# On Fast FIR Filters Implemented as Tail-Canceling IIR Filters

Avery Wang, Julius O. Smith III

**Abstract**

We have developed an algorithm based on synthetic division for deriving the transfer function which cancels the tail of a given arbitrary rational (IIR) transfer function after a desired number of time steps. Our method applies to transfer functions with repeated poles, whereas previous methods of tail-subtraction cannot [1]. We use a parallel state-variable technique with periodic refreshing to induce finite memory in order to prevent accumulation of quantization error in cases where the given transfer function has unstable modes. We present two methods for designing linear-phase Truncated IIR (TIIR) filters: one based on the addition and the other on the cascading of anti-phase filters. We explore finite-register effects for unstable modes and provide bounds on the maximum TIIR filter length. In particular, we show that for unstable systems the available dynamic range of the registers must be three times that of the data. Considerable computational savings over conventional FIR filters are attainable for a given specification of linear-phase filter. We provide examples of filter design. We show how to generate finite-length polynomial impulse responses using TIIR filters. We list some applications of TIIR filters, including uses in digital audio and an algorithm for efficiently implementing Kay's optimal high-resolution frequency estimator [2].

## I. Introduction

Infinite impulse response (IIR) recursive linear digital filters are widely used because of their low computational cost and low storage overhead requirements. Finite impulse response (FIR) filters, on the other hand, allow the possibility of implementing linear-phase linear digital filters which have constant group delay across all frequencies. The tradeoff is that to achieve similar magnitude transfer functions, FIR filters usually require much larger filter orders than their IIR counterparts. For example, a general $N$-th order FIR filter requires $N+1$ multiplies and $N$ adds. In certain cases, however, FIR filters may be designed which have an operation count comparable to that of an IIR filter while maintaining the linear phase property.

The set of finite impulse responses which may be efficiently implemented includes those which are truncated IIR (TIIR) sequences of low order. Although the following results were derived independently of Saramäki and Fam [1, 3], the idea of using truncated IIR filters to generate linear-phase filters was originally introduced by Fam [3]. Fam and Saramäki [1, 3] deal with unstable hidden modes due to pole-zero cancellations outside of the unit circle by employing a switching and resetting algorithm to reduce the effects of quantization error buildup. We introduce a slightly more efficient version of this idea, as well as an error analysis.

In this paper, we describe an algorithm for the efficient implementation of certain classes of FIR filters. We introduce an extension of the TIIR algorithm which allows the truncation of arbitrary IIR filter tails, and provides a way to implement polynomial impulse responses. Additionally, we present an analysis of the effects of limited numerical precision and provide design guidelines for designing systems with acceptable noise tolerance.

## II. Definitions

A general causal $N$-th order FIR filter consists of a tapped delay line $N$ elements long and a table of $N+1$ impulse response coefficients $\{h_0, \ldots, h_N\}$ such that at each time step $n$ the output

$$y[n] = \sum_{k=0}^{N} h_k x[n-k] \tag{1}$$

is formed. The transfer function has the simple form

$$H_{\mathrm{FIR}}(z) \triangleq h_0 + h_1 z^{-1} + \ldots + h_N z^{-N} \tag{2}$$
$$= z^{-N} C(z), \tag{3}$$

where $C(z)$ is the $N$-th degree polynomial formed by the $h_k$. On the other hand, a causal $P$-th order IIR filter has the relation

$$y[n] = -\sum_{k=1}^{P} a_k y[n-k] + \sum_{\ell=0}^{P} b_\ell x[n-\ell]. \tag{4}$$

This work was supported in part by a National Science Foundation Graduate Fellowship

The corresponding transfer function is

$$H_{\mathrm{IIR}}(z) \triangleq \frac{b_0 + b_1 z^{-1} + \ldots + b_P z^{-P}}{1 + a_1 z^{-1} + \ldots + a_P z^{-P}} \tag{5}$$

$$\triangleq \frac{B(z)}{A(z)} \tag{6}$$

$$\triangleq h_0 + h_1 z^{-1} + h_2 z^{-2} + \ldots, \tag{7}$$

where both

$$A(z) = z^P + a_1 z^{P-1} + \ldots + a_P \tag{8}$$

and

$$B(z) = b_0 z^P + b_1 z^{P-1} + \ldots + b_P \tag{9}$$

may be assumed to be relatively prime $P$-th degree polynomials in $z$, and $A(z)$ is monic by construction.

*Group delay* [4] is defined by

$$\delta(\omega) \triangleq -\frac{d \arg\{H(e^{j\omega})\}}{d\omega}. \tag{10}$$

The group delay at normalized frequency $\omega = 2\pi f/f_s$, where $f_s$ is the sampling frequency, is the number of samples of delay experienced by the amplitude envelope of a narrow-band input signal centered at $\omega$.

A *linear-phase filter* is one such that the phase response at a given frequency is a linear function of frequency, i.e., $\arg\left\{H(e^{j\omega})\right\} = K_1\omega + K_2$ for some constants $K_1$ and $K_2$. From this property we see immediately that the group delay is constant for all frequencies. Filters with linear phase response are often desirable because they have no frequency-dependent temporal distortion. A stable IIR filter with non-zero poles cannot have linear phase. However, an FIR filter with coefficients $\{h_0, \ldots, h_N\}$ has linear phase if there exists a $\psi$ such that for all $k \in \{0, \ldots, N\}$

$$h_{N-k} = e^{j\psi} h_k^*, \tag{11}$$

i.e., if the reversed coefficients are the complex conjugates of the forward sequence plus a constant phase shift [5]. The group delay is then

$$\delta(f) = \frac{N}{2}. \tag{12}$$

## III. Truncated IIR (TIIR) Filters

Consider an FIR filter having a truncated geometric sequence $\{h_0, h_0 p, \ldots, h_0 p^N\}$ as an impulse response. This filter has the same impulse response for the first $N+1$ terms as the one-pole IIR filter with transfer function

$$H_{\mathrm{IIR}}(z) = \frac{h_0}{1 - pz^{-1}}. \tag{13}$$

If we subtract off the tail of the impulse response we obtain

$$H_{\mathrm{FIR}}(z) = h_0 + h_0 p z^{-1} + \ldots + h_0 p^N z^{-N} \tag{14}$$

$$= h_0 \frac{1 - p^{N+1} z^{-(N+1)}}{1 - pz^{-1}}. \tag{15}$$

The time-domain recursion for this filter is

$$y[n] = \sum_{k=0}^{N} h_0 p^k x[n-k] \tag{16}$$

$$= py[n-1] + h_0 \left(x[n] - p^{N+1} x[n-(N+1)]\right). \tag{17}$$

We see that the first formulation, Eqn. (16), requires $N+1$ multiplies and $N$ adds to implement directly, whereas the second formulation, Eqn. (17), requires only 3 multiplies and 2 adds, independent of $N$. Thus we see that if we can represent an FIR sequence as a truncated exponential sequence, a tremendous savings in computation can be achieved. Note that the $x[n-(N+1)]$ term in Eqn. (17) still requires a delay line to be maintained, and thus there are no savings in storage. With modern digital signal processing (DSP) chips, however, ring buffers may be implemented with virtually no computational overhead and the full savings in computational cost may be achieved.

Notice that there is a pole-zero cancellation in the representation given by Eqn. (15). If $|p| < 1$ there is no problem since the system is inherently stable. If $|p| \geq 1$, however, then there is a potential problem due to the hidden mode. We will deal with this in Section IV where we will see how to run TIIR filters with unstable modes.

The idea of this section was used by Fam [3], where a partial fraction expansion of a transfer function is taken and each mode is truncated separately using Eqn. (15). This method works only for cases in which the multiplicity of each pole is one, so that the each mode exhibits a simple exponential decay.

### *A. Extension to Higher-Order TIIR Sequences*

We may extend the idea illustrated in the previous section for the one-pole case to computing the TIIR sequence of any rational $H(z)$. The general procedure is to find the "tail" IIR transfer function

$$\begin{aligned} H'_{\mathrm{IIR}}(z) &= h'_0 z^{-1} + h'_1 z^{-2} + \dots \\ &\triangleq h_{N+1} z^{-1} + h_{N+2} z^{-2} + \dots \end{aligned} \tag{18}$$

whose impulse response, except for a time shift of $N$ steps, matches the tail of the transfer function $H_{\mathrm{IIR}}(z)$ which we would like to truncate after time step $N$.

We multiply Eqn. (7) by $z^N$ and obtain

$$\begin{aligned} z^N H_{\mathrm{IIR}}(z) &= h_0 z^N + \dots + h_{N-1} z + h_N \\ &\quad + h_{N+1} z^{-1} + h_{N+2} z^{-2} + \dots && (19) \\ &= C(z) + H'_{\mathrm{IIR}}(z) && (20) \\ &= \frac{z^N B(z)}{A(z)} && (21) \\ &= C(z) + \frac{B'(z)}{A(z)}, && (22) \end{aligned}$$

where $\mathrm{Deg}\,\{B'(z)\} < \mathrm{Deg}\,\{A(z)\} = P$. We may assume that $\mathrm{Deg}\,\{B'(z)\} = P-1$. $B'(z)$ is unique and may be obtained by performing synthetic division on $z^N B(z)$ by $A(z)$ and finding the remainder. Thus, $z^N B(z) \equiv B'(z) \pmod{A(z)}$.

Once we have obtained $B'(z)$, we have $H'_{\mathrm{IIR}}(z) = B'(z)/A(z)$ and we may write

$$\begin{aligned} H_{\mathrm{FIR}}(z) &= H_{\mathrm{IIR}}(z) - z^{-N} H'_{\mathrm{IIR}}(z) && (23) \\ &= \frac{B(z) - z^{-N} B'(z)}{A(z)}. && (24) \end{aligned}$$

The corresponding system is

$$\begin{aligned} y[n] = &-\sum_{k=1}^{P} a_k y[n-k] + \sum_{\ell=0}^{P} b_\ell x[n-\ell] \\ &-\sum_{m=0}^{P-1} b'_m x[n-m-(N+1)], \end{aligned} \tag{25}$$

using the representation in Eqn. (24).

The fact that the denominators of the transfer functions $H_{\mathrm{IIR}}(z)$ and $H'_{\mathrm{IIR}}(z)$ are the same allows additional savings in computational cost due to the fact that the original IIR and tail IIR dynamics are the same and do not need to be performed twice. The term $z^{-(N+1)} z^{-(P-1)} B'(z)$ serves to zero out the dynamics at the end of the delay line and requires only an additional $P$ multiplies and $P-1$ adds. The computational cost of this general truncated $P$-th order IIR system is $3P+1$ multiplies and $3P-2$ adds, independent of $N$. Thus, a net computational savings with this class of FIR filters is achieved if $N > 3P$.

The storage costs for this filter are $P$ output samples for the IIR feedback dynamics, $N$ input samples of the FIR filter, and an additional $P$ input samples for the tail-cancellation dynamics, yielding $N+P$ input delay samples, of which only the first and last $P$ are used, and $P$ output delay samples. Thus, the fast FIR algorithm requires $2P$ more storage samples than a direct FIR implementation.

As in the previous section, we observe cautiously that the effect of subtracting the tail IIR response $z^{-N} H'_{\mathrm{IIR}}(z)$ from $H_{\mathrm{IIR}}(z)$ is to cancel all the poles in Eqn. (24), which is to be expected since an FIR filter is an all-zero filter.

### *B. Other Architectures*

The direct implementation specified by Eqn. (25) may not be desirable for various reasons. For example, one may choose to use a factored structure such as the cascaded biquad or the parallel partial fraction form. The latter form is given as

$$H(z) = \sum_{k=1}^{N_p} \sum_{\ell=1}^{M_k} \frac{C_{k,\ell}}{(1 - p_k z^{-1})^{\ell}}, \tag{26}$$

where $N_p$ is the number of distinct poles, and $M_k$ is the multiplicity of the $k$-th pole. The $(k, \ell)$ term of the partial fraction expansion corresponds to a filter with impulse response

$$h_{k,\ell}[n] = C_{k,\ell} \binom{n + \ell - 1}{\ell - 1} p_k^n. \tag{27}$$

To form the TIIR filter, a tail IIR filter is derived for each partial fraction using synthetic division as outlined in Section III-A. Each TIIR response is calculated separately, and the results are added together to form the complete response. The factorization need not be as complete as outlined in Eqn. (26). One may choose an intermediate level of factorization, leaving some factors lumped together and others separated from each other. For example, one may want to group complex-conjugate pairs together to avoid complex arithmetic. Alternatively, one may wish to leave terms with the same poles together as in

$$H(z) = \sum_{k=1}^{N_p} \frac{B_k(z)}{(z - p_k)^{M_k}}, \tag{28}$$

since calculating the tail IIR response for each $n$-th order multiplicity term yields a degree $n - 1$ polynomial numerator anyway. The impulse response of the $k$-th partial fraction in this case is

$$h_k[n] = \sum_{\ell=0}^{M_k} b_{k,\ell} \binom{n - \ell + M_k - 1}{M_k - 1} p_k^{n-\ell} \tag{29}$$

where the $b_{k,\ell}$, $\ell = 0, \ldots, M_k$ are the coefficients of $B_k(z)$.

Another example is to group together the stable factors, i.e., those with poles $p_k$ such that $|p_k| < 1$, and implement separately the unstable poles for which $|p_k| > 1$. These strategies will become useful in Section V.

## IV. Unstable Hidden Modes

| $n \equiv 0 \pmod N$ | $n \not\equiv 0 \pmod N$ |
|---|---|
| Primary TIIR Filter | |
| $w_1[N+k] \leftarrow 0, \quad k = 1, \ldots, P-1$<br>$w_1[k] \leftarrow w_1[k-1], \quad k = N, \ldots, 1$<br>$w_1[0] \leftarrow x[n]$<br>$q_1[k] \leftarrow q_2[k-1], \quad k = P, \ldots, 1$<br>$q_1[0] \leftarrow -\sum_{k=1}^{P} a_k q_1[k] + \sum_{\ell=0}^{P} b_\ell w_1[\ell]$<br>$y[n] \leftarrow q_1[0]$ | $w_1[k] \leftarrow w_1[k-1], \quad k = N+P, \ldots, 1$<br>$w_1[0] \leftarrow x[n]$<br>$q_1[k] \leftarrow q_1[k-1], \quad k = P, \ldots, 1$<br>$q_1[0] \leftarrow -\sum_{k=1}^{P} a_k q_1[k] + \sum_{\ell=0}^{P} b_\ell w_1[\ell] - \sum_{m=0}^{P-1} b'_m w_1[m+N+1]$<br>$y[n] \leftarrow q_1[0]$ |
| Auxiliary TIIR Filter | |
| $w_2[k] \leftarrow 0, \quad k = 1, \ldots, P-1$<br>$w_2[0] \leftarrow x[n]$<br>$q_2[k] \leftarrow 0, \quad k = 1, \ldots, P-1$<br>$q_2[0] \leftarrow b_0 w_2[0]$ | $w_2[k] \leftarrow w_2[k-1], \quad k = P, \ldots, 1$<br>$w_2[0] \leftarrow x[n]$<br>$q_2[k] \leftarrow q_2[k-1], \quad k = P, \ldots, 1$<br>$q_2[0] \leftarrow -\sum_{k=1}^{P} a_k q_2[k] + \sum_{\ell=0}^{P} b_\ell w_2[\ell]$ |

TABLE I

Fast FIR algorithm for an unstable Truncated IIR (TIIR) filter.

We now address the issue of pole-zero cancellation and the resulting hidden modes in the fast FIR algorithm. Although the naïve Fast FIR algorithm of Eqn. (25) works in theory, there is the practical matter of quantization error due to finite register lengths when dealing with truncated unstable IIR systems, i.e., those with poles $p_k$ such that $|p_k| > 1$. Consider the system in equation (17). We may model quantization error as an independent, identically distributed (IID)

| $n \equiv 0 \pmod{N}$ | $n \not\equiv 0 \pmod{N}$ |
|---|---|
| Primary TIIR Filter | |
| $\begin{aligned} w_1[N+k] &\leftarrow 0, \quad k=1,\ldots,P-1 \\ w_1[k] &\leftarrow w_1[k-1], \quad k=N,\ldots,1 \\ w_1[0] &\leftarrow x[n] \\ q_1[k] &\leftarrow q_2[k-1], \quad k=P,\ldots,1 \\ q_1[0] &\leftarrow -\sum_{k=1}^{P-1} \frac{a_k^* q_1[P-k]}{a_P^*} - \frac{q_1[P]}{a_P^*} \\ &\quad + \sum_{m=0}^{P-1} \frac{b_m'^* w_1[P-1-m]}{a_P^*} \end{aligned}$ | $\begin{aligned} w_1[k] &\leftarrow w_1[k-1], \quad k=N+P,\ldots,1 \\ w_1[0] &\leftarrow x[n] \\ q_1[k] &\leftarrow q_1[k-1], \quad k=P,\ldots,1 \\ q_1[0] &\leftarrow -\sum_{k=1}^{P-1} \frac{a_k^* q_1[P-k]}{a_P^*} - \frac{q_1[P]}{a_P^*} \\ &\quad + \sum_{m=0}^{P-1} \frac{b_m'^* w_1[P-1-m]}{a_P^*} \\ &\quad - \sum_{\ell=0}^{P} \frac{b_\ell^* w_1[N+P-\ell]}{a_P^*} \\ y[n] &\leftarrow q_1[0] \end{aligned}$ |
| Auxiliary TIIR Filter | |
| $\begin{aligned} w_2[k] &\leftarrow 0, \quad k=1,\ldots,P-1 \\ w_2[0] &\leftarrow x[n] \\ q_2[k] &\leftarrow 0, \quad k=1,\ldots,P-1 \\ q_2[0] &\leftarrow \frac{b_{P-1}'^* w_2[0]}{a_P^*} \end{aligned}$ | $\begin{aligned} w_2[k] &\leftarrow w_2[k-1], \quad k=P,\ldots,1 \\ w_2[0] &\leftarrow x[n] \\ q_2[k] &\leftarrow q_2[k-1], \quad k=P,\ldots,1 \\ q_2[0] &\leftarrow -\sum_{k=1}^{P-1} \frac{a_k^* q_2[P-k]}{a_P^*} - \frac{q_2[P]}{a_P^*} \\ &\quad + \sum_{m=0}^{P-1} \frac{b_m'^* w_2[P-1-m]}{a_P^*} \end{aligned}$ |

TABLE II

Fast FIR algorithm for a reversed unstable Truncated IIR (TIIR) filter derived from a stable TIIR filter.

noise signal $\nu[n]$ with zero mean and variance $\sigma_\nu^2$ input to the system of Eqn. (17).[1] The output equation is then

$$y[n] = py[n-1] + h_0\left(x[n] - p^{N+1}x[n-(N+1)]\right) + \nu[n] \tag{30}$$

$$= h_0 \sum_{k=0}^{N-1} p^k x[n-k] + \sum_{k=0}^{n} p^k \nu[n-k]. \tag{31}$$

Thus, the accumulated noise has zero mean, but its variance grows as

$$\sigma_{\mathrm{acc}}^2[n] = \sum_{k=0}^{n} |p|^{2k} \sigma_\nu^2 \tag{32}$$

$$= \sigma_\nu^2 \frac{1-|p|^{2(n+1)}}{1-|p|^2} \tag{33}$$

so that

$$\lim_{n\to\infty} \sigma_{\mathrm{acc}}^2[n] = \begin{cases} +\infty, & |p| \geq 1 \\ \dfrac{\sigma_\nu^2}{1-|p|^2}, & |p| < 1. \end{cases} \tag{34}$$

For a higher-order system, as described in Section III-A, this analysis gives an estimate of the noise accumulation behavior if $p$ is the largest-magnitude pole, since its dynamics dominates the behavior of the system. A direct implementation of an FIR system, such as in equation (16) does not have noise accumulation problems because of its finite memory. We trade computational cost for noise sensitivity in TIIR systems.

[1]Finite-register effects may be modeled as the sum of $q$ uniformly distributed IID $[-\epsilon, +\epsilon]$, where $\epsilon > 0$ is the smallest quantity such that $x + \epsilon \neq x$ in machine arithmetic, and $q$ is the number of additive terms. If $q$ is relatively large, the total noise may be modeled as a Gaussian distribution. For fixed-point arithmetic, $\epsilon$ is constant $\forall x$. Floating point arithmetic presents difficulties since the effective $\epsilon$ varies proportionally to the magnitude of $x$. Assuming independence and fixed-point arithmetic, we may model quantization noise as a Gaussian random variable with variance $\sigma_\nu^2 = q\epsilon^2/3$. A thorough analysis of quantization noise statistics is beyond the scope of this paper.

Recalling the discussion in Section III-B, we may factor the unstable modes of a system apart from the transfer function so that

$$H(z) = H^{\downarrow}(z) + H^{\uparrow}(z) \tag{35}$$
$$= \frac{B^{\downarrow}(z)}{A^{\downarrow}(z)} + \frac{B^{\uparrow}(z)}{A^{\uparrow}(z)}, \tag{36}$$

where the "$\downarrow$" and "$\uparrow$" superscripts denote "stable hidden modes" and "unstable hidden modes," respectively. Thus, we may implement the two filters of orders $P^{\downarrow}$ and $P^{\uparrow}$, where $P^{\downarrow} + P^{\uparrow} = P$, and add together the outputs. $H^{\downarrow}(z)$ may be implemented as described in Section III-A, but $H^{\uparrow}(z)$ must be handled carefully.

Alternatively, we may factor the transfer function into stable and unstable parts

$$H(z) = H^{\downarrow}(z)H^{\uparrow}(z) \tag{37}$$
$$= \frac{B^{\downarrow}(z)}{A^{\downarrow}(z)}\frac{B^{\uparrow}(z)}{A^{\uparrow}(z)}, \tag{38}$$

in which case the stable and unstable filters are cascaded. Note that in this factorization $B^{\downarrow}(z)$ and $B^{\uparrow}(z)$ are not uniquely determined.

A possible algorithm for stabilizing an unstable TIIR filter using two parallel sets of state variables is given in Table I. The algorithm cycles with a period of $N$ time steps. For cycle number $k$, at time step $n = kN$, we clear the auxiliary filter's input delay line $w_2[\cdot]$ and vector of state variables $q_2[\cdot]$, each of length $P$. Otherwise the primary and auxiliary systems evolve with the same dynamics. We have taken into account the fact that during the first $N$ time steps the tail canceling terms due to $z^{-N}B'(z)$ in the auxiliary filter are identically zero, hence those terms are not included. Since the FIR impulse response has length $N+1$ we see that at time $n = kN + N$ the two systems have identical output values, except for the amount of accumulated noise. The state variables $q_1[\cdot]$ and $q_2[\cdot]$ are not identical because the input histories are slightly different. However, since we are only concerned with the output, this does not matter. We then copy the auxiliary system's state variables to the primary system and repeat the cycle. This algorithm may be optimized by realizing that the values for $w_2[m]$ are zero for $m > N$ during each cycle. Thus the last summation in the dynamics for the auxiliary system shown in Table I is omitted for the cases where $n \equiv 1, \ldots, N \pmod{N}$.

Under this scheme the amount of time that quantization errors can accumulate varies from a minimum of $N$ time steps to a maximum of $2N$, depending on when each error occurs. The minimum is due to the amount of time necessary to allow the auxiliary system to evolve to having the same output value as the primary system. For noise accumulation analysis in this system, we may assume that the unstable system dynamics are dominated by the mode with the largest eigenvalue $p_{\max}$ such that $|p_{\max}| > 1$. We assume for the present analysis that the multiplicity of $p_{\max}$ is one. Thus we must reckon for tolerable performance with

$$\sigma^2_{\max} \geq \sigma^2_{\nu}\frac{1 - |p_{\max}|^{4N}}{1 - |p_{\max}|^2}. \tag{39}$$

If the maximum tolerable noise variance is $\sigma^2_{\mathrm{tol}}$ then we must have $\sigma^2_{\max} < \sigma^2_{\mathrm{tol}}$ so that, approximately,

$$N < \frac{\log\left(1 + \dfrac{(|p_{\max}|^2 - 1)\sigma^2_{\mathrm{tol}}}{\sigma^2_{\nu}}\right)}{4\log(|p_{\max}|)}. \tag{40}$$

Thus, given $p_{\max}$, $\sigma^2_{\mathrm{tol}}$, and quantization noise variance $\sigma^2_{\nu}$, we have a fundamental limitation on the length of the effective impulse response. A refinement of this bound is given below in Section V-C.

The computational cost for implementing an unstable TIIR filter is at most twice the cost of a stable TIIR filter. The extra cost is due to the cost of implementing the auxiliary TIIR filter, which costs about $2P$ operations per sample. Thus the computational cost of the unstable TIIR filter is approximately $5P$ multiplies per input sample. The number of adds is about the same. If we take into account the fact that the feed-forward terms due to $B(z)$ are the same (except for cases where $w_2[k]$ has been set to zero) in both the primary and auxiliary systems, we may optimize by avoiding duplicate calculation of these terms and thus reduce the total number of multiplies and adds to about $4P$ per input sample. The shifts of the state variables and input delay lines may be simulated by using pointer arithmetic, and need not actually be performed.

## V. Fast Linear-Phase FIR Filters

The implementation of filters as TIIR systems is rather pointless unless there is an advantage that is unavailable to untruncated IIR filters. That advantage is that it is possible to implement exactly linear-phase filters. We present two basic strategies for designing filters based on TIIR filters. The first uses the factorization given in Eqn. (35), and the second uses the factorization used in Eqn. (37).

### *A. Additive Factorization Design Method*

We saw in Section II that an FIR filter has linear phase if Eqn. (11) holds. It is possible to attain such a relation using TIIR filters. Let $H^+_{\mathrm{FIR}}(z)$ be a TIIR transfer function such that

$$H^+_{\mathrm{FIR}}(z) = \sum_{k=0}^{N} h^+_k z^{-k} \tag{41}$$

$$= \frac{B^+(z) - z^{-N}B'^+(z)}{A^+(z)}. \tag{42}$$

We form the time-reversed truncated transfer function

$$H^-_{\mathrm{FIR}}(z) = \sum_{k=0}^{N} h^-_k z^{-k} \tag{43}$$

$$= \sum_{k=0}^{N} h^{+*}_k z^{k-N} \tag{44}$$

$$= z^{-N}\left\{H^+_{\mathrm{FIR}}\left(1/z^*\right)\right\}^* \tag{45}$$

$$= \frac{z^{-N}\left\{B^+\left(1/z^*\right)\right\}^* - \left\{B'^+\left(1/z^*\right)\right\}^*}{\left\{A^+\left(1/z^*\right)\right\}^*} \tag{46}$$

$$= \frac{-zB'^-(z) + z^{-N}B^-(z)}{A^-(z)}, \tag{47}$$

where the "+" and "−" superscripts denote "forward" and "reverse-conjugate" filter respectively, and the "*" superscript denotes complex conjugation, as usual. Thus, comparing with Eqns. (8) and (9), we have

$$A^-(z) = 1 + a_1^* z + \ldots + a_P^* z^P, \tag{48}$$

$$B^-(z) = b_0^* + b_1^* z + \ldots + b_P^* z^P, \tag{49}$$

and

$$B'^-(z) = b_0'^* + b_1'^* z + \ldots + b_{P-1}'^* z^{P-1}. \tag{50}$$

We have assumed that $B^+(z)$ and $B'^-(z)$ have degrees $P$ and $P-1$, respectively. If we assume that $H^+_{\mathrm{FIR}}(z)$ is a stable TIIR filter then $H^-_{\mathrm{FIR}}(z)$ is an unstable TIIR filter whose hidden modes are conjugate-reciprocals of those of $H^+_{\mathrm{FIR}}(z)$. An example implementation for $H^-_{\mathrm{FIR}}(z)$ is given in Table II. There we have normalized the numerator and denominator of the filter by dividing through by $a_P^*$.

Using the arbitrary time shift $M \geq 0$ and phase shift $\psi$, we define the filter

$$H_{\mathrm{LPFIR}}(z) \triangleq H^+_{\mathrm{FIR}}(z) + e^{j\psi}z^{-M}H^-_{\mathrm{FIR}}(z) \tag{51}$$

$$= \sum_{k=0}^{N} h^+_k z^{-k} + e^{j\psi}\sum_{k=0}^{N} h^{+*}_k z^{k-N-M}. \tag{52}$$

We note that this new FIR filter of length $M+N+1$ is invariant with respect to reversing the order, conjugating the coefficients, and multiplying by the phase factor $\exp(j\psi)$. Eqn. (11) holds, and thus $H_{\mathrm{LPFIR}}(z)$ is a linear-phase FIR filter. We may see the linear-phase property more directly by first noticing that, on the unit circle,

$$H^-_{\mathrm{FIR}}\left(e^{j\omega}\right) = e^{-jN\omega}\left\{H^+_{\mathrm{FIR}}\left(e^{j\omega}\right)\right\}^*, \tag{53}$$

so that

$$H_{\mathrm{LPFIR}}\left(e^{j\omega}\right) = H^+_{\mathrm{FIR}}\left(e^{j\omega}\right) + e^{j\psi}e^{-jM\omega}H^-_{\mathrm{FIR}}\left(e^{j\omega}\right) \tag{54}$$

$$= e^{-j((N+M)\omega+\psi)/2}\mathrm{Re}\left\{e^{j((N+M)\omega+\psi)/2}H^+_{\mathrm{FIR}}\left(e^{j\omega}\right)\right\}. \tag{55}$$

Eqn. (12) gives us the result

$$\delta = \frac{M+N}{2}. \tag{56}$$

It is rather difficult to design fast FIR filters to a given set of specifications using this additive factorization design method because it is not intuitively obvious how to control the magnitude $\left|H_{\mathrm{LPFIR}}\left(e^{j\omega}\right)\right|$ due to Eqn. (55). Nonetheless, for certain impulse response waveforms which are well characterized, this technique provides a useful tool for designing corresponding fast FIR filters to realize them. Some examples are given in Section VI-B.

*B. Magnitude-Squared Design Method*

We begin by choosing a desired non-negative real-valued transfer function $H^2(z) > 0$ for which there exists a stable transfer function $H(z)$ such that

$$H^2\left(e^{j\omega}\right) \triangleq \left|H\left(e^{j\omega}\right)\right|^2 \tag{57}$$
$$= H\left(e^{j\omega}\right)^* H\left(e^{j\omega}\right), \tag{58}$$

and polynomials $A^+(z)$ and $B^+(z)$ such that

$$H(z) = \sum_{k=0}^{\infty} h_k z^{-k} \tag{59}$$
$$= \frac{B^+(z)}{A^+(z)}. \tag{60}$$

If the coefficients of $A^+(z)$ and $B^+(z)$ are real then Eqn. (57) yields

$$H^2\left(e^{j\omega}\right) = H\left(e^{j\omega}\right) H\left(e^{-j\omega}\right). \tag{61}$$

We form the TIIR transfer function

$$H^+_{\text{FIR}}(z) = \sum_{k=0}^{N} h_k z^{-k} \tag{62}$$
$$= \frac{B^+(z) - z^{-N} B'^+(z)}{A^+(z)}, \tag{63}$$

as in Eqn. (42). Similarly, we form $H^-_{\text{FIR}}(z)$, as in Eqn. (47). We see immediately, in light of Eqn. (53), that the filter

$$H^2_{\text{FIR}}(z) \triangleq H^+_{\text{FIR}}(z) H^-_{\text{FIR}}(z) \tag{64}$$

has the property that

$$H^2_{\text{FIR}}\left(e^{j\omega}\right) = e^{-jN\omega} \left|H^+_{\text{FIR}}\left(e^{j\omega}\right)\right|^2, \tag{65}$$

and this is obviously a linear-phase filter with group delay

$$\delta = N. \tag{66}$$

This filter may be simply implemented as a cascade of $H^+_{\text{FIR}}(z)$ and $H^-_{\text{FIR}}(z)$, which may be implemented by Eqn. (25) and Table II, respectively.

The relationship between $H^2_{\text{FIR}}(z)$ and $H^2(z)$ is seen by considering the cyclic convolution [6]

$$H^+_{\text{FIR}}\left(e^{j\omega}\right) = \frac{1}{2\pi} \int_0^{2\pi} H\left(e^{j\theta}\right) W_N\left(e^{j(\omega-\theta)}\right) d\theta \tag{67}$$

where

$$W_N\left(e^{j\omega}\right) = \frac{\sin\{(N+1/2)\omega\}}{\sin(\omega/2)}. \tag{68}$$

The filter length $N$ for $H^+_{\text{FIR}}(z)$ and $H^-_{\text{FIR}}(z)$ must be chosen long enough so that the blurring induced by the periodic convolution of $H(e^{j\omega})$ by $W_N(e^{j\omega})$ does not induce too much distortion in the frequency response. Indeed, disregarding quantization noise for the moment, as $N$ grows to infinity, $W_N\left(e^{j\omega}\right)$ approaches an impulse centered at $\omega = 0$, and $H^+_{\text{FIR}}(z)$ converges to $H(z)$. There are constraints due to Eqn. (40) on the filter length which are considered in the next section.

We notice that the phase of the filter $H(z)$ utilized in the design of $H^2_{\text{FIR}}(z)$ is irrelevant, and thus IIR filters, which are often considered to have excessive phase distortion near sharp cutoffs, may be used. Thus, any stable discrete-time IIR filter design such as elliptic, Chebyshev, and Butterworth filters may be chosen as $H(z)$ in Eqn. (57) and transformed into a magnitude-squared filter with linear phase. Additionally, the fact that the magnitude response for $H^2(z)$ is twice the distance from 0 dB compared to the magnitude response of $H(z)$ implies that constraints in the stop band of $H(z)$ need only be half the desired dB design specification, whereas, on the other hand, the ripples in the passband of $H(z)$ must be better than half the desired dB design specification.

### *C. A Refined Truncation Algorithm*

In both the additive and multiplicative designs of the previous two sections, the length $N$ is constrained by the growth of quantization error in the unstable, reverse-conjugate filter $H^{-}_{\mathrm{FIR}}(z)$. This bound is given in Eqn. (40). We note that it is the most *stable* hidden mode of $H^{+}_{\mathrm{FIR}}(z)$ which gives rise to the most unstable hidden mode of $H^{-}_{\mathrm{FIR}}(z)$ due to the conjugate-reciprocal correspondence between their modes. These problematic conjugate-reciprocal modes may restrict the direct implementation in Table II to undesirably short lengths. One way around this problem is to set a significance floor $\lambda_S$ above the floor for numerical precision $\lambda_P$ so that errors with magnitude less than $\lambda_S$ are insignificant, and quantization errors are bounded by $\lambda_P$. Let $H(z)$ be the untruncated IIR impulse response associated with $H^{+}_{\mathrm{FIR}}(z)$ as in Eqn. (60). Since $H(z)$ is stable, we have $|p_k| < 1$, where the $p_k$'s are the poles of $H(z)$, and consequently the hidden modes of $H^{+}_{\mathrm{FIR}}(z)$. The hidden modes of $H^{-}_{\mathrm{FIR}}(z)$ are thus $1/p_k^*$ and are unstable. We perform a partial fraction expansion on $H(z)$ as in Eqns. (28) and (29). We then observe the impulse response for each partial fraction. We set the cutoff point $N_k$ for the $k$-th partial fraction response to be the smallest time after which the maximum impulse response becomes insignificant, i.e., $\forall n > N_k,\ |\mu h_{k,n}| \le \lambda_S$, where $\mu$ is the largest magnitude input. We may solve

$$|h_{k,N_k}| = \lambda_S/\mu \tag{69}$$

numerically for $N_k$ by using Eqn. (29) or by using the approximation

$$\begin{aligned} h_{k,n} &\approx \sum_{\ell=0}^{M_k} \frac{b_{k,\ell}(n-\ell)^{M_k-1}p_k^{n-\ell}}{(M_k-1)!} && (70)\\ &\approx n^{M_k-1}p_k^n \sum_{\ell=0}^{M_k} \frac{b_{k,\ell}p_k^{-\ell}}{(M_k-1)!} && (71)\\ &= B_k n^{M_k-1}p_k^n && (72)\end{aligned}$$

for large $n$, where $B_k$ is implicitly defined by Eqns. (71) and (72). For $M_k = 1$, we have exactly

$$N_k = \frac{\log\left(\left|\frac{\lambda_S}{\mu B_k}\right|\right)}{\log(|p_k|)}. \tag{73}$$

Thus, if $N_k < N$ we may truncate the $k$-th partial fraction response at $N_k$ instead of $N$ without losing significance. However, since $H(z)$ is stable and responses due to the $k$-th partial fraction beyond the $N_k$-th time step are below the significance floor, this does not make any difference. The refinement comes from implementing $H^{-}_{\mathrm{FIR}}(z)$ as a sum of reversed partial fractions based on Eqn. (28) and Eqn. (47), but with the $k$-th partial fraction truncated after $n = N_k$ samples instead of $n = N$ if $N_k < N$. Thus, the unstable mode of $H^{-}_{\mathrm{FIR}}(z)$ due to the pole at $1/p_k^*$ only needs to grow from having an initial magnitude above the significance floor $\lambda_S$ and has less time to accumulate exponentially growing quantization noise. We have

$$H^{-}_{\mathrm{FIR}}(z) = \sum_{k=1}^{N_p} \frac{-zB_k'^{-}(z) + z^{-N_k'}B_k^{-}(z, N_k')}{(1-p_k^*z)^{M_k}}, \tag{74}$$

where

$$N_k' = \begin{cases} N_k, & N_k \le N \\ N, & N_k > N. \end{cases} \tag{75}$$

Eqn. (39) indicates how much quantization noise will accumulate due to the truncated response of length $N = N_k$ of the $k$-th partial fraction of $H^{-}_{\mathrm{FIR}}(z)$ . Assuming that quantization error occurs on the order of the precision floor $\lambda_P$, we have $\sigma_\nu^2 \approx \lambda_P^2$. Also, the significance floor sets a convenient level of noise tolerance, so that may set $\sigma^2_{\mathrm{tol}} = \lambda_S^2$. We may recast Eqn. (39) as

$$\begin{aligned} \lambda_S^2 &\ge \lambda_P^2 \frac{1-|p_k|^{-4N_k'}}{1-|p_k|^{-2}} && (76)\\ &\approx \lambda_P^2 \frac{|p_k|^{-4N_k'}}{|p_k|^{-2}-1}, && (77)\end{aligned}$$

and thus

$$\lambda_P \le \lambda_S |p_k|^{2N_k'-1}\sqrt{1-|p_k|^2}. \tag{78}$$

Combining Eqns. (78) and (73), assuming $M_k = 1$, we arrive at

$$\lambda_P \leq \frac{\lambda_S^3 \sqrt{1-|p_k|^2}}{\mu^2 |p_k| |B_k|^2}. \tag{79}$$

We see that the precision, in bits, for the state variables must be about 3 times greater than the precision of the data in order to prevent significant noise accumulation. This result is intuitive since the significant part of the impulse response must span the dynamic range specified by the largest magnitude output (which may be assumed to be normalized to 1) and $\lambda_S$ over a period of $N_k'$ samples, whereas the noise in $H^-_{\mathrm{FIR}}(z)$ may grow over a period of up to $2N_k'$ samples using the same dynamics. The preceding analysis is not generally applicable when dealing with floating point quantities because the range of significance varies with the exponent. Assuming a constant-energy input, we see that the analysis still holds since the significance and precision floors are approximately constant at steady-state.

C.1 Unit-Magnitude Mode TIIR filters

A third class of linear-phase filters consisting of pure TIIR responses is possible. If we choose $H_u(z)$ to be a TIIR filter of length $N$ such that all of its (hidden) modes are on the unit circle and each mode has multiplicity one, then we are assured that $H_u(z)$ will have linear phase. This may be seen by observing that such a filter must satisfy Eqn. (11). If there are modes with multiplicity greater than one, then Eqn. (11) is not generally satisfied. However, care may be taken to place the zeros so that Eqn. (11) holds.

Applying the analysis of the previous section, we see from Eqn. (73) that the $N_k$'s for this type filter will all be infinity. In this case, it is not necessary to reset the state variables very often because the growth of quantization error will be additive, not exponential.

From a worst-case point of view, the number of samples in which error in the least significant bit may accumulate before reaching the significance floor $\lambda_s$ is approximately

$$N = \lambda_s / \lambda_p, \tag{80}$$

assuming that there is one bit's worth of error in the LSB per sample processed. If the noise is zero-mean, then the accumulated error is a random walk whose standard error grows proportionally to the square root of the number of samples processed. We would like the standard error to be some distance away from the least significant bit of the output signal. For an error accumulation such that the 3-sigma error is below the significance floor, we have the relation

$$3\lambda_p \sqrt{N} < \lambda_s, \tag{81}$$

giving

$$\begin{aligned} N &< (\lambda_s / 3\lambda_p)^2 && (82) \\ &\approx 4^G / 10, && (83) \end{aligned}$$

where $G$ is the number of guard bits. Hence, we see that TIIR filters with hidden modes only on the unit circle have desirable stability properties over the filters described in the previous two sections.

Since unit-magnitude mode TIIR filters are quasi-stable and do not need to be reset every $N$ cycles, the number of computations necessary to stabilize them may be reduced by performing the state-variable reset infrequently. Since an FIR filter has finite memory, it is sufficient to run two parallel filters starting only $N$ steps before the state-variable transfer.

The impulse responses of TIIR filters with unit-magnitude hidden modes are sums of modal responses of the form

$$\begin{aligned} h_k[n] &= P_k(n) p_k^n && (84) \\ &= P_k(n) \exp(j\omega_k n) && (85) \end{aligned}$$

for $n = 0, \ldots, N$, where $p_k$ is the $k$-th unique pole and $P_k(n)$ is an $m_k - 1$-th degree polynomial, where $m_k$ is the multiplicity of the $k$-th pole; these are truncated polynomials times complex exponentials.

Some examples of TIIR filters of the unit-magnitude-modes class are given in Section VI-B.

### *D. Computational Cost*

Assuming that $H^+_{\mathrm{FIR}}(z)$ is a stable filter, we may implement it with $3P+1$ multiplies and $3P$ adds per sample. $H^-_{\mathrm{FIR}}(z)$ is then unstable and must be calculated with parallel state variables as outlined in Section IV, Table II, and the previous section. Table II reflects a few changes from the algorithm outlined in Table I which are necessary to implement the reversed filter $H^+_{\mathrm{FIR}}(z)$. The cost for $H^-_{\mathrm{FIR}}(z)$ is about $4P$ adds and multiplies per sample, as discussed in Section IV. We see, then, that the number of operations is about $7P$ multiplies and adds per sample.

It is interesting to note that implementing a partial fraction (parallel) structure for a transfer function as opposed to a direct implementation, as in Eqn. (25) does not significantly alter the number of multiply-accumulates used as long as the characteristic polynomials are monic and the numerators have degree one less than the denominator. However, each separate factor requires an additional input delay line of length $N_k + M_k$, which might be avoided by clever indexing. Also, the algorithmic overhead may increase, not to mention a greater amount of design effort. The advantage is that greater numerical stability is achieved, and that $N$ may then be chosen arbitrarily large. Also, each mode is present only where it is above the significance threshold $\lambda_S$. Of course, a full factorization is not necessary: it is sufficient to group together modes with the same magnitude, or to implement only the troublesome modes separately.

The unit-magnitude-mode, linear-phase filters described in Section V-C.1 are particularly computationally efficient. Since such filters are intrinsically self-Hermitian, there is no need to implement a mirror filter as in the additive and magnitude-squared design methods. Additionally, the resetting only needs to be implemented infrequently, and may be absorbed into the overhead cost. Hence, the computational cost should be approximately $3P$ operations per sample, plus the refresh overhead.

## VI. Simulations and Examples

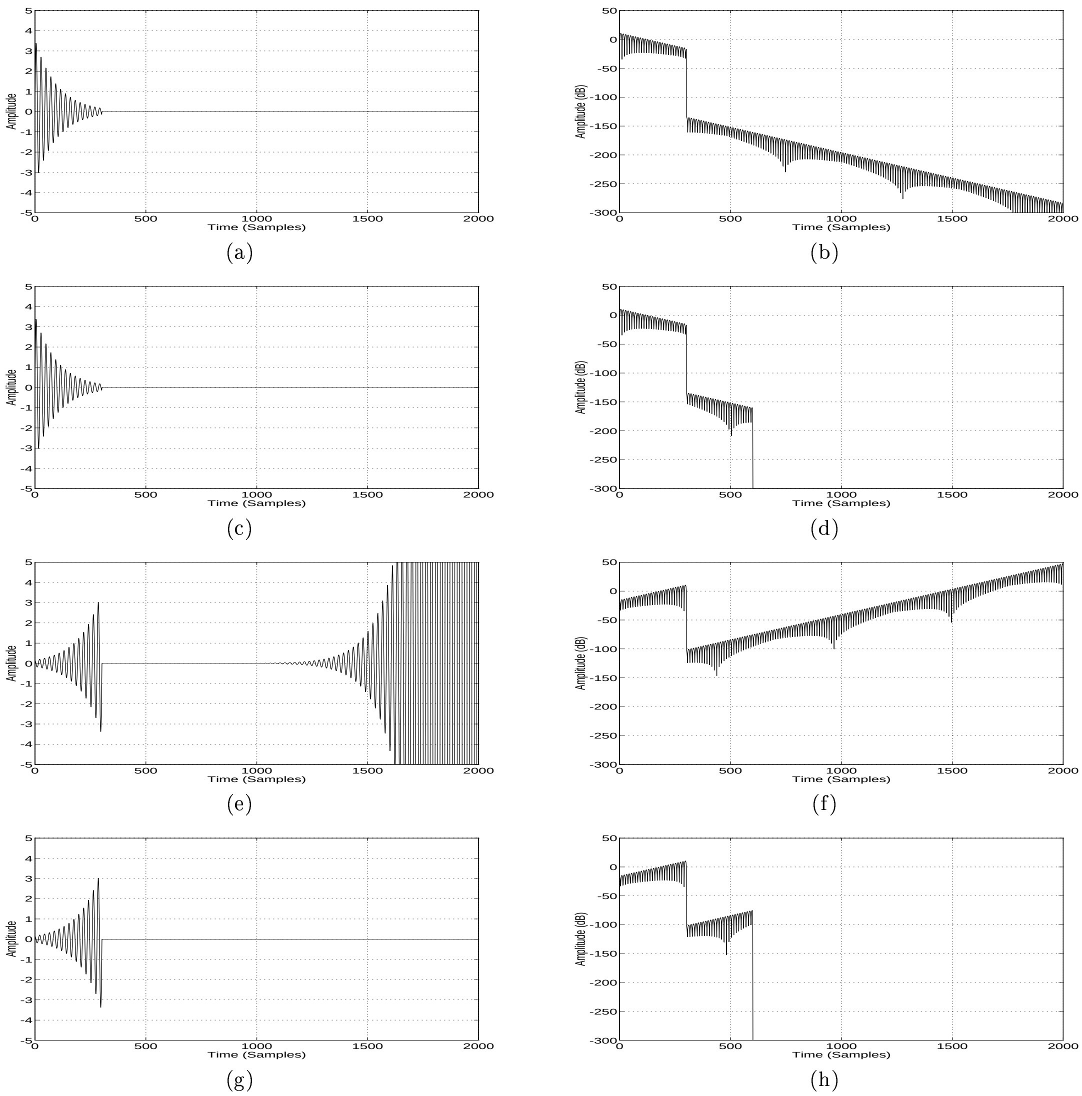


Fig. 1. **(a)** Impulse response of TIIR system of Eqn. (86) truncated after 300 steps using stable implementation. **(c)** Same as (a) except using unstable implementation–state variables refreshed every 300 steps; thus the system has finite memory. **(e)** TIIR impulse response of reversed system of Eqn. (90) using the stable implementation. Note that the quantization error grows without bound. **(g)** Same as (e) except using unstable implementation. Again, state variables refreshed every 300 steps. **(b,d,f, and h)** Magnitude of (a,c,e, and g), respectively, on a decibel scale. The arithmetic was done using double-precision floating point, but with 32-bit single-precision floating-point state variables.

To illustrate the basic idea of a truncated IIR (TIIR) system, we examine the system

$$H(z) = \frac{z^2}{z^2 - 1.9z + 0.98} = \frac{B^+(z)}{A^+(z)} \tag{86}$$

which we wish to truncate after $N = 300$ samples; thus we wish to have an FIR response of 301 steps. We first form the tail-canceling polynomial as in Eqn. (24). We perform synthetic division on $z^{300}B(z)$ by $A(z)$ and obtain the remainder

$$B'^+(z) = -0.162126z + 0.139770 \tag{87}$$

so that $H^+_{\mathrm{FIR}}(z)$ of Eqn. (42) is defined. We plot the impulse response of the direct implementation given in Eqn. (25) in Figures 1(a) and 1(b). We see that at step $n = 301$ the tail of the response has been subtracted off and the response magnitude drops by about 115 dB. Due to quantization errors, there is a residual response. Figures 1(c) and 1(d) show the same system, but implemented using the algorithm presented in Section IV and Table I. Here, the system is forced to have a finite memory and thus the residual response is completely canceled at time step $n = 600$ because the state variables are refreshed every $N = 300$ time steps.

We next form the reverse-conjugate system $H^-_{\mathrm{FIR}}(z)$ as described by Eqn. (47) and arrive at

$$H^-_{\mathrm{FIR}}(z) = \frac{-zB'^-(z) + z^{-N}B^-(z)}{A^-(z)} \tag{88}$$

$$= \frac{-0.139770z^2 + 0.162126z - z^{-N}}{0.98z^2 - 1.9z + 1} \tag{89}$$

$$= \frac{-0.142622z^2 + 0.165435z - 1.020408z^{-N}}{z^2 - 1.938776z + 1.020408}, \tag{90}$$

where, in the last equation, we have normalized by 0.98 in order to have a monic characteristic polynomial. This system has unstable hidden modes and its impulse response is plotted in Figures 1(e) and 1(f). The tail is canceled with about 125 dB attenuation, but the quantization noise grows without bound to overwhelm the signal eventually. Figures 1(g) and 1(h) show how the state variable refresh technique completely cancels the quantization noise before it becomes significant.

### *A. Elliptic-Filter-Based FFIR Design Example*

| Elliptic Filter Coefficients | | | |
|---|---|---|---|
| $a_0$ | 1.00000000 | $b_0$ | 0.05149489 |
| $a_1$ | −5.20086294 | $b_1$ | −0.25706694 |
| $a_2$ | 11.46455205 | $b_2$ | 0.57645267 |
| $a_3$ | −13.68525876 | $b_3$ | −0.74102189 |
| $a_4$ | 9.32002688 | $b_4$ | 0.57645267 |
| $a_5$ | −3.43103178 | $b_5$ | −0.25706694 |
| $a_6$ | 0.53331689 | $b_6$ | 0.05149489 |

TABLE III
ELLIPTIC FILTER COEFFICIENTS FOR THE EXAMPLE IN SECTION VI-A.

We illustrate here the design of a lowpass linear-phase FFIR filter using the magnitude-squared technique outlined in Section V-B. We wish to design a filter which meets the following criteria:

(1) Passband $(0.00, 0.10)$, in fractions of $f_s/2$, with at most 0.080 dB maximum peak-to-peak ripple, where $f_s$ is the sampling frequency
(2) Stopband $(0.11, 1.00)$ with at least 50 dB attenuation
(3) Linear phase
(4) 15 bits of fixed-point significance in the mantissa.
(5) Maximum input amplitude is $\mu = 1.0$

Because of the sharp transition band in the interval $(0.10, 0.11)$, we select an elliptic filter as the basis for our design [7]. We realize that our choice of $H(z)$ only needs to meet half the specifications of (1) and (2) above, and that (3) is irrelevant. Using Matlab, we find that an order 6 filter suffices to give us a filter $H(z)$ with the desired properties such that $H^2(z)$ satisfies the above criteria. The arguments to Matlab's **ellip** function are

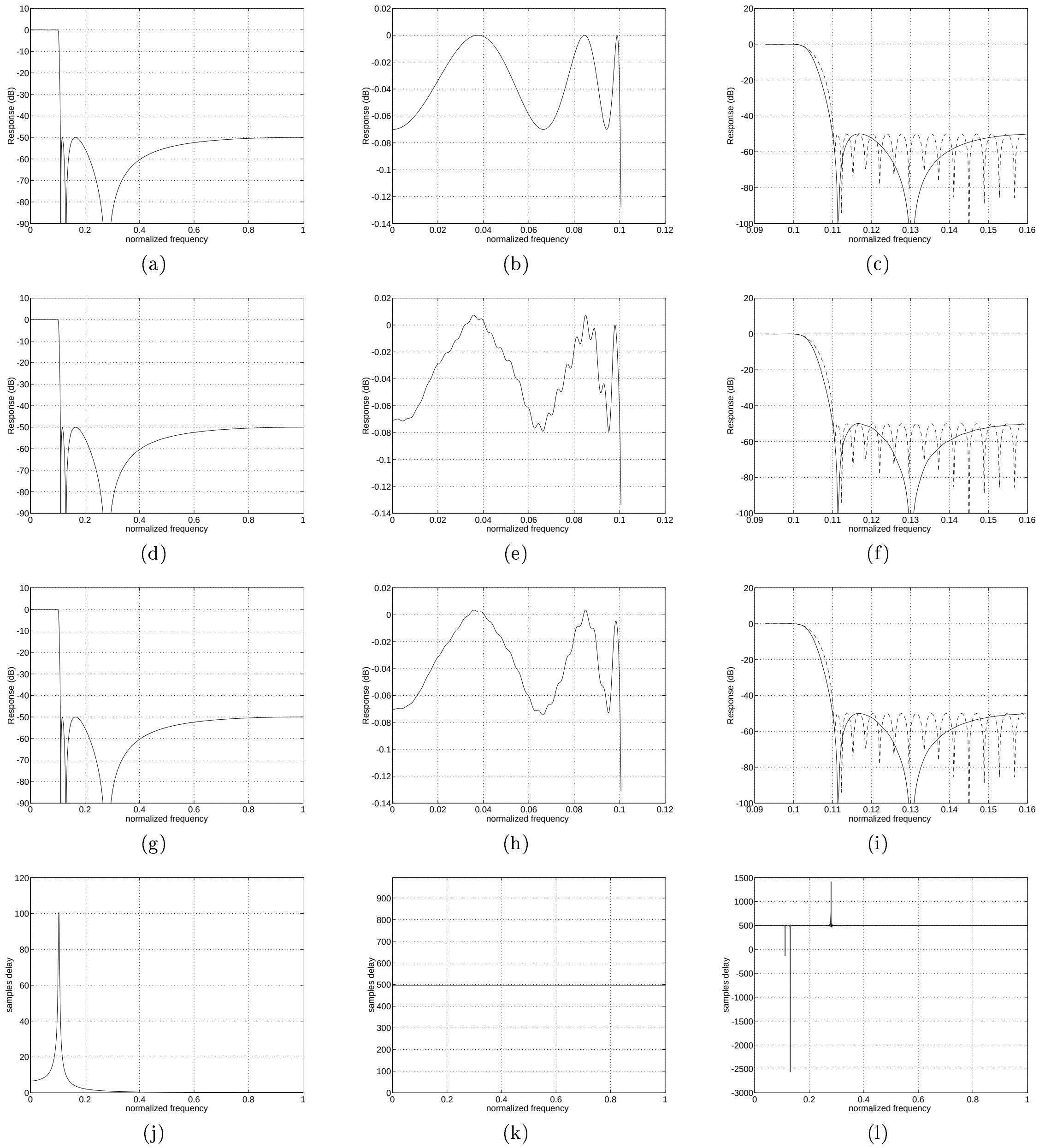


Fig. 2. Filter responses from the linear-phase elliptic filter design example in Section VI-A. **(a-c)** Views of the untruncated response of $H^2\left(e^{j\omega}\right)$. **(d-f)** Views of the response of $H^2_{\mathrm{FIR}}\left(e^{j\omega}\right)$, with truncation at $N = 497$. Note the ringing in the passband in **(e)** due to Gibb's effect. **(g-i)** Views of the hybrid filter $H\left(e^{j\omega}\right) H^-_{\mathrm{FIR}}\left(e^{j\omega}\right)$. Views **(c), (f)**, and **(i)** have portions of an order 500 FIR filter designed by the Parks-McClellan algorithm to have the same specs as the target filter, superimposed for comparison. **(j)** Group delay of the forward filter $H(z)$ used as the design basis of $H^2_{\mathrm{FIR}}(z)$. **(k)** Group delay of the $H^2_{\mathrm{FIR}}\left(e^{j\omega}\right)$. It is exactly 497 time steps. **(l)** Group delay of the hybrid filter $H\left(e^{j\omega}\right) H^-_{\mathrm{FIR}}\left(e^{j\omega}\right)$. (The large spikes occur at the zeros of transmission where the phase is discontinuous and the group delay is technically infinity.)

| $k$ | Poles | Residue | $N_k$ | $\lambda_P(dB)$ |
|---|---|---|---|---|
| 1 | $0.93560805 + j0.31706231$ | $0.00710587 + j0.01100914$ | 497 | $-211.6943$ |
| 2 | $0.93560805 - j0.31706231$ | $0.00710587 - j0.01100914$ | 497 | $-211.6943$ |
| 3 | $0.88941945 + j0.28954430$ | $-0.06609826 + j0.02150355$ | 116 | $-233.0556$ |
| 4 | $0.88941945 - j0.28954430$ | $-0.06609826 - j0.02150356$ | 116 | $-233.0556$ |
| 5 | $0.77540397 + j0.15290736$ | $0.06436785 - j0.21546997$ | 38 | $-247.2166$ |
| 6 | $0.77540397 - j0.15290736$ | $0.06436785 + j0.21546997$ | 38 | $-247.2166$ |
| Direct term: $h_0 = 0.05149489$ | | | | |

TABLE IV

ELLIPTIC FILTER POLES, MAGNITUDES AND DECAY TIMES TO SIGNIFICANCE LEVEL $\lambda_S$ FOR THE EXAMPLE IN SECTION VI-A. THE $N_k$'S ARE CALCULATED USING EQN. (73) WITH $\lambda_S = 2^{-15}$ AND MAXIMUM INPUT AMPLITUDE $\mu = 1$. THE $\lambda_P$'S DENOTE THE PRECISION FLOOR NECESSARY TO IMPLEMENT THIS FILTER.

| | $\ell = 1$ | $\ell = 2$ |
|---|---|---|
| $a_{1,\ell}$ | $-1.87121609$ | $0.97589092$ |
| $b_{1,\ell}$ | $0.01421174$ | $-0.02027779$ |
| $b'_{1,\ell}$ | 6.08847078e–05 | –5.74152962e–05 |
| $a_{2,\ell}$ | $-1.77883890$ | $0.87490286$ |
| $b_{2,\ell}$ | $-0.13219652$ | $0.10512570$ |
| $b'_{2,\ell}$ | –3.88583695e–06 | –1.38234153e–05 |
| $a_{3,\ell}$ | $-1.55080795$ | $0.62463198$ |
| $b_{3,\ell}$ | $0.12873570$ | $-0.03392829$ |
| $b'_{3,\ell}$ | 5.80618310e–05 | –4.35418795e–05 |

TABLE V

COEFFICIENTS FOR THE MODAL DECOMPOSITION OF THE ELLIPTIC FILTER IN EQN. (91) AND EQN. (92). THE $b'_{k,\ell}$ ARE CALCULATED USING SYNTHETIC DIVISION AND $N_k$ FROM TABLE IV.

- $N = 6$
- Passband ripple = 0.035 dB
- Stopband attenuation = 25 dB
- Cutoff frequency = 0.10 (normalized units).

We have allowed an additional 0.005 dB margin in the passband to anticipate Gibbs ringing due to sequence truncation expressed in Eqn. (67). For the form given in Eqn. (5) the coefficients are given in Table III. Since the multiplicity of each pole is one, we may use Eqn. (73) to calculate the decay times $N_k$ to the significance floor. We round up in each case. The poles are given in Table IV as well as the magnitude and $N_k$ for each pole. Because the coefficients come in complex-conjugate pairs, we may rearrange $H(z)$ into three 2nd-order real terms so that

$$\begin{aligned} H(z) = h_0 &+ \frac{b_{1,1}z^{-1} + b_{1,2}z^{-2}}{1 + a_{1,1}z^{-1} + a_{1,2}z^{-2}} \\ &+ \frac{b_{2,1}z^{-1} + b_{2,2}z^{-2}}{1 + a_{2,1}z^{-1} + a_{2,2}z^{-2}} + \frac{b_{3,1}z^{-1} + b_{3,2}z^{-2}}{1 + a_{3,1}z^{-1} + a_{3,2}z^{-2}}, \end{aligned} \tag{91}$$

where the coefficients are given in Table V. We then implement each term separately in the truncated form given in Eqn. (25) to form $H^+_{\rm FIR}(z)$ using each term's decay-to-significance times $N_1$, $N_3$, and $N_5$ for the cutoffs. The coefficients are calculated by using synthetic division and are listed in Table V. We choose $N = 497$ as the cutoff point for truncating the filter since this is the largest $N_k$. Thus we may implement the forward filter as

$$\begin{aligned} H^+_{\rm FIR}(z) = h_0 &+ \frac{b_{1,1}z^{-1} + b_{1,2}z^{-2} - b'_{1,1}z^{-(N_1+1)} - b'_{1,2}z^{-(N_1+2)}}{1 + a_{1,1}z^{-1} + a_{1,2}z^{-2}} \\ &+ \frac{b_{2,1}z^{-1} + b_{2,2}z^{-2} - b'_{2,1}z^{-(N_3+1)} - b'_{2,2}z^{-(N_3+2)}}{1 + a_{2,1}z^{-1} + a_{2,2}z^{-2}} \end{aligned} \tag{92}$$

$$+\frac{b_{3,1}z^{-1}+b_{3,2}z^{-2}-b'_{3,1}z^{-(N_5+1)}-b'_{3,2}z^{-(N_5+2)}}{1+a_{3,1}z^{-1}+a_{3,2}z^{-2}}.$$

Using Eqn. (47) we form

$$\begin{aligned}H^-_{\mathrm{FIR}}(z) = h_0 &+ z^{-N-N_1}\frac{-b'_{1,2}-b'_{1,1}z^{-1}+b_{1,2}z^{-N_1}+b_{1,1}z^{-(N_1+1)}}{a_{1,2}+a_{1,1}z^{-1}+z^{-2}}\\ &+ z^{-N-N_3}\frac{-b'_{2,2}-b'_{2,1}z^{-1}+b_{2,2}z^{-N_3}+b_{2,1}z^{-(N_3+1)}}{a_{2,2}+a_{2,1}z^{-1}+z^{-2}}\\ &+ z^{-N-N_5}\frac{-b'_{3,2}-b'_{3,1}z^{-1}+b_{3,2}z^{-N_5}+b_{3,1}z^{-(N_5+1)}}{a_{3,2}+a_{3,1}z^{-1}+z^{-2}}\end{aligned} \tag{93}$$

from the corresponding parts in Eqn. (92). Notice that each term is delayed by an appropriate $N-N_k$ steps in order to align the phases of the response properly. To implement this filter, we use the algorithm of Table II and Section IV due to the unstable hidden modes of each dynamic term. We may wish to normalize the coefficients by $a_{k,2}$ so that the characteristic polynomials are monic in each term. At this point we form $H^2_{\mathrm{FIR}}(z)$ by cascading the output of $H^-_{\mathrm{FIR}}(z)$ into the input of $H^+_{\mathrm{FIR}}(z)$. This order may be preferable since the state-variable reinitialization scheme used in Section IV may introduce some discontinuities near or below the significance floor $\lambda_S$. Thus, postfiltering by $H^+_{\mathrm{FIR}}(z)$ attenuates any such additional noise, however insignificant.

We show in Figure 2 the results of our filter implementation. Figures 2(a)-(c) show the ideal untruncated $H^2(z)$ response from our design. Its decibel response is simply double that of $H(z)$ from Eqn. (91). Figures 2(d)-(f) show the truncated response of $H^2_{\mathrm{FIR}}(z)=H^-_{\mathrm{FIR}}(z)H^+_{\mathrm{FIR}}(z)$. Notice the presence of Gibbs overshoot due to the truncation of the impulse response. Figures 2(g)-(i) show the implementation $H^-_{\mathrm{FIR}}(z)H(z)$ in which the tail of the forward response is not truncated. The response is smoother, but there are some slight deviations from phase linearity due to the asymmetry. The error due to the tail response of $H(z)$ should be small, however, since the tail of each mode for $H^+_{\mathrm{FIR}}(z)$ is cut off at the significance floor $\lambda_S$. The group delays for $H(z)$, $H^2_{\mathrm{FIR}}(z)$, and $H^-_{\mathrm{FIR}}(z)H(z)$ are shown in Figures 2(j)-($l$).

### *B. Windows and Polynomials*

Unit-magnitude mode TIIR responses are convenient for generating well-known impulse response functions for computing weighted averages of an input sequence. Often in signal processing applications, an exponential window is undesirable for averaging highly non-stationary data because of the long tail. Finite-length windows with polynomial terms may be generated using multiple poles at $z=1$ and using Eqn. (29). The terms due to the binomial coefficients generate the polynomial coefficients. The FFIR algorithm provides a cost-effective method for implementing such weighting sequences. For example, the simple $N$-point rectangular window may be generated as a quasi-stable TIIR filter with filter dynamics

$$H_{\mathrm{rectangular}}(z)=\frac{1-z^{-N}}{1-z^{-1}}. \tag{94}$$

The $2N$-point Bartlett window is

$$H_{\mathrm{Bartlett}}(z)=\frac{1}{N}\frac{1-z^{-N}-z^{-(N+1)}+z^{-(2N+1)}}{1-2z^{-1}+z^{-2}}. \tag{95}$$

The $N$-point Hanning window, also known as the *Hann* window, is

$$H_{\mathrm{Hanning}}(z)=\frac{1}{2}\left(\frac{1-z^{-N}}{1-z^{-1}}-\frac{1-\cos\left(\frac{2\pi}{N-1}\right)(z^{-1}+z^{-N})+z^{-(N+1)}}{1-2\cos\left(\frac{2\pi}{N-1}\right)z^{-1}+z^{-2}}\right). \tag{96}$$

The $N$-point Hamming window is

$$H_{\mathrm{Hamming}}(z)=0.54\frac{1-z^{-N}}{1-z^{-1}}-0.46\frac{1-\cos\left(\frac{2\pi}{N-1}\right)(z^{-1}+z^{-N})+z^{-(N+1)}}{1-2\cos\left(\frac{2\pi}{N-1}\right)z^{-1}+z^{-2}}. \tag{97}$$

An example of a useful polynomial impulse response is the Kay window [2] used for statistically efficient frequency estimation based on averaged phase differences with weighting coefficients

$$w_{\mathrm{Kay}}[n]=\frac{6N}{N^2-1}\left\{\frac{n}{N}-\left(\frac{n}{N}\right)^2\right\} \tag{98}$$

for $1 \le n \le N-1$.

$$H_{\mathrm{Kay}}(z) = \frac{6}{N(N^2-1)} \times \frac{(N-1)z^{-1} - (N+1)z^{-2} + (N+1)z^{-N-1} - (N-1)z^{-N-2}}{(1-z^{-1})^3}. \tag{99}$$

This implementation requires only six adds and two multiplies, independent of $N$.

## VII. Applications

We list here some of the many possible applications for TIIR filters.

### A. Digital speaker crossover networks and mixers for digital audio applications

In the past, commercial digital audio systems designers have had to choose between efficient, nonlinear-phase IIR filters and linear-phase, but computationally expensive, FIR filters. Mixing the outputs from nonlinear-phase filters may result in unwanted nulls in the combined frequency response due to cancellation of signals with different frequency-dependent delays. The linear-phase TIIR algorithm combines the best of these features from both types of filter.

### B. Multirate Filtering

One disadvantage of the fast FIR filtering methods is that, although they are computationally efficient, they are limited to hop sizes of one. Multirate [5] or STFT techniques, which downsample the data stream to account for a reduction in bandwidth, can be quite efficient, thus reducing the relative advantage of TIIR filtering over FIR filters. However, polyphase interpolating filters for upsampling in multirate systems may be easily implemented as TIIR filters, providing a great savings in computation, thus gaining the best features of both filter design approaches. The development of applications of TIIR theory to multirate filter design is beyond the scope of this paper.

### C. High-Resolution, Optimal Frequency Estimator

Kay derived the optimal weighting sequence for estimating the frequency of a pure-frequency sinusoid in Gaussian noise from finite phase differences [8]. The window given in Eqn. (98) attains the Cramér-Rao bound for moderate SNR. Eqn. (99) shows how a running average may be computed using only 2 multiplies and 6 adds per sample, independent of the window length. Such a system may be used for computing a running estimate of instantaneous frequency.

### D. Polynomial Impulse Responses

Sometimes a finite-length window is desired for processing non-stationary signals where long exponential tails are undesirable. Running windows such as the Bartlett, Hamming, or Hanning Window give a more localized and symmetric average than a decaying exponential. Additionally, polynomial responses with sinsuoidal modulation may be constructed easily, as seen in Section V-C.1.

## VIII. Summary

The power of the fast FIR algorithm stems from the ability to form complicated FIR sequences as the quotient of two polynomials with relatively few non-zero coefficients. The resulting cancellation of all the dynamic modes yields the desired FIR sequence. This may be thought of as "filter compression," in which low-entropy redundancies in certain kinds of FIR sequences are encoded using generating functions based on IIR dynamics. We utilize unstable as well as stable IIR dynamics in creating our TIIR responses by exploiting a state-variable-resetting technique which takes advantage of the fact that FIR responses have finite memory. Another technique introduced here is the use of synthetic division to generate a tail-canceling complement filter. Finally, the technique described in Section V-B allows the straightforward use of conventional IIR filter design methods for creating linear-phase fast FIR filters.

## References


[1] T. Saramäki and A. T. Fam, "Properties and structures of linear-phase FIR filters based on switching and resetting of IIR filters," in *Proc. IEEE International Symposium on Circuits and Systems*, (Piscataway, NJ), pp. 3271–3274, IEEE Service Center, 1990.

[2] S. Kay, "Statistically/computationally efficient frequency estimation," in *Proc. Int. Conf. Acoustics, Speech, and Sig. Proc.*, (Piscataway, NJ), pp. 2292–2295, IEEE Service Center, 1988.

[3] A. T. Fam, "FIR filters that approach IIR filters in their computational efficiency," in *Twenty-First Asilomar Conference on Signals, Systems, and Computers, Pacific Grove*, pp. 28–30, Nov. 1987.

[4] A. Papoulis, *Signal Analysis.* New York: McGraw-Hill, 1977.

[5] P. P. Vaidyanathan, *Multirate Systems and Filter Banks.* Englewood Cliffs, NJ: Prentice-Hall, 1993.

[6] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing.* Englewood Cliffs, N.J.: Prentice Hall, 1989.

[7] A. Gray and J. Markel, "A computer program for designing elliptic filters," *IEEE Tr. Acoustics, Speech, and Signal Processing*, vol. 24, pp. 529–538, Dec. 1976.

[8] S. Kay, “A fast and accurate single frequency estimator,” *IEEE Tr. Acoustics, Speech, and Signal Processing*, vol. 37, pp. 1987–1990, Dec. 1989.

**Avery Wang** received the B.S. degree in Mathematics, with distinction, the M.S. degree in Mathematics, and the M.S. degree in Electrical Engineering, all in 1988 from Stanford University. He won a National Science Foundation Graduate Fellowship Award in 1988, and a Fulbright Graduate Scholarship in 1990, with which he spent two years at the Ruhr-Universität Bochum, Germany, studying neurosciences and computational models of auditory perception. He returned to Stanford to study at the Center for Computer Research in Music and Acoustics and received his Ph.D. in Electrical Engineering in 1994. He is currently at Chromatic Research doing algorithm design and implementation. His research interests include auditory source separation, statistical signal processing, and music synthesis. Other interests include molecular biology, neurosciences, mathematics, and languages.

**Julius O. Smith** received the B.S.E.E. degree from Rice University, Houston, TX, in 1975. He received the M.S. and Ph.D. degrees in E.E. from Stanford University, Stanford, CA, in 1978 and 1983, respectively. His Ph.D. research involved the application of digital signal processing and system identification techniques to the modeling and synthesis of the violin, clarinet, reverberant spaces, and other musical systems. From 1975 to 1977 he worked in the Signal Processing Department at ESL, Sunnyvale, CA, on systems for digital communications. From 1982 to 1986 he was with the Adaptive Systems Department at Systems Control Technology, Palo Alto, CA, where he worked in the areas of adaptive filtering and spectral estimation. From 1986 to 1991 he was employed at NeXT Computer, Inc., responsible for sound, music, and signal processing software for the NeXT computer workstation. Since then he has been an Associate Professor at the Center for Computer Research in Music and Acoustics (CCRMA) at Stanford teaching courses in signal processing and music technology, and pursuing research in signal processing techniques applied to music and audio. For more information, see http://www-ccrma.stanford.edu/~jos/.