Quantization Noise Shaping in Oversampled Filter Banks

Tania Leppert

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

April 2005

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of Master's of Engineering.

2005/04/22

Abstract

The use of a noise-shaping system in oversampled filter banks has been shown to improve the effective resolution of subband coders. Although the filter banks directly determine the noise-shaping coefficients, a comparison between theoretical and simulated results has not been done, while the effect of the selection of the filter banks on the performance of the noise-shaping system has not yet been evaluated. Therefore, an algorithm for the generation of cosine-modulated perfect-reconstruction filter banks is presented, such that the generated filters could be used as a test bed. The optimal noise-shaping coefficients are then derived, and the noise-shaping system is inserted into the subband coder.

It is found that the theoretical results agree with the simulations, but that the performance of the noise-shaping system is limited by ill-conditioning at higher system orders. An increase in filter length and an increase in the degree of overlap between neighbouring channels contribute independently to a better performance. Also, it is seen that near-perfect reconstruction filter banks are limited by their reconstruction error but yield good results at low bitrates.

Sommaire

Il a été démontré que l'utilisation d'un système de mise en forme du bruit dans les bancs de filtres sur-échantillonés améliore la résolution des codeurs de sous-bandes. Bien que les coefficients des filtres dédiés à la mise en forme du bruit dépendent directement des bancs de filtres sélectionnés, la comparaison entre les résultats théoriques et ceux provenant de simulations n'ont pourtant pas encore été éffectuées. Il en est de même pour l'étude des répercussions du choix de bancs de filtres. Conséquemment, l'élaboration d'un algorithme générant des bancs de filtres modulés en cosinus et à reconstruction parfaite est présentée, pour qu'ensuite ces bancs filtres puissent être utilisés en tant que banc d'essai. Par la suite, les signaux de sous-bande sont quantifiés, les coefficients optimaux sont dérivés et puis introduits dans le codeur de sous-bandes.

Il est remarqué que les résultats théoriques correspondent aux résultats des simulations, mais que le conditionnement limite la performance du système de mise en forme du bruit. De plus, il est démontré qu'un allongement des filtres ainsi qu'un plus grand chevauchement entre les sous-bandes adjacentes contribuent indépendemment à une amélioration de la performance du système de mise en forme du bruit. En outre, les bancs de filtres approximant la reconstruction parfaite sont limités par leurs erreurs de reconstruction. Cependant, à un bas taux de débit, ils réussissent pourtant assez bien.

Acknowledgments

Firstly, I would like to express my gratitude to both Professor Fabrice Labeau and Professor Peter Kabal for their support and their helpful advice and guidance. Further, I would like to thank my parents and my brother and sister for their unconditional love: countless times, I have relied on them and they were always more than happy to oblige.

Moreover, for their challenging discussions and various helpful contributions, I would like to thank, in no particular order, Karim Ali, Roberto Rotili, Nikolaos Gryspolakis, Eugene Nicolov, Martin Cudnoch, Alex Wyglinski, and last, but not least, Valérie Paquin and Andrea Towstuk.

Contents

List of Figures

Chapter 1

Introduction

A filter is defined as being any system that modifies certain frequencies relative to other frequencies [1]. Thus, if it is desired to focus attention on a particular interval of frequencies, a filter that attenuates all the frequency content outside of that interval would be useful. Indeed, if the frequency distribution of a class of signals were nonuniform, it stands to reason that there is interest in treating different intervals in a different manner. For example, if a class of signals generally has more energy content at low frequencies than at high frequencies, it would make sense to use a better code on the first interval, while perhaps cutting some corners on the second interval, in an attempt to compress the signal by removing non-essential content or to reduce the complexity of a system. This is in fact the motivation for the development of subband coders for speech and image processing, since the frequency distribution of these signals is indeed quite nonuniform [2].

1.1 Subband coders

In order to analyze the different frequency intervals separately, a uniform digital filter bank is used: a digital filter bank is defined as a collection of digital filters with a common input or a common output [3]. Thus a filter bank with a common input separates the input signal into, say, N signals, whose frequency content is mostly limited to the intervals $k\pi/N \leq \omega \leq (k+1)\pi/N$, for $0 \leq k \leq N-1$, where ω is in radians. This filter bank is then termed an *analysis* filter bank. On the other hand, a filter bank with a common output combines, say, N input signals with limited frequency content into one output signal and is therefore termed a *synthesis* filter bank.

Figure 1.1 illustrates a basic subband coder and decoder, where the $H_k(z)$ are the component filters of the analysis bank, while the $F_k(z)$ are the component filters of the synthesis bank.

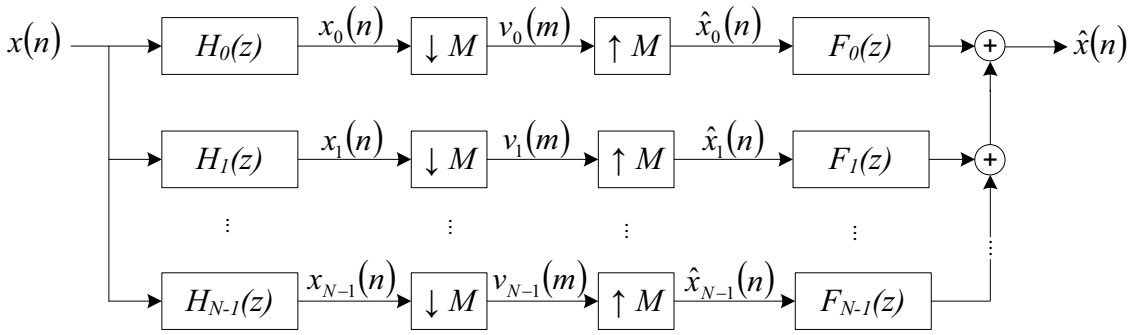


Fig. 1.1 General subband coder and decoder.

Recapitulating, the analysis filter bank splits the input signal $x(n)$ into N signals $x_k(n)$, while the synthesis filter bank combines the $\hat{x}_k(n)$ into the output signal $\hat{x}(n)$.

The use of filter banks is by no means confined to subband coders: they find applications in areas such as digital audio coding and voice privacy systems [3] and transmultiplexers [4]. Further, as will be discussed in Chapter 2, a judicious design of the analysis filter banks allows for a reduction by a factor of M of the sampling rate of the outputs of the filter banks, denoted by $\downarrow M$ in Figure 1.1. The signals $v_k(n)$ at the reduced rate are called *subband* signals, from which an approximation $\hat{x}(n)$ of the input signal $x(n)$ can be obtained, by increasing their rate by a factor of M and by combining them using an appropriate synthesis filter bank. Thus, a subband coder is a specific form of the wider class of *multirate systems* [4].

In most cases, the advantage of subband coding is seen during the process of quantizing the subband signals $v_k(n)$: due to the nonuniform distribution of the content of the input signals in different frequency bands, it is possible to use quantizers with different degrees of precision for the various $v_k(n)$. The degree of precision of a given quantizer is governed by the number of bits it uses to represent a certain signal. This strategy then provides a means by which to either reduce the number of bits needed to represent a signal (compression), or alternatively a way in which to represent a signal more accurately given a certain fixed total number of bits [3].

Finally, as will be dealt with in more depth shortly, in the case that $M = N$, the

system is referred to as critically sampled, since there is precisely the sufficient amount of information needed to reconstruct the original signal $x(n)$, given a judicious choice of analysis filters. On the other hand, if $M > N$ there will be a loss of information and the original signal will be corrupted, while if $M < N$, the system is referred to as *oversampled* as there are more samples than absolutely necessary in the subband signals.

1.2 Oversampling the subband signals

The main motivation for oversampling the subband signals comes from the benefits achieved in oversampled analog-to-digital conversion. In this case, the conversion accuracy can be improved in two ways: either by using quantizers with finer resolution or by decreasing the sampling period, equivalently increasing the sampling rate [5]. In modern A/D conversion, the accuracy is improved by oversampling the input signal, in order to avoid the high costs involved in the construction of high-resolution quantizers. This then suggests that lower resolution quantizers could be used in the subband coder by oversampling the subband signals.

Furthermore, oversampling provides a redundant representation — which can be interpreted as an overcomplete expansion — of the input signal. It is then possible to take advantage of this redundancy in a variety of ways: for example, it was noted that a sophisticated selection of information from this redundancy could yield good compression schemes [6]. Indeed, as shown in [7], although the full potential of the compression schemes based on overcomplete expansions has not yet been explored, they show results on par with standard compression schemes. Another example is the robustness to erasure demonstrated by overcomplete expansions [8], [9], suggesting that oversampling is useful for packet-based communication systems, where packet loss may be inevitable.

Moreover, the oversampling of the subband signals can be exploited in the design of the synthesis filter banks. While in the case of critically sampled filter banks only one synthesis filter bank provides the perfect reconstruction[1] of the input signal given a specific analysis filter bank, oversampled filter banks provide more freedom in the design of the synthesis filter banks, yielding the opportunity to design filter banks with added desirable characteristics [10], [11], [12], [13].

Another way to exploit the inherent redundancy is through insightful processing of

[1] More on this in Chapter 2.

the subband signals such as linear prediction of the subband signals or quantization noise shaping [14].

1.3 Noise-shaping in oversampled filter banks

The introduction of a noise-shaping system into oversampled filter banks stems once more from the use of a such a system in oversampled A/D conversion. In fact, single-bit codewords obtained from artificially high sampling rates were achieved soon after delta modulation was introduced by Cutler in 1946 [15]. The idea was simple: the overall error of the system was reduced by measuring the quantization error in one sample and then subtracting this quantity from the next sample. Subsequently, more sophisticated systems were designed, yielding an improved performance. In the context of oversampled filter banks, the strategy is to shape the quantization noise in the subband signals $v_k(n)$ in such a way that it will be attenuated by the synthesis filter bank. Indeed, the optimal noise-shaping system given a certain filter bank was derived in [14] and were shown to improve the effective resolution of the quantizers.

However, the manner in which the choice of the filter banks affects the performance of the noise-shaping system has yet to be studied. Firstly, the aim of this work is to correlate the results of simulations with the theoretically predicted performance and secondly to explore the effect of the different characteristics of various filters on the performance of the noise-shaping filters.

1.4 Outline

Since it was desired to study the effect of the selected filter banks on the performance of the noise-shaping system, there are two distinct topics to be covered: the design of the filter banks and the introduction of a quantization noise-shaping system into a subband coder. However, they are interrelated and will thus be discussed in a concurrent fashion.

Chapter 2 first introduces fundamental notions in multirate systems, permitting the development of the constraints on filter bank design for perfect reconstruction of the input signal as well as an introduction to the frame-theoretic approach to oversampled filter banks.

Chapter 3 then focuses on the design method selected for the filter banks, followed by

an analysis of the quantization noise in oversampled filter banks. Subsequently, the optimal noise-shaping coefficients are derived.

Chapter 4 turns to the explicit implementation of the design algorithm, followed by a demonstration of the obtained filter banks. Next, the subband signals are quantized with uniform quantizers, the noise-shaping system is introduced into the subband coder and a discussion of the theoretical and simulation results ensues. Finally, the effect of varying different filter bank characteristics on the performance of the noise-shaping system is evaluated. The effects of varying such characteristics as filter lengths, degree of overlap between the subbands and perfect versus near-perfect reconstruction filter banks are shown.

Chapter 2

Multirate systems

Multirate systems contain both linear filters and time-varying operations and are therefore part of the class of linear time-varying systems (LTV). As they form the background for the present work, this chapter will deal first with some underlying concepts to facilitate subsequent discussions, followed by a deeper look at digital filter banks along with their design and finally an introduction to oversampled filter banks from a frame-theoretic point of view.

2.1 Basic notions in multirate systems

In order to appreciate the possible advantages of multirate systems, some essential operations pertaining to them must be defined beforehand. First, the decimation and interpolation operators, which allow for a reduced bit rate in the subband signals, will be explained and an elucidation of the polyphase decomposition, which allows for more elegant solutions and leads to lower computational complexity, will ensue.

2.1.1 Decimation and interpolation

Decimation and interpolation are the two most fundamental operations in multirate digital signal processing [3]. In this section, they are first explained in the time domain for an intuitive approach, followed by a frequency domain interpretation leading to the basics of aliasing.

Time Domain

The explanation of decimation and interpolation requires the definition of two new building blocks: the *decimator* and the *expander*.

Simply put, the decimator can be viewed purely as sampling the signal at an M-times lower rate [1], where M is constrained to be an integer. This leads to the coining of new terms, such as *downsampling* and *subsampling*. The decimator block, as shown in Figure 2.1, transforms the input sequence $x_d(n)$ into the output sequence $y_D(n)$ by retaining only

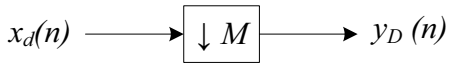


Fig. 2.1 Decimator.

the values that are at time indices which are a multiple of M:

$$y_D(n) = x_d(Mn).$$

Upon closer inspection, it should be obvious that a simple unit delay at the input of the decimator will not lead to a delay of the output by one sample. Indeed,

$$\begin{aligned} y_D(n-1) &= x_d(M(n-1)) \\ &\neq x_d(Mn-1). \end{aligned}$$

It is concluded that downsampling is a time-varying operation.

As opposed to the decimator block, the expander block, (Figure 2.2), inserts $L-1$ zeros

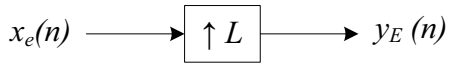


Fig. 2.2 Expander.

between the samples, yielding a signal that has been effectively *upsampled*; that is there are more samples in the expanded signal $y_E(n)$ that in the input signal $x_e(n)$:

$$y_E(n) = \begin{cases} x_e(n/L), & \text{if } n \text{ is an integer-multiple of } L \\ 0, & \text{otherwise.} \end{cases}$$

To complete the interpolation process, a lowpass filter is appended, in order to convert the inserted zero-valued samples of $y_E(n)$ into interpolated samples $y_I(n)$; that is, samples that are the approximation to the original underlying analog signal – assuming a band-limited signal– sampled at a higher rate, as shown in Figure 2.3.

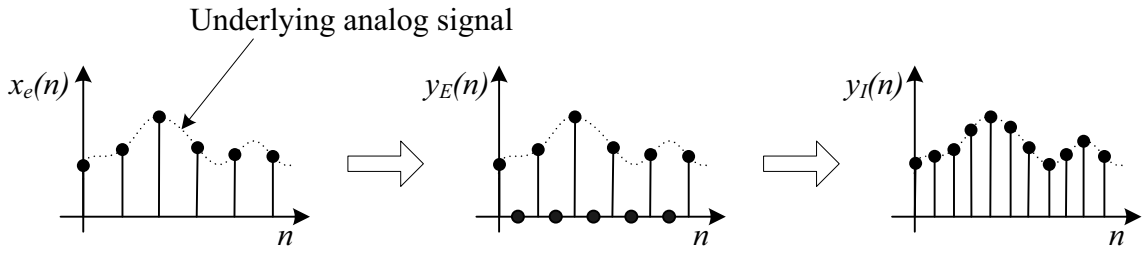


Fig. 2.3 Interpolation process.

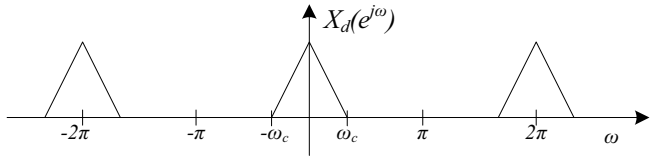
Frequency Domain

In the case of the decimator, it can be shown that the expression in the frequency domain for $Y_D(e^{j\omega})$ as a function of $X_d(e^{j\omega})$ is [3]:

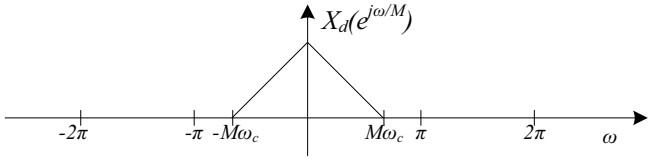
$$Y_D(e^{j\omega}) = \frac{1}{M}\sum_{k=0}^{M-1} X_d(e^{j(\omega-2\pi k)/M}). \tag{2.1}$$

For a more insightful discussion, a graphical interpretation of this equation is given. The first step is to obtain $X_d(e^{j\omega/M})$ by stretching $X_d(e^{j\omega})$ by a factor of M (see Figure 2.4(b)). The second is to create $M-1$ shifted copies of this stretched version, resulting in $X_d(e^{j(\omega-2\pi k)/M})$ for $k = 1, \ldots, M-1$; and the third and final step is summing these stretched copies and dividing by M, such that there is a copy every 2π. The result is $Y_D(e^{j\omega})$ (see Figure2.4(c)).

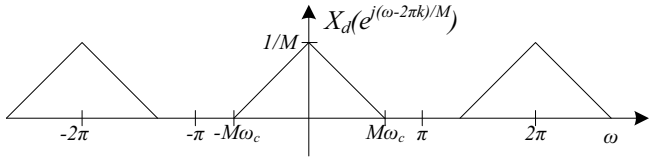
Taking a closer look at these figures, it becomes clear that in order to retain all the information in the frequency spectrum of $X_d(e^{j\omega})$, it is imperative that the frequency content of the signal be limited to π/M; otherwise, during the summation of the copies of $X_d(e^{j\omega/M})$, there will be overlap, resulting in a loss of information (see Figures 2.4(d) and 2.4(e)) and the impossibility of recovering the original signal $x_d(n)$ from $y_D(n)$. This overlap is a phenomenon known as aliasing. In order to ensure that aliasing will not occur, or at least to minimize its effects, an *anti-aliasing filter* is inserted before the decimation



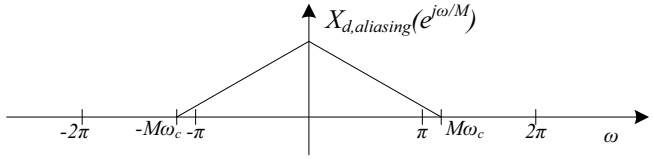
(a) Original signal.



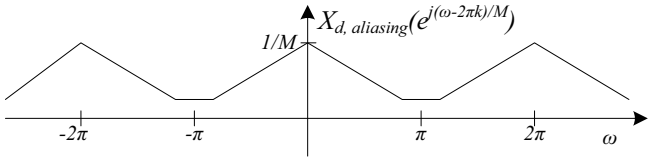
(b) Stretched signal.



(c) Downsampled signal $Y_D(e^{j\omega})$.



(d) Stretched signal with $\omega_c > \pi/M$.



(e) Aliased signal.

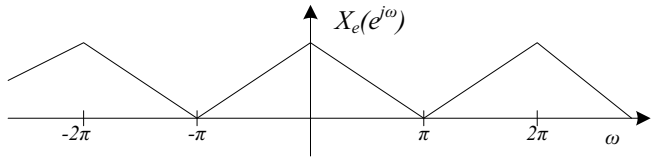
Fig. 2.4 Illustration of the downsampling process.

process, which is essentially a lowpass filter that strongly attenuates the frequency content above π/M.

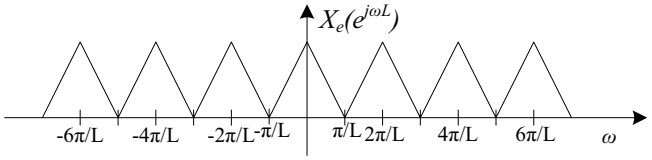
Next, in the case of the expander, the frequency domain expression for $Y_E(e^{j\omega})$ can be shown to be [3]:

$$Y_E(e^{j\omega}) = X_e(e^{j\omega L}). \tag{2.2}$$

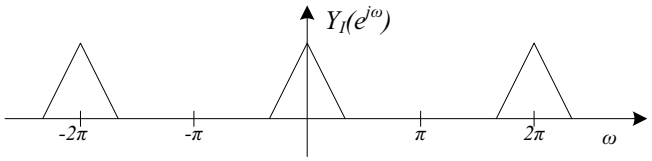
This represents an L-fold compression of $X_e(e^{j\omega})$, and, in addition, images at every $2\pi/L$, called the *imaging effect*(see Figure2.5(b)).



(a) Original signal.



(b) Upsampled signal $Y_E(e^{j\omega})$.



(c) Interpolated signal $Y_I(e^{j\omega})$.

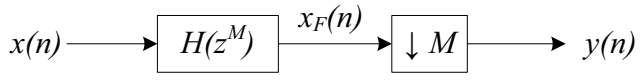
Fig. 2.5 Illustration of the upsampling process.

Finally, considering a case in which a signal (say $X_d(e^{j\omega})$ from Figure 2.4(a)) were to be adequately downsampled (yielding $X_e(e^{j\omega})$ from Figure 2.5(a))and subsequently upsampled ($X_e(e^{j\omega L})$), it is evident that in order to reproduce the original signal $X_d(e^{j\omega})$, the expanded signal should be lowpass filtered. This shows once more the use of the interpolation filter; indeed, the interpolated signal $Y_I(e^{j\omega})$ is identical to $X_d(e^{j\omega})$. This phenomenon is referred to as *perfect reconstruction*. It should be noted that had the downsampling process entailed any aliasing, perfect reconstruction would not be possible, as the high-frequency content would have been compromised.

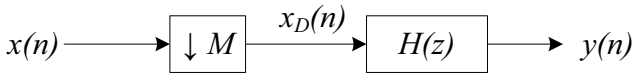
2.1.2 Polyphase decomposition

Another of the basic techniques used to derive more efficient implementation structures for linear filters is the polyphase decomposition. Not only does it reduce the computational complexity of multirate systems, it allows for a greater simplification of theoretical results, and will thus be used in subsequent sections.

As it was seen previously, in a subband coder where it is desired to manipulate different frequency intervals independently, the input signal is first divided into the various frequency subdivisions by means of an analysis filter bank. Then, the output signals of the filter bank are downsampled, and processing — such as quantization, for example — may be done at this point. After the intended operations are completed, these subband signals are then upsampled and are passed through the synthesis filter bank, completing the interpolation and reconstruction procedure. Therefore, originally, the decimator and the expander were cradled between two filters in each subband. However, there are two identities, one in the case of downsampling and the other in the case of upsampling, that can be derived [1] that are helpful in the manipulation and understanding of the polyphase decomposition: they are the interchangeability of the order of the filtering operation and the decimator and expander blocks.



(a)



(b)

Fig. 2.6 Interchangeability of the filter and the decimator.

Considering Figure 2.6(a), it is clear that:

$$X_F(e^{j\omega}) = H(e^{j\omega M})X(e^{j\omega}). \tag{2.3}$$

Next, recalling Eq. (2.1):

$$Y(e^{j\omega}) = \frac{1}{M}\sum_{k=0}^{M-1} X_F(e^{j(\omega-2\pi k)/M}), \tag{2.4}$$

and substituting Eq. (2.3) into Eq. (2.4):

$$Y(e^{j\omega}) = \frac{1}{M}\sum_{k=0}^{M-1} H(e^{j(\omega-2\pi k)})X(e^{j(\omega-2\pi k)/M}). \tag{2.5}$$

Finally, since $H(e^{j(\omega-2\pi k)}) = H(e^{j\omega})$ due to the periodicity of the Fourier transform, Eq. (2.5) becomes:

$$\begin{aligned} Y(e^{j\omega}) &= H(e^{j\omega})\frac{1}{M}\sum_{k=0}^{M-1} X(e^{j(\omega-2\pi k)/M}) \\ &= H(e^{j\omega})X_D(e^{j\omega}), \end{aligned}$$

proving that the filtering and decimation order can be interchanged.

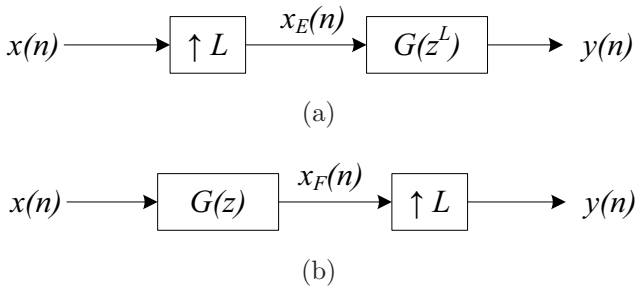


Fig. 2.7 Interchangeability of the filter and the expander.

Similarly, referring to Figure 2.7,

$$X_F(e^{j\omega}) = G(e^{j\omega})X(e^{j\omega}),$$

and therefore, from Eq. (2.2):

$$\begin{aligned} Y(e^{j\omega}) &= X_F(e^{j\omega L}) \\ &= G(e^{j\omega L})X(e^{j\omega L}), \end{aligned}$$

leading to (also following from Eq. (2.2)):

$$Y(e^{j\omega}) = G(e^{j\omega L})X_E(e^{j\omega}),$$

proving that the order of filter and the expander may be reversed. These identities (illustrated in Figures 2.6 and 2.7) are referred to as the *Noble identities*.

In the original implementation of a subband coder (Figure 1.1), the analysis filter bank outputs a value for every time sample n and then the downsampler discards $M-1$ samples for every sample it retains. This suggests that there should be a manner by which only the retained samples are computed, rather than wasting resources by computing useless samples.

Considering the decomposition of the impulse response of a filter $h(n)$ into the M subsequences $h_i(n)$:

$$h_i(n) = \begin{cases} h(n+i), & \text{if } n \text{ is an integer-multiple of } M \\ 0, & \text{otherwise,} \end{cases}$$

it is straightforward to see that $h(n)$ can be recovered through:

$$h(n) = \sum_{i=0}^{M-1} h_i(n-i). \tag{2.6}$$

The $h_i(n)$ are, in fact, an equivalent M-parallel-filter implementation of the original filter $h(n)$.

Next, if the subsequences $h_i(n)$ are downsampled by M, the resulting sequences $e_i(n)$

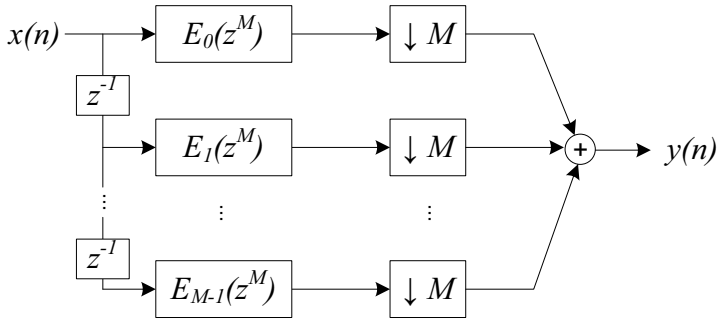
$$e_i(n) = h_i(nM) = h(nM+i), \tag{2.7}$$

are called the polyphase components of $h(n)$. Combining Eq. (2.6) and Eq. (2.7) the frequency domain expression relating the polyphase components to the original filter is

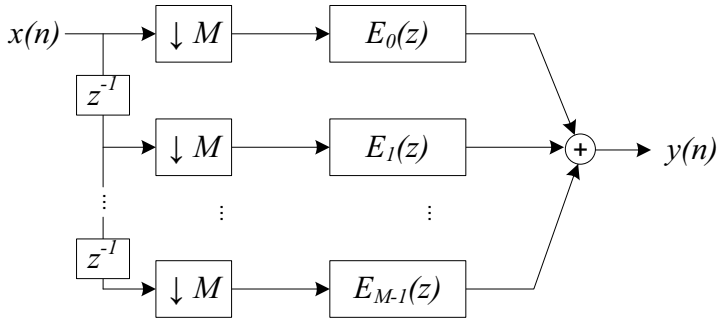
$$H(z) = \sum_{i=0}^{M-1} E_i(z^M)z^{-i}.$$

This equation corresponds to the system shown in Figure 2.8(a) and is an equivalent im-

plementation to the one shown in Figure 2.6(a).



(a) Filtering process using the polyphase decomposition.



(b) More efficient implementation.

Fig. 2.8 Polyphase implementation of $H(z)$.

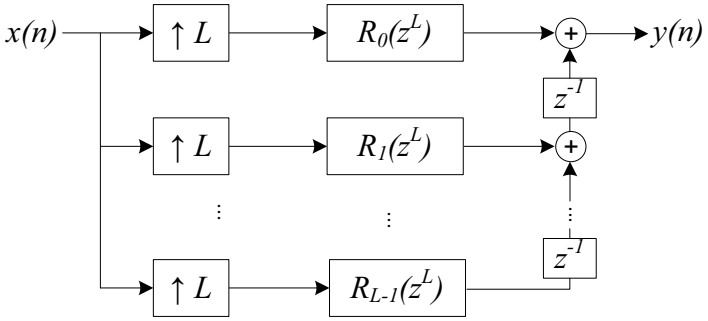
Recalling the previously enunciated identity (Figure 2.6), the order of the polyphase component filters and the decimator can be reversed yielding the system in Figure 2.8(b). Because the downsampling process now occurs before the filtering, the unused samples are no longer computed, resulting in an economy on the number of computations [1].

Similarly, in the case of the reconstruction process, where the upsampling process precedes the synthesis filters $F_k(z)$ (see Figure 1.1), the zero-valued samples inserted by the expander are operated on by the filters. Therefore, savings will be incurred if the filters can be modified such that they deal only with samples containing pertinent information.

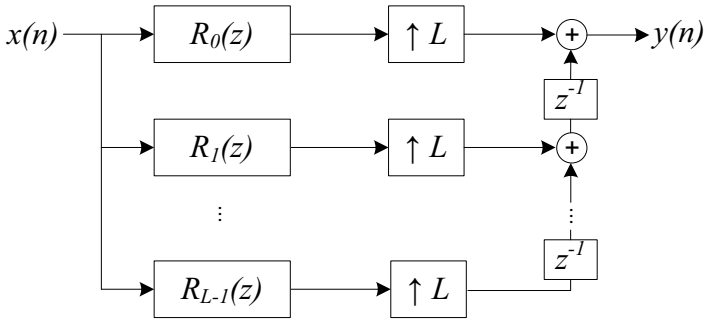
Through an identical manipulation, a filter $f(n)$ can be decomposed into its polyphase components:

$$F(z) = \sum_{i=0}^{L-1} R_i(z^L) z^{-i},$$

and can be graphically depicted as shown in Figure 2.9(a). Again, applying the identity



(a) Filtering process using the polyphase decomposition.



(b) More efficient implementation.

Fig. 2.9 Polyphase implementation of $F(z)$

for the case of the expander, the system in Figure 2.9(a) can be rearranged to yield the system depicted in Figure 2.9(b).

As in the case of the downsampling operation, the economy in computation results from the fact that the filtering is done at the lower sampling rate, rather than the higher one [1].

Finally, it is convenient for theoretical and experimental manipulations to introduce the expression of the polyphase components of a filter bank in matrix form. The $N \times M$ analysis polyphase matrix $\mathbf{E}(z)$ has elements $[\mathbf{E}(z)]_{k,n}$ defined as [14]:

$$[\mathbf{E}(z)]_{k,n} = \sum_{m=-\infty}^{\infty} h_k(mM - n)z^{-m}, \tag{2.8}$$

where $k = 0, ..., N-1$ and $n = 0, ..., M-1$. Here, N corresponds to the number of subbands, $h_k(n)$ is the filter corresponding to the k-th subband, while M is the downsampling factor.

Similarly, the $M \times N$ synthesis polyphase matrix $\mathbf{R}(z)$ has elements $[\mathbf{R}(z)]_{k,n}$ defined as [14]:

$$[\mathbf{R}(z)]_{k,n} = \sum_{m=-\infty}^{\infty} f_k(mM + n)z^{-m}, \tag{2.9}$$

where $k = 0, ..., N-1$ and $n = 0, ..., M-1$.

By replacing the filters with their polyphase components and reversing the order of the filtering and downsampling/upsampling blocks, the subband coder from Figure 1.1 can be redrawn as shown in Figure 2.10.

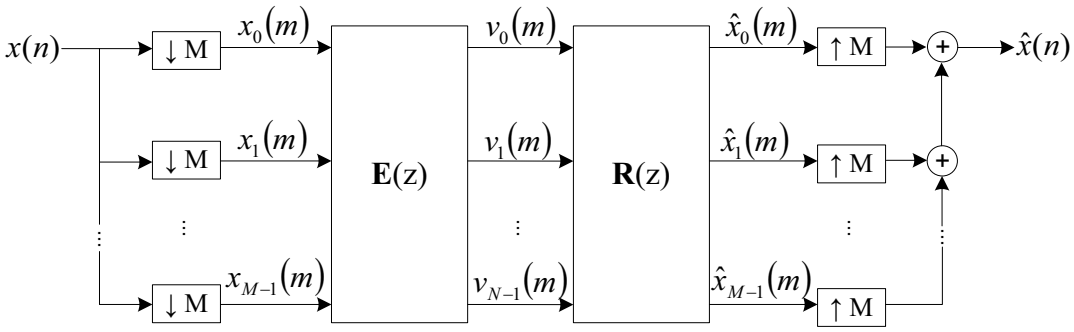


Fig. 2.10 Matrix representation of a subband coder using the polyphase decomposition.

2.2 Filter bank design

One of the advantages of multirate systems is that the sampling theorem need only be satisfied on the sum of the channels, rather than on each one individually [4]. Consequently, unrealizable ideal bandpass filters are no longer needed in the analysis bank, while simultaneous design of the bank is now required. This section takes a closer look at the parallel design of the filters contained in the filter bank. First, the possible distortions that may be caused by a filter bank are discussed, followed by a presentation of the conditions guaranteeing perfect reconstruction of the input signal. Finally, the lossless lattice structure is introduced, which will be used in the design of the filter banks in Chapter 3.

2.2.1 Distortion

There are three types of distortion that a signal passing through a filter bank may be subjected to: Aliasing Distortion, Amplitude Distortion and Phase Distortion.

Aliasing Distortion (ALD)

Referring to Figure 1.1, the overall relation that governs the system, using Eq. (2.1) and Eq. (2.2), is

$$\hat{X}(z) = \frac{1}{M} \sum_{k=0}^{N-1} \sum_{n=0}^{M-1} F_k(z) H_k(zW^n) X(zW^n), \tag{2.10}$$

where W^n replaces $e^{-j\frac{2\pi}{M}n}$ for simplicity [2]. From this equation, it is clear that the output $\hat{X}(z)$ contains the original signal $X(Z)$ and $M-1$ aliasing components $X(zW^n)$ where $n > 0$. It stands to reason that if it were possible to choose $H_k(z)$ and $F_k(z)$ appropriately such that

$$\sum_{k=0}^{N-1} \sum_{n=0}^{M-1} F_k(z) H_k(zW^n) X(zW^n) = 0,$$

the overall system would be alias-free. Indeed, such choices do exist and the conditions to be met are set out in the next section.

Amplitude Distortion (AMD)

Assuming that Aliasing Distortion is eliminated, the transfer function relating $\hat{X}(z)$ and $X(z)$ is given by:

$$T(z) = \frac{\hat{X}(z)}{X(z)} = \frac{1}{M}\sum_{k=0}^{M-1} F_k(z)H_k(z),$$

which is a linear time invariant system (LTI). Now, if $|T(e^{j\omega})|$ is not constant for all ω, then the output signal $\hat{x}(n)$ will suffer from amplitude distortion. Therefore the overall transfer function must be constrained to be all-pass in order to eliminate AMD [2].

Phase Distortion (PHD)

Again, if ALD is canceled, the output signal $\hat{x}(n)$ will still suffer from phase distortion if the transfer function $T(z)$ does not have linear phase. Consequently, the transfer function $T(z)$ must be FIR and have linear phase [2].

2.2.2 Conditions on perfect reconstruction

During the design of filter banks, FIR filters are often desirable, since they are always stable, their numerical properties are good and they can achieve linear phase behaviour [4]. In particular, if the filters that compose the filter bank are linear phase, then the overall transfer function of the system will be linear phase [2] and PHD is eliminated. Furthermore, FIR filters do not require pole-zero cancelation between distinct filters during the reconstruction process [4], which could cause instability in the case where the pole-zero cancelation is imperfect due to the precision of the coefficients. Consequently, FIR filters were chosen for the simulations in this study and so only the conditions on FIR filter banks will be discussed here.

In [4], two fundamental properties of subband coders are stated and proven. However before they are reproduced here, the matrix $\mathbf{H}_m(z)$ is introduced:

$$\mathbf{H}_m(z) = \begin{bmatrix} H_0(z) & H_0(zW) & \cdots & H_0(zW^{N-1}) \\ H_1(z) & H_1(zW) & \cdots & H_1(zW^{N-1}) \\ \vdots & \vdots & \ddots & \vdots \\ H_{M-1}(z) & H_{M-1}(zW) & \cdots & H_{M-1}(zW^{N-1}) \end{bmatrix}.$$

This matrix is alternatively called the *modulated filter matrix*, due to the W^k factors, or the *alias-component matrix*, whose meaning is obvious when recalling Eq. (2.10).

With this tool in hand, the following properties can be proven [4]

i) *Aliasing-free output* is achieved if

$$[\underline{g}(z)]^T \mathbf{H}_m(z) = \begin{bmatrix} T(z) & 0 & 0 & \cdots & 0 \end{bmatrix},$$

where $\underline{g}(z)$ is the vector of synthesis filters, and $T(z)$ is an arbitrary transmission filter.

ii) *Perfect reconstruction* is obtained if

$$[\underline{g}(z)]^T \mathbf{H}_m(z) = \begin{bmatrix} z^{-k} & 0 & 0 & \cdots & 0 \end{bmatrix},$$

where z^{-k} is an arbitrary delay. This ensures that AMD is eliminated as the overall transfer function is now all-pass and that PHD is eliminated since the phase distortion is now linear (simply a delay). Further, it can be shown that for perfect reconstruction it is sufficient that the determinant of $\mathbf{E}(z^M)$ be a pure delay, and that the delay constraint becomes necessary for the case of a downsampling factor of 2 and when the filters are modulated [4]. The role of the determinant of the analysis filter matrix is analogous, in the case of a single filter, to the minimum phase requirement to achieve reconstruction [4].

A theorem is proven in [4]:

> *Theorem 1.* Aliasing-free reconstruction in a subband coder is possible if and only if the analysis filter matrix $\mathbf{H}_m(z)$ has rank M (the downsampling factor).

Although the details of the proof are excluded here, an intuitive reasoning is given. If the matrix $\mathbf{E}(z^M)$ has rank M, then each input signal will have a distinct output signal, representing a one-to-one transformation (injection). On the other hand, if the rank is less than M, then groups of signals will yield the same output, making the original signal unrecoverable [4].

In the derivations in [4], non-linear effects such as quantization were not considered, for they cannot be completely eliminated. However, the manner in which they can be reduced is reserved for a later discussion.

2.2.3 Lossless matrices

Recalling Figure 2.10, it is clear that if the analysis filters and synthesis filters were designed in such a way that they would "cancel" each other's effects in some way, the result would be a perfect reconstruction system. Indeed, it was proven in [16] that a necessary and sufficient condition for perfect reconstruction is that the overall response $\mathbf{P}(z)$ of the cascade of the analysis and synthesis polyphase matrices $\mathbf{R}(z)\mathbf{E}(z)$ have the following form

$$\mathbf{P}(z) = dz^{-K}\mathbf{I}_M,$$

where d is an arbitrary nonzero constant and K is the overall delay through the system. Next, if this condition is satisfied the synthesis filters will, in general, be IIR, since , given $\mathbf{E}(z)$, the determination of $\mathbf{R}(z)$ will involve the inversion of $\mathbf{E}(z)$. However, in order to obtain a linear phase response, all the filters employed must be FIR filters, as mentioned previously. Referring to the previous section, it is required that the determinant of $\mathbf{E}(z)$ be a delay. Fortunately, there exists a family of FIR filters for which the determinant is a delay: lossless matrices [2].

If a transfer matrix $\mathbf{L}(z)$ describing the input-output relationship of a system whose input vector is $\underline{x}(z)$ and output vector $\underline{y}(z)$ is such that

$$E_x = cE_y,$$

where $E_v \equiv \sum_n \underline{v}^\dagger(n)\underline{v}(n)$, and $c > 0$ holds for any input $\underline{x}(z)$, the system is said to be lossless [2]. Equivalently, the transfer matrix $\mathbf{L}(z)$ is lossless if it is stable and

$$\widetilde{\mathbf{L}}(z)\mathbf{L}(z) = c\mathbf{I} \quad \text{, for all } z, \tag{2.11}$$

where $\widetilde{\mathbf{L}}(z)$ denotes conjugation of the coefficients, transposition of the matrix and replacement of z by z^{-1}. This property implies that

$$\mathbf{L}^\dagger(e^{j\omega})\mathbf{L}(e^{j\omega}) = c\mathbf{I} \quad \text{, for all } \omega,$$

where $\mathbf{L}^\dagger(e^{j\omega})$ denotes transpose conjugation. This further indicates that $\frac{1}{\sqrt{c}}\mathbf{L}(z)$ is unitary on the unit circle.

If $\mathbf{E}(z)$, the analysis polyphase matrix, is chosen to be lossless, $\mathbf{E}^{-1}(z)$ is simply $\widetilde{\mathbf{E}}(z)$,

inversion is avoided and the synthesis polyphase matrix $\mathbf{R}(z)$ can easily be FIR by choosing $\mathbf{R}(z) = z^{-K}\widetilde{\mathbf{E}}(z)$. Thus, if $\mathbf{E}(z)$ can be chosen to be lossless, the objective is thus accomplished: all filters are FIR, the phase is linear and the filter bank satisfies the perfect reconstruction property. Finally, choosing $\mathbf{R}(z)$ in this manner leads to the following choice of filter coefficients:

$$f_k(n) = \alpha h_k^*(n_0 - 1 - n) \quad 0 \leq k \leq N-1, \tag{2.12}$$

where n_0 is the length of the longest analysis filter and α is an arbitrary non-zero constant [2].

The manner in which a lossless $\mathbf{E}(z)$ can be achieved and the design of the filter bank are reserved for discussion in Chapter 3.

2.3 Oversampled filter banks

Attention is now turned to the oversampling of the subband signals in a filter bank. Figure 2.3 reproduces the subband coder with N subbands and a downsampling factor of M, from Figure 1.1 for sake of continuity. (It is also recalled that oversampled filter banks are implemented by choosing $M < N$.) In this section, in order to take a more formal approach

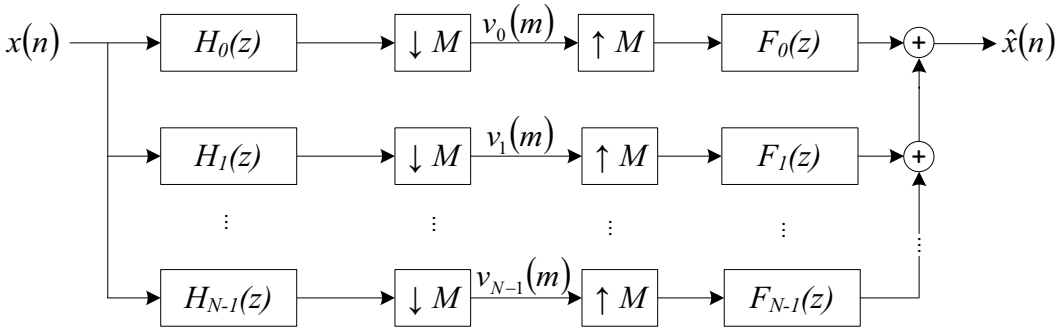


Fig. 2.11 Subband Coder.

to the problem, a few results on frame expansions in the context of filter banks are first presented[1]. Subsequently, their relation to subband signals and oversampled filter banks is drawn.

[1]For a more in depth treatment of frame theory applied to oversampled filter banks, one is referred to [6]

2.3.1 Frame expansions

It is well-known from linear algebra that given an N-dimensional vector space $\mathbf{R}^N$, any vector $\mathbf{x}$ in the space can be represented as a linear combination of a set of vectors $\{\underline{v}_i\}$, $i = 0, 1, \ldots, K$, given that this set of vectors span the space $\mathbf{R}^N$. The scalar weights of the linear combination are given by the inner product (denoted by $\langle \cdot, \cdot \rangle$) of the vector $\underline{x}$ with the spanning vectors $\underline{v}_i$, yielding

$$\underline{x} = \sum_{i=0}^{K} \langle \underline{x}, \underline{v}_i \rangle \underline{v}_i.$$

This is analogous to frame theory. Although an in-depth analysis of frame theory is beyond the scope of this work, a few results pertinent to filter banks are presented here.

It can be shown [6] that a given signal $x(n)$ can be expanded as

$$x(n) = \sum_{i=0}^{N-1} \sum_{j=-\infty}^{\infty} \langle x, \phi_{i,j} \rangle \psi_{i,j}(n). \tag{2.13}$$

where the $\phi_{i,j}$ and the $\psi_{i,j}$ are members of the families of vectors Φ and Ψ respectively, as defined in [6]:

$$\Phi = \{\ \phi_{i,j} : \phi_{i,j}(n) = \phi_i(n - jM) \quad i = 0, \ldots N-1, j \in \mathbf{Z}\ \},$$

$$\Psi = \{\ \psi_{i,j} : \psi_{i,j}(n) = \psi_i(n - jM) \quad i = 0, \ldots N-1, j \in \mathbf{Z}\ \},$$

where $M \leq N$. Thus, the analysis of $x(n)$ is performed through a sliding window, using N elementary waveforms $\phi_i(n)$, while the synthesis is done using the $\psi_i(n)$.

In order for Eq. (2.13) to hold for any $x(n)$ that is an element of the space of square summable series $\ell^2(\mathbf{Z})$ (i.e. finite energy signals) and be implemented in a numerically stable way, the families Φ and Ψ must constitute frames in $\ell^2(\mathbf{Z})$. A frame is defined as the family of vectors Φ that satisfy the condition

$$A\|x\|^2 \leq \sum_{i=0}^{N-1} \sum_{j=-\infty}^{\infty} |\langle x, \phi_{i,j} \rangle|^2 \leq B\|x\|^2, \tag{2.14}$$

for some constants $A > 0$ and $B < \infty$ and for any $x \in \ell^2(\mathbf{Z})$. Further, it is emphasized

that Φ and Ψ can only be frames if $M \leq N$.

Summarizing, if a signal $x(n)$ is decomposed using the frame Φ, it can be recovered through another frame Ψ. It should be noted that given a certain frame Φ, the reconstruction frame Ψ is not necessarily unique [6]. However, there is one solution which is of interest: the dual frame to Φ [17]. It is the only synthesis frame that leads to maximal noise reduction as the orthogonal component of the additive noise with respect to the range of the expansion will be projected to zero (more on this in Chapter 3).

2.3.2 Filter bank frames

It can be shown that for a given frame Φ, the filter bank with N subbands whose impulse responses $h_i(n)$ are related to the members of Φ through

$$h_i(n) = \phi_i^*(-n)$$

for $i = 0, 1, \ldots, N-1$, implements a frame expansion, and Φ is termed a *filter bank frame* [6]. Indeed, recalling Figure (2.3), the subband signals $v_i(m)$ can be written as the inner products

$$v_i(m) = \langle x, h_{i,m} \rangle, \tag{2.15}$$

where $h_{i,m}(n) = h_i^*(mM - n)$ [14].

Next, if Φ does in fact constitute a frame, then the signal $x(n)$ can be recovered from the subband signals when the synthesis filters are given by the members of the synthesis frame Ψ:

$$f_i(n) = \psi_i(n).$$

If the filter bank satisfies the perfect reconstruction condition, the reconstructed signal $\hat{x}(n)$ is equal to the input signal $x(n)$ and can be expressed as

$$\hat{x}(n) = \sum_{i=0}^{N-1} \sum_{m=-\infty}^{\infty} \langle x, h_{i,m} \rangle f_{i,m}(n),$$

where $f_{i,m}(n) = f_i^*(n - mM)$. Comparing this result with Eq. 2.14, it is seen that the filter bank expands the input signal $x(n)$ as a function of the set $\{f_{i,m}(n)\}$ [14].

Next, a few theorems on filter bank frame expansions are given; for the proofs the reader

is referred to [6].

Theorem 2. A filter bank implements a frame expansion if and only if its polyphase analysis matrix is of full rank on the unit circle.

Theorem 3. A filter bank implements a tight[2] frame expansion if and only if its polyphase analysis matrix is paraunitary $\widetilde{\mathbf{E}}(z)\mathbf{E}(z) = c\mathbf{I}$.

It is noted that these theorems share some common points with the conditions on perfect reconstruction for filter banks enunciated in the previous section. This is expected, since the existence of an analysis frame Φ implies the existence of a frame Ψ that will yield the recovery of the original signal $x(n)$.

Finally, if the filter bank does implement a frame expansion, the subband signals $v_i(m)$ satisfy (recalling Eqs. 2.14 and 2.15)

$$A\|x\|^2 \leq \sum_{i=0}^{N-1} \sum_{j=-\infty}^{\infty} |v_i(j)|^2 \leq B\|x\|^2$$

for any signal $x(n)$ in $\ell^2(\mathbf{Z})$.

2.3.3 Oversampled filter bank frames

In the critically sampled case ($M = N$), the subband signals $v_i(m)$ yield orthogonal or biorthogonal expansions of the input signal $x(n)$ to the filter bank. On the other hand, in an oversampled filter bank ($M < N$) the $v_i(m)$ form a redundant representation of the signal $x(n)$ [14]. Further, by defining a filter bank analysis operator $\mathcal{T}$ that assigns the set of subband signals $v_i(m)$ to an input signal $x(n)$, it is shown in [14] that the range space $\mathcal{R}$ of the operator $\mathcal{T}$ is only a subspace of the codomain $[\ell^2(\mathbf{Z})]^N$ of $\mathcal{T}$.

Similarly, a synthesis filter bank operator $\mathcal{U}$ can be defined that maps the set of subband signals $v_i(m)$ to the reconstructed signal $\hat{x}(n)$. Because the subband signals $v_i(m)$ are contained in a subspace of $[\ell^2(\mathbf{Z})]^N$, $\mathcal{U}$ is not unique. This is instrumental in the justification of the freedom in the design of the synthesis filters, inherent to oversampled filter banks [14]. While only one synthesis filter bank will have the maximal noise reduction property, others might have desirable design characteristics.

[2] A tight frame corresponds to $A\|x\|^2 = \sum_{i=0}^{N-1} \sum_{j=-\infty}^{\infty} |\langle x, \phi_{i,j}\rangle|^2 = B\|x\|^2$

Again, the details of this theory is beyond the scope of this text, so only an intuitive overview was given because the result is pertinent to the design of the noise-shaping system used in oversampled filter banks, which will be treated in the following chapter.

2.4 Summary

In this chapter, notions instrumental to multirate signal processing such as downsampling, upsampling and the polyphase decomposition were first introduced. Subsequently, topics in filter design were discussed in order to facilitate the obtention of filter banks. In particular, the condition for perfect reconstruction was enunciated and a manner in which it is satisfied while keeping all filters at a finite length was described. Finally, oversampled filter banks were presented in the context of frame theory, which was briefly touched upon. The conclusion was that oversampled filter banks not only permit a certain design freedom for the synthesis filter bank, but also yield a redundant representation of the signal input to the analysis filter bank. The manner in which this redundancy can be used is the main focus of Chapter 3.

Chapter 3

Oversampled filter banks with quantization noise shaping

In order to investigate the performance of an oversampled subband coder, the design of the filters used in the signal decomposition according to the conditions enunciated in the previous chapters is essential. The first topic discussed in this chapter will thus be the method used in the design of these filters. Then, attention is turned to the derivation of the noise-shaping filters, based on the previously obtained analysis and synthesis filters.

3.1 Iterative filter design

In this section, it is first discussed how the design method for the filter banks was chosen, followed by an outline of the selected method and finally a description of the optimization algorithm.

3.1.1 Justification of the choice of the design method

As mentioned previously, there is a certain amount of design freedom inherent to oversampled filter banks: because the subband signals $v_i(m)$ are contained within a subspace of the codomain $[\ell^2(\mathbf{Z})]^N$, the reconstruction frame is not unique. Hence, there are many synthesis filter banks that will lead to the recovery of the original signal $x(n)$. Indeed, there is an emerging exploration of this freedom [10], [12], [13], [11]. However, of all the possible reconstruction frames, there is one frame that has a maximal noise reduction property: the

dual to the analysis frame.

The para-pseudo-inverse is defined as:

$$\hat{\mathbf{R}}(z) = [\widetilde{\mathbf{E}}(z)\mathbf{E}(z)]^{-1}\widetilde{\mathbf{E}}(z).$$

Recalling that perfect reconstruction is obtained if the analysis and synthesis polyphase matrices satisfy $\mathbf{R}(z)\mathbf{E}(z) = \mathbf{I}_M$, the para-pseudo-inverse $\hat{\mathbf{R}}(z)$ of $\mathbf{E}(z)$ is the minimum-norm least-squares solution [18]. Applied to filter banks, this means that $\hat{\mathbf{R}}(z)$ is the particular solution that minimizes the reconstruction error variance due to the quantization process[1], when compared to all other perfect reconstruction synthesis filter banks. This is in fact equivalent to the frame dual to the analysis filter bank [14]. Further, recalling the definition of a lossless matrix (Eq. (2.11)), if $\mathbf{E}(z)$ is chosen to be lossless, $\hat{\mathbf{R}}(z)$ is then given by

$$\hat{\mathbf{R}}(z) = [c\mathbf{I}]^{-1}\widetilde{\mathbf{E}}(z) = \frac{1}{c}\widetilde{\mathbf{E}}(z).$$

And so the choice of a lossless $\mathbf{E}(z)$ is once more justified: not only does it produce FIR synthesis filters as the inversion of $\mathbf{E}(z)$ is avoided, but it also yields the synthesis filter bank corresponding to the frame dual of the analysis filter bank, minimizing the reconstruction error (more on this in Section 3.2).

Because this work is focused on noise-reduction, it is not desired to take advantage of the design freedom of the filter banks. Consequently, filter banks that satisfy the perfect reconstruction condition for a critical sampling rate are adequate. In fact, if $M < N$ for an N-channel filter bank designed for the critically sample case, the reconstructed signal $\hat{x}(n)$ will not be affected, except for a scale factor [19], given that the oversampling ratio $K = N/M$ is an integer. Furthermore, perfect reconstruction filter banks allow for a comparison in performance between critically and oversampled filter banks.

Cosine-modulated filter banks were chosen for this work, as their design involves the construction of only one filter: the prototype filter is then modulated in order to obtain the remaining member filters of the filter bank. Moreover, cosine-modulated filter banks were chosen since their subband signals are real-valued, given that the input signal is also real-valued, as opposed to discrete Fourier transform filter banks whose resulting subband signal will be complex [10].

[1] the reconstruction error variance due to the quantization process will be defined in the following section

The selected design method was proposed by Koilpillai and Vaidyanathan [20]. This method was chosen because the resulting filter banks satisfy the perfect reconstruction condition, the number of channels can be arbitrarily selected and all the analysis and synthesis filters are of equal length. Furthermore, the objective function to be optimized is relatively simple and requires a small number of parameters to be optimized, while it can be shown that the perfect reconstruction property is maintained, even when the coefficients are quantized [20].

Finally, the optimization method for the objective function was based on [21]. The minimization is achieved using the modified Newton method, which allows for the selection of the search direction and the size of the steps used to find a solution. This selection is critical as a larger step size will increase the speed with which the algorithm converges to a solution, while if it is too large the optimal solution may never be reached [22].

3.1.2 Cosine-modulated filter banks using a lossless lattice structure

Since the perfect reconstruction property of the filter bank is ensured by the choice of a lossless analysis filter bank $\mathbf{E}(z)$ [20], conditions on the design of $\mathbf{E}(z)$ such that it is lossless must first be enunciated. To do so, the polyphase representation of the prototype filter $P_0(z)$ is first rewritten as

$$\begin{aligned} P_0(z) &= \sum_{n=0}^{2mN-1} p_0(n) z^{-n} \\ &= \sum_{q=0}^{2N-1} \sum_{p=0}^{m-1} p_0(q+2pM) z^{-q+2pN} \\ &= \sum_{q=0}^{2N-1} z^{-q} W_q(z^{2N}). \end{aligned} \tag{3.1}$$

It is noted that this equation appears different than the previously defined polyphase decomposition, but upon closer inspection it is clear that it is essentially the same. This new expression is possible because the length of the prototype filter is constrained to be $L_h = 2mN$, as the polyphase components are constrained to a length of $2m$. It will be seen shortly that this re-indexing simplifies the derivation and implementation of the design strategy.

Indeed, since the prototype filter is linear phase (and hence symmetric) and its length is $L_h = 2mN$, the polyphase components $W_q(z)$ are related by

$$W_k(z) = z^{-(m-1)}\widetilde{W}_{2N-1-k}(z) \quad \text{for } 0 \le k \le N-1 \tag{3.2}$$

Next, it can be verified [20] that a necessary and sufficient condition for $\mathbf{E}(z)$ to be lossless is that the appropriate pairs of polyphase components of $P_0(z)$ be power complementary:

$$\widetilde{W}_k(z)W_k(z) + \widetilde{W}_{N+k}(z)W_{N+k}(z) = \tfrac{1}{2N} \quad \text{for } 0 \le k \le N-1. \tag{3.3}$$

Further, it is noted that due to the symmetry of the prototype filter demonstrated by Eq. (3.2), the power complementary condition in Eq. (3.3) is redundant. An equivalent condition is then given by

$$\widetilde{W}_k(z)W_k(z) + \widetilde{W}_{N+k}(z)W_{N+k}(z) = \tfrac{1}{2N} \quad \text{for } 0 \le k \le \tfrac{N}{2}-1,$$

for N even, while

$$\begin{aligned} \widetilde{W}_k(z)W_k(z) + \widetilde{W}_{N+k}(z)W_{N+k}(z) &= \tfrac{1}{2N} \quad \text{for } 0 \le k \le \left\lfloor \tfrac{N}{2} \right\rfloor - 1, \\ \widetilde{W}_k(z)W_k(z) &= \tfrac{1}{2N} \quad \text{for } k = \tfrac{N-1}{2} \end{aligned}$$

for N odd [20].

It is observed that in the case of an odd number of channels N, the component $W_{(N-1)/2}(z)$ (and $W_{N+(N-1)/2}(z)$ by symmetry) is constrained to be a pure delay, determined by the length of the prototype filter L_h.

Next, attention is turned to the manner in which these pairwise power complementary polyphase components can be obtained. It is stated in [20] that any FIR bounded-real[2] pair $\{S(z), T(z)\}$ that satisfies

$$\widetilde{S}(z)S(z) + \widetilde{T}(z)T(z) = 1, \quad \forall z \tag{3.4}$$

can always be realized as the nonrecursive cascade of two-channel lossless lattice structures [20]. Thus, the pair $\{W_k(z), W_{N+k}(z)\}$ can be generated by the cascade of $m-1$ lattice structures, shown in Figure 3.1. In this figure, the superscript p denotes the p^{th} lattice

[2] A stable digital filter $H(z)$ with real coefficients is said to be bounded-real if $|H(e^{j\omega})| \le 1$, $\forall \omega$.

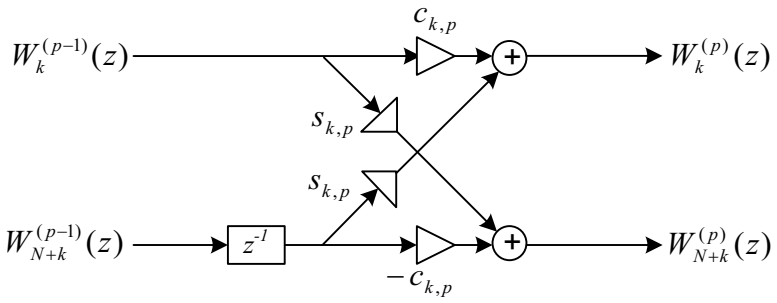


Fig. 3.1 Two-channel lattice section

section ($1 \leq p \leq m-1$), $c_{k,p} = \cos\theta_{k,p}$ and $s_{k,p} = \sin\theta_{k,p}$. Hence, the k $\{W_k(z), W_{N+k}(z)\}$ pairs ($0 \leq k \leq \lfloor \frac{N}{2} \rfloor - 1$) are obtained as the output of k parallel implementations of the cascade of $m-1$ two-channel lattice sections.

Next, given that the transfer functions from the input to the output of the p^{th} section of the k^{th} lattice are $\{W_k^{(p)}(z), W_{N+k}^{(p)}(z)\}$, they can be written as

$$\begin{bmatrix} W_k^{(p)}(z) \\ W_{N+k}^{(p)}(z) \end{bmatrix} = \begin{bmatrix} \cos\theta_{k,p} & \sin\theta_{k,p} \\ \sin\theta_{k,p} & -\cos\theta_{k,p} \end{bmatrix} \begin{bmatrix} W_k^{(p-1)}(z) \\ z^{-1}W_{N+k}^{(p-1)}(z) \end{bmatrix}, \tag{3.5}$$

$$1 \leq p \leq m-1 \ , \ 0 \leq k \leq \left\lfloor \frac{N}{2} \right\rfloor - 1.$$

Finally, the lattice transfer functions are initialized as

$$\begin{bmatrix} W_k^{(0)}(z) \\ W_{N+k}^{(0)}(z) \end{bmatrix} = \begin{bmatrix} \cos\theta_{k,0} \\ \sin\theta_{k,0} \end{bmatrix}, \tag{3.6}$$

$$0 \leq k \leq \left\lfloor \frac{N}{2} \right\rfloor - 1.$$

Summarizing, Eq. (3.3) gives the power complementary condition on the pairs of polyphase components $\{W_k(z), W_{N+k}(z)\}$ of the prototype filter $P_0(z)$ that ensures that the polyphase analysis matrix $\mathbf{E}(z)$ is lossless. This then guaranties that the resulting filter bank will have the perfect reconstruction property [20]. Next, it was noted that any FIR bounded-real pair satisfying Eq. (3.4) can be generated using a cascade of two-channel lattice structures. Consequently, the pairs of polyphase components are generated using the $\lfloor \frac{N}{2} \rfloor$ parallel cascade of $m-1$ lattice sections, while the remaining components are found using the

symmetry defined in Eq. (3.2) and, in the case of N odd, $W_{(N-1)/2}(z)$ and $W_{N+(N-1)/2}(z)$ are forced to be pure delays. Thus, $2N$ polyphase components are generated, each with length m, and the resulting prototype filter $P_0(z)$ has length $L_h = 2mN$.

It now remains to determine the optimal parameters $\theta_{k,p}$ of the two-channel lattice structures. This is the focus of the following section.

3.1.3 Optimization procedure

Because the perfect reconstruction property is inherently satisfied by the two-channel lattice structure, it need not be included as a constraint on the optimization of the parameters $\theta_{k,p}$. The objective function for the minimization is then selected as the stopband energy of the prototype filter, defined as [20]

$$\rho = \int_{\pi/2M+\varepsilon}^{\pi} |P_0(e^{j\omega})|^2 d\omega,$$

where the choice of ε governs the transition bandwidth of the prototype filter.

This equation is then rearranged to facilitate the computation:

$$\rho(\theta_{k,p}) = 4\underline{p}_0'^T(n)\left\{\int_{\pi/2M+\varepsilon}^{\pi} \underline{c}(\omega)\underline{c}^T(\omega)d\omega\right\}\underline{p}_0'(n),$$

where $\underline{c}(\omega) = [\ \cos\big((L_h-1)\omega/2\big)\quad \cos\big((L_h-3)\omega/2\big)\quad \cdots\quad \cos\big(\omega/2\big)\ ]^T$
and $\underline{p}_0'(n) = [\ p_0(0)\quad p_0(1)\quad \cdots\quad p_0(mN-1)\ ]^T$, i.e. the first mN elements of $p_0(n)$ [21]. The dependence of ρ on the $\theta_{k,p}$ stems from the coefficients $p_0(n)$, who themselves are related to the polyphase components $W_k(z^{2N})$ through Eq. (3.1), which are generated with the two-channel lattice structure whose parameters are the $\theta_{k,p}$.

In order to minimize the stopband energy $\rho(\theta_{k,p})$, Newton's method is used. In order to carry out the optimization, the vector Θ is first defined as an $m\lfloor\frac{N}{2}\rfloor \times 1$ vector arranged as $\Theta = \left[\ \theta_{0,0}\quad \cdots\quad \theta_{\lfloor\frac{N}{2}\rfloor,0}\quad \theta_{0,1}\quad \ldots\quad \theta_{\lfloor\frac{N}{2}\rfloor,m}\ \right]^T$. Then, as derived in [22], the recursive adaptive algorithm is given by

$$\Theta^n = \Theta_{k-1} - \lambda\{\nabla^2\rho(\Theta^{n-1})\}^{-1}\nabla\rho(\Theta^{n-1}), \tag{3.7}$$

where Θ^n is the vector of parameters to be optimized, Θ^{n-1} is the vector from the previous

iteration $n-1$ and $\rho(\cdot)$ is the objective function to be minimized.

Further, the gradient of the objective function $\nabla\rho(\Theta)$ is denoted by $\mathbf{D}(\Theta)$ and is determined by [21]

$$[\mathbf{D}(\Theta)]_i = 2\left\{\frac{\partial \underline{p}_0'(n)}{\partial[\Theta]_i}\right\}^T \mathbf{U}_s \underline{p}_0'(n), \tag{3.8}$$

where $\mathbf{U}_s$ is $\mathbf{U}_s = \int_{\pi/2M+\varepsilon}^{\pi} \underline{c}(\omega)\underline{c}^T(\omega)d\omega$ and the subscript i indicates the i^{th} element of a vector ($1 \leq i \leq m\lfloor\frac{N}{2}\rfloor$). The second order gradient $\nabla^2\rho(\Theta)$ is denoted by $\mathbf{H}(\Theta)$ and its elements are determined by [21]

$$[\mathbf{H}(\Theta)]_{i,j} = 2\left\{\frac{\partial^2 \underline{p}_0'(n)}{\partial[\Theta]_i\partial[\Theta]_j}\right\}^T \mathbf{U}_s \underline{p}_0' + 2\left\{\frac{\partial \underline{p}_0'(n)}{\partial[\Theta]_i}\right\}^T \mathbf{U}_s \left\{\frac{\partial \underline{p}_0'(n)}{\partial[\Theta]_j}\right\}. \tag{3.9}$$

The manner in which the vector $\mathbf{D}(\Theta)$ and the matrix $\mathbf{H}(\Theta)$ can be efficiently obtained through various two-channel lattice structures will be presented in Chapter 4.

Applying Eq. (3.7) to the optimization under study, the n^{th} iteration is then given by

$$\Theta^n = \Theta^{n-1} - \lambda\{\mathbf{H}(\Theta)\}^{-1}\mathbf{D}(\Theta). \tag{3.10}$$

Thus, by computing the appropriate matrices, a method has been described through which the lattice parameters $\theta_{k,p}$ can be optimized, based on the algorithm described in [21].

With the filter banks generated through this method in hand, attention is now directed to the manner in which noise can be reduced by introducing a noise-shaping system into the subband coder.

3.2 Noise-shaping system

Although there is no noise injected into the system by the subband coding process given that the filter banks in Figure 2.10 satisfy the perfect reconstruction condition, if quantizers were to be inserted between the polyphase analysis and synthesis filters $\mathbf{E}(z)$ and $\mathbf{R}(z)$ quantization noise would affect the performance of the system. This noise is targeted by the noise-shaping system proposed by Bölcskei and Hlawatsch in [14]. In this section, an analysis of the quantization noise in oversampled filter banks is first given, followed by an explanation of the introduced noise-shaping system.

3.2.1 Quantization noise analysis

In order to facilitate the analysis of the quantization noise, it is convenient to gather the inputs $X_j(z)$, $j = 0, 1, \ldots, M-1$, to the analysis polyphase filter bank $\mathbf{E}(z)$ (see Figure 2.10) into the vector $\underline{x}(z) = \begin{bmatrix} X_0(z) & X_1(z) & \cdots & X_{M-1}(z) \end{bmatrix}^T$ and similarly with the outputs of the polyphase synthesis filter bank $\mathbf{R}(z)$, $\underline{\hat{x}}(z) = \begin{bmatrix} \hat{X}_0(z) & \hat{X}_1(z) & \cdots & \hat{X}_{M-1}(z) \end{bmatrix}^T$. Also, the additive quantization noise $\underline{q}(m)$ represents the vector of the quantization noise $q_i(m)$, $i = 0, 1, \ldots, N-1$, in each subband, yielding

$$\underline{q}(z) = \sum_{m=-\infty}^{\infty} \underline{q}(m) z^{-m}.$$

It is further assumed that $\underline{q}(m)$ is a wide-sense stationary, zero-mean process, with power spectral matrix $\mathbf{S}_q(z)$ given by

$$\mathbf{S}_q(z) = \sum_{l=-\infty}^{\infty} \mathbf{C}_q(l) z^{-l},$$

where the autocorrelation matrix $\mathbf{C}_q(l) = \mathrm{E}\{\underline{q}(m)\underline{q}^H(m-l)\}$.

This notation then leads to a convenient representation of the subband coder, shown in Figure 3.2.

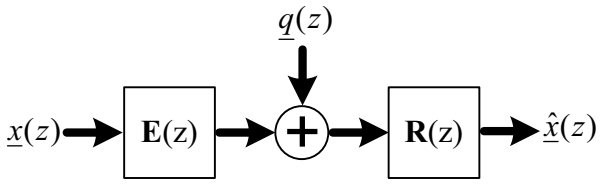


Fig. 3.2 Subband coder with additive quantization noise

It is then straightforward to see that the reconstructed signal $\underline{\hat{x}}(z)$ is simply

$$\underline{\hat{x}}(z) = \mathbf{R}(z)[\mathbf{E}(z)\underline{x}(z) + \underline{q}(z)].$$

Assuming a perfect reconstruction system (i.e. $\mathbf{R}(z)\mathbf{E}(z) = \mathbf{I}_M$), the expression for the

reconstructed signal becomes

$$\underline{\hat{x}}(z) = \underline{x}(z) + \mathbf{R}(z)\underline{q}(z).$$

If the reconstruction error $\underline{e}(z)$ is taken to be the difference between the input and reconstructed signals

$$\begin{aligned}\underline{e}(z) &= \underline{\hat{x}}(z) - \underline{x}(z) \\ &= \mathbf{R}(z)\underline{q}(z),\end{aligned}$$

it can be shown [3] that it is also wide-sense stationary and zero-mean. Also, its power spectral matrix is given by

$$\begin{aligned}\mathbf{S}_e(z) &= \sum_{l=-\infty}^{\infty} \mathrm{E}\{\underline{e}(n)\underline{e}^H(n-l)\}z^{-l} \\ &= \mathbf{R}(z)\mathbf{S}_q(z)\widetilde{\mathbf{R}}(z).\end{aligned}$$

The reconstruction error variance σ_e^2 is then found to be [14]

$$\sigma_e^2 = \frac{1}{2\pi M}\int_{-\pi}^{\pi} \mathrm{Tr}\{\mathbf{R}(e^{j\omega})\mathbf{S}_q(e^{j\omega})\mathbf{R}^H(e^{j\omega})\}d\omega, \tag{3.11}$$

which can be interpreted as the average of the reconstruction error of each subband, a familiar result [3].

Next, if the noise signals $q_i(m)$ are assumed to be uncorrelated, white and with identical variances σ_q^2, the power spectral matrix reduces to [14] $\mathbf{S}_q(z) = \sigma_q^2\mathbf{I}_N$ and Eq. (3.11) is simplified to

$$\sigma_e^2 = \frac{\sigma_q^2}{2\pi M}\int_{-\pi}^{\pi} \mathrm{Tr}\{\mathbf{R}(e^{j\omega})\mathbf{R}^H(e^{j\omega})\}d\omega.$$

Recalling that the filter bank analysis operator $\mathcal{T}$, defined in Chapter 2, mapped the input signal $x(n)$ to a subspace $\mathcal{R}$ of the codomain $[\ell^2(\mathbf{Z})]^N$, it can be shown [14] that (again assuming that the subband noise signals $q_i(m)$ are white and uncorrelated) the reconstruction error can be split into two components: one lying in the range $\mathcal{R}$ of $\mathcal{T}$, say $\underline{e}_{\mathcal{R}}(z)$, and one lying in its orthogonal complement[3], $\mathcal{R}^{\perp}$, say $\underline{e}_{\mathcal{R}^{\perp}}(z)$. Moreover, these two

[3] In general, any subspace has an orthogonal complement, and together they span the entire space [23].

components are uncorrelated due to the orthogonality of $\mathcal{R}$ and $\mathcal{R}^{\perp}$. Consequently, the reconstruction error variance σ_e^2 is the sum of the variances of the two components:

$$\sigma_e^2 = \sigma_{\mathcal{R}}^2 + \sigma_{\mathcal{R}^{\perp}}^2. \tag{3.12}$$

Further, it can be shown [14] that for a paraunitary filter bank with normalized analysis filters, the frame expansion is tight and the frame bounds are $A = B = \frac{1}{K}$, where K is the integer oversampling factor. This in turn leads to [14]

$$\frac{\sigma_e^2}{\sigma_q^2} = \frac{1}{K}. \tag{3.13}$$

From Eq. 3.13 it is clear that in the critically sampled case ($K = 1$), the reconstruction error variance σ_e^2 is simply the quantization noise variance σ_q^2. This is expected since the only source of noise in this system is the quantization process. By renaming the critically sampled reconstruction error variance σ_c^2 and inserting it back into Eq. 3.13, the following equation is obtained:

$$\frac{\sigma_e^2}{\sigma_c^2} = \frac{1}{K},$$

which is consistent with the results in [15] for oversampled analog to digital conversion. This equation indicates that there is a reduction in the overall reconstruction error variance proportional to $1/K$ simply due to oversampling. This can be explained intuitively by the fact that, in general, the range subspace $\mathcal{R}$ becomes "smaller" relative to the codomain[4], as the oversampling factor increases. This, in turn, leads to a reduction of the in-range noise component $\sigma_{\mathcal{R}^{\perp}}^2$ [14]. Therefore, the redundancy injected by the oversampling of the subband signals induces an improvement in the subband coder's performance: a reduction in error variance represents a gain in output signal-to-noise-ratio (SNR)

$$\mathrm{SNR} = 10 \log_{10}\left(\frac{\sigma_s^2}{\sigma_e^2}\right),$$

where σ_s^2 is the signal variance. However, this of course comes at the cost of a rate increase in the subband signals by a factor of K.

[4]The codomain of a function is the set within which the values of a function lie [23].

3.2.2 Goal of the noise-shaping system

As mentioned previously, noise-shaping was first developed in the context of oversampled analog to digital conversion. When the band-limited analog signal is oversampled, the resulting digital signal is then bandlimited to $0 \leq \omega \leq \pi/K$ as opposed to occupying the entire frequency range ($0 \leq \omega \leq \pi$). However, the power spectral density of the quantization noise does occupy the entire range. It then stands to reason that if this noise could be somehow transformed such that it were constrained to the $\pi/K \leq \omega \leq \pi$ interval, it could be completely eliminated by subsequently applying a lowpass filter with cutoff frequency $\omega_c = \pi/K$. Indeed, the practice is to first estimate the error in the interval $0 \leq \omega \leq \pi/K$ and then subtract a quantity containing this prediction from the quantization error [15].

This is analogous to the objective of noise-shaping in oversampled filter banks. Recalling that the range $\mathcal{R}$ of the previously defined analysis filter bank operator $\mathcal{T}$ is a subspace of the codomain $[\ell^2((Z))]^N$, the goal is to use the redundancy in the subband signals $v_i(n)$ to effectively "push" the quantization noise into the orthogonal complement $\mathcal{R}^\perp$ [14]. This concept is illustrated in Figure 3.3: the subband signal $V_i(z)$ is limited to

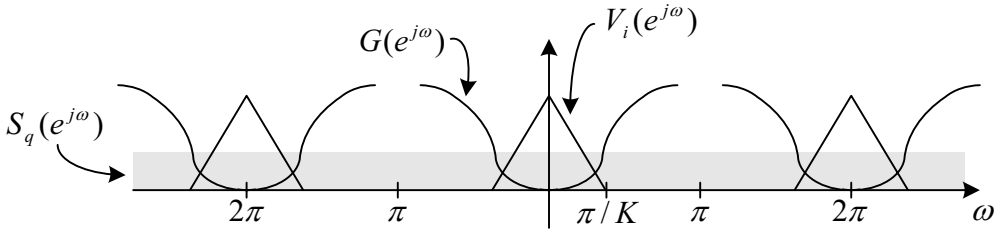


Fig. 3.3 Illustration of a noise-shaping filter

$\omega \leq \pi/K$, while the power spectral density of the noise extends over the full frequency range. Recalling that the reconstruction error can be split into two components (as well as their respective variances, Eq. (3.12)), the purpose of the noise-shaping filter $\mathbf{G}(z)$ is to predict the component of $\underline{q}(z)$ than will cause $\underline{e}_{\mathcal{R}}(z)$. This quantity can then be subtracted from $\underline{q}(z)$ and thus attenuate $\underline{e}_{\mathcal{R}}(z)$ or, ideally, remove it completely. Subsequently, the synthesis filter bank will attenuate $\underline{e}_{\mathcal{R}^\perp}(z)$. Again, if the synthesis filters are chosen as the dual frame, the noise in $\mathcal{R}^\perp$ will be completely removed. Recalling that the para-pseudo-inverse $\hat{\mathbf{R}}(z)$ corresponds to the dual frame, its use as the synthesis filter bank is once more justified: it can be shown [14] that $\hat{\mathbf{R}}(z)$ removes the component of the noise lying in $\mathcal{R}^\perp$.

Thus, it is theoretically possible to *completely* remove the reconstruction error through a judicious choice of both the noise-shaping filters and the synthesis filters.

3.2.3 System design

The system proposed by Bölcskei and Hlawatsch [14] is illustrated in Fig. 3.4. The block

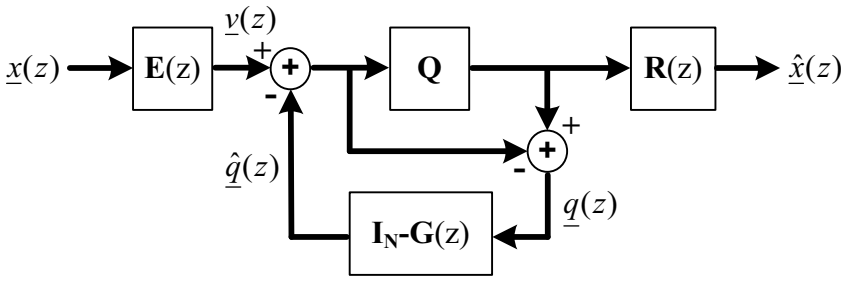


Fig. 3.4 Oversampled filter bank with noise-shaping

labeled **Q** represents the quantizers introduced in each subband and are the source of the quantization error $\underline{q}(z)$, referred to previously. The noise-shaping system $\mathbf{G}(z)$ is inserted between the analysis and synthesis filters such that it may act directly upon the quantization noise. Further, the noise-shaping system is an $N \times N$ multiple-input-multiple-output (MIMO) system. It has been shown that although single-input-single-output (SISO) subband linear prediction systems yield good performances ([24], [25]), better performances are achieved using a MIMO system [26]. Intuitively this makes sense: the subband signals cannot in practice be constrained to a certain frequency band, since the analysis filters themselves have a certain transition bandwidth and imperfect attenuation in the stopband. Hence, there will necessarily be residual information from neighboring subbands in the particular subband under consideration. Consequently, the noise-shaping system — which is also a linear predictor as will be seen shortly — is expected to perform better if a MIMO system is used. This will indeed be shown in Chapter 4.

Returning to the system in Figure 3.4, the quantization error $\underline{q}(z)$, taken as the difference between the input and the output of the quantizers **Q**, is fed through the noise-shaping filters to produce the estimate of the component lying in $\mathcal{R}$, $\underline{\hat{q}}(z)$. This estimate is then subtracted from the subband signals $\underline{v}(z)$ and the result is subsequently quantized. The

reconstructed signal $\underline{\hat{x}}(z)$ is thus given by:

$$\begin{aligned}\underline{\hat{x}}(z) &= \mathbf{R}(z)[\underline{v}(z) - \underline{\hat{q}}(z) + \underline{q}(z)] \\ &= \mathbf{R}(z)[\mathbf{E}(z)\underline{x}(z) - \{\mathbf{I}_N - \mathbf{G}(z)\}\underline{q}(z) + \underline{q}(z)] \\ &= \mathbf{R}(z)\mathbf{E}(z)\underline{x}(z) + \mathbf{R}(z)\mathbf{G}(z)\underline{q}(z).\end{aligned}$$

Then, if perfect reconstruction filters are used ($\mathbf{R}(z)\mathbf{E}(z) = \mathbf{I}_M$) the reconstruction error $\underline{e}(z)$ will be given by

$$\begin{aligned}\underline{e}(z) &= \underline{\hat{x}}(z) - \underline{x}(z) \\ &= \mathbf{R}(z)\mathbf{E}(z)\underline{x}(z) + \mathbf{R}(z)\mathbf{G}(z)\underline{q}(z) - \underline{x}(z) \\ &= \mathbf{R}(z)\mathbf{G}(z)\underline{q}(z),\end{aligned}$$

the power spectral density matrix of the reconstruction error is

$$\begin{aligned}S_e(z) &= \sum_{l=-\infty}^{\infty} \mathrm{E}\{\underline{e}(n)\underline{e}^H(n-l)\}z^{-l} \\ &= \mathbf{R}(z)\mathbf{G}(z)S_q(z)\widetilde{\mathbf{G}}(z)\widetilde{\mathbf{R}}(z),\end{aligned}$$

and the reconstruction error variance is

$$\sigma_e^2 = \frac{1}{2\pi M}\int_{-\pi}^{\pi} \mathrm{Tr}\{\mathbf{R}(e^{j\omega})\mathbf{G}(e^{j\omega})\mathbf{S}_q(e^{j\omega})\mathbf{G}^H(e^{j\omega})\mathbf{R}^H(e^{j\omega})\}d\omega.$$

Again, if the quantization error is assumed to be white and uncorrelated and with equal variance in all subbands ($S_q(z) = \sigma_q^2\mathbf{I}_N$), the variance of the reconstruction error reduces to [14]

$$\sigma_e^2 = \frac{\sigma_q^2}{2\pi M}\int_{-\pi}^{\pi} \mathrm{Tr}\{\mathbf{R}(e^{j\omega})\mathbf{G}(e^{j\omega})\mathbf{G}^H(e^{j\omega})\mathbf{R}^H(e^{j\omega})\}d\omega. \tag{3.14}$$

The objective now is to find the noise-shaping filters $\mathbf{G}(z)$ that minimize the reconstruction error. It was demonstrated in [14] that, using the dual frame for the synthesis filter bank, the ideal noise-shaper does indeed project the noise onto $\mathcal{R}^{\perp}$, thus eliminating the noise completely, since the synthesis filter bank subsequently suppresses this noise. However, the ideal noise-shaper cannot be implemented because it is not causal. Non-causality indicates that the estimate of the current noise sample depends on both past and future

noise samples. This system could therefore not be used in a feedback loop, since the future values of the noise samples are not available to the estimator for the simple reason that they have not yet been computed. Consequently, the MIMO noise-shaper is constrained to be FIR, causal and of the form:

$$\mathbf{I_N} - \mathbf{G}(z) = \sum_{l=1}^{L} \mathbf{G}_l z^{-l},$$

where L is the order of the noise-shaping system.

Hence, the quantization noise estimate $\underline{\hat{q}}(n)$ is given by

$$\underline{\hat{q}}(n) = \sum_{l=1}^{L} \mathbf{G}_l \underline{q}(n-l).$$

It is now obvious how the noise-shaping filter can be interpreted as a linear predictor [22]: the estimate of the current noise sample $\underline{\hat{q}}(n)$ is a linear combination of the L previous noise samples $\underline{q}(n)$.

3.2.4 Derivation of the noise-shaping system coefficients

With all these tools in hand, it is now possible to calculate the coefficients $\mathbf{G}_i$. First, the manner in which the full MIMO system can be obtained will be described, followed by two alternatives that are less computationally demanding.

Complete Interchannel Noise-Shaping System

As shown in [14], the reconstruction error variance σ_e^2 under the assumption $S_q(z) = \sigma_q^2 \mathbf{I}_N$ as given by Eq. (3.14), can be rewritten

$$\sigma_e^2 = \frac{\sigma_q^2}{M} \mathrm{Tr}\{\mathbf{\Gamma}_0 - \sum_{l=1}^{L} [\mathbf{\Gamma}_l \mathbf{G}_l^T + \mathbf{\Gamma}_l^T \mathbf{G}_l] + \sum_{m=1}^{L} \mathbf{G}_m^T \sum_{l=1}^{L} \mathbf{\Gamma}_{m-l} \mathbf{G}_l\}, \tag{3.15}$$

where the $\mathbf{\Gamma}_l$ are defined as:

$$\begin{aligned}\mathbf{\Gamma}_l &= \sum_{m=-\infty}^{\infty} \mathbf{R}_m^T \mathbf{R}_{m+l} \\ &= \sum_{m=-\infty}^{\infty} \mathbf{R}_{m-l}^T \mathbf{R}_m \\ &= \sum_{m=-\infty}^{\infty} (\mathbf{R}_m^T \mathbf{R}_{m-l})^T \\ &= \mathbf{\Gamma}_{-l}^T,\end{aligned} \tag{3.16}$$

and the $\mathbf{R}_m$ are defined by $\mathbf{R}(z) = \sum_{m=-\infty}^{\infty} \mathbf{R}_m z^{-m}$. Moreover, for causal filters of finite length $L_h = JM$ ($J \in \mathbf{N}$), $\mathbf{R}_m = \mathbf{0}$ for $m < 0$ and $m > J$. This results in $\mathbf{\Gamma}_l = \mathbf{0}$ for $|l| > J$, lending itself well to matrix computations.

Next, it is desired to determine the $\mathbf{G}_i$ that minimize Eq. (3.15). Therefore, the partial derivatives $\frac{\partial \sigma_e^2}{\partial \mathbf{G}_i}$ are set to $\mathbf{0}$, for $i = 1, 2, \ldots, L$, and the linear set of equations

$$\begin{bmatrix} \mathbf{\Gamma}_0 & \cdots & \mathbf{\Gamma}_{-(L-1)} \\ \mathbf{\Gamma}_1 & \cdots & \mathbf{\Gamma}_{-(L-2)} \\ \vdots & \ddots & \vdots \\ \mathbf{\Gamma}_{(L-1)} & \cdots & \mathbf{\Gamma}_0 \end{bmatrix} \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \vdots \\ \mathbf{G}_L \end{bmatrix} = \begin{bmatrix} \mathbf{\Gamma}_1 \\ \mathbf{\Gamma}_2 \\ \vdots \\ \mathbf{\Gamma}_L \end{bmatrix} \tag{3.17}$$

is obtained [14]. Recalling the definition of the $\mathbf{\Gamma}_l$, they are essentially the product of time-shifted versions of $\mathbf{R}(z)$ and so could be seen as a type of autocorrelation of the synthesis filters. With this interpretation in mind, Eq. (3.17) is reminiscent of the normal equations [22], [26], which, when solved, yield the linear least squares estimator. Once more, it is observed that the noise-shaping filters are akin to predictors. Although the noise itself is assumed to be white, the noise-shaping filters essentially predict the components of the quantization noise samples that will be passed by the synthesis filter bank $\mathbf{R}(z)$.

Finally, solving Eq. (3.17) yields the complete interchannel noise-shaping coefficients. Rewriting Eq. (3.17) with $\mathbf{G}_{i,\text{opt}}$ as the solution to Eq. (3.17)

$$\sum_{l=1}^{L} \mathbf{\Gamma}_{i-l} \mathbf{G}_{l\text{opt}} = \mathbf{\Gamma}_i \quad \text{for } i = 1, \ldots, L$$

and inserting into Eq. (3.15), the minimum reconstruction error variance $\sigma^2_{e,\min}$ is found to be [14]

$$\sigma^2_{e,\min} = \frac{\sigma_q^2}{M}\mathrm{Tr}\left\{\mathbf{\Gamma}_0 - \sum_{l-=1}^{L} \mathbf{\Gamma}_l^T \mathbf{G}_{l,\mathrm{opt}}\right\} \tag{3.18}$$

As a last note, for a lossless filter bank with normalized analysis filters $\mathrm{Tr}\{\mathbf{\Gamma}_0\} = \frac{M}{K}$. Inserting this result into Eq. (3.18), $\sigma^2_{e,\min} = \frac{\sigma_q^2}{K}$ corroborating the result in Eq. (3.13).

Local interchannel noise-shaping system

In order to reduce the computational complexity of the system, the number of interchannel dependencies considered is reduced. Here, the use of dependencies on only the adjacent channels is explored. This strategy is expected to yield only a slight degradation in performance when compared to the complete interchannel noise-shaping system since the transition bandwidth of the synthesis filters cause the most significant overlap of the frequency responses to be in adjacent subbands. Therefore, the correlation between adjacent synthesis filters is greater than that between the remaining filters and so the local interchannel noise-shaping system exploits most of the useful information available to the MIMO system.

In this case, the $\mathbf{\Gamma}_l$ are modified such that only the main and off-diagonal elements of the $\mathbf{\Gamma}_l$ are selected, while the remaining entries are set to 0. They are then re-inserted into Eq. (3.17), which is solved to yield the local interchannel noise-shaping system. The only nonzero elements of the resulting $N \times N$ noise shaping filter $\mathbf{G}(z)$ are then also located on the main and first off-diagonals. Finally, the minimum reconstruction error variance is [14]:

$$\sigma^2_{e,\min} = \frac{\sigma_q^2}{M}\sum_{i=0}^{N-1}\left\{\gamma_{i,i}^{(0)} - \sum_{l=1}^{L}\left[\gamma_{i,i-1}^{(l)} g_{i,i-1;\mathrm{opt}}^{(l)} + \gamma_{i,i}^{(l)} g_{i,i;\mathrm{opt}}^{(l)} + \gamma_{i,i+1}^{(l)} g_{i,i+1;\mathrm{opt}}^{(l)}\right]\right\}, \tag{3.19}$$

where the $g_{i,i}^{(l)}$ correspond to $[\mathbf{G}_l]_{i,j}$ and the $\gamma_{i,i}^{(l)} = [\mathbf{\Gamma}_l]_{i,j}$.

Intrachannel Noise-Shaping System

Because the solution to Eqs. (3.17) involves the inversion of a matrix whose dimensions grow linearly with the order of the noise-shaper and the length L_h of the synthesis filters,

performance is once more traded for reduced complexity and (as will be seen in Chapter 4) better conditioning of the matrices. Here, the system under consideration is SISO. Although it cannot perform as well as the MIMO system [26], fair results can still be expected [24], [25], especially for a small system order.

The noise-shaping system is therefore constrained to take advantage of *intra*channel dependencies only, meaning that there are separate noise-shaping systems in each channel, and thus a diagonal $\mathbf{G}(z)$ [14]. Consequently, Eq. (3.15) is reduced to:

$$\sigma_e^2 = \frac{\sigma_q^2}{M}\mathrm{Tr}\Big\{\gamma_{i,i}^{(0)} - \sum_{l=1}^{L}[\gamma_{i,i}^{(l)} g_{i,i}^{(l)} + \gamma_{i,i}^{(l)} g_{i,i}^{(l)}] + \sum_{m=1}^{L} g_{i,i}^{(m)} \sum_{l=1}^{L} \gamma_{i,i}^{(m-l)} g_{i,i}^{(l)}\Big\}. \tag{3.20}$$

Again, by setting the derivatives $\frac{d\sigma_e^2}{d\gamma_{i,i}} = \mathbf{0}$, for $i = 0, 1, \ldots, N-1$ and $l = 1, \ldots, L$, the N $L \times L$ linear sets of equations:

$$\textstyle\sum_{m=1}^{L} \gamma_{i,i}^{(l-m)} g_{i,i}^{(m)} = \gamma_{(i,i)}^{(l)} \quad \text{for } l = 1, \ldots, L \text{ and } i = 0, \ldots, N-1, \tag{3.21}$$

are obtained. Further, the obtaining of the noise-shaping coefficients is simplified by arranging Eq. (3.21) as

$$\mathbf{A}_i \underline{g}_i = \underline{\gamma}_i \quad \text{for } i = 0, 1, \ldots, N-1, \tag{3.22}$$

where $[\mathbf{A}_i]_{l,m} = \gamma_{i,i}^{(l-m)}$, $[\underline{g}_i]_l = g_{i,i}^{(l)}$ and $[\underline{\gamma}_i]_l = \gamma_{i,i}^{(l)}$. Finally, using $\gamma_{i,i}^{(m-l)} = \gamma_{i,i}^{(m-l)}$ from Eq. (3.16) and inserting the solution to Eq. (3.22) into Eq. (3.20), the minimum reconstruction error variance in this case is given by [14]

$$\sigma_{e,\min}^2 = \frac{\sigma_q^2}{M} \sum_{i=0}^{N-1} \left[\gamma_{i,i}^{(0)} - \sum_{l=1}^{L} \gamma_{i,i}^{(l)} g_{i,i;\mathrm{opt}}^{(l)}\right]. \tag{3.23}$$

3.3 Summary

In this chapter, the use of cosine-modulated filter banks using a lossless lattice structure was first justified: all the filters are FIR — allowing for an overall linear phase — and the resulting synthesis filter bank corresponds to the frame dual to the analysis filter bank, which eliminates the noise in $\mathcal{R}^{\perp}$. Further, perfect reconstruction cosine-modulated filter banks permit a comparison between critically sampled and oversampled filter banks. Then, the two-channel lattice structure was discussed, leading into the optimization of the lattice

parameters. An analysis of the noise in a subband coder introduced by the quantization process was then given, followed by the goal of the noise-shaping system: predicting the noise component in $\mathcal{R}$ in order to be able to remove the corresponding estimate from the subband signals. Only the noise component in $\mathcal{R}$ is of concern, seeing as the synthesis filters employed subsequently remove the component lying in $\mathcal{R}^{\perp}$. Finally, a manner in which to obtain the optimal noise-shaping coefficients was described, including more computationally efficient noise-shapers whose reduction in complexity came at the cost of an expected decrease in performance.

Chapter 4

Investigation

Although, the optimal noise-shaping system was derived in [14], there was only a limited amount of experimental results. It was thus desired to study the effect of using different filters on the performance of the noise-shaping system. In this chapter, the setup of the investigation is first presented: in order to form an adequate test bed, it was required to generate filter banks with a different number of subbands and different filter lengths. Subsequently, the noise-shaping system in Figure 3.4 was implemented, using the designed filter banks. It was then possible to firstly discuss the theoretical and simulation results, and secondly to evaluate the effects of different characteristics of the filter banks, such as filter length, degree of overlap between subbands and perfect-versus near-perfect reconstruction.

4.1 Iterative filter design

This section first describes the manner in which the generated filter banks were obtained, followed by the presentation of the resulting filter banks. It is then verified that they satisfy perfect reconstruction by introducing them into a subband coder.

4.1.1 Design algorithm

An algorithm based on [21] using the tools described in sections 3.1.2 and 3.1.3 was implemented using MATLAB, in order to obtain filter banks pertinent to the discussion. An outline of this algorithm is presented here.

Step 1

First, the vector Θ is initialized by setting the $m\lfloor\frac{N}{2}\rfloor$ elements to

$$\theta_{k,p} = \begin{cases} \frac{\pi}{4}, & p = 0, \quad 0 \leq k \leq \lfloor\frac{N}{2}\rfloor - 1 \\ \frac{\pi}{2}, & 1 \leq p \leq m-1, \quad 0 \leq k \leq \lfloor\frac{N}{2}\rfloor - 1. \end{cases} \tag{4.1}$$

By substituting these values into Eqs. (3.5) and (3.6) and subsequently using the determined polyphase components to generate $p_0(n)$, it ca be deduced that [20]

$$p_0(n) = \begin{cases} 1, & (mN - N) \leq k \leq (mN + N - 1) \\ 0, & \text{else.} \end{cases} \tag{4.2}$$

It is stated in [20] that this prototype generally has a stopband attenuation of approximately 13 dB and a cutoff frequency $\omega_c < \frac{\pi}{M}$ rad. Indeed, Figure 4.1 shows an example of the prototype filter[1] generated using the initial conditions given by Eq. (4.1) and $N = 16$ subbands: it is observed that the stopband attenuation is approximately 14 dB, while the 3 dB cutoff frequency $\omega_c \approx \frac{\pi}{32}$, well below the desired cutoff frequency of $\frac{\pi}{16}$. The optimization procedure is thus begun with a valid filter.

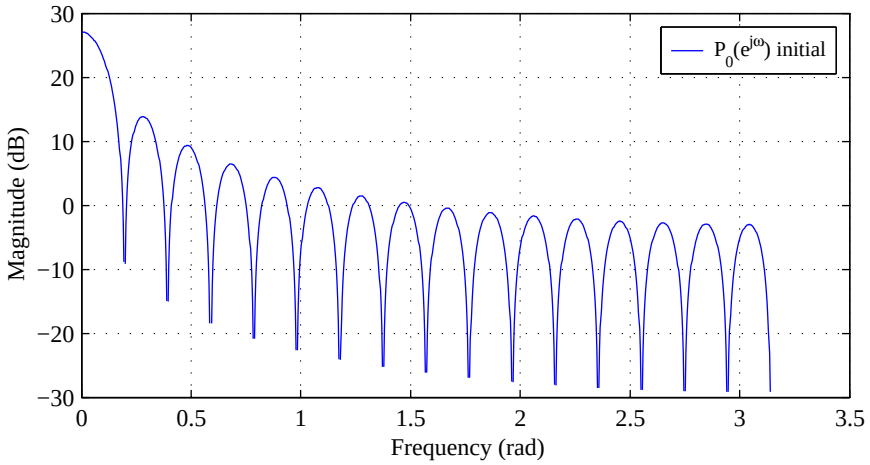


Fig. 4.1 Magnitude of the frequency response of the prototype filter ($|P_0(e^{j\omega})|$) using the $\theta_{k,p}$ given by the initial conditions in Eq. (4.1), with $N = 16$ and $m = 2$.

[1]Here and in all subsequent figures, the phase of the frequency response is omitted as it is always linear and provides no further insight.

It is further noted that it is appropriate to compute the matrix $\mathbf{U}_s$ (see Eqs. (3.8) and (3.9)) at this point, as it is does not depend on Θ.

Step 2

The elements of the vector $\mathbf{D}(\Theta)$ and the matrix $\mathbf{H}(\Theta)$ are now computed for the $(n-1)^{th}$ iteration of Θ.

Recalling Eq. (3.8), the partial derivatives of the first mN coefficients prototype filter with respect to the $[\Theta]_i$ must be obtained. Using Eq. (3.1),

$$\sum_{n=0}^{2mN-1} \frac{\partial p_0(n)}{\partial [\Theta]_i} z^{-n} = \sum_{q=0}^{2N-1} z^{-q} \frac{\partial W_q(z^{2N})}{\partial [\Theta]_i} \tag{4.3}$$

such that $\mathbf{D}(\Theta)$ can be directly obtained from the partial derivatives of the polyphase components. Further, it was noted that these partial derivatives could be obtained efficiently through the cascade of two-channel lattice structures (see Appendix A):

$$\begin{bmatrix} \frac{\partial W_k^{(i)}(z)}{\partial \theta_{k,p}} \\ \frac{\partial W_{N+k}^{(i)}(z)}{\partial \Theta_{k,p}} \end{bmatrix} = \begin{bmatrix} \cos\theta_{k,i} & \sin\theta_{k,i} \\ \sin\theta_{k,i} & -\cos\theta_{k,i} \end{bmatrix} \begin{bmatrix} \frac{\partial W_k^{(i-1)}(z)}{\partial \theta_{k,p}} \\ z^{-1}\frac{\partial W_{N+k}^{(i-1)}(z)}{\partial \theta_{k,p}} \end{bmatrix}, \quad i > p$$

and the lattices are initialized as:

$$\begin{bmatrix} \frac{\partial W_k^{(i)}(z)}{\partial \theta_{k,p}} \\ \frac{\partial W_{N+k}^{(i)}(z)}{\partial \Theta_{k,p}} \end{bmatrix} = \begin{bmatrix} -\sin\theta_{k,i} & \cos\theta_{k,i} \\ \cos\theta_{k,i} & \sin\theta_{k,i} \end{bmatrix} \begin{bmatrix} W_k^{(i-1)}(z) \\ z^{-1}W_{N+k}^{(i-1)}(z) \end{bmatrix}, \quad i = p \neq 0$$

and

$$\begin{bmatrix} \frac{\partial W_k^{(0)}(z)}{\partial \theta_{k,0}} \\ \frac{\partial W_{N+k}^{(0)}(z)}{\partial \theta_{k,0}} \end{bmatrix} = \begin{bmatrix} \sin\theta_{k,0} \\ -\cos\theta_{k,0} \end{bmatrix}, \quad i = p = 0 \ , \tag{4.4}$$

for $0 \leq k \leq \left\lfloor \frac{N}{2} \right\rfloor - 1$, while the remaining partial derivatives are all 0 (Appendix A).

The elements of the vector $\mathbf{D}(\Theta)$ are then obtained by rearranging the polyphase components via Eq. (4.3) to yield the $\partial p_0(n)/\partial\theta_{k,p}$ for this iteration, while $\partial \underline{p}_0'(n)/\partial[\Theta]_i$ is taken as the first $\partial p_0(0)/\partial\theta_{k,p} \ldots \partial p_0(mN-1)/\partial\theta_{k,p}$ and substituting the appropriate values into Eq. (3.8).

Next, recalling Eq. (3.9), the second order partial derivatives of the first mN coefficients

prototype filter with respect to the $[\Theta]_i$, $[\Theta]_j$ must be obtained. From Eq. (3.1),

$$\sum_{n=0}^{2mN-1} \frac{\partial^2 p_0(n)}{\partial[\Theta]_i \partial[\Theta]_j} z^{-n} = \sum_{q=0}^{2N-1} z^{-q} \frac{\partial^2 W_q(z^{2N})}{\partial[\Theta]_i \partial[\Theta]_j} \tag{4.5}$$

it is seen that $\mathbf{H}(\Theta)$ can be directly obtained from the second order partial derivatives of the polyphase components. Again, the elements of $\mathbf{H}(\Theta)$ can be obtained through the cascade of two-channel lattice structures (see Appendix A). This time, there are 4 cases for which the second order partial derivatives are non-zero:

1. $i > p$ and $i > q$

$$\begin{bmatrix} \frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}} W_k^{(i)}(z) \\ \frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}} W_{N+k}^{(i)}(z) \end{bmatrix} = \begin{bmatrix} \cos\theta_{k,i} & \sin\theta_{k,i} \\ \sin\theta_{k,i} & -\cos\theta_{k,i} \end{bmatrix} \begin{bmatrix} \frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}} W_k^{(i-1)}(z) \\ z^{-1}\frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}} W_{N+k}^{(i-1)}(z) \end{bmatrix};$$

2. $i = p < q$

$$\begin{bmatrix} \frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}} W_k^{(p)}(z) \\ \frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}} W_{N+k}^{(p)}(z) \end{bmatrix} = \begin{bmatrix} -\sin\theta_{k,p} & \cos\theta_{k,p} \\ \cos\theta_{k,p} & \sin\theta_{k,p} \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial\theta_{k,q}} W_k^{(p-1)}(z) \\ z^{-1}\frac{\partial}{\partial\theta_{k,q}} W_{N+k}^{(p-1)}(z) \end{bmatrix};$$

3. $i = q < p$

$$\begin{bmatrix} \frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}} W_k^{(q)}(z) \\ \frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}} W_{N+k}^{(q)}(z) \end{bmatrix} = \begin{bmatrix} -\sin\theta_{k,q} & \cos\theta_{k,q} \\ \cos\theta_{k,q} & \sin\theta_{k,q} \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial\theta_{k,p}} W_k^{(q-1)}(z) \\ z^{-1}\frac{\partial}{\partial\theta_{k,p}} W_{N+k}^{(q-1)}(z) \end{bmatrix};$$

4. $i = p = q$

$$\begin{bmatrix} \frac{\partial^2}{\partial^2\theta_{k,p}} W_k^{(p)}(z) \\ \frac{\partial^2}{\partial^2\theta_{k,p}} W_{N+k}^{(p)}(z) \end{bmatrix} = \begin{bmatrix} -\cos\theta_{k,p} & -\sin\theta_{k,p} \\ -\sin\theta_{k,p} & \cos\theta_{k,p} \end{bmatrix} \begin{bmatrix} W_k^{(p-1)}(z) \\ z^{-1} W_{N+k}^{(p-1)}(z) \end{bmatrix}.$$

Finally, the second order partial derivatives are initialized as

$$\begin{bmatrix} \frac{\partial^2}{\partial^2\theta_{k,0}} W_k^{(0)}(z) \\ \frac{\partial^2}{\partial^2\theta_{k,0}} W_{N+k}^{(0)}(z) \end{bmatrix} = \begin{bmatrix} -\cos\theta_{k,0} \\ -\sin\theta_{k,0} \end{bmatrix}.$$

Step 3

Recalling Eq. (3.10), the search step size λ is then set and the vector Θ^n can now be computed. With this new set of parameters, Step 2 and 3 are repeated, until an appropriate number of iterations have been made. The appropriate number of iterations is determined by the rate of convergence of the algorithm: by monitoring the stopband energy ρ, it is possible to determine the number of iterations needed for the stopband energy to stabilize at a minimum. It was found to be in the vicinity of 10 iterations for the employed filter banks.

Increasing the filter lengths $L_h = 2mN$, $m > 2$

Recalling the initialization of the vector Θ (Eq. (4.1)), it is clear that the initial prototype filter is independent of the length factor m. Indeed, this corresponds to only $2N$ coefficients being non-zero (Eq. (4.2)). Consequently, this approach works well for small values of m [20], but breaks down at higher values. This phenomenon was observed in the implemented algorithm and was resolved using a method suggested in [20]. An m_{init} for which the algorithm works well is selected and the first $m_{init}\left\lfloor \frac{N}{2} \right\rfloor$ values of Θ are obtained. These are then used as the initial conditions, while setting the remaining $\theta_{k,m_{init}+1} = \pi/2$. The algorithm is then run again, with the new initial conditions. This can then be repeated for bigger m.

4.1.2 Obtained filter banks

Because of the filters available for comparison purposes, perfect reconstruction filters of lengths $N = 8, 16, 32$ were generated using the described algorithm.

N=16

An analysis filter with $N = 16$ subbands was first generated, such that the results presented in [14] could be verified.

Figure 4.2 shows the magnitude of the frequency response of the generated 16-subband prototype filter, with $L_h = 2mN = 64$, along with the initial prototype filter from Figure 4.1. The corresponding filter coefficients $p_0(n)$ are given in appendix B.2.

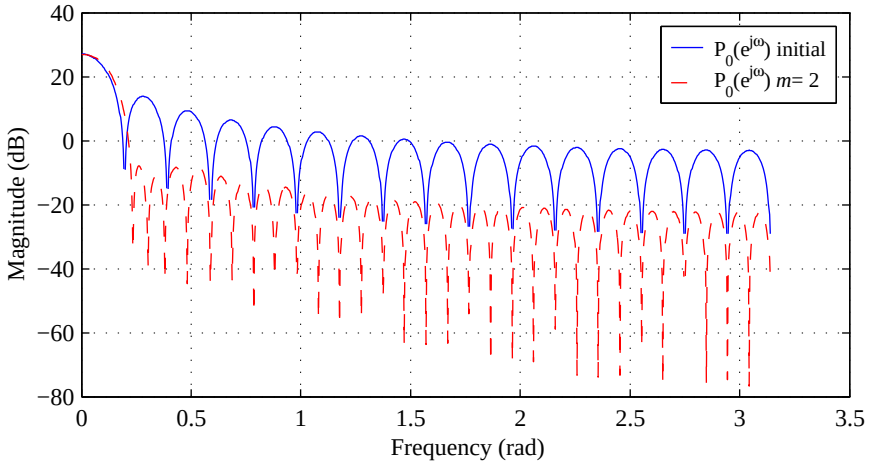


Fig. 4.2 Comparison of the magnitude of the frequency responses of the prototype filter using the initial conditions for Θ and after 20 iterations ($\lambda = 0.5$).

It is clear that the iteration was successful: it decreased the stopband attenuation by approximately 25 dB.

N=8

Next, as will be seen shortly, it was desired to evaluate the effect of variations in filter length and degree of subband filter overlap on the noise-shaping system. Thus, Lapped Orthogonal Transform (LOT) filters [27] were used for their wider transition bandwidth, with $N = 8$ and $L_h = 16$. Due to their relatively short lengths, this type of filter bank can be alternatively interpreted as block transforms, whose basis functions overlap adjacent blocks by 50%. The LOT was developed with the aim of reducing the discontinuities in the reconstructed signal at block boundaries. Consequently, their design involves the constraint of the basis functions being both orthogonal within the same block as well as with the basis functions of the two neighboring blocks, meaning that the design of all the filters of the filter bank must be done simultaneously, as opposed to the cosine-modulated case, where only the prototype filter is designed, as discussed in section 3.1.1.

Thus, for an appropriate comparison, it was then required to generate filters with 8 subbands, since the employed LOT filter bank had also been designed for 8 subbands. The frequency responses of the designed filters are shown in Figure 4.3, and their corresponding coefficients are given in appendix B.1.

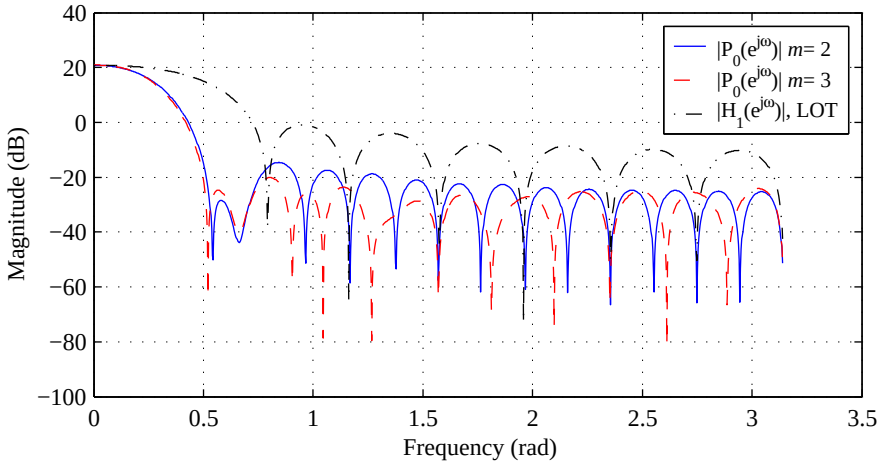


Fig. 4.3 Magnitudes of the frequency responses of the two designed prototype filters ($N = 8$, $m = 2, 3$) and of the first filter of the LOT filter bank, $h_1(n)$.

It is seen that increasing the length factor m does increase the stopband attenuation: by taking m from 2 to 3 (effectively increasing the length L_h from 32 to 48) there is approximately a 5 dB stronger attenuation. The frequency response of the first filter in the filter bank from [27] is also shown, and it is clear that the transition bandwidth is much wider, inducing a larger overlap between neighboring subbands. Again, the effect of the degree of overlap on the system will be discussed in section 4.2.

N=32

Next it was desired to compare the performance of perfect versus near-perfect reconstruction filter banks. It might seem counterintuitive to produce *near*-perfect reconstruction when perfect reconstruction has been shown to be achievable. However, as was seen previously, there is a highly non-linear relation between the prototype filter $p_0(n)$ and the lattice coefficients. Thus, the optimization procedure is very sensitive to changes in the lattice coefficients [28]. Consequently, it is difficult to design perfect reconstruction filters with a high stopband attenuation. By relaxing the perfect reconstruction condition, it is actually possible to design near-perfect reconstruction filters with a high stopband attenuation [28].

Indeed, Figure 4.4 shows the frequency responses of the perfect reconstruction prototype filter $p_0(n)$ ($L_h = 128$, $m = 2$) and that of the first filter of a near-perfect reconstruction filter bank ($N = 32$, $L_h = 256$), from [28]. The stopband attenuation achieved by the near-

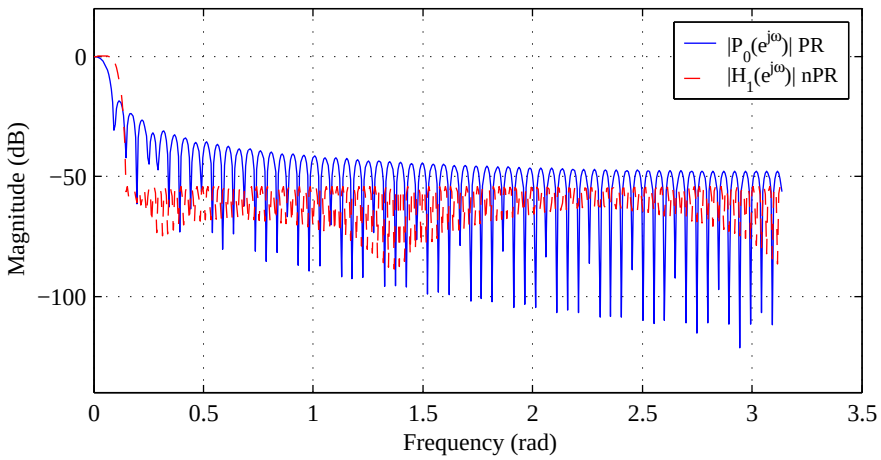


Fig. 4.4 Magnitude of the frequency responses of the designed prototype filter ($N = 32$ and $m = 2$) and the first filter of the near-perfect reconstruction filter bank, $h_1(n)$

perfect reconstruction filter is almost 55 dB, while that the perfect-reconstruction filter is barely 15 dB. Further, it is shown in [28] that although perfect reconstruction is not achieved, the most significant aliasing terms are canceled and so *near*-perfect reconstruction is accomplished, with the reconstruction error being of the same order as the stopband attenuation of the filters.

4.1.3 Filter bank implementation

Next, the obtained prototype filters were cosine modulated to obtain the member filters of the filter banks. It can be verified [20] that through this procedure the analysis and synthesis filters are then related as in Eq. (2.12). Subsequently, the filters were normalized (such that $|h_k(n)| = 1$, for $k = 0, \ldots, N-1$), since it was shown in [14] that paraunitary[2] filter banks with normalized analysis filters corresponded to a tight frame expansion. Next, the analysis and synthesis polyphase matrices were obtained by using Eqs. (2.8) and (2.9) and it was verified that $\mathbf{R}(z) = \frac{1}{K}\widetilde{\mathbf{E}}(z)$ for paraunitary filter banks with normalized analysis filters, as shown in [14]. Finally, the subband coder in Figure 2.10 was implemented such that the reconstruction of the input signal could be confirmed. With all obtained filter

[2]Lossless matrices are paraunitary, and thus the resulting filters will also be paraunitary

banks, the signal-to-noise-ratio (SNR) was obtained as

$$\mathrm{SNR} = 10 \log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right), \tag{4.6}$$

where σ_x^2 is the variance of the input signal, while σ_e^2 is the variance of the reconstruction error (taken as the difference between the input and the output). For all the perfect reconstruction filter banks, the SNR was found to be in the vicinity of 280 dB (a good approximation to perfect reconstruction within the confines of finite-precision arithmetic). On the other hand, the reconstruction error for the near-perfect reconstruction filter banks is only 53.5 dB, which confirms that the reconstruction error is of the same order as the stopband attenuation of the analysis and synthesis filters. Moreover, it was observed that, without noise-shaping (i.e. $L = 0$) Eq. 3.13 held and is verified by all of the figures in the following section that show the normalized reconstruction error variance. Indeed, an oversampling factor of $K = 2$ yields a normalized reconstruction error variance of $\frac{\sigma_{e,\min}^2}{\sigma_q^2} \approx -3$ dB, $K = 4$ yields $\frac{\sigma_{e,\min}^2}{\sigma_q^2} \approx -6$ dB, etc.

4.2 Noise-shaping system

The noise-shaping system proposed by Bölcskei in [14] and reproduced in Figure 3.4 was then implemented. The quantizers used had equal stepsizes in all subbands and an infinite dynamic range. Also, unless otherwise specified, the input $x(n)$ was taken to be a randomly generated auto-regressive (AR(1)) process, with a correlation factor $\rho_g = 0.9$. In this section, the theoretical performance given by the equations for the minimum reconstruction error variance is first studied and compared with simulation results, followed by a comparative study of the various filters.

4.2.1 Discussion of theoretical and simulation results

Using Eqs. (3.17) and (3.22), the optimal coefficients for the noise-shaping system were found, for the intrachannel, complete and adjacent interchannel noise-shaping systems. These were then substituted into Eqs. (3.18), (3.19) and (3.23) to find the theoretical reduction in reconstruction error variance due to the noise-shaping system, with respect to the quantization noise variance. The coefficients were then inserted in the feedback loop

and the simulated reconstruction error variance was obtained. It is noted at this point that only a selected set of results are shown in this section that exemplify the observed behavior of the various systems studied.

Ill-conditioning of the matrices

It was found that the simulation results coincided with the predicted values, up to a certain order of prediction. This is exemplified in Figure 4.5, where the normalized error variance — taken as $10\log_{10}(\frac{\sigma^2_{e,\min}}{\sigma^2_q})$ — is shown as a function of the noise-shaping system order L.

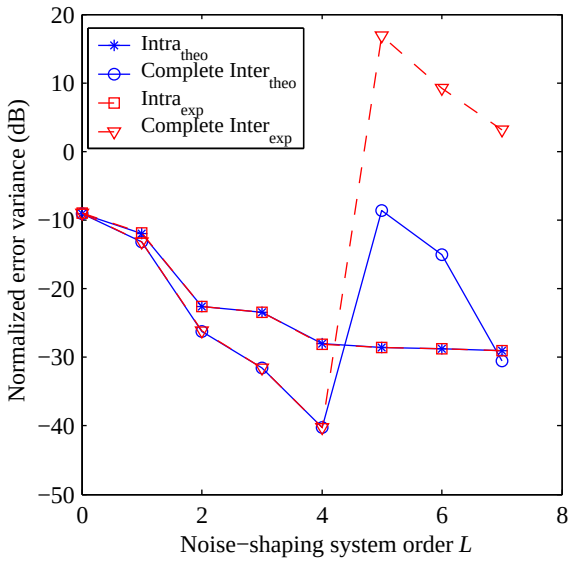


Fig. 4.5 Comparison of theoretical and simulation results of the complete interchannel and the intrachannel noise-shaping systems, demonstrating the deviation in performance of the complete interchannel noise shaping system from the projected result ($N = 16$, $L_h = 64$ and $K = 8$).

It is clear that although the intrachannel noise-shaping system performs consistently at higher system orders, the complete interchannel system does not. In addition, the theoretical prediction of the reduction of the reconstruction error deviates from expectation: while an increase in system order should yield a better estimate of the reconstruction error lying in $\mathcal{R}^{\perp}$ (and thus a better reduction in reconstruction error), it seems that the performance actually deteriorates beyond a system order of $L = 4$. Upon further investigation, it was found that this deterioration in performance coincided with an increase in the condition number of the matrices in Eqs. (3.17) and (3.22) (see Figure 4.6).

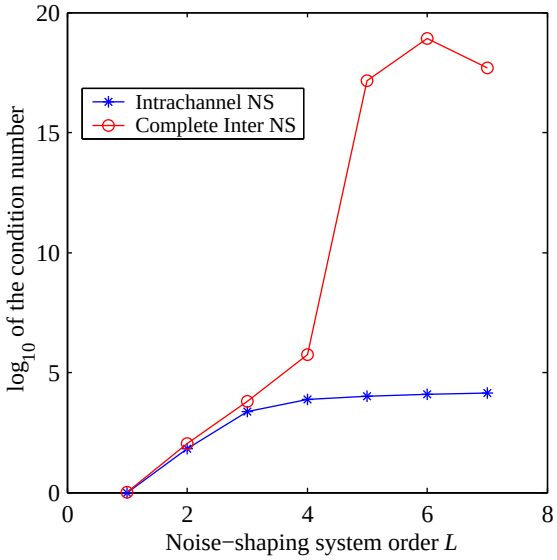


Fig. 4.6 Logarithm of the condition number of the matrices used in solving the linear equations for the complete interchannel and intrachannel noise-shaping systems, corresponding to those of Figure 4.5.

For an invertible real or complex square matrix $\mathbf{A}$, the condition number $\kappa(\mathbf{A})$ is defined as [29]:

$$\kappa(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\|,$$

where $\|\cdot\|$ is the operator norm. The operator norm of a linear operator $\mathcal{A} : \mathcal{X} \rightarrow \mathcal{U}$ is defined as $\|\mathcal{A}\| = \sup_{\|\underline{x}\|=1} \|\mathcal{A}(\underline{x})\|$ [18]. In the case that the linear operator $\mathcal{A}$ is given by a matrix $\mathbf{A}$ — as is the case here — the $\|\mathbf{A}\|$ is given by the square root of the largest eigenvalue of the matrix (equivalently the largest singular value) [23]. Thus, the condition number of an matrix is given by the ratio of the largest singular value of a matrix to the smallest singular value of the matrix.

As shown in [29], condition numbers estimate the relative error in the solution of a linear set of equations $\mathbf{A}\underline{x} = \underline{u}$, due to the relative error both in $\underline{x}$ and $\mathbf{A}$. Further, it was stated that taking the base 10 logarithm of the condition number yields the loss of precision in numbers of digits. Thus, when a matrix has a very large condition number, it is termed *ill-conditioned* since there is a considerable loss of accuracy in the solution of the set of linear equations. Returning to Figure 4.6, it is clear that the matrices used in the solution for the complete interchannel noise-shaping system are ill-conditioned for a system order

$L \geq 4$, and thus the obtained results beyond this point are increasingly inaccurate.

Finally, it is noted that in subsequent figures, where the matrices are ill-conditioned the corresponding performances of the noise-shapers were omitted.

Comparison of the intrachannel, complete and local interchannel noise-shaping systems

Although the minimum reconstruction error variance was derived in [14] for the three different noise-shaping systems, no comparison between the obtained system was given. Further, as it was observed that the matrices for the intrachannel noise-shaping system were better conditioned than those of the complete and local interchannel systems, it was desired to compare the performances of these three cases, to establish whether or not the intrachannel system could produce better results by increasing the system order, when compared to the point at which the remaining two failed, due to ill-conditioning.

Figure 4.7 compares the performance of all three noise-shaping systems for a filter bank with $N = 16$ subbands. As expected, the complete interchannel noise-shaping system consistently outperforms the other two systems, while the local interchannel system does show an impressive performance.

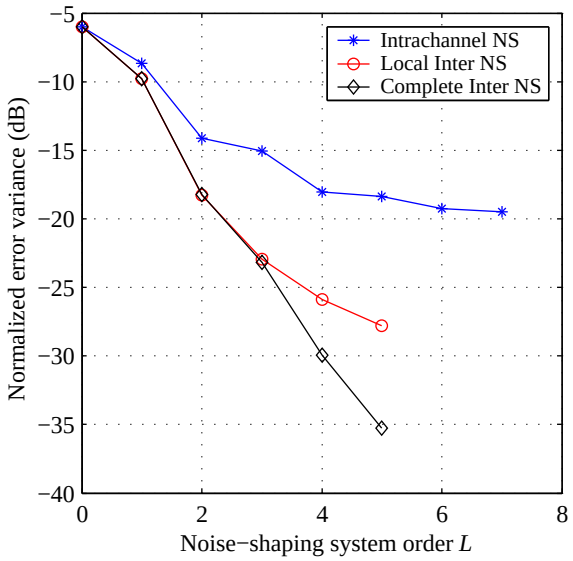


Fig. 4.7 Comparison of the intrachannel, complete and local interchannel noise-shaping systems ($N = 16$, $L_h = 64$ and $K = 4$).

The superiority of the complete interchannel system is expected: as mentioned previously, the MIMO system takes advantage of the most possible information from all the other subbands. On the other hand, for low system orders ($L \leq 3$), the performance of local interchannel noise-shaping system is almost identical to that of the complete interchannel system. As discussed in chapter 3, the good performance of the local system is due to the fact that the most significant overlap of the analysis filters is in adjacent bands, such that this MIMO system exploits most of the information useful to the noise-shaping process. Indeed, Figure 4.8 shows the logarithm of the magnitude of the entries of the $NL \times NL$ matrix in Eq. (3.17), for the complete interchannel noise-shaping system.

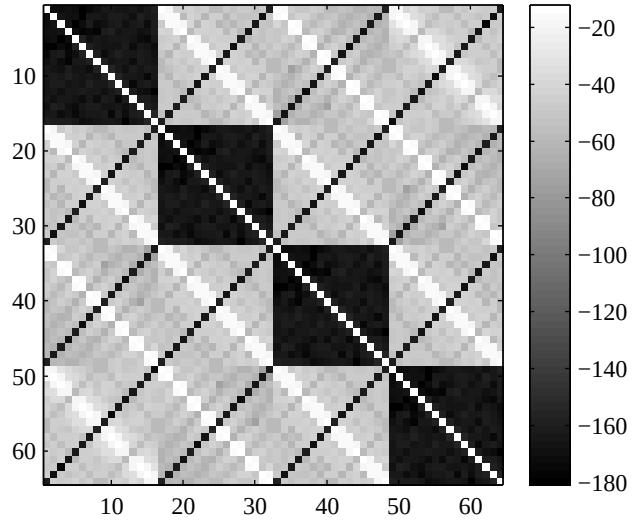


Fig. 4.8 Graphical representation of the matrix of Γ_i's of Eq. (3.17) for a system order $L = 4$, $N = 16$ subbands and $K = 4$, where brightness is proportional to the logarithm of the magnitude of the entries of the Γ_i

Recalling the expression for the elements of the synthesis polyphase component matrix in Eq. (2.9) and that for the $N \times N$ Γ_i in Eq. (3.16), it is clear that the $\gamma_{j,k}^{(l)}$ represent the correlation between the synthesis filters $F_i(z)$ and $F_j(z)$, for different time-shifts l. Thus, the diagonal elements of the Γ_l are the autocorrelations of the synthesis filters, while the first off-diagonal elements are the cross-correlations of the synthesis filters in adjacent subbands[3]. Examination of Figure 4.8 reveals that the cross-correlations between adjacent subbands is much greater than that of the remaining channels, confirming the postulation

[3]This justifies the derivation of the local interchannel noise-shaping system by setting all but the main and first off-diagonal elements of the Γ_i to zero.

that most of the information useful to the MIMO noise-shaping system is inherent to the adjacent subbands.

Further, the local interchannel system comes at a reduced complexity: it only has $3N-2$ noise-shaping filters, while the complete interchannel system has N^2 filters.

Unfortunately, the intrachannel noise-shaping system performs rather poorly, when compared to the other two systems. Thus, although it is generated using better conditioned matrices, it is still advantageous to use both the complete or local interchannel systems.

4.2.2 Performance evaluation

Attention is now turned to the comparison of the the performance of the noise-shaping system with the use of filters with different characteristics.

Length of filters and degree of overlap between subbands

As mentioned previously, the LOT filter bank was chosen as a basis for comparison on the topics of length of filters and degree of overlap between subbands, due to their short length and their significant overlap between subbands (recall Figure 4.3).

In order to evaluate the significance of the length of the filters, the performance of the three filters — the LOT and the designed cosine-modulated perfect reconstruction filter banks, denoted by $\mathrm{CM_{PR}}$ — was compared using the intra-channel noise-shaping filters. Their performance is shown in Figure 4.9.

It is clearly observed that increasing the length of the filters improves the performance of the noise-shaping filters. This is attributed to the fact that longer filters induce longer-term dependencies, which in turn lead to a greater gain.

The results become more interesting when the performance of the complete interchannel noise-shaping system is studied. Referring to Figure 4.10, it is observed that the LOT filter banks slightly outperform the $\mathrm{CM_{PR}}$. This is due to the greater amount of overlap between adjacent subbands of the LOT filter bank: the effectiveness of using a MIMO system is fully appreciated here.

Moreover, it is noted that increasing the length of the $\mathrm{CM_{PR}}$ filters still leads to an improvement in the performance. In fact, as the length of the $\mathrm{CM_{PR}}$ increases, it is observed that its performance approaches that of the LOT filter bank. In addition, the $\mathrm{CM_{PR}}$ benefit

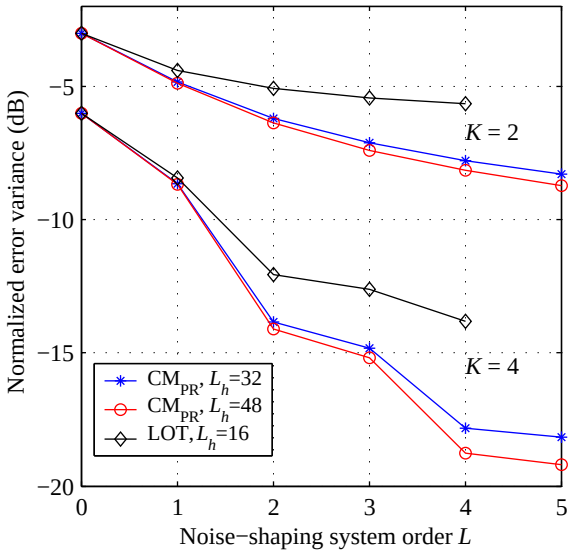


Fig. 4.9 Performance of the intrachannel noise-shaping systems using the LOT filter bank ($N = 8$, $L_h = 16$) and the designed $\mathrm{CM_{PR}}$ ($N = 8$, $L_h = 32, 48$).

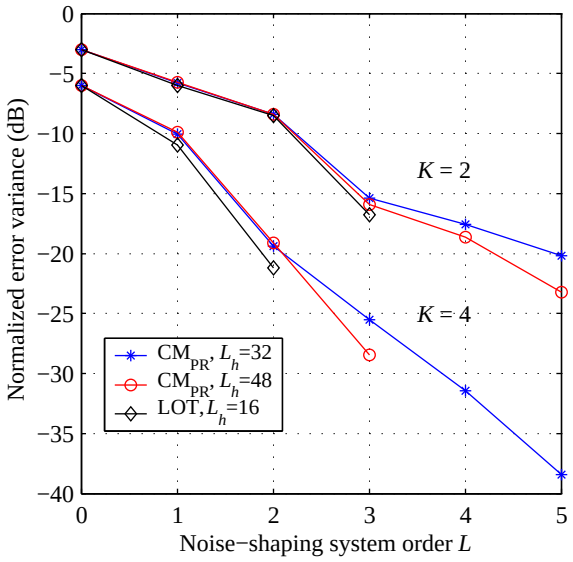


Fig. 4.10 Performance of the complete interchannel noise-shaping systems using the LOT filter bank ($N = 8$, $L_h = 16$) and the designed $\mathrm{CM_{PR}}$ ($N = 8$, $L_h = 32, 48$).

from better conditioning of the matrices. Indeed, in the case of $K = 4$ with $L_h = 48$ a gain of approximately 8 dB is achieved by using a noise-shaping system of order $L = 3$, as opposed to using a system order of $L = 2$ for the LOT. Also, a further gain of approximately 10 dB is possible by using the CM_{PR} with a shorter length ($L_h = 32$), but a greater system order $L = 5$.

Perfect reconstruction vs. near-perfect reconstruction filter banks

The focus is now shifted to the performance of the noise-shaping system for near-perfect reconstruction filter banks, denoted by CM. The theoretical performance was first generated using Eq. (3.18) and is shown in Figure 4.11. It is seen that the near-perfect reconstruc-

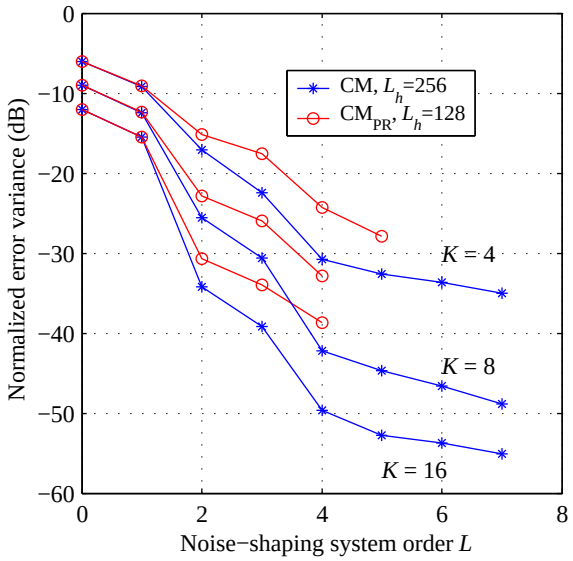


Fig. 4.11 Performance of the complete interchannel noise-shaping system for the designed CM_{PR} filter bank of length $L_h = 128$ and the near-perfect reconstruction filter bank CM of length $L_h = 256$ ($N = 32$, $K = 4, 8, 16$).

tion CM filter bank outperforms the perfect reconstruction filter bank and is once more explained by the fact that the CM filters have a greater length. However, when the theoretical performance was compared to the simulation results a different behavior was observed (Figure 4.12). While the theoretical results predict a further improvement with increasing noise-shaping order L, the experimental results show that the performance flattens out after a certain prediction order, depending on the oversampling factor.

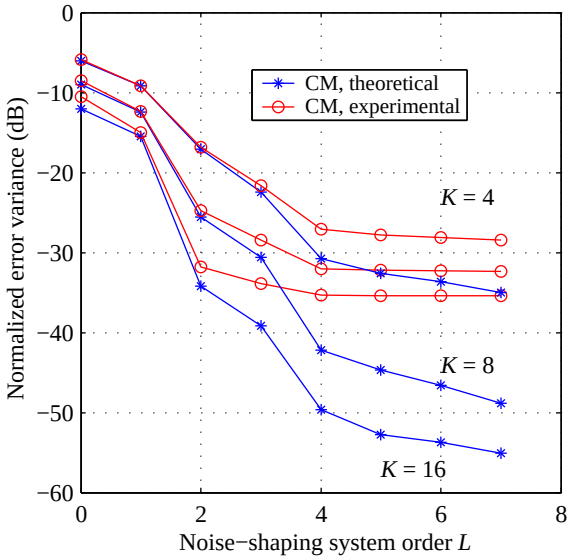


Fig. 4.12 Comparison of the theoretical and experimental performances for the near-perfect reconstruction filter bank CM ($N = 32$, $K = 4, 8, 16$) using the complete interchannel noise-shaping system with quantizer stepsizes $s = 1$.

The explanation of this behavior becomes apparent when the resulting SNR is studied. Figure 4.13 shows the output SNR of the system in Figure 3.4 for different quantizer stepsizes $s = 0.1, 0.25, 1$ which lead to quantization error variances, $\sigma_q^2 = -31, -23, -11$ dB respectively, in all subbands.

This figure demontrates that beyond an SNR of 53.5 dB, the noise-shaping system no longer provides an improvement in the performance of the near-perfect reconstruction filter bank. This point coincides exactly with the observed SNR at the output of the filter bank in the case that there are no quantizers present. Thus, the performance of the noise-shaping system for near-perfect reconstruction filter banks is limited by the reconstruction error of the filter bank itself. This is expected since the noise-shaping filters operate on the quantization noise and so do not affect the reconstruction error.

Finally, it is noted that there is motivation to use the noise-shaping system when the stepsizes of the quantizers are large (i.e. a large quantization error) since in this case the noise-shapers provide a good improvement in SNR. For example, with a stepsize $s = 1$, an improvement of over 20 dB is achieved by using a noise-shaper of order $L = 4$.

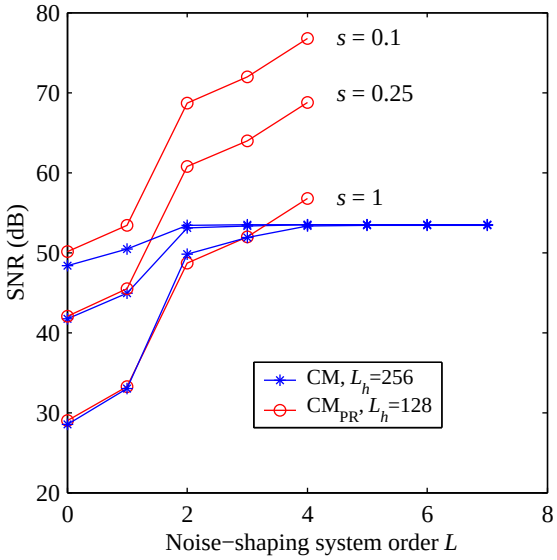


Fig. 4.13 Output SNR for the $\mathrm{CM_{PR}}$ and the CM ($N = 32$, $K = 8$) using the complete interchannel noise-shaping system and varying quantizer stepsizes $s = 0.1, 0.25, 1$.

Rate-distortion characteristic

Although a large improvement in performance has been shown by the introduction of the noise-shaping system, it comes at the cost of a rate increase, proportional to the oversampling factor. It then stands to reason that the performance of the noise-shaping system should be studied from a rate-distortion point of view. The rate-distortion curves show the SNR as a function of the bitrate required to transmit the signal.

It can be shown [30] that the minimum rate needed for reliable transmission of data is given by the entropy of said data. The entropy of an n-dimensional set X of possible symbols is defined as

$$H(X) = \sum_{i=1}^{n} P(x_i) \log_2 P(x_i),$$

where x_i is the i^{th} possible symbol, $P(x_i)$ is the probability of occurrence of that symbol and $H(X)$ is in bits per sample. The entropy of each subband signal was estimated experimentally: the probability of symbol was taken as

$$P(x_i) = \frac{\text{number of occurences of } x_i}{\text{total number of samples}}.$$

Then, the total entropy of the subband signals was taken as the sum of their entropies, divided by the downsampling factor M, since there are M-times less samples in each subband signal than in the input signal. It is finally noted that the entropy gives the minimum *achievable* rate and can only be approached by means of sophisticated source coding[4].

Unfortunately, the results in [14] were confirmed: although an increase in system order did provide a better rate-distortion characteristic, actually decreasing the oversampling factor provided a better gain.

However, it was observed that using longer filters did improve the rate-distortion characteristic at high bitrates. Indeed, Figure 4.14 shows that using the $\mathrm{CM}_{\mathrm{PR}}$ with a longer

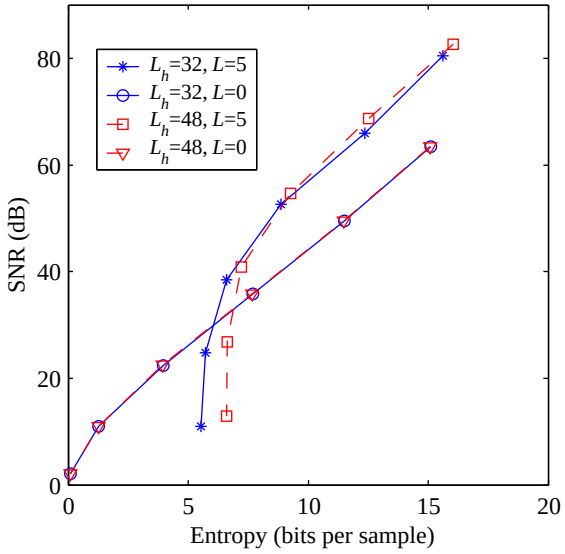


Fig. 4.14 Rate-distortion characteristic of the complete interchannel noise-shaping system using the generated filters, with $N = 8$, $K = 2$ and $L_h = 32, 48$.

length $L_h = 48$ resulted in an increase in SNR without a corresponding rate increase, at high bitrates (i.e. greater than 10 bits per sample).

4.3 Summary

This chapter began with the outline of the algorithm used in the design of cosine-modulated perfect reconstruction filter banks, followed by a presentation of the obtained filters along

[4]For a more in-depth analysis of source coding, one is referred to [30].

with a discussion of the other filters used as a comparison basis for the noise-shaping system. Subsequently, an experimental approach to the study of the effects of the varying filter characteristics was undertaken. The problem of ill-conditioning of the matrices used in the derivation of the noise-shaping filters was reported and a comparison of the intra-channel, complete and local interchannel noise-shaping systems ensued. It was then seen that increasing the length of the filters in the filter bank resulted in an increase in effectiveness of the noise-shaping system, while the advantage of using a MIMO system was evident when comparing performances of filters with a different degree of overlap between subbands. Next, it was shown that the near-perfect-reconstruction filter banks were limited by their reconstruction error. Finally, it was observed that although the improvement in performance due to the noise-shaping filters was not justified by the rate increase, longer filters did improve the rate-distortion characteristic.

Chapter 5

Conclusion

5.1 Summary

The aim of this work was to evaluate the impact of the selection of the filter banks on the performance of a noise-shaping system inserted into an oversampled subband coder. Thus, some basic notions pertinent to subband coders and filter banks were first developed, laying the groundwork to introduce some concepts in filter bank design, such as the condition for perfect reconstruction and the use of lossless matrices to satisfy this condition.

In addition, these concepts allowed the relation of filter banks to frame theory, reiterating that perfect reconstruction is possible by using the dual to the analysis frame as the synthesis frame. Further, the use of frame theory yielded a formulation of the redundancy inherent to oversampling the subband signals: the range space $\mathcal{R}$ of the the analysis operator $\mathcal{T}$ is a subspace of the codomain $\ell^2(\mathbf{Z})^N$, and thus the synthesis operator $\mathcal{U}$ is not unique. However, only the synthesis frame dual to the analysis frame has the maximal noise-reduction property, as it is the only one that projects the noise-components lying in $\mathcal{R}^\perp$ to the zero vector, $\mathbf{0}$.

Subsequently, the choice of critically sampled cosine-modulated perfect reconstruction filter banks as the test bed for the noise-shaping system was justified: not only do they simply require the design of one prototype filter, but the corresponding synthesis filter bank also yields the dual to the analysis frame. It was also shown that the prototype filter could be efficiently obtained using a lossless two-channel lattice structure, ensuring that the perfect reconstruction condition was met; the coefficients of the lattice structure were to be optimized using Newton's method.

Attention was then focused on the noise-shaping system, whose goal is to push the quantization noise to $\mathcal{R}^{\perp}$, such that it would be attenuated by the synthesis filter bank. The MIMO noise-shaping filters were constrained to be FIR and their derivation led to the theoretical calculation of the reconstruction error variance for three cases: the complete interchannel, the local interchannel and the intrachannel (SISO) noise-shaping systems.

The implementation of the algorithm used in the design of the filter banks was then outlined, and it was shown that increasing the filter length improved the stopband attenuation of the component filters, while this attenuation was even better for near-perfect reconstruction filter banks. The perfect reconstruction of the input signal was verified, while it was seen that the reconstruction error of the near-perfection reconstruction filter was on the same order as the attenuation of the stopbands of the filters.

Next, quantization was performed on the subband signals and the noise-shaping system was implemented: simulation results were compared with the theoretically computed values for the normalized reconstruction error variance. It was found that the simulations agreed with the predicted results, except when the matrices used in the derivation of the noise-shaping coefficients were ill-conditioned, leading to inaccurate results. Thus, ill-conditioning limited the order of the noise-shaping filters. While the intrachannel noise-shaping system remained better conditioned for higher noise-shaping orders, both the complete and local interchannel systems outperformed it, even with their lower implementable orders. Further, the local interchannel noise-shaping system provided a good approximation to the complete interchannel noise-shaping systems — for small orders — at a reduced computational load.

It was finally possible to compare the performance of the noise-shaping filters using filters with different characteristics. First, it was found that longer filters could achieve a greater reduction on reconstruction error. Secondly, it was observed that filters with a larger overlap between subbands could yield further reduction in reconstruction error and thirdly it was found that near-perfect reconstruction filter banks were limited by their reconstruction error, which depended, in turn, on the stopband attenuation of the component filters of the filter bank.

The performance of the noise-shaping filters were then examined from a rate-distortion point of view. Previous results [14] were confirmed: critically sampled filter banks outperform oversampled noise-shaping filter banks, and thus the rate increase proportional to the oversampling ratio is not justified by the use of noise-shaping filters. However, it was noted

that an increase in filter length did improve the rate-distortion characteristic.

5.2 Future Work

The immediate extension to this work would be to design filter banks with the objective of maximizing the performance of their corresponding noise-shaping filters, given the characteristics that improved their behaviour: increased filter length, larger overlap between subbands and stronger attenuation in the filter banks.

Of course, these design characteristics come at a cost. For example, increased filter lengths lead to a longer delay through the system: the delay — in number of samples — is equal to the length of the filters in the filter bank. Thus, this is a problem in real-time systems, where it is desired to restrict processing delay to a minimum. Moreover, larger overlap between subbands leads to a poorer frequency resolution in the subband signals. If it had been desired to accomplish some other type of processing on the subbands separately, the larger degree of overlap would contaminate the subband content with that of neighboring subbands. Finally, the design of filter banks with higher stopband attenuation requires the elaboration of sophisticated design algorithms.

While it is doubtful that it would then be possible to achieve a better rate-distortion characteristic than in the critically sampled case, these filters could be useful in systems that are already oversampled, for other purposes. For instance, as mentioned previously, it might be desirable to trade an increase in rate for simpler quantizers as the introduction of a noise-shaping system on the subbands improves the effective resolution of the quantizers. Another example is the use of oversampled filter banks in audio signal processing: because it is desired to analyze an audio signal in a manner emulating the human ear, the frequency bands will be non-uniform [31]. This then leads to aliasing due to unequal processing of the subbands, which can be reduced below the level of human hearing by using oversampled filter banks [11].

This leads then to the question of whether or not there are perceptual advantages to using noise-shaping in oversampled filter banks. Indeed, perceptual audio coding is well-documented and used in many international and commercial standards [32]. Here, the coding noise is shaped such that it is below the masking threshold: thus, compression is achieved by allowing a certain degree of error, as long as it is imperceptible to humans. Applying this to the situation at hand, not only might there be perceptual advantages

to the investigated noise-shaping system, but perhaps a more effective noise-shaper could be developed combining the knowledge of the filter bank used and the knowledge of the perception of sound.

As a final note, noise-shaping in oversampled filter banks need not be constrained to one dimension: it would be interesting to evaluate the performance of the noise-shaping system for images. In this case, the phase bears more pertinent information to the recovery of the original signal than for speech. It would thus be interesting to once more evaluate the benefits of noise-shaping, and in addition, perhaps develop a noise-shaper making use of the additional information inherent to images.

Appendix A

Lattice structure for partial derivatives of the polyphase components

A.1 First-order partial derivatives

Recalling Eq. (3.5), the first order partial derivatives of the polyphase components with respect to the lattice parameters $\theta_{k,p}$ can be computed as follows:

$$\begin{aligned}\frac{\partial W_l^{(i)}}{\partial \theta_{k,p}} = &\left[\left(\frac{\partial}{\partial \theta_{k,p}} \cos\theta_{l,i}\right) W_l^{i-1} + \left(\frac{\partial}{\partial \theta_{k,p}} W_l^{(i-1)}\right) \cos\theta_{l,i}\right] \\ &+ z^{-1}\left[\left(\frac{\partial}{\partial \theta_{k,p}} \sin\theta_{l,i}\right) W_{N+l}^{i-1} + \left(\frac{\partial}{\partial \theta_{k,p}} W_{N+l}^{(i-1)}\right) \sin\theta_{l,i}\right].\end{aligned} \tag{A.1}$$

for $0 \le k, l \le \left\lfloor \frac{N}{2} \right\rfloor$ and $1 \le p, i \le m-1$. A similar equation can be written for the remaining $\frac{\partial W_{N+l}^{(i)}}{\partial \theta_{k,p}}$, but is omitted here in the interest of conciseness. (Further, the dependance of the polyphase components on z is considered to be implicit: as it does not affect the partial derivatives, it is dropped for clarity.)

In the case $k \neq l$, $\frac{\partial W_l^{(i)}}{\partial \theta_{k,p}} = 0$, since k, l are the indices of parallel lattice structures and thus the polyphase components $W_k^{(i)}$ are generated only by those parameterized by $\theta_{k,p}$.

If $k = l$, the partial derivatives can be separated into three cases:

Case 1. $i < p$

Here, $\frac{\partial W_l^{(i)}}{\partial \theta_{k,p}} = 0$, since the parameters $\theta k, p$ are not yet included in the generation of $W_l^{(i)}$, when $i < p$. Consequently, $\frac{\partial W_l^{(i-1)}}{\partial \theta_{k,p}} = 0$ also for all $i \leq p$.

Case 2. $i = p$

Since $\frac{\partial W_l^{(i-1)}}{\partial \theta_{k,p}} = 0$ for $i = p$, Eq. (A.1) can be reduced to

$$\frac{\partial W_k^{(p)}}{\partial \theta_{k,p}} = -\sin\theta_{k,p} W_k^{(p-1)} + z^{-1}\cos\theta_{k,p} W_{N+l}^{(p-1)},$$

while similarly,

$$\frac{\partial W_{N+k}^{(p)}}{\partial \theta_{k,p}} = \cos\theta_{k,p} W_k^{(p-1)} + z^{-1}\sin\theta_{k,p} W_{N+l}^{(p-1)},$$

for $p \neq 0$, which can be easily obtained using a two-channel lattice structure. Next, for $p = 0$, (see Eq. (4.4))

$$\begin{aligned}\frac{\partial W_k^{(0)}}{\partial \theta_{k,0}} &= \sin\theta_{k,0} \\ \frac{\partial W_{N+k}^{(0)}}{\partial \theta_{k,0}} &= -\cos\theta_{k,0},\end{aligned}$$

yielding the initial conditions for these lattices.

Case 3. $i > p$

In this case, $\frac{\partial}{\partial \theta_{k,p}}\cos\theta_{k,i} = \frac{\partial}{\partial \theta_{k,p}}\sin\theta_{k,i} = 0$ and Eq. (A.1) is reduced to

$$\frac{\partial W_k^{(i)}}{\partial \theta_{k,p}} = \cos\theta_{k,i}\frac{\partial}{\partial \theta_{k,p}} W_k^{(i-1)} + z^{-1}\sin\theta_{k,p}\frac{\partial}{\partial \theta_{k,p}} W_{N+k}^{(i-1)},$$

and similarly,

$$\frac{\partial W_{N+k}^{(i)}}{\partial \theta_{k,p}} = \sin\theta_{k,i}\frac{\partial}{\partial \theta_{k,p}} W_k^{(i-1)} - z^{-1}\cos\theta_{k,p}\frac{\partial}{\partial \theta_{k,p}} W_{N+k}^{(i-1)}.$$

Once again, these can be obtained easily using a two-channel lattice structure, whose initial conditions are obtained from the output of the lattice structures for the case $i = p$.

A.2 Second order partial derivatives

From Eq. (A.1), the second order partial derivatives with respect to the $\theta_{k,p}$ are written as:

$$\begin{aligned}\frac{\partial^2}{\partial\theta_{j,q}\partial\theta_{k,p}}W_l^{(i)} = &\left[\left(\frac{\partial^2}{\partial\theta_{j,q}\partial\theta_{k,p}}\cos\theta_{l,i}\right)W_l^{(i-1)} + \left(\frac{\partial}{\partial\theta_{j,q}}\cos\theta_{l,i}\right)\frac{\partial}{\partial\theta_{k,p}}W_l^{(i-1)}\right.\\ &\left.+\left(\frac{\partial}{\partial\theta_{k,p}}\cos\theta_{l,i}\right)\frac{\partial}{\partial\theta_{j,q}}W_l^{(i-1)} + \cos\theta_{l,i}\left(\frac{\partial^2}{\partial\theta_{j,q}\partial\theta_{k,p}}W_l^{(i-1)}\right)\right]\\ &+z^{-1}\left[\left(\frac{\partial^2}{\partial\theta_{j,q}\partial\theta_{k,p}}\sin\theta_{l,i}\right)W_{N+l}^{(i-1)} + \left(\frac{\partial}{\partial\theta_{j,q}}\sin\theta_{l,i}\right)\frac{\partial}{\partial\theta_{k,p}}W_{N+l}^{(i-1)}\right.\\ &\left.+\left(\frac{\partial}{\partial\theta_{k,p}}\sin\theta_{l,i}\right)\frac{\partial}{\partial\theta_{j,q}}W_{N+l}^{(i-1)} + \sin\theta_{l,i}\left(\frac{\partial^2}{\partial\theta_{j,q}\partial\theta_{k,p}}W_{N+l}^{(i-1)}\right)\right],\end{aligned} \tag{A.2}$$

for $0 \le j,k,l \le \left\lfloor\frac{N}{2}\right\rfloor$ and $1 \le p,q,i \le m-1$. Once more, a similar equation can be written for the remaining $\frac{\partial^2 W_{N+l}^{(i)}}{\partial\theta_{j,q}\partial\theta_{k,p}}$, but is omitted here in the interest of conciseness.

Again, for $k \neq j \neq l$ the second order partial derivatives $\frac{\partial^2 W_l^{(i)}}{\partial\theta_{j,q}\partial\theta_{k,p}} = 0$, since k,j,l denote the separate two-channel lattice structures. Also, for $k = l = j$, in both cases that $i < p$ or $i < q$, the partial derivatives are 0, since either $\theta_{k,p}$ or $\theta_{k,q}$ are not included in the parametrization of the polyphase component $W_k^{(i)}$. The non-zero cases are then:

Case 1. $i > p$ **and** $i > q$

In this case, $\frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}}\cos\theta_{k,i} = 0$ for $i \neq q$ or $i \neq p$, $\frac{\partial}{\partial\theta_{k,s}}\cos\theta_{k,i} = 0$ for $s \neq i$ and $\frac{\partial}{\partial\theta_{k,s}}W_k^{(i-1)} = 0$ for $s \neq i-1$. Hence, Eq. (A.2) is reduced to

$$\frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}}W_k^{(i)} = \cos\theta_{k,i}\frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}}W_k^{(i-1)} + z^{-1}\sin\theta_{k,i}\frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}}W_{N+k}^{(i-1)}$$

and through similar steps

$$\frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}}W_{N+k}^{(i)} = \sin\theta_{k,i}\frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}}W_k^{(i-1)} - z^{-1}\cos\theta_{k,i}\frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}}W_{N+k}^{(i-1)}$$

which can be obtained efficiently using the cascade of two-channel lattices.

Case 2. $i = p$ **and** $i > q$

Here, $\frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}}\cos\theta_{k,p} = 0$ as $i \neq q$, $\frac{\partial}{\partial\theta_{k,q}}W_k^{(p-1)} = 0$ for $q \neq p-1$ and $\frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,i}}W_k^{(p-1)} = 0$ for $q \neq p-1$. Thus, Eq. (A.2) is reduced to

$$\frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}}W_k^{(p)} = -\sin\theta_{k,p}\frac{\partial}{\partial\theta_{k,q}}W_k^{(p-1)} + z^{-1}\cos\theta_{k,p}\frac{\partial}{\partial\theta_{k,q}}W_{N+k}^{(p-1)},$$

and similarly for the remaining polyphase components are given by

$$\frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}}W_{N+k}^{(p)} = \cos\theta_{k,p}\frac{\partial}{\partial\theta_{k,q}}W_k^{(p-1)} + z^{-1}\sin\theta_{k,p}\frac{\partial}{\partial\theta_{k,q}}W_{N+k}^{(p-1)}.$$

Again, the implementation using the lattice structure is apparent.

Case 3. $i > p$ **and** $i = q$

Through similar manipulations as in case 2, Eq. (A.2) is reduced to

$$\frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}}W_k^{(q)} = -\sin\theta_{k,q}\frac{\partial}{\partial\theta_{k,p}}W_k^{(q-1)} + z^{-1}\cos\theta_{k,q}\frac{\partial}{\partial\theta_{k,p}}W_{N+k}^{(q-1)},$$

and

$$\frac{\partial^2}{\partial\theta_{k,q}\partial\theta_{k,p}}W_{N+k}^{(q)} = \cos\theta_{k,q}\frac{\partial}{\partial\theta_{k,p}}W_k^{(q-1)} + z^{-1}\sin\theta_{k,q}\frac{\partial}{\partial\theta_{k,p}}W_{N+k}^{(q-1)},$$

which are again amenable to the lattice structure.

Case 4. $p = q = i$

This time, $\frac{\partial}{\partial\theta_{k,p}}W_k^{(p-1)} = 0$ and $\frac{\partial^2}{\partial^2\theta_{k,p}}W_k^{(p-1)} = 0$, such that Eq. (A.2) is reduced to

$$\frac{\partial^2}{\partial^2\theta_{k,p}}W_k^{(p)} = -\cos\theta_{k,p}W_k^{(i-1)} - z^{-1}\sin\theta_{k,p}W_{N+k}^{(i-1)},$$

while

$$\frac{\partial^2}{\partial^2\theta_{k,p}}W_{N+k}^{(p)} = -\sin\theta_{k,p}W_k^{(i-1)} + z^{-1}\cos\theta_{k,p}W_{N+k}^{(i-1)},$$

once more implementable through a the appropriate lattice structure.

Finally, the second order partial derivatives are initialized as

$$\frac{\partial^2}{\partial^2 \theta_{k,0}} W_k^{(0)} = -\cos\theta_{k,0},$$

and

$$\frac{\partial^2}{\partial^2 \theta_{k,0}} W_{N+k}^{(0)} = -\sin\theta_{k,0}.$$

Appendix B

Generated coefficients

The coefficients of the prototype filters generated by using the algorithm are presented. Only the first mN coefficients are given, since the remaining mN are symmetric.

B.1 $N = 8$

B.1.1 $m = 2$, $L_h = 32$

$$p_0(n) = [\ \begin{matrix} -0.0524 & -0.0433 & -0.0282 & -0.0103 & 0.0120 & 0.0457 & 0.0992 & 0.1768 \\ 0.2790 & 0.3978 & 0.5241 & 0.6488 & 0.7608 & 0.8499 & 0.9110 & 0.9424 \end{matrix}\ ] \tag{B.1}$$

B.1.2 $m = 3$, $L_h = 48$

$$p_0(n) = [\ \begin{matrix} -0.0051 & -0.0052 & -0.0031 & 0.0014 & -0.0012 & -0.0062 & -0.0174 & -0.0288 \\ -0.0360 & -0.0366 & -0.0298 & 0.0000 & 0.0000 & 0.0598 & 0.1224 & 0.2039 \\ 0.3030 & 0.4155 & 0.5349 & 0.6524 & 0.7579 & 0.8423 & 0.9004 & 0.9297 \end{matrix}\ ] \tag{B.2}$$

B.2 $N = 16$, $m = 2$, $L_h = 64$

$$
\begin{aligned}
p_0(n) = [\ & -0.0755 \quad -0.0729 \quad -0.0673 \quad -0.0592 \quad -0.0491 \quad -0.0372 \quad -0.0237 \quad -0.0084 \\
& 0.0090 \quad 0.0291 \quad 0.0526 \quad 0.0801 \quad 0.1124 \quad 0.1496 \quad 0.1919 \quad 0.2391 \\
& 0.2915 \quad 0.3475 \quad 0.4047 \quad 0.4624 \quad 0.5198 \quad 0.5761 \quad 0.6306 \quad 0.6824 \\
& 0.7309 \quad 0.7752 \quad 0.8148 \quad 0.8491 \quad 0.8775 \quad 0.8996 \quad 0.9149 \quad 0.9231\ ]
\end{aligned}
\tag{B.3}
$$

References

[1] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. New Jersey: Prentice-Hall, second ed., 1999.

[2] P. P. Vaidyanathan, "Multirate digital filters, filter banks, polyphase networks, and applications: A tutorial," *Proceedings of the IEEE*, vol. 78, pp. 56–93, January 1990.

[3] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. New Jersey: Prentice-Hall, 1993.

[4] M. Vetterli, "A theory of multirate filter banks," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, pp. 356–372, March 1987.

[5] W. Chen, B. Han, and R.-Q. Jia, "On simple oversampled A/D conversion in shift-invariant spaces," *IEEE Transactions on Information Theory*, vol. 51, pp. 648–657, February 2005.

[6] Z. Cvetković and M. Vetterli, "Oversampled filter banks," *IEEE Transactions on Signal Processing*, vol. 46, pp. 1245–1255, May 1998.

[7] V. K. Goyal, M. Vetterli, and N. T. Thao, "Quantized overcomplete expansions in R^N: Analysis, synthesis and algorithms," *IEEE Transactions on Information Theory*, vol. 44, pp. 16–31, January 1998.

[8] V. K. Goyal, J. Kovačević, and M. Vetterli, "Multiple description transform coding: robustness to erasures using tight frame expansions," in *Proceedings of IEEE International Symposium on Information Theory*, (Cambridge MA), p. 408, August 1998.

[9] V. K. Goyal, J. Kovačević, and M. Vetterli, "Quantized frame expansions as source-channel codes for erasure channels," in *Proceedings of Data Compression Conference*, (Snowbird UT), pp. 326–335, March 1999.

[10] H. Bölcskei and F. Hlawatsch, "Oversampled cosine modulated filter banks with perfect reconstruction," *IEEE Transactions on Circuits and Systems - II: Analog and Digital Sgnal Processing*, vol. 45, pp. 1057–1071, August 1998.

[11] Z. Cvetković and J. D. Johnston, "Nonuniform oversampled filter banks for audio signal processing," *IEEE Transactions on Speech and Audio Processing*, vol. 11, pp. 393–399, September 2003.

[12] R. F. von Borries, R. L. de Queiroz, and C. S. Burrus, "On filter banks with rational oversampling," *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001 (ICASSP '01). Proceedings.*, vol. 6, pp. 3657–3660, May 2001.

[13] R. F. von Borries and C. S. Burrus, "Linear phase oversampled filter banks," *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004 (ICASSP '04). Proceedings.*, vol. 2, pp. 961–964, May 2004.

[14] H. Bölcskei and F. Hlawatsch, "Noise reduction in oversampled filter banks using predictive quantization," *IEEE Transactions on Information Theory*, vol. 47, pp. 155–172, January 2001.

[15] J. C. Candy and G. C. Temes, *Oversampling Delta-Sigma Data Converters*. IEEE Press, first ed., 1992.

[16] P. P. Vaidyanathan, "Theory and design of M-channel maximally decimated quadrature mirror filter with arbitrary M, having perfect reconstruction property," *IEEE Transactions on Acoustic Speech Signal Processing*, vol. 35, pp. 476–492, April 1987.

[17] R. J. Duffin and A. C. Schaeffer, "A class of nonharmonic Fourier series," *Transactions of the American Mathematics Society*, vol. 72, pp. 341–366, March 1952.

[18] G. Zames, "Course notes – linear systems, ECSE-501." McGill University, 2002.

[19] Z. Doğanata, P. P. Vaidyanathan, and T. Q. Nguyen, "General synthesis procedures for FIR lossless transfer matrices, for perfect reconstruction multirate filter bank applications," *IEEE Transactions on Acoustic and Speech Signal Processing*, vol. 36, pp. 1561–1574, October 1988.

[20] R. D. Koilpillai and P. P. Vaidyanathan, "Cosine-modulated FIR filter banks satisfying perfect reconstruction," *IEEE Transactions on Signal Processing*, vol. 40, pp. 770–783, April 1992.

[21] O. G. Ibarra-Manzano and G. Jovanovic-Dolecek, "Cosine modulated FIR filter banks satisfying perfect reconstruction: An iterative algorithm," *42nd Midwest Symposium on Circuits and Systems*, vol. 2, pp. 1061–1064, August 1999.

[22] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing*. Boston: McGraw-Hill Higher Education, first ed., 2000.

[23] E. Weisstein, "Mathworld — a Wolfram web ressource." URL: `http://mathworld.wolfram.com`, 2005.

[24] S. Rao and W. A. Pearlman, "Analysis of linear prediction, coding and spectral estimation from subbands," *IEEE Transactions on Information Theory*, vol. 42, pp. 1160–1178, July 1996.

[25] S.-L. Tan and T. R. Fischer, "Linear prediction of subband signals," *IEEE Journals on Selected Areas in Communications*, vol. 12, pp. 1576–1583, December 1994.

[26] B. Maison, *Adaptive operators and higher order statistics for digital image compression*. PhD thesis, Université catholique de Louvain, 1998.

[27] H. S. Malvar, "Lapped transforms for efficient transform/subband coding," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, pp. 969–978, June 1990.

[28] T. Q. Nguyen, "Near-perfect-reconstruction pseudo-QMF banks," *IEEE Transactions on Signal Processing*, vol. 42, pp. 65–76, January 1994.

[29] L. Blum, F. Cucker, M. Schub, and S. Smale, *Complexity and Real Computation*. New York: Springer-Verlag, first ed., 1998.

[30] J. Proakis, *Digital Communications*. Boston: McGraw-Hill Higher Education, fourth ed., 2001.

[31] D. O'Shaughnessy, *Speech Communications — Human and Machine*. New Jersey: IEEE Press, second ed., 1998.

[32] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proceedings of the IEEE*, vol. 88, pp. 451–513, April 2000.