very wide passband, the relaxation of  $\delta_1$  to  $2\delta_1$  dominates, with the result that  $N_1 > \hat{N}$  and  $E_R < \frac{1}{2}$ . One would expect the same principle to apply to multipassband filters.

Because of the symmetry  $h_k = h_{N-k}$  in the coefficients  $h_k (k=0, \dots, N)$  of an LP filter of order  $N$  (and the lack of symmetry in the coefficients of an NLP filter), an LP filter of order  $2N$  can be implemented through the adjustment of  $N$  parameters, as can an NLP filter of order  $N$ . Thus, when  $E_R < \frac{1}{2}$ , NLP implementation is more efficient than LP implementation. When  $E_R > \frac{1}{2}$ , LP implementation is more efficient than NLP, provided the LP coefficient symmetry is exploited. In this case the "complexity" of an LP filter of order  $2N$  compares with that of an NLP filter of order  $N$ . However, in certain cases, this coefficient symmetry cannot effectively be exploited. For example, when an LP filter is utilized with a delta-modulated signal [11] or with ADPCM [12], there is no term-by-term multiplication as in the above convolution, but, rather, sums of coefficients are added, corresponding to the same incremental term ( $\pm 1$ ). Also, in implementing a CCD filter one is concerned with minimizing the filter length so that the effect of charge transfer inefficiency may be reduced. In such cases the efficiency of an LP filter of order  $N$  compares with that of an NLP filter of order  $N$ , and from this point of view the NLP implementation is more efficient so long as  $E_R < 1$ .

The conclusion is that with regard to filter length, given gain specifications can be realized more efficiently by LP filters than by NLP filters in most cases. However, for very wide passband filters or filters in which coefficient symmetry is absent or cannot be exploited, such as CCD filters or special purpose implementations for delta-modulated or ADPCM signals, NLP filters are more efficient.

## V. COEFFICIENT QUANTIZATION ERROR IN LP VERSUS NLP FILTERS

The NLP design algorithm provides no intrinsic guidelines for choosing among the zeros of nonunity modulus. In this section the coefficient quantization error as a function of the choice of zeros is analyzed and compared to that of the optimal LP filter for the same gain response specifications.

Since the direct form realization is the most commonly used for FIR filters, it will also be assumed when making the comparison. In this case, what is of main interest is the distortion in gain response caused by coefficient quantization. In [10] it was shown that for each frequency, the change in gain response is a random value (approximately Gaussian) whose maximal value, as well as its variance, is similar for LP and NLP filters having the same order and quantization step. What is left to be checked is mainly the dependence of this error on the specific zero selection. In order to do this, it is convenient to define a measure which characterizes this error for each filter. This is done as follows.

Let  $D(e^{j2\pi f})$  be the desired amplitude response of an FIR

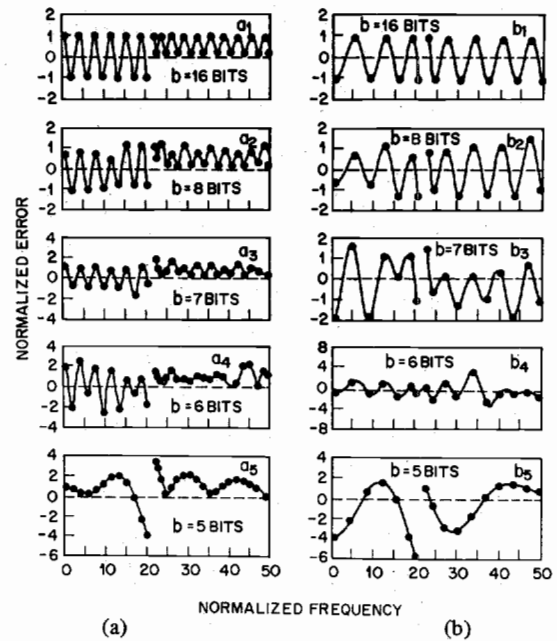


Fig. 6. Normalized error as function of normalized frequency for optimal FIR filters with varying coefficient wordlength  $b$  and the following parameters:  $N = 31$ ,  $F_p = 0.2$ ,  $\delta_1 = \delta_2 = 0.1$ . (a) Nonlinear-phase filter. (b) Linear-phase filter.

filter, and  $H(e^{j2\pi f})$  its frequency response, which can be described for the case of an LP filter as in [10]

$$H(e^{j2\pi f}) = \bar{H}(e^{j2\pi f}) e^{-j2\pi f(N/2)} \quad (20)$$

where  $\bar{H}$  is a positive function. The following  $\bar{H}$  can then be defined.

$$E(e^{j2\pi f}) = \begin{cases} |H(e^{j2\pi f})| - D(e^{j2\pi f}) & \text{for NLP filters} \\ \bar{H}(e^{j2\pi f}) - D(e^{j2\pi f}) & \text{for LP filters.} \end{cases} \quad (21)$$

For each value of  $f$ ,  $E(e^{j2\pi f})$  describes the error in gain response caused by the fact that the filter's order is finite. In the case of a multiband filter, the error function  $E(e^{j2\pi f})$  has an equiripple form. The normalized error in the  $i$ th band is defined as

$$E_i^*(e^{j2\pi f}) = \frac{1}{\delta_i} \left\{ \hat{E}_i^*(e^{j2\pi f}) \right\}, \quad f \in B_i \quad (22)$$

where  $B_i$  is the region of frequencies in the  $i$ th band,  $\hat{E}_i^*(e^{j2\pi f})$  is the error function when the coefficients are represented with finite wordlength, and  $\delta_i$  is the maximal deviation in the  $i$ th band for infinite coefficient wordlength.

Fig. 6 shows the normalized error as a function of normalized frequency for optimal LP and NLP low-pass filters having the same gain specifications, for varying value of coefficient wordlength  $b$ . From Fig. 6 it seems that the effect of coefficient quantization is similar for LP and NLP filters. It can also be seen that the maximal normalized error in the  $i$ th band, which will be defined as  $E_i^*$ , tends to increase when  $b$  decreases. Moreover, the frequency at which the normalized error attains its highest value is randomly located in the region  $B_i$ , and as  $E_i^*$  increases, all of the peaks of the normalized error tend to increase.

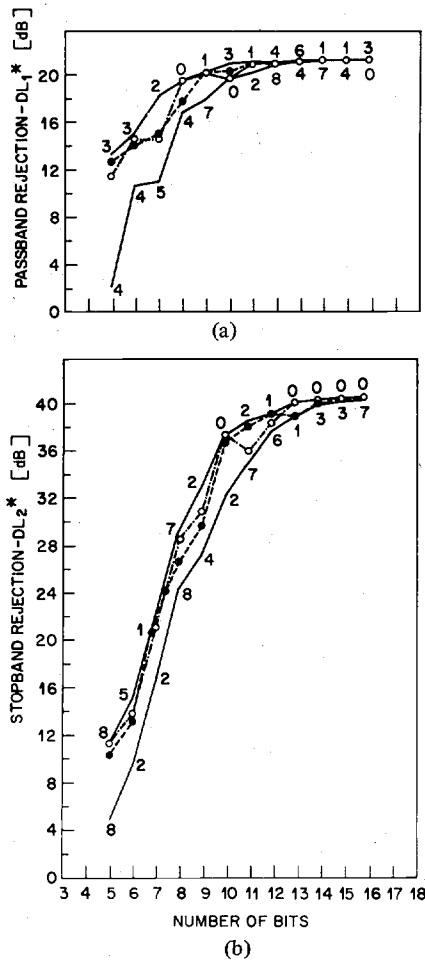


Fig. 7. Actual passband and stopband rejections as a function of coefficient wordlength for optimal linear-phase and nonlinear-phase filters with various zero selections. Here,  $DL_1 = 20.95$  dB,  $DL_2 = 40.44$  dB,  $F_p = 0.2$ ,  $\Delta F = 0.045$ ,  $N_{LP} = 29$ ,  $N_{NLP} = 26$ , the open circles show linear phase, and the closed circles show minimum phase.

These observations are found to be valid when repeating Fig. 6 for a wide range of filter specifications. It may therefore be concluded that the maximal normalized error  $E_i^*$  can serve as a measure of the loss in performance in the  $i$ th band due to coefficient quantization. Another measure which is related to  $E_i^*$  is the actual in-band rejection, defined as

$$DL_i^* = -20 \log_{10} \delta_i^* \quad (23)$$

where  $\delta_i^*$  is the actual (after coefficient quantization) maximal deviation in the  $i$ th band. Since  $\delta_i^*/\delta_i = E_i^*$ , it is clear that

$$DL_i^* = DL_i - 20 \log_{10} E_i^* \quad (24)$$

where  $DL_i$  is the ideal in-band rejection, defined in (18).

Fig. 7 shows  $DL_1^*$  and  $DL_2^*$  as functions of the coefficient wordlength for optimal LP and NLP low-pass filters. Four curves are depicted for each band. The inner two correspond to linear-phase and minimum-phase filters. The outer two describe the maximal and minimal value of  $DL_i^*$ , respectively, for each wordlength, when all of the possibilities of zero selection (from off unit circle zeros) are taken into account (practically, only half of the combinations had to be considered, since when each zero of the NLP filter is replaced by its re-

iprocal, the actual normalized error remains unchanged). The indexes attached to these two curves define the filters which correspond to the maximal (or minimal) value of  $DL_i^*$  for each wordlength. An index "0" defines a linear phase filter, "1"—a minimum-phase filter (all zeros inside the unit circle), "2"—all of the zeros except one are inside the unit circle, and "k" for  $k > 0$  defines a filter all but  $k - 1$  of whose zeros are inside the unit circle.

From Fig. 7 it can easily be seen that for each value of  $b$ , the filter which corresponds to the highest value of  $DL_i^*$  can randomly be an LP or an NLP filter. Moreover, in the case of an NLP filter, any combination of zero selection can be the one which corresponds to the highest  $DL_i^*$ . This means that in order to achieve the filter with best performance, it is necessary to design  $2^{n-1}$  NLP filters (where  $n$  is the number of pairs of zeros of the prototype LP filter off the unit circle), and to compute the actual gain response of each one of them for a given wordlength. This, of course, is not practical. It is also unnecessary, since for each  $b$  the difference in actual rejection between the best choice and the poorest one is not significant. This is not mentioning the fact that for each  $b$  the filter which corresponds to the best performance in the passband does not necessarily have the best performance in the stopband, and vice versa. This can also be seen from another point of view—if a given value of actual rejection has to be met, then the difference between the smallest and the largest wordlength is not more than one or two bits.

The conclusions above are verified when Fig. 7 is repeated for a wide range of low-pass filter parameters. Moreover, the shape of the curves for each band does not depend on whether it is a passband or a stopband, but only on the value of  $DL_i$ . This fact, which is consistent with the results of [10], implies that the same conclusions are also valid for a multiband filter having more than two bands. No substantial advantage is therefore gained from this point of view by having the freedom of selecting the NLP filter's zeros.

## VI. PRACTICAL CONSIDERATIONS IN THE DESIGN OF NLP FIR DIGITAL FILTERS

When designing optimal multiband LP filters using the Remez exchange algorithm, it turns out that for high in-band rejections, the resulting error function [defined in (21)] does not have an exact equiripple form. This phenomenon is caused by the fact that the design procedure is executed by a processor whose multiplication and division operations have a finite precision. As a result, the designed in-band rejections are smaller than the desired ones.

Table I shows an example of the actual passband and stopband rejections of an optimal low-pass filter,  $DL_1^*$  and  $DL_2^*$ , respectively, as a function of the desired rejections,  $DL_1$  and  $DL_2$ . In this example, an IBM 370/168 computer was used for executing the design process. An asterisk alongside the value of the rejection indicates that there were some difficulties in the convergence of the Remez algorithm. (The algorithm did not converge when the error function was close to equiripple.) A double asterisk indicates that the algorithm failed to converge. From Table I it seems that for rejections above 90 dB, the actual rejections were smaller than the desired ones.

TABLE I  
ACTUAL REJECTIONS (IN dB) AS FUNCTION OF DESIGNED REJECTIONS FOR AN  $N = 26$ ,  $F_p = 0.15$  OPTIMAL LINEAR-PHASE LOW-PASS FILTER

DL <sub>1</sub>	DL <sub>2</sub>		20	30	40	50	60	70	80	90	100	110	120
20	PASSBAND		19.97	30.01	39.96	50.00	60.00	69.86	**	**	**	29.78*	**
	STOPBAND		19.97	20.01	19.96	20.02	20.04	19.99			20.22		
30	PASSBAND		20.00	30.03	40.06	50.07	59.97	69.91	79.54	88.79	95.68	98.56	**
	STOPBAND		30.00	30.03	30.06	30.07	30.03	30.03	29.95	30.07*	29.93*	29.97*	
40	PASSBAND		19.99	30.02	39.98	49.97	60.01	69.85	79.35	88.65	95.40	100.97	94.07
	STOPBAND		39.99	40.02	39.98	39.98	40.06	39.93	40.04	40.07*	39.95*	39.95*	39.98*
50	PASSBAND		19.97	29.98	40.07	50.04	60.02	69.86	79.53	88.55	95.87	99.44	92.72
	STOPBAND		49.96	49.98	50.07	50.05	50.06	49.97	49.97	49.93*	49.91*	49.99*	49.90*
60	PASSBAND		20.02	30.02	39.92	50.07	59.95	69.86	79.66	88.90	96.13	100.52	**
	STOPBAND		59.98	60.01	59.92	60.07	59.98*	59.96	60.05*	59.98	60.04*	59.95*	
70	PASSBAND		19.98	30.02	39.96	49.95	59.95	69.85	79.56	88.60	95.28	99.78	**
	STOPBAND		69.94	70.00	69.87	69.94	69.98	69.94	69.93*	69.92*	69.75*	69.91*	
80	PASSBAND		20.06	30.00	40.04	50.05	59.91	69.81	79.70	88.67	96.23	97.57	101.51
	STOPBAND		79.66	79.87	79.94	80.04	79.86	79.88	79.97	79.97*	79.92*	79.88*	80.01
90	PASSBAND		19.99	29.96	39.91	50.02	59.95	69.79	79.62	88.54	96.47	99.49	95.28
	STOPBAND		89.03	89.61	89.67	89.97*	89.70	89.72	89.91	89.32	89.79*	89.67*	89.29
100	PASSBAND		20.01	30.01	39.94	50.07	59.94	69.81	79.72	89.02	96.50	99.29	**
	STOPBAND		97.14	98.37	97.14	99.40	99.69	97.07	99.17	98.65*	97.00*	99.32*	
110	PASSBAND		20.02	30.04	40.08	49.92	59.96	69.66	79.66	87.64	96.03	**	**
	STOPBAND		107.65	107.47	106.93	106.35	107.62	102.16	107.47	102.13*	107.11*		
120	PASSBAND		19.99	29.96	39.94	49.99	59.95	69.31	79.55	89.14	94.12	96.03	**
	STOPBAND		79.73	*88.88	85.89	94.56	113.52	95.28	112.30	111.98*	113.09*	104.65*	

TABLE II  
ACTUAL REJECTIONS (IN dB) AS FUNCTION OF DESIGNED REJECTIONS FOR AN  $N = 26$ ,  $F_p = 0.15$  OPTIMAL NONLINEAR-PHASE LOW-PASS FILTER

DL <sub>1</sub>	DL <sub>2</sub>		20	30	40	50	60	70	80
20	PASSBAND		20.03	29.94	40.06	49.99	59.99	69.86	79.58
	STOPBAND		20.01	19.97	20.03	20.00	20.00	19.96	19.99
30	PASSBAND		19.92	30.06	40.07	50.02	59.94	69.91	79.80
	STOPBAND		29.92	30.02	30.03	30.01	29.97	29.99	30.02
40	PASSBAND		20.03	30.08	40.00	49.93	59.96	69.94	79.79
	STOPBAND		39.94	39.89	39.73	39.74	39.94	39.92	39.98
50	PASSBAND		20.06	29.99	39.98	50.03	59.95	69.88	79.78
	STOPBAND		47.87	46.48	48.56	47.17	49.08	49.06	49.17*
60	PASSBAND		20.04	29.96	39.93	50.02	59.91	69.88	78.79
	STOPBAND		41.28	36.50	48.62	37.26	54.82	49.22*	43.95*
70	PASSBAND		19.95	29.94	39.95	49.99	59.73	69.51	**
	STOPBAND		30.13*	39.28	35.89	37.62	50.29	45.68	
80	PASSBAND		**	30.00	39.84	46.96	**	**	**
	STOPBAND			30.76*	33.96	26.57*			

Moreover, it was impossible to design a filter having higher rejections than 100-110 dB.

Table II shows the actual rejections for optimal NLP filters having the same orders and passband cutoff frequencies as in Table I. It is not difficult to see that while the limitations on the passband rejection are similar for LP and NLP filters, the stopband rejection is much more restricted for NLP filters. In fact, it was not possible to design an NLP filter having higher stopband rejection than 60 dB. This is not surprising since the design of an optimal NLP filter whose passband and stopband deviations are  $\delta_1$  and  $\delta_2$ , respectively, is based upon designing an optimal LP filter whose deviations are  $2\delta_1$  and  $(\frac{1}{2})\delta_2^2$ .

Tables similar to I and II were derived for various filter parameters, resulting in the same limitations. This raises the problem of designing FIR filters with higher rejections than those permitted in Tables I and II. In the case of LP filters, this problem can be solved by using the method mentioned

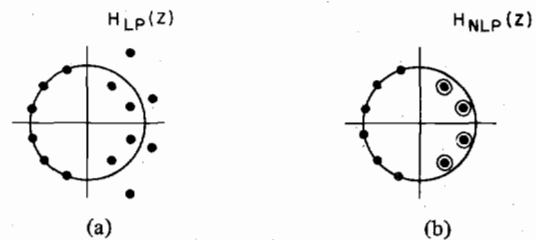


Fig. 8. Design of a low-pass minimum-phase filter.

in [13], that is, to design an optimal LP filter with relatively low rejections, and making a proper transformation on it. However, this does not solve the problem for NLP filters. In this case, the following algorithm may be used for designing an NLP filter with high rejections.

- a) Design an LP filter with the desired specifications.
- b) Find the zeros of the designed filter.
- c) Replace the set of zeros off the unit circle with a new one, by replacing each undesired zero with its conjugate reciprocal.

Fig. 8 gives a demonstration of the algorithm for a minimum phase filter, where a circled dot represents a double zero. The proposed algorithm is based on the well-known fact that the amplitude response remains unchanged when replacing a zero with its conjugate reciprocal. The resulting NLP filter has by definition the same order as an optimal LP filter with the same specifications, which, of course, is greater than the order of the corresponding optimal NLP filter. However, this NLP filter has some advantages which make it sometimes more attractive than an optimal NLP filter. First, the limitations on the rejections are identical, by definition, to those of optimal LP filters. In addition, the designer has more flexibility in defining the filter's specifications. Finally, the implementation of the resulting NLP



filter can be simplified, using the fact that its transfer function can be factored in the following manner.

$$H_{\text{NLP}}(z) = H_1(z) \cdot H_2(z) \quad (25)$$

where  $H_1(z)$  contains all of the zeros on the unit circle, and  $H_2(z)$ , all of the zeros off the unit circle. Clearly,  $H_1(z)$  is a linear phase filter and, therefore, has symmetrical coefficients, and its implementation can generally be simplified.

Although in principle this simplification can be applied to optimal NLP filters, practically this is not the case. The reason for this is the fact that even for relatively small values of stopband rejections, the gain response of the prototype LP filter is no longer equiripple, and it intersects the frequency axis instead of being tangent to it [see Fig. 2(b)]. In order to achieve the desired stopband rejection, (7) must be replaced by the following equation.

$$\begin{aligned} \hat{D}_i &= \frac{1}{2} \delta_i^2 \\ \hat{\delta}_i &= \frac{1}{2} \delta_i^2 - \epsilon_i, \quad \epsilon_i \approx 0.1\delta_i. \end{aligned} \quad (26)$$

In this case, the gain response curve does not touch the frequency axis, which means that the filter has no zeros on the unit circle, and the factorization of (25) cannot be applied to it.

It thus appears that the alternative design method presented in this section can be preferable to the original one for optimal NLP filters. This is especially true when the stopband region becomes wider (for example, a narrow-band filter), for two reasons. The first is that the order of  $H_1(z)$  in (25) becomes greater, resulting in more efficiency in implementation. The second one is that in this case the difference between the orders of optimal and nonoptimal NLP filters becomes smaller (see Fig. 5). Notice that in the case of a minimum phase filter, further simplification may be achieved using the fact that the zeros inside the unit circle are doubled.

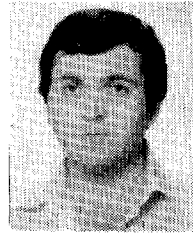
#### ACKNOWLEDGMENT

We wish to express our thanks to J. F. Kaiser for his helpful comments.

#### REFERENCES

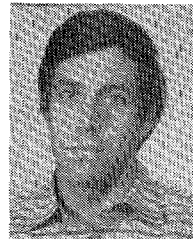
- [1] L. R. Rabiner, J. H. McClellan, and T. W. Parks, "FIR digital filter design techniques using weighted Chebyshev approximation," *Proc. IEEE*, vol. 63, pp. 595-610, Apr. 1975.
- [2] D. Y. Sze, "An algorithm for extremal polynomials of Chebyshev systems with applications in digital filter design," unpublished manuscript.
- [3] A. G. Holt, J. Attikiouzel, and R. Bennet, "Iterative technique for designing nonrecursive digital filters to non-linear phase characteristics," *Radio Electron. Eng.*, vol. 46, pp. 589-592, 1976.
- [4] M. A. Razzak and L. G. Cuthbert, "Performance comparison for nonrecursive digital filters with nonlinear phase characteristics," *Electron. Lett.*, vol. 14, pp. 370-372, June 8, 1978.

- [5] O. Herrmann and H. W. Schuessler, "Design of nonrecursive digital filters with minimum phase," *Electron. Lett.*, vol. 6, pp. 329-330, May 28, 1970.
- [6] F. E. Burris, "Two design methods for minimum phase FIR digital filters," in *Proc. 13th Annu. Allerton Conf. Circuits Syst.*, Oct. 1-3, 1975, pp. 485-497.
- [7] T. W. Parks and J. H. McClellan, "Chebyshev approximation for nonrecursive digital filters with linear phase," *IEEE Trans. Circuit Theory*, vol. CT-19, pp. 189-194, Mar. 1972.
- [8] L. R. Rabiner, "Approximate design relationship for low-pass FIR digital filters," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 456-460, Oct. 1973.
- [9] O. Herrmann, L. R. Rabiner, and D. S. K. Chan, "Practical design rules for optimum finite impulse response low-pass digital filters," *Bell Syst. Tech. J.*, vol. 52, pp. 769-799, July-Aug. 1973.
- [10] D. K. Chan and L. R. Rabiner, "Analysis of quantization errors in the direct form for finite impulse response digital filters," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 354-366, Aug. 1973.
- [11] A. Peled and B. Liu, "A new approach to the realization of nonrecursive digital filters," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 477-484, Dec. 1973.
- [12] L. H. Goldstein and B. Liu, "An ADPCM realization of nonrecursive digital filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 312-320, Aug. 1976.
- [13] J. F. Kaiser and R. W. Hamming, "Sharpening the response of symmetric nonrecursive filter by multiple use of the same filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 415-422, Oct. 1977.
- [14] J. F. Kaiser, personal communication.



Eli Goldberg was born in Lod, Israel, on December 24, 1949. He received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel, in 1971 and 1979, respectively.

During the years 1976 to 1981, he was a Teaching Assistant and Instructor in the Department of Electrical Engineering of the Technion. His current research interests are in digital filtering and speech processing.



Robert Kurshan (M'79) received the Ph.D. degree in mathematics from the University of Washington, Seattle, WA, in 1968.

Since 1968 he has been with Bell Laboratories, Murray Hill, NJ, where he is presently engaged in mathematics research, with an emphasis on algebra, in the Mathematics and Statistics Research Center. He held the Krantzberg Chair for Visiting Scientists at the Technion-Israel Institute of Technology, Haifa, Israel, from March through August 1976.

David Malah (S'67-M'71), for a photograph and biography, see p. 283 of the April 1981 issue of this TRANSACTIONS.