

# A Practical Approach for the Design of Nonuniform Lapped Transforms

Byung-Jun Yoon, *Student Member, IEEE*, and Henrique S. Malvar, *Fellow, IEEE*

**Abstract**—We propose a simple method for the design of lapped transforms with nonuniform frequency resolution and good time localization. The method is a generalization of an approach previously proposed by Princen, where the nonuniform filter bank is obtained by joining uniform cosine-modulated filter banks (CMFBs) using a transition filter. We use several transition filters to obtain a near perfect-reconstruction (PR) nonuniform lapped transform with significantly reduced overall distortion. The main advantage of the proposed method is in reducing the length of the transition filters, which leads to a reduction in processing delay that can be useful for applications such as real-time audio coding.

**Index Terms**—Cosine-modulated filter banks (CMFBs), extended lapped transforms (ELTs), filter banks, lapped transforms, modulated lapped transform (MLT), nonuniform filter banks.

## I. INTRODUCTION

LAPPED transforms such as the lapped orthogonal transform (LOT) and the modulated lapped transform (MLT) have been widely used in various applications, such as image processing and audio coding [1]. For example, the MLT is a special kind of cosine-modulated filter bank (CMFB) that has been very popular in audio processing. It is well known for its computational efficiency and good frequency discrimination, and most modern audio coders such as the MPEG-2 Layer III (MP3), Dolby AC-3, and MPEG-4 AAC are based on the MLT or similar constructions [2].

Even though the MLT has many advantages, there are cases where a different time-frequency resolution may be more desirable. One such example is audio coding at low bit rates. When there are high-frequency transients in the original signal, the poor time resolution of the basis functions gives rise to ringing artifacts. Those are especially noticeable when they precede the signal (the pre-echoes in audio coding [2]). In order to alleviate this problem, many modern audio coders adopt a window switching strategy, which uses a shorter window when high-frequency transient sounds are detected. An alternative approach is to use a nonuniform filter bank or a lapped transform with a nonuniform subband decomposition, in which the higher frequency subband has a wider bandwidth. The corresponding shorter impulse responses reduce the pre-echo effect [4] and remove the need to look ahead in the signal buffer in order to determine the window switching points, thus reducing processing delay. With a nonuniform filter bank using

wider bandwidths for high-frequency subbands, those subbands are always being processed with much shorter block lengths, thus easily allowing for encoder adaptation to high-frequency transients. Furthermore, with nonuniform filter banks, low-frequency components are not subject to additional distortion due to window switching.

Various methods have been proposed for designing nonuniform filter banks [4]–[11]. Many are based on the subband merging approach [4]–[9], while others are based on a tree-structured design [1], [3] or constructed from joining uniform filter banks by a transition filter [10]. In applications such as real-time audio coding, we need transforms with a large number of subbands, nonuniform (or variable) frequency resolution, and low system delay. In this letter, we propose a new method for designing a lapped transform with the aforementioned properties. The method is a generalization of that proposed by Princen [10], with the main advantage of reducing system delay. In Section II, we review briefly Princen's approach and present our proposed generalization. Section III shows design examples that demonstrate the advantages of the proposed approach.

## II. DESIGN OF NONUNIFORM LAPPED TRANSFORMS

In [10], Princen proposed an interesting approach for designing nonuniform filter banks (NUFBs). Two uniform sections obtained from CMFBs with different decimation ratios are joined, and one of the filters in the edge subband is replaced by a transition filter. The transition filter is derived from a complex (hence asymmetric) prototype, which is cosine-modulated in such a way to produce a real-valued transition filter, whose passband is located between the two uniform sections. This is illustrated in Fig. 1(a). The prototype filter is optimized such that it minimizes the aliasing between itself and the adjacent filters. Aliasing between other filters is assumed to be negligible during this optimization procedure. This approach provides an effective way of designing NUFBs with good frequency responses [10].

However, as the transition filter is designed with the goal of near cancellation of frequency aliasing from both sides, it generally needs to be much longer than the adjacent filters. That increases the overall system delay. Another problem is that it assumes that the original CMFBs have large stopband attenuation; hence, nonnegligible aliasing exists only between adjacent filters. Thus, the method does not lead to satisfactory results when that assumption is not met, for example, when we join two MLTs of order  $M$  and  $M/2$ . We can avoid these problems by using several transition filters, whose lengths do not exceed those of the adjacent filters. For example, let us consider the case when we want to join two CMFBs, where the filter bank in the low-frequency region has a larger decimation ratio than the

Manuscript received December 10, 2005; revised January 23, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dimitri Van De Ville.

B.-J. Yoon is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA.

H. S. Malvar is with Microsoft Research, Redmond, WA 98052 USA.

Digital Object Identifier 10.1109/LSP.2006.873145

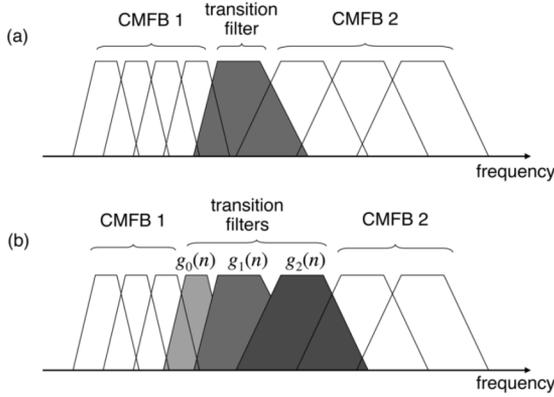


Fig. 1. Joining two uniform CMFBs using transition filters. (a) Princen [10]. (b) Proposed.

filter bank in the high-frequency region. This gives us better frequency resolution for low-frequency signals and better time localization for high-frequency signals. As shown in Fig. 1(b), we assume that the filters in the original CMFBs do not have sharp cutoffs, giving rise to aliasing between nonadjacent bands. This is the case when we use a CMFB with relatively short filters such as the MLT. In such a situation, using a single optimized filter  $g_1(n)$  in the transition region is not enough, and using additional filters  $g_0(n)$  and  $g_2(n)$  on both sides significantly reduces the overall distortion of the filter bank.

Another advantage of using several transition filters is that the two additional filters  $g_0(n)$  and  $g_2(n)$  relieve the burden on the center filter  $g_1(n)$ , which makes it possible to use shorter transition filters compared to the original design [10]. In particular, we can use transition filters that are no longer than the longest of the two CMFB filters. Thus, there is no delay penalty in using transition filters, which is an important aspect for many practical applications.

The transition filters have to be optimized such that the resulting transform becomes as close as possible to a perfect-reconstruction (PR) system. For this purpose, we first define an error metric that can serve as an indicator of how close the given transform is to a PR system. Consider a  $2MK \times M$  transform matrix  $\mathbf{P}$  with  $M$  subbands,

$$\mathbf{P} = [\mathbf{P}_0^T \quad \mathbf{P}_1^T \quad \cdots \quad \mathbf{P}_{2K-1}^T]^T \quad (1)$$

where  $\mathbf{P}_j$  is a  $M \times M$  square matrix. We define

$$\mathbf{E}_k = \sum_{i=0}^{2K-1-k} \mathbf{P}_i \mathbf{P}_{i+k}^T - \delta(k) \mathbf{I} \quad (2)$$

where  $\mathbf{I}$  is the  $M \times M$  identity matrix. Now, the error metric can be defined as the sum of the Frobenius norm<sup>1</sup> of  $\mathbf{E}_k$

$$\eta = \sum_{k=0}^{2K-1} \|\mathbf{E}_k\|^2. \quad (3)$$

<sup>1</sup>The Frobenius norm of a matrix  $\mathbf{A}$  is defined as  $\|\mathbf{A}\|^2 = \sum_{i,j} |a_{ij}|^2 = \text{trace}(\mathbf{A}\mathbf{A}^H)$ .

Note that if  $\eta = 0$ , the lapped transform  $\mathbf{P}$  is PR [1], [11]. Thus, we can design good near-PR nonuniform lapped transforms by minimizing  $\eta$ , e.g., by using unconstrained minimization techniques such as the quasi-Newton method [12].

In the proposed design, we choose the same decimation ratio for the subband that uses  $g_0(n)$  as the CMFB in the left-hand side. For  $g_1(n)$  and  $g_2(n)$ , the decimation ratios are chosen to be identical to that of the CMFB in the right-hand side. Then, the three transition filters are optimized sequentially, through the following steps.

- 1) Combine the two uniform CMFB sections, using filters from the first uniform CMFB section up to the frequency closest to the transition frequency, and filters from the second uniform CMFB section afterward.
- 2) Initialize the transition filters  $g_0(n)$  to  $g_2(n)$ ;  $g_0(n)$  is set equal to the last filter of the first section, and  $g_1(n)$  and  $g_2(n)$  are set equal to the first and second filters of the last section, respectively.
- 3) Optimize  $g_1(n)$  only, with  $g_0(n)$  and  $g_2(n)$  fixed.
- 4) Optimize  $g_0(n)$  only, with  $g_1(n)$  and  $g_2(n)$  fixed.
- 5) Optimize  $g_2(n)$  only, with  $g_0(n)$  and  $g_1(n)$  fixed.
- 6) If  $\eta < \eta_0$ , terminate. Otherwise, go back to Step 3).

The optimization routine terminates once  $\eta < \eta_0$  is satisfied, indicating that the transform  $\mathbf{P}$  is close to PR. Because at each of Steps 3)–5) the error metric  $\eta$  is a quadratic function of the filter coefficients to be optimized, the minimum obtained in each of those steps is a global one (under the constraint that the other two filters are fixed). Therefore, the whole sequential optimization (also called coordinate descent) procedure is guaranteed to converge to at least a local unconstrained minimum [12]. As is usually the case with coordinate descent methods, convergence is slow, typically taking dozens of iterations. By choosing  $\eta_0$  small enough, we can obtain a good nonuniform lapped transform with nearly flat overall transfer function and negligible aliasing components, whose magnitudes are comparable to the stopband attenuation of the filters in the original CMFBs.

### III. DESIGN EXAMPLES

Let us first consider the case where we join two MLTs. This is especially interesting for applications such as real-time audio coding, where CMFBs with long filters are not desired, because they increase processing delay. In this example, we join an MLT matrix  $\mathbf{R}_1$  of order  $M_1 = 32$  and a smaller MLT matrix  $\mathbf{R}_2$  of order  $M_2 = M_1/2 = 16$ , according to the design procedure described in Section II. Using these transforms, we construct a  $2M_1 \times M_1$  transform matrix  $\mathbf{P}$  as follows. The first  $M_1/2 - 1$  column vectors ( $k = 0, \dots, M_1/2 - 2$ ) of  $\mathbf{P}$  are obtained from the first  $M_1/2 - 1$  column vectors of  $\mathbf{R}_1$ . The last  $M_1/2 - 4$  column vectors ( $k = M_1/2 + 4, \dots, M_1 - 1$ ) are obtained from the last  $M_2/2 - 2$  column vectors of  $\mathbf{R}_2$  and the same vectors shifted by  $M_2$ . The transition filters are located at  $M_1/2 - 1 \leq k \leq M_1/2 + 3$ . The transition filter  $g_0(n)$  is located at  $k = M_1/2 - 1$  and has length  $2M_1$ . The transition filter  $g_1(n)$  is located at  $k = M_1/2$  and  $M_1/2 + 1$ , where the vector at  $k = M_1/2 + 1$  is a shifted version of  $g_1(n)$  by  $M_2$ . The length of  $g_1(n)$  is set to  $3M_2 = 48$ . Similarly, the filter  $g_2(n)$  is located at  $k = M_1/2 + 2$  and  $k = M_1/2 + 3$ . The structure of the transform matrix is shown in Fig. 2.

Fig. 3 shows the magnitude response of some of the filters in the filter bank. Those in the far left and far right correspond to

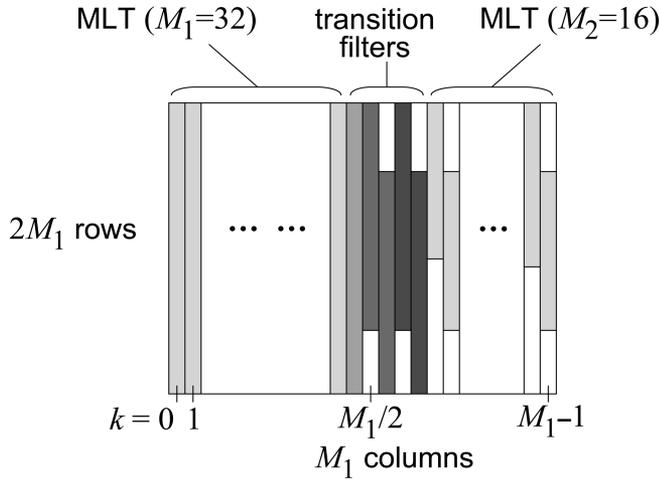


Fig. 2. Graphical depiction of the structure of the transform matrix; each column represents one filter. Regions with nonzero coefficients are shaded.

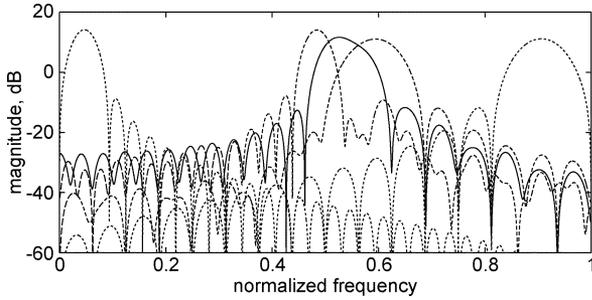


Fig. 3. Magnitude response of selected filters of a 24-subband nonuniform lapped transform based on joining MLTs using the proposed approach. Dotted lines: responses of two of the original MLT filters. Dashed and solid lines: responses of the transition filters.

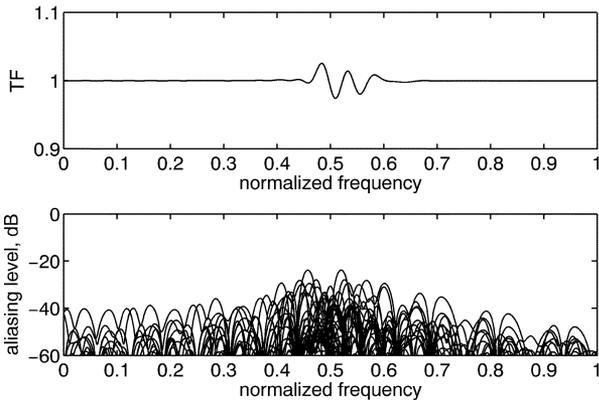


Fig. 4. Proposed approach. Top: analysis-synthesis transfer function. Bottom: aliasing components.

the two original CMFBs for  $M_1 = 32$  and  $M_2 = 16$ , respectively. In the middle, we show the responses of the three transition filters. We see that their stopband performances are comparable, which is an advantage of using several transition filters. Fig. 4 shows the overall transfer function and the magnitudes of the aliasing components (in the same scale as in Fig. 3) of the nonuniform lapped transform that was obtained using the proposed method. The corresponding plots for the use of a single transition filter  $g_1(n)$  are shown in Fig. 5. We see that the peak distortion of the overall transfer function and the magnitude of the aliasing components have been reduced.

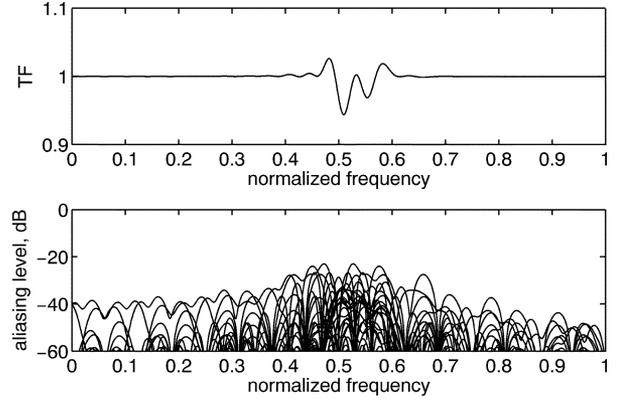


Fig. 5. Princen's approach. Top: analysis-synthesis transfer function. Bottom: aliasing components.

The characteristics of the two transforms are summarized in Table I, which shows this improvement quantitatively. The value  $\bar{\eta} = \eta/M^2$  is a measure that shows how close the transform  $\mathbf{P}$  is to a PR system. As mentioned earlier,  $\bar{\eta} = 0$  implies that  $\mathbf{P}$  is PR. The overall transfer function and the aliasing terms can be computed from

$$T_\ell(z) = \frac{1}{M} \sum_{k=0}^{M-1} H_k(z) W_M^\ell F_k(z) \quad (4)$$

where  $H_k(z)$  and  $F_k(z)$  are, respectively, the analysis filter and the synthesis filter in the  $k$ th band [3]. We define the maximum ripple size of the overall transfer function  $T_0(z)$  as

$$\delta_T = \max_{\omega} |T_0(e^{j\omega}) - 1|. \quad (5)$$

The energy of the aliasing components is defined as

$$\mathcal{E}_{\text{alias}} = \sum_{\ell=1}^{M-1} \left( \frac{1}{\pi} \int_0^\pi |T_\ell(z)|^2 d\omega \right). \quad (6)$$

Finally, we define the signal-to-noise ratio (SNR)  $\gamma$  as the ratio between the input signal variance  $\sigma^2$  and the reconstruction error variance, i.e.,

$$\gamma = 10 \log_{10} \frac{\sigma^2}{\sigma_\epsilon^2} \quad (7)$$

where we assume that the input signal is a white Gaussian noise. We see in Table I that the proposed approach considerably reduces all three distortion metrics  $\bar{\eta}$ ,  $\delta_T$ , and  $\mathcal{E}_{\text{alias}}$ . For example, for the MLT-based design, all metrics are reduced by about 50%.

The proposed method can also be used for joining CMFBs with longer filters. For example, let us consider joining two extended lapped transforms (ELTs [1]) with  $M_1 = 32$ ,  $M_2 = M_1/2 = 16$ , and  $K = 4$  to construct a  $2M_1K \times M_1$  transform matrix  $\mathbf{P}$ . The length of the filters in the CMFB in the low-frequency region is  $2M_1K = 256$ , and the length of the filters in the CMFB in the high-frequency region is  $2M_2K = 128$ . As in the first example, the transition filters  $g_0(n)$ ,  $g_1(n)$ , and  $g_2(n)$  are located in  $M/2 - 1 \leq k \leq M/2 + 3$ . The length of the

TABLE I  
ERROR METRICS FOR SEVERAL DESIGN CHOICES

	$\bar{\eta}$	$\delta_T$	$\mathcal{E}_{\text{alias}}$	$\gamma$ (dB)
MLT (Princen)	2.84 E-5	5.60 E-2	1.31 E-3	28.51
MLT (proposed)	1.40 E-5	2.60 E-2	7.15 E-4	31.38
ELT (Princen)	3.49 E-7	4.73 E-4	2.22 E-5	46.50
ELT (proposed)	2.23 E-7	3.00 E-4	1.42 E-5	48.53
near PR (Princen)	7.18 E-8	1.12 E-3	4.56 E-6	53.38
near PR (proposed)	5.30 E-8	8.51 E-4	3.37 E-6	54.68

transition filter  $g_0(n)$  is  $2M_1K = 256$ , and the length of  $g_1(n)$  and  $g_2(n)$  is  $2M_1K - M_2 = 240$ . The properties of the resulting transforms are summarized in Table I. We see that using higher order CMFBs and longer transition filters improves the overall performance of the transform significantly, as expected. We note also that when the original CMFBs have better stopband characteristics, the improvement of our proposed approach (using several transition filters) over Princen's approach (using a single transition filter) is reduced. In other words, for a given number of subbands, the longer the filters of the original CMFB sections, the less the need for using more than one transition filter.

It is also possible to use near-PR CMFBs for designing nonuniform lapped transforms, instead of PR transforms such as MLTs and ELTs. Such a design may have advantages in terms of an increased stopband attenuation, and thus better frequency separation and reduced overall distortion, at the expense of a small increase in the maximum distortion level. We designed two CMFBs using the procedure proposed in [13] and [14], with the same number of subbands and identical filter lengths as the ELTs that were previously used. Simulation results are summarized in Table I, which confirms that NUFBs designed from near-PR CMFBs have better stopband and aliasing characteristics, at the expense of a small increase in the ripple of the overall transfer function. Because that penalty is small, near-PR-based designs may be preferable in many practical applications.

#### IV. COMPUTATIONAL COST

We see from the design examples that the use of multiple transition filters leads to improved performance when compared to using a single one [10]. That was naturally to be expected, because there are more degrees of freedom in optimizing multiple filters. This improvement comes at a small cost, though: increased computational complexity. For example, let us consider the case when two MLTs were joined to obtain the nonuniform transform. The transition filters must be applied to the input signal frame in a direct form, that is, they cannot be obtained from the fast transforms that are used to compute the MLT subbands [1]. However, this computational overhead becomes less significant as the number of subbands  $M_1$  and  $M_2$  gets larger. Consider the parameters in the first example:  $M_1 = 32$ ,  $M_2 = 16$ . For each input frame, the longer MLT can be computed via the fast Fourier transform (FFT) with 160 multiplications and 288 additions [1], and the shorter MLT with 72 multiplications and 120 additions, for a total of 232 multiplications and 408 additions. The transition filters take a total of 112 multiplications and 112 additions. Thus, the transition filters lead to

a 35% increase in computational cost. This overhead increases if ELTs are used, as that requires longer transition filters. For applications such as audio coding, where  $M_1 = 2,048$  is typical, the computational overhead of the MLT-based design is reduced to about 25%. Considering that computing the filter bank takes usually a small fraction of the total computational load in such applications, the 25% overhead on the filter bank computation would typically be acceptable.

#### V. CONCLUDING REMARKS

We have proposed an effective approach for designing nonuniform lapped transforms. The proposed approach can be used to construct lapped transforms with a large number of bands, nonuniform frequency resolution, and good time localization of the high-frequency basis functions. The proposed approach is an extension of Princen [10], using several transition filters to join uniform CMFBs with different decimation ratios. As shown in the examples, this method reduces the distortion and the aliasing components in the original design significantly, by up to 50%. It has the advantages that it can be effectively used with CMFBs with relatively short filters (such as the MLT) and that the overall system delay is identical to that of the longest delay of the original CMFBs that are joined together. These advantages make the proposed design approach especially useful in practical applications such as audio coding, as long as the small increase in computational complexity can be tolerated—which is usually the case.

#### REFERENCES

- [1] H. S. Malvar, *Signal Processing With Lapped Transforms*. Norwood, MA: Artech House, 1992.
- [2] S. Shlien, "The modulated lapped transform, its time-varying forms, and its applications to audio coding standards," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 4, pp. 359–366, Jul. 1997.
- [3] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [4] H. S. Malvar, "Biorthogonal and nonuniform lapped transforms for transform coding with reduced blocking and ringing artifacts," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 1043–1053, Apr. 1998.
- [5] M. Purat and P. Noll, "Audio coding with a dynamic wavelet packet decomposition based on frequency-varying modulated lapped transforms," in *Proc. ICASSP*, May 1996, vol. 2, pp. 1021–1024.
- [6] O. A. Niamut and R. Heusdens, "Subband merging in cosine-modulated filter banks," *IEEE Signal Process. Lett.*, vol. 10, no. 4, pp. 111–114, Apr. 2003.
- [7] R. L. de Queiroz, "Uniform filter banks with nonuniform bands: post-processing design," in *Proc. ICASSP*, May 1998, vol. 3, pp. 1341–1344.
- [8] S. C. Chan, X. M. Xie, and T. I. Yuk, "Theory and design of a class of cosine-modulated nonuniform filter banks," in *Proc. ICASSP*, Jun. 2000, vol. 1, pp. 504–507.
- [9] J. Li, T. Q. Nguyen, and S. Tantarata, "A simple design method for nonuniform multirate filter banks," in *Proc. 28th Asilomar Conf. Circuits, Systems, Computers*, Oct. 1994, vol. 2, no. 31, pp. 1015–1019.
- [10] J. Princen, "The design of nonuniform modulated filterbanks," *IEEE Trans. Signal Process.*, vol. 43, no. 11, pp. 2550–2560, Nov. 1995.
- [11] K. Nayebi, T. P. Barnwell, III, and M. J. T. Smith, "Nonuniform filter banks: A reconstruction and design theory," *IEEE Trans. Signal Process.*, vol. 41, no. 3, pp. 1114–1127, Mar. 1993.
- [12] D. G. Luenberger, *Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1984.
- [13] T. Q. Nguyen, "Digital filter bank design quadratic-constrained formulation," *IEEE Trans. Signal Process.*, vol. 43, no. 9, pp. 2103–2108, Sep. 1995.
- [14] T. Q. Nguyen, "Near-perfect-reconstruction pseudo-QMF banks," *IEEE Trans. Signal Process.*, vol. 42, no. 1, pp. 65–76, Jan. 1994.