Robust IIR Digital Filter Sharpening

Gerald D. Cain
DSP Creations Ltd.
London, UK
cain@dspcreations.com

Anush Yardim
Dept. of Electronic Engineering
Royal Holloway, University of London
Surrey, UK
anush.yardim@rhul.ac.uk

Fredric J. Harris
Dept. of Electrical and Computer Engineering, UC San Diego
California, USA
fjharris@eng.ucsd.edu

***Abstract*—Digital filter sharpening aims to improve the performance of a prototype filter by cascading it with a "partial compensator" incorporating multiple uses of that same coefficient set. Although the Kaiser & Hamming FIR sharpening structure has long enjoyed popularity, there has been a widespread mistaken impression that no corresponding structure exists for IIR sharpening. In this paper we review the features of three available methods and also introduce a new sharpener that operates with no restrictions on phase conditions for prototypes and without requiring root-finding. The use of conjugate-reversal of numerator coefficient vectors, along with prototype pairing and allpass filtering, combine to facilitate robust internal sharpener delay alignment. Performance of the new structure is highlighted here in one complex and three real filter sharpening examples.**

Keywords—IIR filter sharpening, complex filters, frequency response, sharpening polynomials.

I. Introduction

Digital filter sharpening is most appropriate in the context of flat-top multiband filtering (typified by the trusty lowpass filter) with desired passband and stopband gain values unity and zero respectively. The goal of sharpening is to reduce peak errors in passbands and to increase peak attenuation in stopbands, thereby compacting transition bands. Other considerations – such as specification of resulting impulse response symmetry patterns – cannot in general be guaranteed in sharpening processes.

Kaiser and Hamming in the seminal paper [1] pioneered their process of filter improvement where the prototype's output is further filtered by a cascade of two subfilters that employ prototype-specific elements. The initial subfilter is just a replica of the prototype and the second is a specially-configured Two-path subfilter:

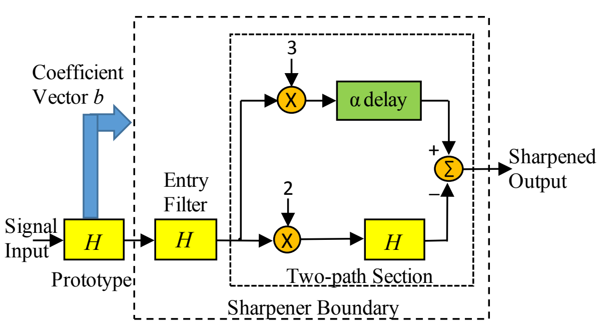

Fig.1. Kaiser & Hamming ("K&H") Sharpening

The effect of the repeated "Entry Filter" is of course that the stopband's deviation from zero becomes smaller but the passband deviation from unity becomes larger. It is the task of the "Two-path Section" to reduce the passband ripple levels without significant reduction in the stopband improvement caused by the cascaded prototype.

K&H sharpening offers a very useful, easily-applied, "quick-look" facility for exploring the effect of band-flattening gain improvements. This capability comes at a cost of increased computational burden, amounting to $3N$–2 equivalent coefficients for any N-coefficient filter explored. Nevertheless, sharpening earns its place as "a valued and natural addition to the DSP person's toolbox" [2].

The K&H approach has a trio of severe shortcomings however: the prototype must be FIR, must be of the even-symmetric linear-phase FIR variety, and must further have an odd-length impulse response. This restricts K&H sharpening to just a small corner of the digital filter landscape and presents a huge barrier to wider uptake of sharpening. Much effort is expended on modifications of the basic structure to accommodate special situations and classes of filters, such as minimum-phase FIR filters [3], [4].

There is a common misconception in the DSP literature that there is no method available for sharpening of IIR prototype filters [5]. Our chief objectives in this paper are to highlight the characteristics of three existing IIR sharpeners, to introduce a new IIR sharpening structure, and to demonstrate its versatility and enhanced applicability compared to the earlier approaches.

II. Existing IIR Sharpening Methods

We first set out our filter notation to clarify discussion. The z-domain expression for our general IIR filters is

$$H(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z^{-1} + \cdots + b_{N-1} z^{-(N-1)}}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_M z^{-M}} \qquad (1)$$

which can be expressed in terms of M poles and (N-1) zeros in the z domain. Evaluation on the z-plane unit circle gives us the frequency response $H(e^{j2\pi\nu})$, where normalized frequency ν - in its primary period - ranges from –0.5 to +0.5 and "gain" is $|H(e^{j2\pi\nu})|$. We speak of filter coefficient vectors b and a, composed of the numerator and denominator polynomial coefficients. In some cases the numerator will be linear-phase with group delay

$$\tau_b(\nu) = (N-1)/2 = \alpha \,. \qquad (2)$$

The only IIR sharpening method known by us to have led to published performance results is that by one of the current authors in [2]. Reference [2] treated, to very good effect, sharpening of Elliptic filters and set the scene for other IIR filter types which have linear-phase numerators (by constructing a "super-prototype" consisting of two prototypes joined in cascade). This sharpening method (which we refer to here as "*harris*" sharpening) was specifically developed to retain the "3, –2" pathway multipliers, but in a more expanded rendition of the general structure we see in Fig. 1.

The reader can study the details in [2], where each single *H* block is replaced by a prototype pair and the upper pathway of the Two-path Section features an (IIR) allpass filter as well as an adjustable delay block. We enthusiastically employ *harris* for benchmarking comparisons in this paper.

Soon after the K&H sharpening method came to prominence, Claasen and Mecklenbrauker introduced a different sharpening scheme for IIR filters [6] which we call "C&M". They too observed the critical need for **prototype pairs** (for reasons we elaborate in Section III) in conjunction with "2, –1" pathway multipliers. This ushered in a "Sharpening Polynomial" quite different from that of [1]. As is also true for [1], a family of such polynomials (involving higher-order Multi-path Sections) was devised.

Then in [7] Jalaludeen, Mathew, and Chander provided some much-needed further detail to this theoretical setting, breaking free of the limitation of linear-phase numerators. They arrived at what appears to be the first all-filter sharpening arrangement. We have found no report of practical testing of their "JMC" theory prior to our own work in [8].

To give a feel for the style of the multiband, two-level compensation action that is our sharpening goal, Fig. 2 displays the resultants of all three techniques mentioned above. In addition, we have also included the "CYH" method which will be explained in Section III below. The prototype is a crude IIR having 12 poles and 12 zeros. Crucially, its numerator coefficient set is linear-phase.

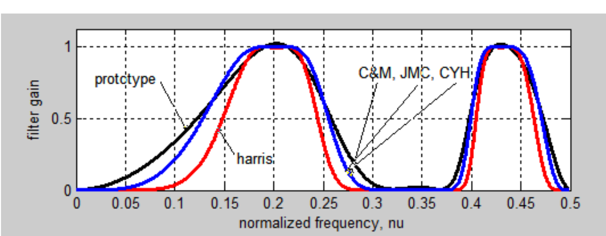

Fig. 2. Crude IIR order 12 prototype multiband filter subjected to 4 sharpeners

Notice that three methods yield perfectly coincident gain curves. This has been repeated in hundreds of sharpening trials, with prototypes ranging from the crude to the initially excellent. Although we have not formulated any equivalence proof as such, our conjecture is that, while individual implementation details may vary greatly, the ultimate determiner of "sharpening shape" is the Sharpening Polynomial. This is an appealing position to adopt, because "C&M" is incapable of operation outside the linear-phase numerator situation (and moreover requires modification from the original if *M* does not equal *N*–1).

On the other hand, "JMC" warrants attention because of its broad scope. JMC requires a total of three ancillary allpass filters to operate and we have found it cumbersome to use. Much worse is JMC's total reliance on zero-finding and coefficient subvector re-assembly. This threatens numerical inaccuracies.

In summary, we cannot recommend C&M and JMC for general usage, leaving our enthusiasm only for *harris* and the new sharpener we are about to describe.

III. Prototype Pairing and the New CYH Sharpening Structure

An underlying problem in the sharpening structures we have treated comes in the Two-path Section, where it is vital that both pathways are in time alignment. Our two most trustworthy components for ensuring alignment are unit delay blocks and allpass filters. Every successful sharpener is obliged to effectively manage the deployment of delay as an "antidote" to the nonlinear group delay injected by the remaining pathway elements. Claasen and Mecklenbrauker in [6] put forward a pivotal observation making it obvious that prototype pairing was advantageous in dealing with poles:

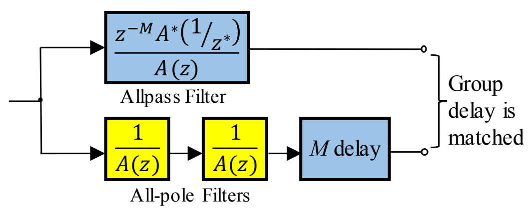

Fig. 3. A key two-path delay relationship

For Fig. 3, we denote the group delay of the All-pole filter pair (shown yellow) as $\tau_{pair}(\nu)$. A separate allpass filter (which is formed by the denominator of **one** of the cascaded pair's transfer function) is named $\tau_{Ap}(\nu)$. Then, following on from [9, p.252], it is straightforward to derive this expression:

$$\tau_{Ap}(\nu) = \tau_{pair}(\nu) + M \qquad (3)$$

where *M* is the order of the pole-only transfer function in one of the prototypes. This guiding relationship has proven to be pivotal in assigning delay assets in the paths of the Two-path and higher Sections. (A corresponding detailed argument cast as phase equalization is given in [2]). Incorporating cascaded prototype pairs forces the use of even powers of transfer function magnitude in sharpening polynomials. The lowest order sharpening candidate polynomials that we favor as we move into consideration of our new structure are:

K&H: $P_{K\&H} = |H|^2(3-2|H|)$ from Fig.1 (4)

harris: $P_{harris} = |H|^4(3-2|H|^2)$ from [2] (5)

CYH: $P_{CYH} = |H|^2(2-|H|^2)$ see Fig. 4 (6)

(Here we suppress mention of frequency considerations by defining the symbol $|H| = |H(e^{j2\pi\nu})|$.

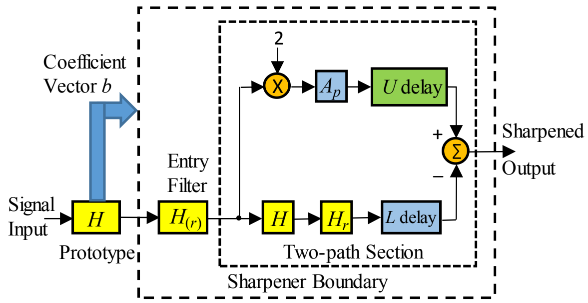

Fig. 4. The CYH Conjugate-Reversal Sharpening Processor

Though invisible in (4)-(6), phase plays a pivotal role in sharpening (for numerators as well as denominators) and we have embedded shorthand information about that in the block symbols given in Fig. 4:

H: The original prototype filter supplying all coefficient data to the process, as well passing on its signal output; one such duplicate is also located inside the Two-path Section.

H_r: A duplicate which has undergone conjugate-reversal of its "b" coefficient vector.

$H_{(r)}$: An "Entry Filter" duplicate which is software-switchable between H and H_r states.

It is easy to appreciate that the numerator portion of the pairing in the Two-path Section is locked into a matched filter coupling, hence constituting a linear-phase combination possessing group delay 2α (which is integer-valued for any prototype). Therefore, taking into account the pole compensation due to the allpass block labelled "A_p", both Two-path branches are perfectly delay-aligned, and none of sizing difficulties of K&H occur. If there are no prototype poles and we wish to deliver linear-phase overall, then we also switch the Entry Filter to the conjugate-reversal state [in MATLAB, this is achieved by the simple command conj(fliplr(b))]. Otherwise our style of sharpening delivers a mixed-phase resultant.

Table 1 compares our two sharpeners in terms of overall equivalent coefficients for sharpened resultants.

TABLE I. NUMBER OF EQUIVALENT COEFFICIENTS

Equivalent	Harris	CYH
Numerator	$(4N–4)+C$	$(2N–2)+C$
Denominator	$(7M+1)$	$(5M+1)$
where term C is max[$(2N+M–1)$, $(3M+1)$]		

In Fig. 5 are two typical sharpening outcomes for a lowpass prototype, demonstrating that CYH is better at flattening the passband, while *harris* does a phenomenal job of pulling down stopband sidelobes. Below Fig. 5 and elsewhere here we use clusters of four-element MATLAB cell arrays as "quality measurement strips" to compactly quantify the measured merit of the prototype lowpass filters. For instance, the first entry for Qproto is the value for the prototype's peak passband error δ_p = 2.64e-5, the second is its measured transition width separating pass and stop bands in normalized frequency (equalling 0.08), the third is stopband attenuation in dB=32.79, and the fourth is peak stopband (linear) error δ_s = 0.02293 for the prototype.

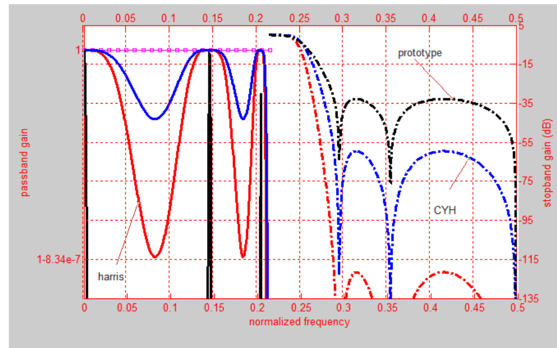

Fig. 5. Gain plots for a 5th order Elliptic prototype from [2]
(*harris* equiv. coeffs: 36 numerator and 36 denominator)
(CYH equiv. coeffs.: 26 numerator and 26 denominator)

Qproto	={2.64e-5;	0.08 Hz;	32.79 dB;	0.02293}
harris: Q	={833.0e-5;	0.08 Hz;	121.64 dB;	8.28e-7 }
harris: deltaQ	= { 316.21;	1 (Hz);	88.84 dB;	27676.0}
CYH: Q	= {278e-5;	0.08 Hz;	59.57 dB;	0.00105}
CYH: deltaQ	= {948.29;	1 (Hz);	26.78 dB;	21.82}

Meanwhile, three of the deltaQ array entries indicate sharpening changes as the before-to-after sharpening ratios. So deltaQ{1}=316.21 is the "peak passband improvement ratio" for *harris*, while deltaQ{4} indicates its "peak stopband reduction factor" is over 27 thousand! But deltaQ{3} differs in that it is not a ratio, rather giving the additional dB stopband attenuation level attained (88.84 dB) via the sharpening. Of course, all deltaQ entries being greater than unity will signify a thoroughly successful sharpening event (as happened for Fig. 5).

As noted in [2] this filter was not really in initial need of passband improvement. Although off the vertical scale in Fig. 5, its δ_p was already small, at some –2 millidB. But the improvements were almost immeasurably good, dropping to some –7 microdB for *harris* and even –2 microdB for CYH.

IV. INFLUENCE OF THE SHARPENING POLYNOMIALS

Initial examination of the somewhat flattened passband gains in Fig. 5 raises these queries: Why is *harris*'s excursion more pronounced than the one for CYH? Why are all sharpened passband gain excursions clipped to unity level? All such questions are answered by studying the sharpening polynomials (5) and (6) (which in turn have dictated the structures presented in Fig. 1 and Fig. 4).

Many examples of sharpening polynomials can be found [3], [10], [11]. Such polynomials act on filter transfer functions to cause gain values to flatten in the region of zero and unity, so have the common requirement of equalling unity at argument value one, and zero at the origin. Also there must be a leading multiplier term with $|H|^2$ (representing the prototype and the Entry Filter) as a factor for all the structures we take as "standard" for purposes of this paper.

Following the hint in [6], we take the opportunity to elevate the terminology from the popularly used term "Amplitude Change Function" [1] to the more appropriate "Magnitude Change Function" ("MCF") when referring to the type of sharpening polynomial we use. In Fig. 6 we show K&H only as historical reference (and to see how dramatically the horizontal square-magnitude scaling for (4) has horizontally compressed the magenta curve into becoming *harris*'s red curve). The two Magnitude Change Functions (5) and (6) actually utilized in this paper provide us with much practical information.

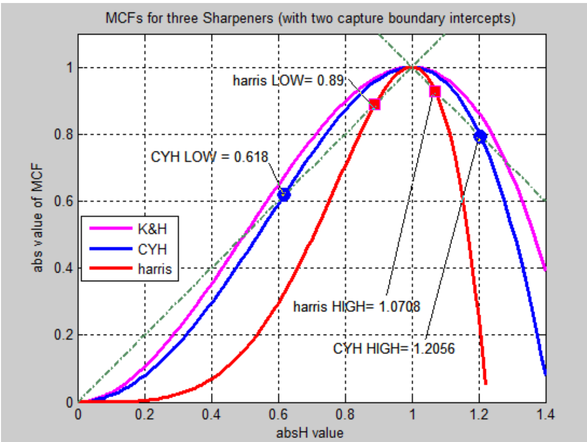

Fig. 6. Magnitude Change Functions for three sharpeners

The downturn in both blue and red plots beyond their peaks explains why our sharpened passbands all exhibit down-going ripples. Since the CYH curve mostly lies above that for *harris*, any passband ripple excursion permitted by CYH will nestle closer to the unity gain plateau than *harris* will allow, making CYH more potent for passband flattening. But the argument for stopband allowance reverses the story: *harris* will always significantly outperform CYH in stopband attenuation improvement.

Important MCF aspects are the four specially-labelled coordinates. These show where the dotted lines delineating linear ramp crossovers occur and can be interpreted as critical points where gain values will be driven either to perhaps-intended or perhaps-unintended gain plateaus. All this brings home strongly the message that **sharpening is a nonlinear operation**. It can be overdriven, if the crudity of the prototype is too extreme. The truth of this assertion now stands out starkly: "…sharpening makes good filters better, but bad filters worse" [12].

Thanks to the intercepts given in Fig. 6, we are in a position to gauge whether or not a given prototype will be receptive to worthwhile sharpening. We proceed to sharpening only if the prototype fits within such a <u>Passband Acceptance Corridor</u>.

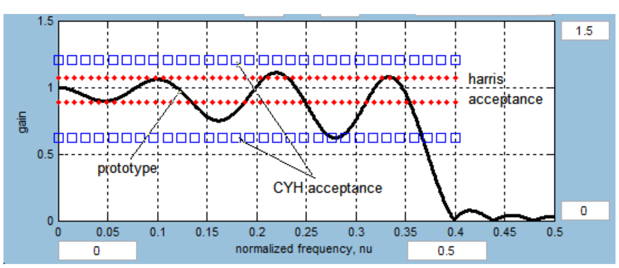

Fig. 7 *harris* and CYH Passband Acceptance Corridors
harris: (0.89 to 1.0708); CYH: (0.618 to 1.2056)

Once inside the acceptance corridor, improvement is certain, achieving error compression by distorting shapes in both stop and pass bands. This is disconcerting in the passband, where (for equiripple prototypes) we can no longer expect to see equal gain oscillations - but instead see vertical contortions. Visual inspection rarely does justice to just how gigantic and **disproportionate** the taming of peaks can be, in relative terms, for filters already of good quality.

V. EXAMPLES

Example 1:

In Fig. 2 we saw sharpening of a very low-quality prototype, while in Fig. 5 the initial quality was very high. In this Example we want to take a more middle-ground problem: a bandpass filter of intermediate quality, safely inside the acceptance corridors of both CYH and *harris*.

We generate our prototype by the MATLAB command [b, a] = ellip(4, 0.37, 20, [0.15 0.35]*2). Since elliptic filters have linear-phase numerators we can apply *harris*. This gives us superb stopband suppression as we see in the second subplot of Fig. 8, but very modest passband improvement. CYH gives us the moderate stopband, but very good passband, effects seen (in blue) in Fig. 8. Below the figure caption we have a measurement strip where single stopband and single transition metrics occupy the same locations as we have seen earlier for lowpass filter measurement strips.

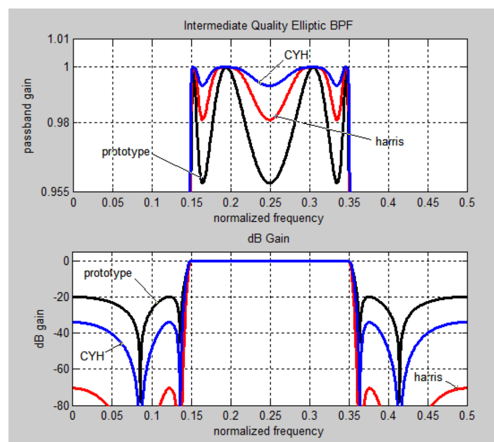

Fig. 8. Sharpened 4th-order Elliptic bandpass filter

Qproto	= {0.0417;	0.011	Hz;	20 dB;	0.1}
harris: Q	= {0.01892;	0.011	Hz;	70.52 dB;	0.0003}
harris: deltaQ	= {2.20;	1	(Hz);	50.52 dB;	333.33}
CYH: Q	= {0.00667;	0.011	Hz;	34.02 dB;	0.0199}
CYH: deltaQ	= {6.25;	1	(Hz);	14.02 dB;	5.03}

Example 2 [13]-[15]:

In [13] there is a complex-valued lowpass equiripple minimum-phase FIR design with N=26 coefficients which is a reduced-length filter arising from a 50-coefficient parent. The authors' robust Discrete Hilbert Transform method can be used to retrieve the impulse response sequence values. (Alternatively, they can be copied from the tabulated value in their Table 1). This parent is given (plotted red) in Fig. 9. The minimum-phase nature of this filter makes it a ready candidate for conversion to a (much smaller) IIR approximant [14], [15].

Using the Brandenstein-Unbehauen algorithm we arrive at the black plot in Fig. 9, which will serve as our prototype as regards sharpening. Incidentally, the FIR-to-IIR conversion process has slightly improved the stopband attenuation level (while smearing the equirriple structure in the stopband!).

We see that this end-to-end three-step process improved the passband peak error by about a factor of 93 and an increase of about 16 dB in the stopband attenuation was achieved. Most importantly of all, CYH is obviously able to play a useful role in sharpening any complex-valued IIR coefficient sets that might be encountered.

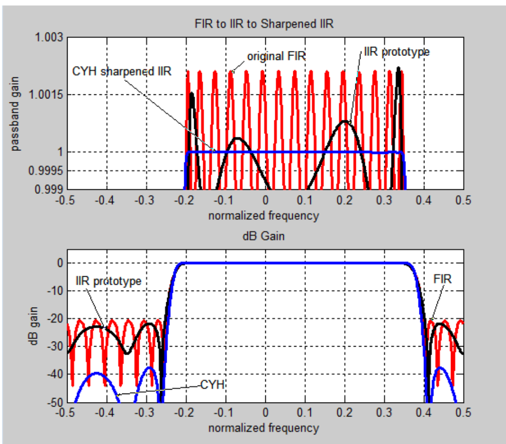

Fig.9 *N*=26 Complex FIR filter (in red) converted to IIR (order 5 in black) and sharpened with CYH (order 20 in blue)

FIR: Q	= {0.00213;	0.05 Hz;	20.69 dB;	0.09234}
IIR: Q	= {0.00271;	0.057 Hz;	21.75 dB;	0.08162}
Qproto	= {0.00271;	0.057 Hz;	21.75 dB;	0.08162}
CYH: Q	= {2.91e-5;	0.05649 Hz;	37.59 dB;	0.01319}
deltaQ	= 92.75;	1 (Hz);	15.8 dB;	6.1676}

VI. Conclusion

The new CYH structure very effectively combines the transfer function pole handling approach of [2], [6] and [7] while incorporating the zero handling (by matched filter numerator pairing) advocated in [4]. Thus, for the additional cost of including a single allpass filter, the capabilities of CYH have been extended to encompass the general IIR case. This new sharpener is free from past IIR limitations to prototype filters having linear-phase numerators and also from any burden of root-finding. While not being as vigorous in sidelobe level taming as the *harris* sharpener, CYH is considerably more potent at passband flattening, has a much wider acceptance corridor, and has been shown to be capable of complex IIR prototype sharpening.

References

[1] J.F. Kaiser and R.W. Hamming, "Sharpening the response of a symmetric nonrecursive filter by multiple use of the same filter", *IEEE Trans. ASSP*, vol. ASSP-25, no. 5, pp. 415-422, October 1977.

[2] F.J. Harris, "Implementing high performance, low computation, IIR filters with 2-path recursive all-pass filters and the harris-sharpening filter", *DSP 2009*, 16th Int. Conf. on Digital Signal Processing, Santorini, pp. 1-5, July 2009.

[3] G.J. Dolecek and A. Fernandez-Vazquez, "Sharpening minimum-phase filters", *J. Phys.: Conf. Ser.* 012012, vol. 410, no. 1, pp. 1-4, 2013.

[4] G.D. Cain, A.Yardim and F. Harris, "Sharpening minimum-phase FIR digital filters using paired prototypes", in press.

[5] M. Donaldson, "Lost knowledge refound: sharpened FIR filters", *IEEE Signal Processing Magazine*, vol. 20, no.5, pp.61-63, Sept. 2003; also in *Streamlining Digital Signal Processing: A Tricks of the Trade Guidebook*, edited by R.G. Lyons, Wiley-IEEE Press, 2nd Edition, pp. 3-9, 2012.

[6] T.A.C.M. Claasen and W.F.G. Mecklenbrauker, "Extension of a method for sharpening the response of digital filters based upon the amplitude change function", *IEEE Trans.ASSP*, vol. ASSP-27, no. 5, pp. 559-560, October 1979.

[7] C.A. Jalaludeen, M.P. Mathew, and V. Chander, "Extending the use of amplitude change function to stable recursive filters of any pole-zero distribution", IEEE Int. Conf. Signal Proc. *ICSP'96*, vol. 1, pp. 52-54, October 1996.

[8] G.D. Cain, A.Yardim and F. Harris, "Live Demonstration: Comparison of three FIR digital filter sharpening techniques", in press.

[9] Manolakis, D.G. and V.K. Ingle, *Applied Digital Signal Processing*, Cambridge: Cambridge University Press, 2001.

[10] J.O. Coleman, "Equiripple-stopband multiplierless FIR filters by Chebyshev sharpening of two-sample averaging", Proc. 2018 *IEEE Int'l Symp on Circuits and Systems* (*ISCAS* 2018), May 2018.

[11] S. Samadi, "Explicit formula for improved filter sharpening polynomial", *IEEE Trans. Sig. Proc.*, vol. 9, no. 10, pp. 2957-2959, Oct. 2000.

[12] P.P. Vaidyanathan, "Design and implementation of digital FIR filters", pp. 112, in *Handbook of Digital Signal Processing Engineering Applications*, edited by D.F. Elliott, San Diego: Academic Press, 1987.

[13] N. Damera-Venkata, B.L.Evans and S.R. McCaslin, "Design of optimal minimum-phase digital filters using discrete Hilbert transforms", *IEEE Trans. Signal Proc.*, vol. 48, no. 5, pp. 491-1495, May 2000.

[14] Beliczynski, B., I. Kale and G. D. Cain, "A balanced model reduction algorithm for transforming FIR to IIR approximants", *IEEE Trans. Sig. Proc.,* vol. 40, no. 3 , pp. 532-542, March 1992.

[15] Brandenstein, H. and R. Unbehauen,"Least-squares approximation of FIR by IIR digital filters", *IEEE Trans. Sig. Proc.*, vol. 46, no. 1, pp. 21-30, January 1998.