Noise Reduction in Oversampled Filter Banks Using Predictive Quantization

Helmut Bölcskei, *Member, IEEE,* and Franz Hlawatsch, *Member, IEEE*

Abstract—We introduce two methods for quantization noise reduction in oversampled filter banks. These methods are based on predictive quantization (noise shaping or linear prediction). It is demonstrated that oversampled noise shaping or linear predictive subband coders are well suited for subband coding applications where, for technological or other reasons, low-resolution quantizers have to be used. In this case, oversampling combined with noise shaping or linear prediction improves the effective resolution of the subband coder at the expense of increased rate. Simulation results are provided to assess the achievable quantization noise reduction and resolution enhancement, and to investigate the rate-distortion properties of the proposed methods.

Index Terms—Filter banks, frame theory, linear prediction, noise reduction, noise shaping, oversampling, quantization, rate-distortion theory, sigma–delta converter, subband coding.

I. INTRODUCTION AND OUTLINE

RECENTLY, oversampled filter banks (FBs) have received increased attention [1]–[10], which is mainly due to their *noise reducing properties* and *increased design freedom* (i.e., nonuniqueness of the perfect reconstruction synthesis FB for a given analysis FB). In this paper, we introduce two techniques for quantization noise reduction in oversampled FBs. These techniques are based on predictive quantization, specifically, on *noise prediction (noise shaping)* and *signal prediction*. The corresponding oversampled subband coders can be viewed as extensions of oversampled predictive A/D converters [11]–[13] and of critically sampled predictive subband coders [14]–[16].

We show that predictive quantization in oversampled FBs yields significant noise reduction at the cost of increased bit rate. Hence, oversampled predictive subband coders allow to trade bit rate for quantizer accuracy. They are, therefore, well suited for subband coding applications where, for technological or other reasons, quantizers with low accuracy (even single-bit) have to be used. The practical advantages of using low-resolution quantizers at the cost of increased rate are indicated by the popular sigma–delta techniques [11]–[13]. Using low-resolution quantizers increases circuit speed and reduces circuit complexity. 1-bit codewords, for example, eliminate the need for word framing [14]. We, furthermore, study rate-distortion properties of oversampled predictive subband coders. Specifically, we demonstrate by means of simulation results that oversampled predictive subband coders are inferior to critically sampled subband coders from a pure rate-distortion point of view. (An information-theoretic treatment of the rate-distortion properties of one specific class of redundant representations, namely, frames of sinc functions or equivalently oversampled A/D conversion, can be found in [17] and [18].)

As a basis for our development of predictive quantization in oversampled FBs, we provide a subspace-based noise analysis of oversampled FBs. In particular, it is proven that the perfect reconstruction (PR) synthesis FB corresponding to the para-pseudo-inverse of the analysis polyphase matrix minimizes the reconstruction error variance resulting from uncorrelated white noise in the subbands. This result is then generalized to include correlated and/or colored subband noise signals. The fact that other PR synthesis FBs lead to an additional reconstruction error corresponds to a fundamental tradeoff between noise reduction and design freedom.

The paper is organized as follows. In Section II, we develop a subspace-based stochastic noise analysis of oversampled FBs and we calculate the PR synthesis FB minimizing the reconstruction error due to noise. Section III introduces oversampled noise shaping (noise predictive) subband coders. We calculate the optimum noise shaping system and provide simulation results demonstrating the achievable noise reduction. Oversampled signal predictive subband coders are introduced in Section IV. The optimum multichannel prediction system is calculated and the achievable resolution enhancement is demonstrated by simulation results. Finally, Section V concludes our presentation.

II. SUBSPACE-BASED NOISE ANALYSIS OF OVERSAMPLED FBs

In this section, we provide a subspace-based noise analysis of oversampled FBs and demonstrate that oversampled FBs have noise-reducing properties. We start with a brief review of frame theory on which some of our results will be based, followed by a brief discussion of oversampled A/D conversion where the subspaces involved have a particularly simple structure. We then turn to oversampled FBs and study the reconstruction error caused by noisy subband signals. Bounds on the error variance are derived, and the dependence of the error variance on the frame bounds and oversampling factor is discussed. We finally calculate the PR synthesis FB minimizing the reconstruction

error variance due to noise and describe a tradeoff between noise reduction and design freedom in oversampled FBs.

A. Brief Review of Frame Theory

The theory of frames [19]–[21] is a powerful tool for the study of redundant (overcomplete) signal expansions. A function set $\{g_k(t)\}$ with $k \in \mathbb{Z}$ is called a *frame* for $L^2(\mathbb{R})$ if[1]

$$A\|x\|^2 \leq \sum_{k=-\infty}^{\infty} |\langle x, g_k \rangle|^2 \leq B\|x\|^2 \quad \forall\, x(t) \in L^2(\mathbb{R}) \quad (1)$$

with the *frame bounds* $A > 0$ and $B < \infty$. If $\{g_k(t)\}$ is a frame for $L^2(\mathbb{R})$, any signal $x(t) \in L^2(\mathbb{R})$ can be decomposed as [19]–[21]

$$x(t) = \sum_{k=-\infty}^{\infty} \langle x, \tilde{g}_k \rangle g_k(t) = \sum_{k=-\infty}^{\infty} \langle x, g_k \rangle \tilde{g}_k(t).$$

Here, $\tilde{g}_k(t) = (\boldsymbol{S}^{-1} g_k)(t)$ where $\boldsymbol{S}^{-1}$ is the inverse of the *frame operator* $\boldsymbol{S}$ that is defined as

$$(\boldsymbol{S}x)(t) = \sum_{k=-\infty}^{\infty} \langle x, g_k \rangle g_k(t).$$

The function set $\{\tilde{g}_k(t)\}$ is again a frame (the "dual" frame), with frame bounds $\tilde{A} = 1/B$ and $\tilde{B} = 1/A$. The frame bounds determine important numerical properties of the frame [19]–[21]. A frame is called *snug* if $B/A \approx 1$ and *tight* if $B/A = 1$. For a tight frame we have $\boldsymbol{S} = A\boldsymbol{I}$ (where $\boldsymbol{I}$ is the identity operator on $L^2(\mathbb{R})$), and hence there is simply $\tilde{g}_k(t) = \frac{1}{A} g_k(t)$.

B. Noise Analysis and Design Freedom in Oversampled A/D Conversion

As a motivation of our noise analysis of oversampled FBs (to be presented in Section II-C), this subsection provides a frame-theoretic, subspace-based interpretation of noise reduction in oversampled A/D conversion.

We shall first interpret A/D conversion as a frame expansion [19]–[21]. From the sampling theorem [22], [23], we know that a band-limited continuous-time signal $x(t)$ with bandwidth B_0 can be perfectly recovered from its samples $x(kT)$, where $T = \frac{1}{F_s}$ with $F_s \geq 2B_0$, i.e.,

$$x(t) = \frac{1}{K} \sum_{k=-\infty}^{\infty} x(kT)\mathrm{sinc}\,[2\pi B_0(t - kT)]. \quad (2)$$

Here, $\mathrm{sinc}\,(\alpha) = \frac{\sin \alpha}{\alpha}$ and $K = \frac{F_s}{2B_0}$ is the oversampling factor. The samples $x(kT)$ can be written as $x(kT) = \langle x, g_k \rangle$, where $g_k(t) = 2B_0 \mathrm{sinc}\,[2\pi B_0(t - kT)]$. Thus, (2) can be rewritten as
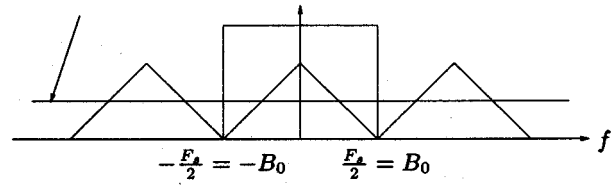
$$x(t) = \frac{1}{F_s} \sum_{k=-\infty}^{\infty} \langle x, g_k \rangle g_k(t) \quad (3)$$

---

[1]Here $L^2(\mathbb{R})$ denotes the space of square-integrable functions $x(t)$. Furthermore,

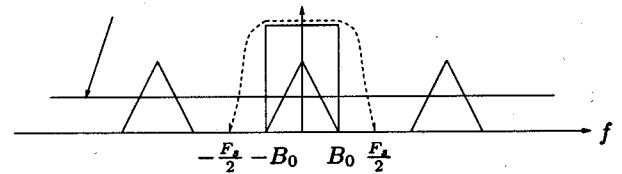$$\langle x, y \rangle = \int_{-\infty}^{\infty} x(t)y^*(t)\, dt$$

(where the superscript $*$ stands for complex conjugation) denotes the inner product of the functions $x(t)$ and $y(t)$, and $\|x\|^2 = \langle x, x \rangle$.

(a)

(b)

Fig. 1. Reconstruction of analog signal by low-pass filtering. (a) Critical case. (b) Oversampled case.

which shows that A/D conversion can be interpreted as an expansion of $x(t)$ into the function set $\{g_k(t)\}$.

The frequency-domain expression of the frame operator $\boldsymbol{S}$ of $\{g_k(t)\}$ is given by[2]

$$(\widehat{\boldsymbol{S}}X)(f) = \sum_{k=-\infty}^{\infty} \langle X, G_k \rangle G_k(f) = F_s \,\mathrm{rect}_{B_0}(f)X(f)$$

where

$$G_k(f) = (\boldsymbol{F}g_k)(t) = \int_{-\infty}^{\infty} g_k(t)e^{-j2\pi ft}\, dt$$

is the Fourier transform of $g_k(t)$ and $\widehat{\boldsymbol{S}} = \boldsymbol{F}\boldsymbol{S}\boldsymbol{F}^{-1}$. Since
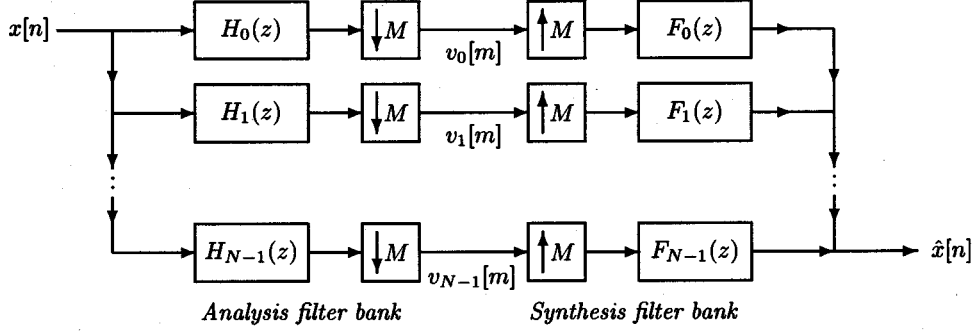
$$(\widehat{\boldsymbol{S}}X)(f) = F_s X(f)$$

for $x(t)$ B_0-band-limited, we can conclude that $\{g_k(t)\}$ is a tight frame for the space of B_0-band-limited functions, with frame bounds $A = B = F_s$. Hence, the dual frame is given by $\tilde{g}_k(t) = \frac{1}{F_s} g_k(t)$. This shows that the interpolation formula (2) or (3) corresponds to a reconstruction using the dual frame. Moreover, it is easily checked that for critical sampling ($K = 1$ or $T = \frac{1}{2B_0}$) the $g_k(t)$ are orthogonal to each other, i.e., $\langle g_k, g_l \rangle = 2B_0\delta[k - l]$. In the oversampled case, the set $\{g_k(t)\}$ is redundant.

The reconstruction of $x(t)$ from its samples $x(kT)$ can alternatively be interpreted as the application of a low-pass filter to the signal $\sum_{k=-\infty}^{\infty} x(kT)\delta(t - kT)$. In the case of critical sampling (i.e., $F_s = 2B_0$), the ideal low-pass filter of bandwidth B_0 is the only filter that provides PR of $x(t)$ [see Fig. 1(a)]. In the oversampled case (i.e., $F_s > 2B_0$), an infinite number of reconstruction low-pass filters will provide PR [see Fig. 1(b)]; the resulting design freedom [12] can be exploited for designing reconstruction filters with desirable filter characteristics like, e.g., rolloff.

Assuming quantization errors modeled as additive white noise, with the quantization error variance per sample held

---

[2]Here, $\mathrm{rect}_{B_0}(f) = 1$ for $|f| \leq B_0$ and $\mathrm{rect}_{B_0}(f) = 0$ else.

Fig. 2. N-channel uniform filter bank.

constant, and employing an ideal low-pass reconstruction filter with bandwidth B_0, it follows from Fig. 1 that the reconstruction error variance σ_o^2 in the oversampled case is given by [12]

$$\frac{\sigma_o^2}{\sigma_c^2} = \frac{1}{K}$$

where σ_c^2 is the reconstruction error variance in the critically sampled case and $K = \frac{F_s}{2B_0}$ is the oversampling factor. Any other reconstruction filter providing PR must pass some of the noise outside the signal band [see Fig. 1(b)] and will thus lead to a larger reconstruction error variance. In this sense, there exists a tradeoff between noise reduction and design freedom in oversampled A/D conversion. Practically desirable (or realizable) reconstruction filters (i.e., filters with rolloff) lead to an additional reconstruction error.

We shall finally provide a frame-theoretic, subspace-based interpretation of these well-known facts. For oversampling factor K, the range space $\mathcal{R}$ of the analysis (sampling) operator $\boldsymbol{T}: x(t) \to x(kT)$ is the space of discrete-time functions bandlimited to the interval $\theta \in [-\frac{1}{2K}, \frac{1}{2K}]$. Reconstruction of $x(t)$ using the ideal low-pass filter of bandwidth B_0 (or, equivalently, in the discrete-time domain, bandwidth $\frac{1}{2K}$) corresponds to an orthogonal projection onto $\mathcal{R}$; on the other hand, we recall that it also corresponds to a reconstruction using the dual frame $\tilde{g}_k(t) = \frac{1}{F_s} g_k(t)$. Hence, it follows that reconstruction using the dual frame involves an orthogonal projection onto $\mathcal{R}$. This projection suppresses the noise component lying in the orthogonal complement $\mathcal{R}^\perp$ of the range space $\mathcal{R}$ (corresponding to the out-of-band region $\frac{1}{2K} \leq |\theta| \leq \frac{1}{2}$). This intuitively explains why reconstruction using the dual frame leads to minimum reconstruction error.

In Section II-C, we shall see that a similar tradeoff between noise reduction and design freedom arises in oversampled FBs. The analysis is less intuitive there, however, since the signal spaces $\mathcal{R}$ and $\mathcal{R}^\perp$ do not correspond to simple frequency bands.

C. Noise Analysis and Design Freedom in Oversampled FBs

After this discussion of oversampled A/D conversion, we now turn to oversampled FBs. In this subsection, we will provide a stochastic noise analysis of oversampled FBs and describe a tradeoff between noise reduction and design freedom. We begin with a brief review of oversampled FBs.

1) Oversampled FBs: We consider an N-channel FB (see Fig. 2) with subsampling by the integer factor M in each channel. The transfer functions of the analysis and synthesis filters are $H_k(z)$ and $F_k(z)$ ($k = 0, 1, \ldots, N - 1$), with corresponding impulse responses $h_k[n]$ and $f_k[n]$, respectively.[3] In a *critically sampled* (or *maximally decimated*) FB we have $N = M$ and thus the subband signals $v_k[m]$ ($k = 0, 1, \ldots, N - 1$) contain exactly as many samples per unit of time as the input signal $x[n]$. In the *oversampled* case $N > M$, the subband signals are redundant in that they contain more samples per unit of time than the input signal. (In a finite-dimensional setting, oversampling would correspond to representing an $M \times 1$ vector using $N > M$ expansion coefficients.)

The $N \times M$ analysis polyphase matrix $\mathbf{E}(z)$ is defined as $[\mathbf{E}(z)]_{k, n} = E_{k, n}(z)$, where [24], [25]

$$E_{k,n}(z) = \sum_{m=-\infty}^{\infty} h_k[mM - n]z^{-m},$$
$$k = 0, 1, \ldots, N - 1, \quad n = 0, 1, \ldots, M - 1.$$

Similarly, the $M \times N$ synthesis polyphase matrix $\mathbf{R}(z)$ is defined as $[\mathbf{R}(z)]_{n, k} = R_{k, n}(z)$, where

$$R_{k,n}(z) = \sum_{m=-\infty}^{\infty} f_k[mM + n]z^{-m},$$
$$k = 0, 1, \ldots, N - 1, \quad n = 0, 1, \ldots, M - 1.$$

We have

$$\mathbf{E}(z) = \sum_{m=-\infty}^{\infty} \mathbf{E}_m z^{-m} \qquad \mathbf{R}(z) = \sum_{m=-\infty}^{\infty} \mathbf{R}_m z^{-m} \quad (4)$$

with

$$[\mathbf{E}_m]_{k, n} = h_k[mM - n]$$

and

$$[\mathbf{R}_m]_{n, k} = f_k[mM + n].$$

For an FB with PR and zero delay, we have $\hat{x}[n] = x[n]$ where $x[n]$ and $\hat{x}[n]$ denote the input and reconstructed signal, respectively. FB analysis and synthesis can here be interpreted

---

[3]Here $H_k(z) = \sum_{n=-\infty}^{\infty} h_k[n] z^{-n}$ denotes the z-transform of $h_k[n]$.

as a signal expansion [24]–[26]. The subband signals $v_k[m]$ can be written as the inner products

$$v_k[m] = \langle x, h_{k,m} \rangle,$$
$$\text{with } h_{k,m}[n] = h_k^*[mM - n], \ k = 0, 1, \ldots, N-1.$$

Furthermore, with the PR property, we have

$$x[n] = \hat{x}[n] = \sum_{k=0}^{N-1} \sum_{m=-\infty}^{\infty} v_k[m] f_{k,m}[n]$$
$$= \sum_{k=0}^{N-1} \sum_{m=-\infty}^{\infty} \langle x, h_{k,m} \rangle f_{k,m}[n],$$
$$\text{with} \quad f_{k,m}[n] = f_k[n - mM].$$

This shows that the FB corresponds to an expansion of the input signal $x[n]$ into the function set $\{f_{k,m}[n]\}$ with $k = 0, 1, \ldots, N-1$ and $-\infty < m < \infty$. Critically sampled FBs correspond to orthogonal or biorthogonal signal expansions [27], [25], whereas oversampled FBs correspond to redundant (overcomplete) expansions [25], [2], [1], [7], [28]. If $\{h_{k,m}[n]\}$ is a frame for $l^2(\mathbb{Z})$, we say that the FB provides a frame expansion. The frame bounds A and B or, equivalently, $\tilde{A} = 1/B$ and $\tilde{B} = 1/A$ determine important numerical properties of the FB [21], [1]. The subband signals $v_k[m] = \langle x, h_{k,m} \rangle$ of an FB providing a frame expansion satisfy [cf. (1)]

$$A\|x\|^2 \leq \sum_{k=0}^{N-1} \sum_{m=-\infty}^{\infty} |v_k[m]|^2 \leq B\|x\|^2 \qquad \forall \, x[n] \in l^2(\mathbb{Z})$$

with $0 < A \leq B < \infty$. It is shown in [1] and [9] that the (tightest possible) frame bounds A and B of an FB providing a frame expansion are given by the essential infimum and supremum, respectively, of the eigenvalues $\lambda_n(\theta)$ of the $M \times M$ matrix[4] $\mathbf{S}(e^{j2\pi\theta}) = \mathbf{E}^H(e^{j2\pi\theta})\mathbf{E}(e^{j2\pi\theta})$

$$A = \operatorname*{ess\,inf}_{\theta \in [0,1), n=0, 1, \ldots, M-1} \lambda_n(\theta)$$
$$B = \operatorname*{ess\,sup}_{\theta \in [0,1), n=0, 1, \ldots, M-1} \lambda_n(\theta). \tag{5}$$

2) Design Freedom in Oversampled FBs: An oversampled FB satisfies the PR condition $\hat{x}[n] = x[n]$ if and only if [1], [2]

$$\mathbf{R}(z)\mathbf{E}(z) = \mathbf{I}_M \tag{6}$$

where $\mathbf{I}_M$ is the $M \times M$ identity matrix. For analysis polyphase matrix $\mathbf{E}(z)$ given, the PR synthesis polyphase matrix $\mathbf{R}(z)$ is not uniquely determined: any solution of (6) can be written as [1], [9] (assuming $\operatorname{rank}\{\mathbf{E}(z)\} = M$ a.e.)

$$\mathbf{R}(z) = \hat{\mathbf{R}}(z) + \mathbf{U}(z)[\mathbf{I}_N - \mathbf{E}(z)\hat{\mathbf{R}}(z)]. \tag{7}$$

Here, $\hat{\mathbf{R}}(z)$ is the para-pseudo-inverse of $\mathbf{E}(z)$, which is a particular solution of (6) defined as[5]

$$\hat{\mathbf{R}}(z) = [\tilde{\mathbf{E}}(z)\mathbf{E}(z)]^{-1}\tilde{\mathbf{E}}(z) \tag{8}$$

and $\mathbf{U}(z)$ is an $M \times N$ matrix with arbitrary elements $[\mathbf{U}(z)]_{k,l}$ satisfying $|[\mathbf{U}(e^{j2\pi\theta})]_{k,l}| < \infty$. Choosing the PR synthesis FB

---

[4]The superscript H denotes conjugate transposition.

[5]Here, $\tilde{\mathbf{E}}(z) = \mathbf{E}^H(1/z^*)$ is the para-conjugate of $\mathbf{E}(z)$.

according to $\mathbf{R}(z) = \hat{\mathbf{R}}(z)$ corresponds to reconstruction using the dual frame [2], [1], [9].

This nonuniqueness of the PR synthesis FB corresponds to an increased design freedom (as compared to critically sampled FBs) [1], [9] that is a major advantage of oversampled FBs. Certain PR synthesis FBs have desirable properties (such as good frequency selectivity, linear phase, etc.) that may not be shared by the PR synthesis FB corresponding to the para-pseudo-inverse $\hat{\mathbf{R}}(z)$. Such properties are especially important in coding applications where the synthesis filters determine the perceptual impact of quantization errors. In the oversampled case, therefore, we can impose additional properties (besides PR) on the synthesis filters and perform an optimization over all PR synthesis FBs. Using the parameterization (7) or related ones [10], [1], [9], this can be done by means of an unconstrained optimization procedure since PR need not be incorporated via a side constraint. We note that this increased design freedom in oversampled FBs is similar to that in oversampled A/D conversion (see Section II-B). In Section II-C4, we shall show that again there exists a tradeoff between noise reduction and design freedom.

3) Noise Analysis in Oversampled FBs: We next investigate the sensitivity of oversampled FBs to (quantization) noise $q_k[m]$ added to the subband signals $v_k[m]$. The N-dimensional vector noise process

$$\mathbf{q}[m] \triangleq [q_0[m] \ q_1[m] \ \cdots \ q_{N-1}[m]]^T$$

is assumed wide-sense stationary and zero-mean. The $N \times N$ power spectral matrix of $\mathbf{q}[m]$ is

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

with the autocorrelation matrix $\mathbf{C}_q[l] = \mathcal{E}\{\mathbf{q}[m]\mathbf{q}^H[m-l]\}$, where $\mathcal{E}$ denotes the expectation operator.

It is convenient to redraw the FB in the polyphase domain as shown in Fig. 3 [24]. Here

$$\mathbf{x}(z) = [X_0(z) \ X_1(z) \ \cdots \ X_{M-1}(z)]^T$$

with

$$X_n(z) = \sum_{m=-\infty}^{\infty} x[mM + n] z^{-m}$$

$$\mathbf{q}(z) = \sum_{m=-\infty}^{\infty} \mathbf{q}[m] z^{-m}$$

and

$$\hat{\mathbf{x}}(z) = [\hat{X}_0(z) \ \hat{X}_1(z) \ \cdots \ \hat{X}_{M-1}(z)]^T$$

with

$$\hat{X}_n(z) = \sum_{m=-\infty}^{\infty} \hat{x}[mM + n] z^{-m}.$$

(Note that $\mathbf{q}(z)$ is just a notational aid since $\sum_{m=-\infty}^{\infty} \mathbf{q}[m] z^{-m}$ may not converge.) Assuming an arbitrary PR synthesis FB $\mathbf{R}(z)$, we have [see Fig. 3 and (6)]

$$\hat{\mathbf{x}}(z) = \mathbf{R}(z)[\mathbf{E}(z)\mathbf{x}(z) + \mathbf{q}(z)] = \mathbf{x}(z) + \mathbf{R}(z)\mathbf{q}(z).$$
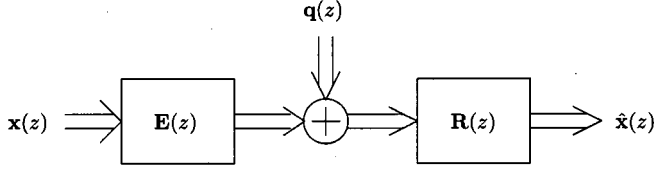
Fig. 3. Adding noise to the subband signals (polyphase-domain representation).

Hence, the reconstruction error vector is given by

$$\mathbf{e}(z) \overset{\Delta}{=} \hat{\mathbf{x}}(z) - \mathbf{x}(z) = \mathbf{R}(z)\mathbf{q}(z).$$

In the time domain, the reconstruction error $\mathbf{e}[n]$ is wide-sense stationary and zero-mean, with $M \times M$ power spectral matrix [24]

$$\mathbf{S}_e(z) = \mathbf{R}(z)\mathbf{S}_q(z)\tilde{\mathbf{R}}(z)$$

and variance[6]

$$\sigma_e^2 = \frac{1}{M} \int_{-1/2}^{1/2} \mathrm{Tr}\{\mathbf{R}(e^{j2\pi\theta})\mathbf{S}_q(e^{j2\pi\theta})\mathbf{R}^H(e^{j2\pi\theta})\}\, d\theta.$$

In particular, for uncorrelated white noise signals $q_k[m]$ (i.e., $q_k[m]$ and $q_{k'}[m']$ are uncorrelated for $k \neq k'$ and also for $m \neq m'$) with identical variances $\sigma_q^2 = \mathcal{E}\{|q_k[m]|^2\}$, i.e., $\mathbf{C}_q[l] = \sigma_q^2 \mathbf{I}_N \delta[l]$ and $\mathbf{S}_q(z) = \sigma_q^2 \mathbf{I}_N$, the reconstruction error variance simplifies to

$$\sigma_e^2 = \frac{\sigma_q^2}{M} \int_{-1/2}^{1/2} \mathrm{Tr}\{\mathbf{R}(e^{j2\pi\theta})\mathbf{R}^H(e^{j2\pi\theta})\}\, d\theta. \qquad (9)$$

This result permits an interesting frame-theoretic interpretation. Assuming reconstruction using the dual frame, i.e., $\mathbf{R}(z) = \hat{\mathbf{R}}(z)$ [see (8)], and using

$$\mathrm{Tr}\{\hat{\mathbf{R}}(e^{j2\pi\theta})\hat{\mathbf{R}}^H(e^{j2\pi\theta})\} = \sum_{n=0}^{M-1} \tilde{\lambda}_n(\theta)$$

with $\tilde{\lambda}_n(\theta)$ denoting the eigenvalues of the matrix

$$\mathbf{S}^{-1}(e^{j2\pi\theta}) = \hat{\mathbf{R}}(e^{j2\pi\theta})\hat{\mathbf{R}}^H(e^{j2\pi\theta})$$

it follows from

$$\tilde{A} = \underset{\theta \in [0, 1), n = 0, 1, \ldots, M-1}{\mathrm{ess\,inf}} \tilde{\lambda}_n(\theta)$$

$$\tilde{B} = \underset{\theta \in [0, 1), n = 0, 1, \ldots, M-1}{\mathrm{ess\,sup}} \tilde{\lambda}_n(\theta)$$

[cf. (5)] that

$$\tilde{A} \leq \frac{\sigma_e^2}{\sigma_q^2} \leq \tilde{B} \qquad (10)$$

i.e., the reconstruction error variance σ_e^2 is bounded in terms of the frame bounds $\tilde{A}$, $\tilde{B}$, and the subband noise variance σ_q^2. For normalized analysis filters, i.e., $\|h_k\| = 1$ for $k = 0, 1, \ldots, N-1$, it can be shown [1], [9] that $\tilde{A} \leq \frac{1}{K} \leq \tilde{B}$ where $K = \frac{N}{M}$ is the oversampling factor. For a paraunitary

[6]Here, $\mathrm{Tr}\{\cdot\}$ denotes the trace of a matrix.

FB (corresponding to a tight frame expansion [2], [1], [9]) with normalized analysis filters, we have $\tilde{A} = \tilde{B} = \frac{1}{K}$ [1], [9] and (10) yields

$$\frac{\sigma_e^2}{\sigma_q^2} = \frac{1}{K}. \qquad (11)$$

Similarly, for an FB corresponding to a snug frame, $\tilde{A} \approx \tilde{B}$ and thus $\frac{\sigma_e^2}{\sigma_q^2} \approx \frac{1}{K}$. Hence, paraunitary FBs or FBs providing snug frame expansions are desirable since it is guaranteed that small errors in the subband signals will result in small reconstruction errors. This is important in signal coding applications involving quantization errors and in signal processing applications involving intentional modifications of the subband signals. Since in the critically sampled case $\sigma_c^2 \overset{\Delta}{=} \sigma_e^2|_{K=1} = \sigma_q^2$, (11) can be rewritten as

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

Thus, for a paraunitary FB, the reconstruction error variance is inversely proportional to the oversampling factor $K = \frac{N}{M}$, which means that more oversampling entails better noise reduction. Such a "$1/K$ behavior" has been observed in Section II-B for oversampled A/D conversion, which was shown to correspond to a tight frame expansion. Since also a paraunitary FB corresponds to a tight frame expansion, its $1/K$ behavior does not come as a surprise. A $1/K$ behavior has furthermore been observed for tight frames in finite-dimensional spaces [21], [29] and for reconstruction from a finite set of Weyl–Heisenberg (Gabor) or wavelet coefficients [21], [30]. Under additional conditions, a $1/K^2$ behavior has been demonstrated for Weyl–Heisenberg frames in [30]. In [5], [31]–[34], based on a deterministic quantization noise model, a nonlinear set-theoretic estimation method is used to achieve a $1/K^2$ behavior for frames of sinc functions (A/D conversion) and for Weyl–Heisenberg frames. In Sections III and IV, we shall propose oversampled predictive subband coders that are based on a stochastic quantization noise model. These subband coders also achieve a $1/K^2$ performance and in some cases can do even better.

Unfortunately, the assumption of uncorrelated white noise is not justified for $K > 1$. For arbitrary (possibly correlated and/or nonwhite) noise with power spectral matrix $\mathbf{S}_q(z)$, a noise whitening approach can be employed. Using the factorization $\mathbf{S}_q(z) = \mathbf{S}_q^{1/2}(z)\tilde{\mathbf{S}}_q^{1/2}(z)$ (which is guaranteed to exist [35], [36]), it is easily seen that the system depicted in Fig. 3 is equivalent to a system with noise power spectral matrix $\mathbf{S}_q(z) = \mathbf{I}_N$ (corresponding to uncorrelated white noise with equal variances $\sigma_q^2 = 1$ in all channels) if $\mathbf{E}(z)$ and $\mathbf{R}(z)$ are replaced by

$$\mathbf{E}^{(q)}(z) = \mathbf{S}_q^{-1/2}(z)\mathbf{E}(z)$$

and

$$\mathbf{R}^{(q)}(z) = \mathbf{R}(z)\mathbf{S}_q^{1/2}(z)$$

respectively. The double inequality (10) continues to hold if the frame bounds in (10) are replaced by the frame bounds of the FB $\mathbf{R}^{(q)}(z)$. Similarly, (11) continues to hold if $\mathbf{R}^{(q)}(z)$ is paraunitary.
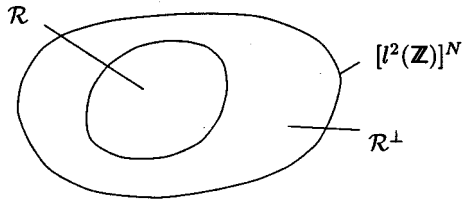
Fig. 4.   Range space of analysis FB and its orthogonal complement.

4) Noise Reduction Versus Design Freedom in Oversampled FBs: We shall now establish a subspace-based interpretation of noise reduction in oversampled FBs that is analogous to the interpretation given for oversampled A/D conversion in Section II-B. Let us define the FB analysis operator $\boldsymbol{T}$ that assigns to each input signal $x[n]$ the subband vector signal $\mathbf{v}[m] = [v_0[m]\ v_1[m]\ \cdots\ v_{N-1}[m]]^T$. The orthogonal projection operator on the range space $\mathcal{R} \subseteq [l^2(\mathbb{Z})]^N$ of $\boldsymbol{T}$ is given by $\boldsymbol{P}_{\mathcal{R}} = \boldsymbol{T}\boldsymbol{S}^{-1}\boldsymbol{T}^*$ [21]. Since the analysis operator $\boldsymbol{T}$, its adjoint $\boldsymbol{T}^*$, and the frame operator $\boldsymbol{S}$ are represented by the matrices $\mathbf{E}(z)$, $\tilde{\mathbf{E}}(z)$, and $\tilde{\mathbf{E}}(z)\mathbf{E}(z)$, respectively [1], [9], the matrix representation of $\boldsymbol{P}_{\mathcal{R}}$ is

$$\mathbf{P}_{\mathcal{R}}(z) = \mathbf{E}(z)[\tilde{\mathbf{E}}(z)\mathbf{E}(z)]^{-1}\tilde{\mathbf{E}}(z) = \mathbf{E}(z)\hat{\mathbf{R}}(z).$$

Similarly,

$$\mathbf{P}_{\perp}(z) = \mathbf{I}_N - \mathbf{P}_{\mathcal{R}}(z) = \mathbf{I}_N - \mathbf{E}(z)\hat{\mathbf{R}}(z)$$

is the matrix representation of the orthogonal projection operator $\boldsymbol{P}_{\mathcal{R}^\perp} = \boldsymbol{I} - \boldsymbol{P}_{\mathcal{R}}$ on the orthogonal complement $\mathcal{R}^\perp$ of $\mathcal{R}$ (see Fig. 4).

Let us consider an oversampled FB with $\mathbf{R}(z)$ chosen as in (7)

$$\mathbf{R}(z) = \hat{\mathbf{R}}(z) + \mathbf{U}(z)[\mathbf{I}_N - \mathbf{E}(z)\hat{\mathbf{R}}(z)] = \hat{\mathbf{R}}(z) + \mathbf{U}(z)\mathbf{P}_{\perp}(z) \tag{12}$$

so that PR is satisfied. With (12), the reconstruction error can be decomposed as

$$\begin{aligned} \mathbf{e}(z) &= \mathbf{R}(z)\mathbf{q}(z) = [\hat{\mathbf{R}}(z) + \mathbf{U}(z)\mathbf{P}_{\perp}(z)]\mathbf{q}(z) \\ &= \mathbf{e}_{\mathcal{R}}(z) + \mathbf{e}_{\perp}(z) \end{aligned}$$

where

$$\mathbf{e}_{\mathcal{R}}(z) = \hat{\mathbf{R}}(z)\mathbf{q}(z) \quad \text{and} \quad \mathbf{e}_{\perp}(z) = \mathbf{U}(z)\mathbf{P}_{\perp}(z)\mathbf{q}(z). \tag{13}$$

Since $\hat{\mathbf{R}}(z)\mathbf{P}_{\mathcal{R}}(z) = \hat{\mathbf{R}}(z)$, we can equivalently write

$$\mathbf{e}_{\mathcal{R}}(z) = \hat{\mathbf{R}}(z)\mathbf{P}_{\mathcal{R}}(z)\mathbf{q}(z)$$

which shows that $\mathbf{e}_{\mathcal{R}}(z)$ is reconstructed from the subband noise component $\mathbf{P}_{\mathcal{R}}(z)\mathbf{q}(z)$ that lies in $\mathcal{R}$. Similarly,

$$\mathbf{e}_{\perp}(z) = \mathbf{U}(z)\mathbf{P}_{\perp}(z)\mathbf{q}(z)$$

is reconstructed from the subband noise component $\mathbf{P}_{\perp}(z)\mathbf{q}(z)$ that lies in $\mathcal{R}^\perp$.

For subband noise signals $q_k[m]$ that are uncorrelated and white (i.e., $\mathbf{S}_q(z) = \sigma_q^2\mathbf{I}_N$), it follows from the orthogonality of the spaces $\mathcal{R}$ and $\mathcal{R}^\perp$ that the error components $\mathbf{e}_{\mathcal{R}}[n]$ and

$\mathbf{e}_{\perp}[n]$ are uncorrelated [37]. Hence, their variances, denoted, respectively, $\sigma_{\mathcal{R}}^2$ and $\sigma_{\perp}^2$, can simply be added to yield the overall reconstruction error variance

$$\sigma_e^2 = \sigma_{\mathcal{R}}^2 + \sigma_{\perp}^2. \tag{14}$$

This relation leads to the following result.

Proposition 1: For an oversampled PR FB with uncorrelated and white subband noise signals with equal variance σ_q^2 in all channels, the synthesis FB $\mathbf{R}(z)$ minimizing the reconstruction error variance among all PR synthesis FBs (i.e., among all $\mathbf{R}(z)$ satisfying $\mathbf{R}(z)\mathbf{E}(z) = \mathbf{I}_M$) is the para-pseudo-inverse $\hat{\mathbf{R}}(z) = [\tilde{\mathbf{E}}(z)\mathbf{E}(z)]^{-1}\tilde{\mathbf{E}}(z)$ of $\mathbf{E}(z)$, and the resulting minimum reconstruction error variance is

$$\sigma_{e,\,\min}^2 = \sigma_{\mathcal{R}}^2 = \frac{\sigma_q^2}{M}\int_{-1/2}^{1/2} \text{Tr}\,\{\hat{\mathbf{R}}(e^{j2\pi\theta})\hat{\mathbf{R}}^H(e^{j2\pi\theta})\}\,d\theta. \tag{15}$$

Proof: According to (13), the variance component $\sigma_{\mathcal{R}}^2$ does not depend on the parameter matrix $\mathbf{U}(z)$, and thus it does not depend on the particular $\mathbf{R}(z)$ chosen. On the other hand, the "orthogonal" variance component $\sigma_{\perp}^2$ in (13) and (14) is an additional variance that is zero for all $\mathbf{q}(z)$ if and only if $\mathbf{U}(z)\mathbf{P}_{\perp}(z) \equiv \mathbf{0}$, which yields $\mathbf{R}(z) = \hat{\mathbf{R}}(z)$. The expression for $\sigma_{e,\,\min}^2$ in (15) is obtained from (9). $\square$

Hence, using $\hat{\mathbf{R}}(z)$ will suppress all noise components orthogonal to the range space $\mathcal{R}$, whereas any other PR synthesis FB (possibly with desirable properties such as improved frequency selectivity, etc.) will lead to an additional error variance $\sigma_{\perp}^2$ since also noise components orthogonal to $\mathcal{R}$ are passed to the FB output. Thus, similar to oversampled A/D conversion (see Section II-B), there exists a tradeoff between noise reduction and design freedom. Even though in the FB case the spaces $\mathcal{R}$ and $\mathcal{R}^\perp$ no longer correspond to frequency bands, the same interpretations and conclusions as in oversampled A/D conversion apply.

In the case of correlated and/or colored noise signals, the above results continue to hold if the matrices $\mathbf{E}(z)$ and $\mathbf{R}(z)$ are replaced by

$$\mathbf{E}^{(q)}(z) = \mathbf{S}_q^{-1/2}(z)\mathbf{E}(z)$$

and

$$\mathbf{R}^{(q)}(z) = \mathbf{R}(z)\mathbf{S}_q^{1/2}(z)$$

respectively (cf. Section II-C3). In particular, for a given analysis FB with polyphase matrix $\mathbf{E}(z)$ and for a given noise power spectral matrix $\mathbf{S}_q(z)$, the synthesis FB minimizing the reconstruction error variance is defined by [cf. (8)]

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

which yields

$$\begin{aligned} \mathbf{R}(z) &= \hat{\mathbf{R}}^{(q)}(z)\mathbf{S}_q^{-1/2}(z) \\ &= [\tilde{\mathbf{E}}(z)\mathbf{S}_q^{-1}(z)\mathbf{E}(z)]^{-1}\tilde{\mathbf{E}}(z)\mathbf{S}_q^{-1}(z). \end{aligned}$$

We finally note that the tradeoff between noise reduction and design freedom discussed above is not restricted to redundant *shift-invariant* signal expansions (such as oversampled A/D conversion and oversampled FBs) but is inherent in general
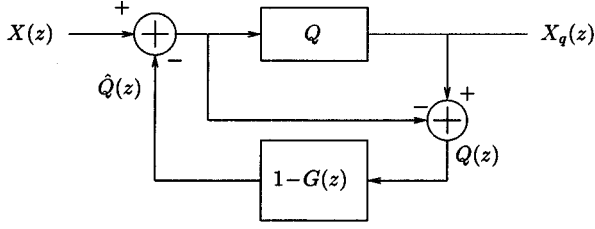
Fig. 5. Noise-shaping A/D converter. The quantizer (box labeled Q) adds quantization noise $q[n] \leftrightarrow Q(z)$.

redundant representations. In general, more oversampling tends to result in better noise reduction since the range space $\mathcal{R}$—and thus also the fixed noise component $\sigma_{\mathcal{R}}^2$—becomes "smaller."

III. OVERSAMPLED NOISE-SHAPING (NOISE PREDICTIVE) SUBBAND CODERS

This section introduces a method for noise reduction in oversampled FBs that is based on noise prediction. The resulting noise-shaping (noise-predictive) subband coders can be viewed as extensions of oversampled noise-shaping A/D converters, which will be reviewed first.

A. Oversampled Noise-Shaping A/D Converters

Noise feedback coding has found widespread use in A/D conversion [11]–[14]. A noise-shaping coder, modeled as an entirely discrete-time system [11], is shown in Fig. 5. Here, $X(z)$ denotes the z-transform of the input signal (the oversampled version of the analog signal $x(t)$) and

$$G(z) = 1 - \sum_{l=1}^{L} g[l] z^{-l}$$

is the noise-shaping filter of order L. The quantization noise estimate is obtained as

$$\hat{q}[n] = \sum_{l=1}^{L} g[l] q[n-l] \qquad \hat{Q}(z) = [1 - G(z)] Q(z)$$

where $Q(z)$ is the z-transform of the quantization noise sequence $q[n]$. The noise-shaping system is designed such that $\hat{q}[n]$ optimally estimates or predicts the *in-band* component of the current quantization noise sample $q[n]$ based on the past L noise samples $q[n-1], q[n-2], \ldots, q[n-L]$ [11]. In this sense, noise-shaping coders can be interpreted as noise-predictive coders. Equivalently, the goal is to minimize the in-band component of $q[n] - \hat{q}[n]$, i.e., the component lying in $\mathcal{R}$, the range space of the analysis/sampling operator $\boldsymbol{T}$ from Section II-B. The out-of-band component (lying in $\mathcal{R}^{\perp}$) is subsequently attenuated by the reconstruction low-pass filter in the decoder (not shown in Fig. 5).

The signal presented to the quantizer is $x[n] - \hat{q}[n]$, which results in an effective noise reduction while leaving the A/D converter's dynamic range unchanged (this is fundamentally different from a signal predictive coder discussed in Section IV-A). Since the in-band noise power is reduced relative to the quantization noise power of the A/D converter, it is possible to increase

the quantization step size and thereby reduce the overall converter complexity. From Fig. 5, it follows that the coder output signal is given by $x_q[n] = x[n] + q[n] - \hat{q}[n]$ or, equivalently, $X_q(z) = X(z) + G(z)Q(z)$. Note that $x[n]$ is not affected by the noise-shaping system, whereas $q[n]$ is passed through $G(z)$. Hence, the reconstruction error is

$$E(z) = X_q(z) - X(z) = G(z)Q(z).$$

We now provide a frame-theoretic, subspace-based interpretation of noise shaping which will motivate our results in Section III-B. In a noise-shaping coder, the quantization noise is effectively moved to a high-frequency band which is then attenuated by the low-pass reconstruction filter (see Fig. 6). Equivalently (recall from Section II-B that the signal band $-\frac{1}{2K} \le \theta \le \frac{1}{2K}$ corresponds to the range space $\mathcal{R}$ of the analysis/sampling operator $\boldsymbol{T}$), the quantization noise is moved to the orthogonal complement $\mathcal{R}^{\perp}$ of $\mathcal{R}$. Reconstruction using the dual frame, i.e., ideal low-pass filtering with minimum bandwidth, then performs an orthogonal projection onto $\mathcal{R}$ that suppresses all noise components in[7] $\mathcal{R}^{\perp}$.

B. Noise Shaping in Oversampled FBs

We recall from Section II-C that the subband signals $v_k[m]$ in an oversampled PR FB constitute a redundant representation of the FB input signal $x[n]$, with the range space $\mathcal{R}$ of the FB analysis operator $\boldsymbol{T}$ being a subspace of $[l^2(\mathbb{Z})]^N$. This analogy to oversampled A/D converters again suggests the application of noise shaping. The goal is to exploit the redundancy of the subband signal samples in order to push the quantization noise to the orthogonal complement space $\mathcal{R}^{\perp}$. The noise-shaping subband coders introduced here combine the advantages of subband coding with those of noise or error feedback coding.

1) The Noise-Shaping Subband Coder: We propose a multi-input multi-output (MIMO) noise-shaping system, represented by an $N \times N$ transfer matrix $\mathbf{G}(z)$, that is cradled between the analysis FB $\mathbf{E}(z)$ and the synthesis FB $\mathbf{R}(z)$ as depicted in Fig. 7. The quantization noise $\mathbf{q}(z)$ is fed back through the noise-shaping system $\mathbf{I}_N - \mathbf{G}(z)$ to yield the quantization noise estimate $\hat{\mathbf{q}}(z) = [\mathbf{I}_N - \mathbf{G}(z)]\mathbf{q}(z)$, which is then subtracted from the subband signal vector $\mathbf{v}(z) = \mathbf{E}(z)\mathbf{x}(z)$. Assuming an FB with PR (i.e., $\mathbf{R}(z)\mathbf{E}(z) = \mathbf{I}_M$), the reconstructed signal is obtained as

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

It follows that the reconstruction error equals $\mathbf{q}(z)$ filtered by $\mathbf{G}(z)$ and then by the synthesis FB $\mathbf{R}(z)$

$$\mathbf{e}(z) \stackrel{\Delta}{=} \mathbf{x}_q(z) - \mathbf{x}(z) = \mathbf{R}(z)\mathbf{G}(z)\mathbf{q}(z). \tag{16}$$

Hence, the $M \times M$ power spectral density matrix of $\mathbf{e}(z)$ is [24]

$$\mathbf{S}_e(z) = \mathbf{R}(z)\mathbf{G}(z)\mathbf{S}_q(z)\tilde{\mathbf{G}}(z)\tilde{\mathbf{R}}(z)$$

---

[7]From this interpretation, it appears that the optimal noise-shaping filter $G(z)$ would be the ideal high-pass filter with passband $\frac{1}{2K} \le |\theta| \le 1/2$, since this filter projects the noise onto $\mathcal{R}^{\perp}$ and after reconstruction no noise would be left. However, this filter is not realizable and would lead to a noncausal system $1 - G(z)$ that cannot operate in a feedback loop.
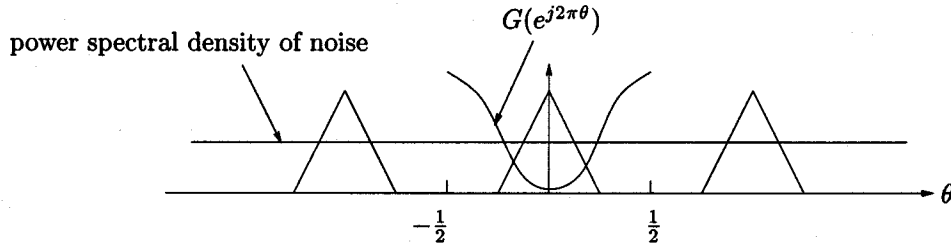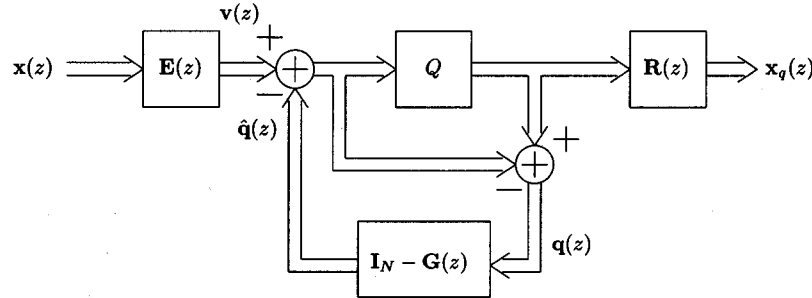
Fig. 6.   Typical noise-shaping filter.

Fig. 7.   Oversampled noise-shaping subband coder (polyphase-domain representation).

and the reconstruction error variance is

$$\sigma_e^2 = \frac{1}{M} \int_{-1/2}^{1/2} \mathrm{Tr}\{\mathbf{R}(e^{j2\pi\theta})\mathbf{G}(e^{j2\pi\theta})\mathbf{S}_q(e^{j2\pi\theta})$$
$$\cdot\mathbf{G}^H(e^{j2\pi\theta})\mathbf{R}^H(e^{j2\pi\theta})\} \, d\theta. \quad (17)$$

The optimum noise-shaping system minimizes σ_e^2. Without further constraints, the noise could be completely removed ($\sigma_e^2 = 0$) using $\mathbf{G}(z) = \mathbf{I}_N - \mathbf{E}(z)\mathbf{R}(z)$. Indeed, inserting this into (16), it follows with $\mathbf{R}(z)\mathbf{E}(z) = \mathbf{I}_M$ that $\mathbf{e}(z) \equiv \mathbf{0}$. In the case of reconstruction using the dual frame, i.e., $\mathbf{R}(z) = \hat{\mathbf{R}}(z)$, this ideal noise shaper is

$$\mathbf{G}(z) = \mathbf{I}_N - \mathbf{E}(z)\hat{\mathbf{R}}(z) = \mathbf{P}_\perp(z)$$

the orthogonal projection operator on the orthogonal complement $\mathcal{R}^\perp$ of the analysis FB's range space $\mathcal{R}$ (see Section II-C4). Thus, the ideal noise shaper projects the noise onto $\mathcal{R}^\perp$, and the projected noise is then suppressed by the synthesis FB $\hat{\mathbf{R}}(z)$ that involves an orthogonal projection onto $\mathcal{R}$. This is similar to oversampled A/D conversion where the theoretically ideal noise-shaping filter was seen to be an ideal high-pass (projection) filter.

Unfortunately, this ideal noise shaper is inadmissible since it is not causal and therefore cannot operate in a feedback loop. Hence, we hereafter constrain $\mathbf{G}(z)$ to be a causal finite-impulse response (FIR) MIMO system of the form

$$\mathbf{G}(z) = \mathbf{I}_N - \sum_{l=1}^{L} \mathbf{G}_l z^{-l} \quad (18)$$

resulting in a strictly causal feedback loop system

$$\mathbf{I}_N - \mathbf{G}(z) = \sum_{l=1}^{L} \mathbf{G}_l z^{-l}.$$

Here L denotes the order of the noise-shaping system. The quantization noise estimate $\hat{\mathbf{q}}[n]$ now becomes

$$\hat{\mathbf{q}}[n] = \sum_{l=1}^{L} \mathbf{G}_l \mathbf{q}[n - l].$$

The purpose of the noise-shaping system $\mathbf{I}_N - \mathbf{G}(z)$ is to estimate or predict the quantization noise component that will be passed by the synthesis FB $\mathbf{R}(z)$, based on the past noise samples $\mathbf{q}[n - 1], \mathbf{q}[n - 2], \ldots, \mathbf{q}[n - L]$. In the case of reconstruction using the dual frame, i.e., $\mathbf{R}(z) = \hat{\mathbf{R}}(z)$, the synthesis FB passes everything in the range space $\mathcal{R}$. In this case, the noise-shaping system has to predict the *in-range* component of $\mathbf{q}[n]$, i.e., the quantization noise component in $\mathcal{R}$. Equivalently, the optimum noise-shaping system pushes the quantization noise to the orthogonal complement space $\mathcal{R}^\perp$ that is subsequently suppressed by $\hat{\mathbf{R}}(z)$.

2) Calculation of the Optimum Noise-Shaping System: We now derive the optimal noise-shaping system, i.e., the matrices $\mathbf{G}_l$ minimizing the reconstruction error variance σ_e^2 in (17). We shall first assume uncorrelated white quantization noise with equal noise variance in all channels, i.e., $\mathbf{C}_q[l] = \sigma_q^2\mathbf{I}_N\delta[l]$ and $\mathbf{S}_q(z) = \sigma_q^2\mathbf{I}_N$. Inserting (18) and (4) into (17), it follows after lengthy but straightforward manipulations that

$$\sigma_e^2 = \frac{\sigma_q^2}{M} \mathrm{Tr}\left\{\boldsymbol{\Gamma}_0 - \sum_{l=1}^{L}[\boldsymbol{\Gamma}_l\mathbf{G}_l^T + \boldsymbol{\Gamma}_l^T\mathbf{G}_l] + \sum_{m=1}^{L}\mathbf{G}_m^T\sum_{l=1}^{L}\boldsymbol{\Gamma}_{m-l}\mathbf{G}_l\right\}$$
$$(19)$$

with the $N \times N$ matrices

$$\boldsymbol{\Gamma}_l = \sum_{m=-\infty}^{\infty} \mathbf{R}_m^T\mathbf{R}_{m+l}$$

that satisfy $\mathbf{\Gamma}_{-l}^T = \mathbf{\Gamma}_l$. Here, the FB has been assumed real-valued and we recall that $\mathbf{R}_m$ was defined in (4). Setting $\frac{\partial \sigma_e^2}{\partial \mathbf{G}_i} = \mathbf{0}$ for $i = 1, 2, \ldots, L$ and using the matrix derivative rules (see [38, Sec. 5.3]),

$$\frac{\partial}{\partial \mathbf{G}_i} \text{Tr}\{\mathbf{A}\mathbf{G}_i^T\} = \frac{\partial}{\partial \mathbf{G}_i} \text{Tr}\{\mathbf{A}^T \mathbf{G}_i\} = \mathbf{A}$$

and

$$\frac{\partial}{\partial \mathbf{G}_i} \text{Tr}\{\mathbf{G}_i^T \mathbf{A} \mathbf{G}_i\} = \mathbf{A}^T \mathbf{G}_i + \mathbf{A} \mathbf{G}_i$$

yields

$$\sum_{l=1}^{L} \mathbf{\Gamma}_{i-l} \mathbf{G}_l = \mathbf{\Gamma}_i, \qquad \text{for} \quad i = 1, 2, \ldots, L \qquad (20)$$

or, equivalently,

$$\begin{bmatrix} \mathbf{\Gamma}_0 & \mathbf{\Gamma}_{-1} & \ldots & \mathbf{\Gamma}_{-(L-1)} \\ \mathbf{\Gamma}_1 & \mathbf{\Gamma}_0 & \ldots & \mathbf{\Gamma}_{-(L-2)} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{\Gamma}_{L-1} & \mathbf{\Gamma}_{L-2} & \ldots & \mathbf{\Gamma}_0 \end{bmatrix} \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \vdots \\ \mathbf{G}_L \end{bmatrix} = \begin{bmatrix} \mathbf{\Gamma}_1 \\ \mathbf{\Gamma}_2 \\ \vdots \\ \mathbf{\Gamma}_L \end{bmatrix}. \quad (21)$$

This linear system of equations has block Toeplitz form and can be solved efficiently using the multichannel Levinson recursion [39]. The maximum possible system order L is determined by the rank of the block matrix in (21), which, in turn, depends on the synthesis filters. Inserting (20) into (19), the minimum reconstruction error variance is obtained as

$$\sigma_{e,\min}^2 = \frac{\sigma_q^2}{M} \text{Tr}\left\{\mathbf{\Gamma}_0 - \sum_{l=1}^{L} \mathbf{\Gamma}_l^T \mathbf{G}_{l,\text{opt}}\right\} \qquad (22)$$

where $\mathbf{G}_{l,\text{opt}}$ denotes the solution of (20) or (21).

The paraunitary case merits special attention. For a paraunitary FB with normalized, real-valued analysis filters, we have $\mathbf{R}(z) = \frac{1}{K} \tilde{\mathbf{E}}(z)$ and thus $\mathbf{R}_m = \frac{1}{K} \mathbf{E}_{-m}^T$, with $\mathbf{E}_m$ defined in (4). This implies

$$\mathbf{\Gamma}_l = \frac{1}{K^2} \sum_{m=-\infty}^{\infty} \mathbf{E}_m \mathbf{E}_{m-l}^T. \qquad (23)$$

If the analysis filters $h_k[n]$ are furthermore causal and of finite length $L_h = JM$ (with some $J \in \mathbb{N}$), we have $\mathbf{E}_m = \mathbf{0}$ for $m < 0$ and $m > J$ and hence

$$\mathbf{\Gamma}_l = \frac{1}{K^2} \sum_{m=0}^{J} \mathbf{E}_m \mathbf{E}_{m-l}^T$$

which implies $\mathbf{\Gamma}_l = \mathbf{0}$ for $|l| > J$. (In the nondecimated case $M = 1$, we have $\mathbf{E}_m = \mathbf{0}$ for $m < 0$ and $m > J - 1$ and hence

$$\mathbf{\Gamma}_l = \frac{1}{K^2} \sum_{m=0}^{J-1} \mathbf{E}_m \mathbf{E}_{m-l}^T$$

which implies $\mathbf{\Gamma}_l = \mathbf{0}$ for $|l| > J - 1$.) Hence, the block Toeplitz matrix in (21) will become increasingly banded for small analysis filter length.

For a paraunitary FB with normalized analysis filters ($\|h_k\| = 1$), we have $\text{Tr}\{\mathbf{\Gamma}_0\} = \frac{M}{K}$. Hence, (22) becomes

$$\sigma_{e,\min}^2 = \frac{\sigma_q^2}{M}\left[\frac{M}{K} - \text{Tr}\left\{\sum_{l=1}^{L} \mathbf{\Gamma}_l^T \mathbf{G}_{l,\text{opt}}\right\}\right]$$

$$= \sigma_e^2|_{L=0} - \frac{\sigma_q^2}{M} \text{Tr}\left\{\sum_{l=1}^{L} \mathbf{\Gamma}_l^T \mathbf{G}_{l,\text{opt}}\right\} \qquad (24)$$

where $\sigma_e^2|_{L=0} = \frac{\sigma_q^2}{K}$ is the reconstruction error variance obtained without noise shaping [see (11)].

We finally extend our results to the general case of correlated and/or colored quantization noise. Inserting the factorization $\mathbf{S}_q(z) = \mathbf{S}_q^{1/2}(z)\tilde{\mathbf{S}}_q^{1/2}(z)$ (cf. Section II-C3) into (17), we obtain

$$\sigma_e^2 = \frac{1}{M} \int_{-1/2}^{1/2} \text{Tr}\Big\{\mathbf{R}(e^{j2\pi\theta})\mathbf{G}'(e^{j2\pi\theta})$$

$$\mathbf{G}'^H(e^{j2\pi\theta})\mathbf{R}^H(e^{j2\pi\theta})\Big\} d\theta$$

$$\text{with} \quad \mathbf{G}'(z) = \mathbf{G}(z)\mathbf{S}_q^{1/2}(z). \qquad (25)$$

Comparing (25) with (17), we see that σ_e^2 is minimized if $\mathbf{G}'(z)$ is the optimum noise-shaping system for $\mathbf{S}_q(z) = \mathbf{I}_N$, i.e., for uncorrelated white noise with equal variances $\sigma_q^2 = 1$ in all channels. This system, denoted $\mathbf{G}'_{\text{opt}}(z)$, can be calculated as explained above. The optimum noise-shaping system for correlated and/or colored quantization noise is then obtained as

$$\mathbf{G}_{\text{opt}}(z) = \mathbf{G}'_{\text{opt}}(z)\mathbf{S}_q^{-1/2}(z).$$

3) Constrained Optimum Noise Shaping: The computational complexity of oversampled noise-shaping subband coders can be reduced by restricting $\mathbf{G}(z)$ to exploit only intrachannel dependencies or to exploit interchannel dependencies only between neighboring channels. Especially the latter strategy can be expected to perform well if the analysis filters are well localized in frequency so that only the transfer functions of neighboring channels are overlapping significantly. In the following, we restrict ourselves to uncorrelated and white noise for simplicity.

We shall first calculate the optimum intrachannel noise-shaping system (i.e., there are *separate* noise-shaping systems in the individual subchannels). Here $\mathbf{G}(z)$ and the matrices $\mathbf{G}_l$ are diagonal. Specializing (19) to diagonal $\mathbf{G}_l$, we obtain

$$\sigma_e^2 = \frac{\sigma_q^2}{M} \sum_{i=0}^{N-1}\left[\gamma_{i,i}^{(0)} - 2\sum_{l=1}^{L} \gamma_{i,i}^{(l)} g_{i,i}^{(l)} + \sum_{m=1}^{L} g_{i,i}^{(m)} \sum_{l=1}^{L} \gamma_{i,i}^{(m-l)} g_{i,i}^{(l)}\right] \qquad (26)$$

where $g_{i,j}^{(l)} \triangleq [\mathbf{G}_l]_{i,j}$ and $\gamma_{i,j}^{(l)} \triangleq [\mathbf{\Gamma}_l]_{i,j}$. Setting the derivatives of σ_e^2 with respect to $g_{i,i}^{(l)}$ ($i = 0, 1, \ldots, N-1, l = 1, 2, \ldots, L$) equal to zero, we obtain the following $NL \times L$ Toeplitz systems of equations:

$$\sum_{m=1}^{L} \gamma_{i,i}^{(l-m)} g_{i,i}^{(m)} = \gamma_{i,i}^{(l)},$$

$$l = 1, 2, \ldots, L, \quad i = 0, 1, \ldots, N-1 \quad (27)$$

or, briefly, $\mathbf{A}_i\mathbf{g}_i = \boldsymbol{\gamma}_i$ for $i = 0, 1, \ldots, N-1$, where $[\mathbf{A}_i]_{l,m} = \gamma_{i,i}^{(l-m)}$, $[\mathbf{g}_i]_l = g_{i,i}^{(l)}$, and $[\boldsymbol{\gamma}_i]_l = \gamma_{i,i}^{(l)}$. Inserting (27) into (26) and using $\gamma_{i,i}^{(m-l)} = \gamma_{i,i}^{(l-m)}$, we obtain the minimum reconstruction error variance as

$$\sigma_{e,\min}^2 = \frac{\sigma_q^2}{M} \sum_{i=0}^{N-1} \left[\gamma_{i,i}^{(0)} - \sum_{l=1}^{L} \gamma_{i,i}^{(l)} g_{i,i;\mathrm{opt}}^{(l)} \right] \qquad (28)$$

where $g_{i,i;\mathrm{opt}}^{(l)}$ denotes the solution of (27).

We next consider the noise-shaping system that exploits only neighboring channel dependencies. Here, only the main, first upper, and first lower diagonals of $\mathbf{G}_l$ may be nonzero. Hence, setting the derivatives of σ_e^2 with respect to $g_{i,i+1}^{(l)}$, $g_{i,i}^{(l)}$, and $g_{i,i-1}^{(l)}$ equal to zero, we obtain the system of equations

$$\gamma_{i-1,i}^{(l)} = \sum_{m=1}^{L} \left[\gamma_{i-1,i-1}^{(l-m)} g_{i-1,i}^{(m)} + \gamma_{i-1,i}^{(l-m)} g_{i,i}^{(m)} + \gamma_{i-1,i+1}^{(l-m)} g_{i+1,i}^{(m)} \right]$$

$$\gamma_{i,i}^{(l)} = \sum_{m=1}^{L} \left[\gamma_{i,i-1}^{(l-m)} g_{i-1,i}^{(m)} + \gamma_{i,i}^{(l-m)} g_{i,i}^{(m)} + \gamma_{i,i+1}^{(l-m)} g_{i+1,i}^{(m)} \right]$$

$$\gamma_{i+1,i}^{(l)} = \sum_{m=1}^{L} \left[\gamma_{i+1,i-1}^{(l-m)} g_{i-1,i}^{(m)} + \gamma_{i+1,i}^{(l-m)} g_{i,i}^{(m)} + \gamma_{i+1,i+1}^{(l-m)} g_{i+1,i}^{(m)} \right]$$

for $i = 0, 1, \ldots, N-1$, $l = 1, 2, \ldots, L$ (here, the γ's and g's with indexes -1 or N are considered to be zero). This can be rewritten as the block Toeplitz system of equations

$$\sum_{m=1}^{L} \mathbf{A}_{l-m}\mathbf{g}_m = \boldsymbol{\gamma}_l, \qquad l = 1, 2, \ldots, L$$

with the $(3N-2)$-dimensional vectors $\mathbf{g}_l$ and $\boldsymbol{\gamma}_l$ shown at the bottom of the page and the $(3N-2) \times (3N-2)$ block diagonal matrices $\mathbf{A}_l = \mathrm{diag}\{\mathbf{A}_i^{(l)}\}_{i=0}^{N-1}$ where

$$\mathbf{A}_i^{(l)} = \begin{bmatrix} \gamma_{i-1,i-1}^{(l)} & \gamma_{i-1,i}^{(l)} & \gamma_{i-1,i+1}^{(l)} \\ \gamma_{i,i-1}^{(l)} & \gamma_{i,i}^{(l)} & \gamma_{i,i+1}^{(l)} \\ \gamma_{i+1,i-1}^{(l)} & \gamma_{i+1,i}^{(l)} & \gamma_{i+1,i+1}^{(l)} \end{bmatrix},$$
$$i = 1, 2, \ldots, N-2$$

and

$$\mathbf{A}_0^{(l)} = \begin{bmatrix} \gamma_{0,0}^{(l)} & \gamma_{0,1}^{(l)} \\ \gamma_{1,0}^{(l)} & \gamma_{1,1}^{(l)} \end{bmatrix}$$

$$\mathbf{A}_{N-1}^{(l)} = \begin{bmatrix} \gamma_{N-2,N-2}^{(l)} & \gamma_{N-2,N-1}^{(l)} \\ \gamma_{N-1,N-2}^{(l)} & \gamma_{N-1,N-1}^{(l)} \end{bmatrix}.$$

The minimum reconstruction error variance is obtained as

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

4) An Example: As a simple example, we consider a paraunitary two-channel FB (i.e., $N = 2$) with $M = 1$ and, hence, oversampling factor $K = 2$. The analysis filters are the Haar filters

$$H_0(z) = \frac{1}{\sqrt{2}}(1 + z^{-1})$$

and

$$H_1(z) = \frac{1}{\sqrt{2}}(1 - z^{-1})$$

and the synthesis filters (corresponding to $\mathbf{R}(z) = \hat{\mathbf{R}}(z)$) are

$$F_0(z) = \frac{1}{2}\tilde{H}_0(z)$$

and

$$F_1(z) = \frac{1}{2}\tilde{H}_1(z).$$

We assume uncorrelated and white quantization noise, i.e., $\mathbf{S}_q(z) = \sigma_q^2 \mathbf{I}_2$.

Without noise shaping, the reconstruction error variance σ_e^2 is obtained from (9) as

$$\sigma_e^2 = \sigma_q^2 \int_{-1/2}^{1/2} [|F_0(e^{j2\pi\theta})|^2 + |F_1(e^{j2\pi\theta})|^2]\, d\theta = \frac{\sigma_q^2}{2} \quad (29)$$

where we used $F_0(z)\tilde{F}_0(z) + F_1(z)\tilde{F}_1(z) = \frac{1}{2}$. This is consistent with our $1/K$ result in (11).

We next calculate the optimum first-order (i.e., $L = 1$) noise-shaping system. The analysis polyphase coefficient matrices are given by $\mathbf{E}_0 = \frac{1}{\sqrt{2}}[1\ 1]^T$ and $\mathbf{E}_1 = \frac{1}{\sqrt{2}}[1\ -1]^T$. With (23), we obtain

$$\mathbf{\Gamma}_0 = \frac{1}{4}[\mathbf{E}_0\mathbf{E}_0^T + \mathbf{E}_1\mathbf{E}_1^T] = \frac{1}{4}\mathbf{I}_2$$
$$\mathbf{\Gamma}_1 = \frac{1}{4}\mathbf{E}_1\mathbf{E}_0^T$$
$$\mathbf{\Gamma}_{-1} = \mathbf{\Gamma}_1^T$$

and

$$\mathbf{\Gamma}_l = \mathbf{0}$$

for $|l| > 1$. Inserting this into (21), it follows that the optimal noise-shaping system of order $L = 1$ is $\mathbf{G}(z) = \mathbf{I}_2 - \mathbf{G}_{1,\mathrm{opt}}z^{-1}$ with

$$\mathbf{G}_{1,\mathrm{opt}} = \mathbf{\Gamma}_0^{-1}\mathbf{\Gamma}_1 = 4\mathbf{\Gamma}_1 = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}.$$

$$\mathbf{g}_l = \left[g_{0,0}^{(l)} g_{1,0}^{(l)} \middle| g_{0,1}^{(l)} g_{1,1}^{(l)} g_{2,1}^{(l)} \middle| g_{1,2}^{(l)} g_{2,2}^{(l)} g_{3,2}^{(l)} \middle| \cdots \middle| g_{N-3,N-2}^{(l)} g_{N-2,N-2}^{(l)} g_{N-1,N-2}^{(l)} \middle| g_{N-2,N-1}^{(l)} g_{N-1,N-1}^{(l)} \right]^T$$

$$\boldsymbol{\gamma}_l = \left[\gamma_{0,0}^{(l)} \gamma_{1,0}^{(l)} \middle| \gamma_{0,1}^{(l)} \gamma_{1,1}^{(l)} \gamma_{2,1}^{(l)} \middle| \gamma_{1,2}^{(l)} \gamma_{2,2}^{(l)} \gamma_{3,2}^{(l)} \middle| \cdots \middle| \gamma_{N-3,N-2}^{(l)} \gamma_{N-2,N-2}^{(l)} \gamma_{N-1,N-2}^{(l)} \middle| \gamma_{N-2,N-1}^{(l)} \gamma_{N-1,N-1}^{(l)} \right]^T$$
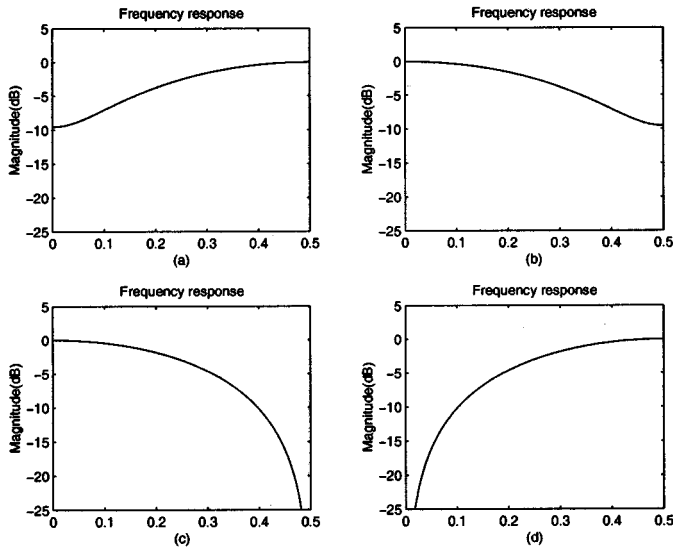
Fig. 8. Noise-shaping filters and synthesis filters in an oversampled two-channel FB. (a) $|G_{0,0;\mathrm{opt}}(e^{j2\pi\theta})|$. (b) $|G_{1,1;\mathrm{opt}}(e^{j2\pi\theta})|$. (c) $|F_0(e^{j2\pi\theta})|$. (d) $|F_1(e^{j2\pi\theta})|$.

The corresponding (minimum) reconstruction error variance is obtained from (24) as

$$\sigma_{e,\min}^2 = \sigma_q^2 \left[\frac{1}{2} - \mathrm{Tr}\{\mathbf{\Gamma}_1^T \mathbf{G}_{1,\mathrm{opt}}\}\right] = \frac{\sigma_q^2}{4}. \qquad (30)$$

Comparing (30) with (29), we see that the first-order noise-shaping system achieves an error variance reduction by a factor of 2. It is instructive to compare this result with the optimum *intrachannel* noise-shaping system $\mathbf{G}(z)$ of order $L = 1$ (see Section III-B3). Here, it follows from (27) that

$$g_{0,0}^{(1)} = \frac{\gamma_{0,0}^{(1)}}{\gamma_{0,0}^{(0)}} = \frac{1}{2} \quad \text{and} \quad g_{1,1}^{(1)} = \frac{\gamma_{1,1}^{(1)}}{\gamma_{1,1}^{(0)}} = -\frac{1}{2}$$

and hence we obtain

$$\mathbf{G}_{1,\mathrm{opt}} = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The corresponding reconstruction error variance is obtained from (28) as $\sigma_{e,\min}^2 = \frac{3}{8}\sigma_q^2$. Thus, as expected, failing to exploit the interchannel redundancy leads to a larger error variance, which, however, is still smaller than the error variance $\sigma_e^2 = \frac{\sigma_q^2}{2}$ obtained without noise shaping.

Fig. 8 shows the transfer functions of the noise-shaping filters in the diagonal of $\mathbf{G}(z)$ (note that these are identical for the general and the intrachannel noise-shaping systems) and the transfer functions of the synthesis filters. We see that the noise-shaping system operating in the low-pass channel $G_{0,0;\mathrm{opt}}(z) = 1 - \frac{1}{2}z^{-1}$ attenuates the noise at low frequencies (note that subsequently $F_0(z)$ attenuates high frequencies), whereas the noise-shaping system operating in the high-pass channel $G_{1,1;\mathrm{opt}}(z) = 1 + \frac{1}{2}z^{-1}$ attenuates the noise at high frequencies (subsequently, $F_1(z)$ attenuates low frequencies). Thus, the noise-shaping system shifts part of the noise to those frequencies that are subsequently attenuated by the synthesis filters.

5) Simulation Study 1: Further insight into the performance of noise-shaping subband coders was obtained by evaluating
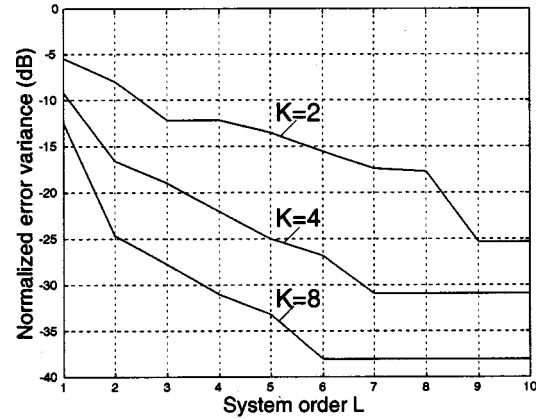

Fig. 9. Normalized reconstruction error variance $10\log(\sigma_{e,\min}^2/\sigma_q^2)$ as a function of the noise-shaping system order L.

(24) for the reconstruction error variance $\sigma_{e,\min}^2$ for three paraunitary, odd-stacked, cosine-modulated FBs [3], [8], [9] with $N = 16$ channels, normalized analysis filters of length $L_h = 81$, and oversampling factors $K = 2, 4$, and 8. The quantization noise was assumed uncorrelated and white with variance σ_q^2 in each channel. Fig. 9 shows the normalized reconstruction error variance $10\log(\sigma_{e,\min}^2/\sigma_q^2)$ as a function of the noise-shaping system's order L. For increasing L, the reconstruction error variance decreases up to a certain point, after which it remains constant. The maximum system order (i.e., the order after which the reconstruction error variance does not decrease any more) depends on the rank of the block Toeplitz matrix in (21), which is determined by the oversampling factor and the analysis filters.

The results in Fig. 9 show that for large values of L, the reconstruction error variance of the proposed noise-shaping subband coders follows a $1/K^2$ behavior. However, for small L, an increase of K is observed to produce a stronger reduction of the reconstruction error variance, i.e., the reconstruction error variance can drop faster than according to $1/K^2$. Specifically, for noise-shaping system order $L = 4$, we can see from Fig. 9 that doubling the oversampling factor results in a reduction of the reconstruction error variance by about 9 dB, and for system order $L = 6$, we even get about 12-dB error-variance reduction.

Next, we investigate the quantization error—redundancy behavior in an implemented noise-shaping subband coder. We coded an audio signal using a paraunitary, 64-channel, odd-stacked, cosine-modulated FB and a noise-shaping system designed under the assumption of uncorrelated and white quantization noise.[8] Uniform quantizers with equal stepsizes in all subbands were employed. Fig. 10 depicts the resulting SNR (defined as $\mathrm{SNR} = \frac{\|x\|^2}{\|x_q - x\|^2}$) as a function of the quantization step size for different oversampling factors K. For each K, we used the maximum possible noise-shaping system order. For K between 4 and 32, we observe a 6-dB SNR increase for each

---

[8]Similar to oversampled A/D conversion [12], the assumption of uncorrelated white noise is not justified in the oversampled case, which causes the performance of implemented coders to be poorer than the theoretical performance observed further above. Nonetheless, we are forced to use this assumption because estimating the actual quantization noise statistics and designing the noise-shaping system accordingly is not possible since the quantizer is placed within a feedback loop [14].
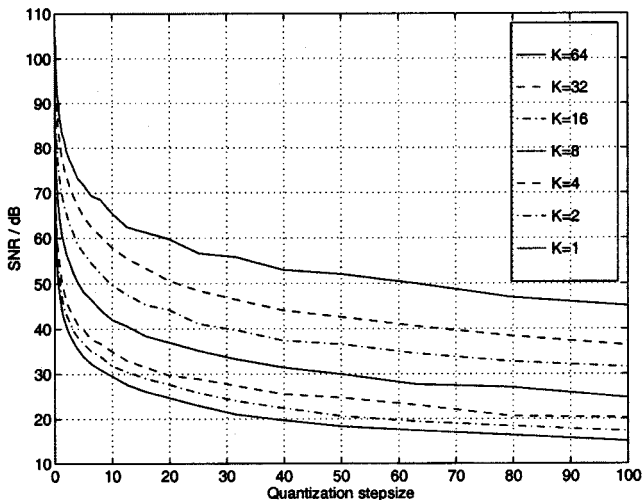
Fig. 10. SNR improvement as a function of the quantization stepsize for various oversampling factors.

TABLE I
IMPROVING THE EFFECTIVE RESOLUTION OF A SUBBAND CODER BY MEANS OF OVERSAMPLING AND NOISE SHAPING (N_Q DENOTES THE NUMBER OF QUANTIZATION INTERVALS REQUIRED)

K	L	SNR/dB	N_Q
1	0	23.80	63
8	2	23.76	15
1	0	39.47	593
64	2	39.49	15

doubling of K, corresponding to a K^2-dependence of the SNR. The SNR increase in going from $K = 32$ to $K = 64$ is about 9 dB, and thus stronger than K^2.

6) Simulation Study 2: Our next experiment demonstrates that noise shaping in oversampled FBs is capable of drastically improving the effective resolution of the resulting subband coder. We coded an audio signal using a paraunitary, 64-channel, critically sampled, odd-stacked, cosine-modulated FB using uniform quantizers with 63 quantization intervals (6-bit quantizers) in each subband. The resulting SNR was 23.80 dB. Then, we coded the same signal using a paraunitary, 64-channel, odd-stacked, cosine-modulated FB with oversampling factor $K = 8$, a noise-shaping system with order $L = 2$ (designed under the assumption of uncorrelated and white quantization noise), and quantizers with 15 quantization intervals (4-bit quantizers). The resulting SNR was 23.76 dB. Thus, in the oversampled case, the same SNR was achieved using a quantization with far lower resolution (corresponding to a reduction of 2 bits in each of the 64 channels) than in the critical case. For oversampling factor 64, quantizers with 15 intervals (4-bit quantizers), and noise-shaping system order 2, we obtained an SNR of 39.49 dB. In order to achieve an SNR of 39.47 dB in the critically sampled case, we had to use quantizers with 593 intervals (10-bit quantizers). Thus, here we were able to save 578 quantization intervals (or, equivalently, 6 bits of quantizer resolution) in each of the 64 channels. Table I summarizes these results.

7) Simulation Study 3: In the previous simulation study, we observed that oversampling and noise shaping drastically improve the effective resolution of a subband coder. However,
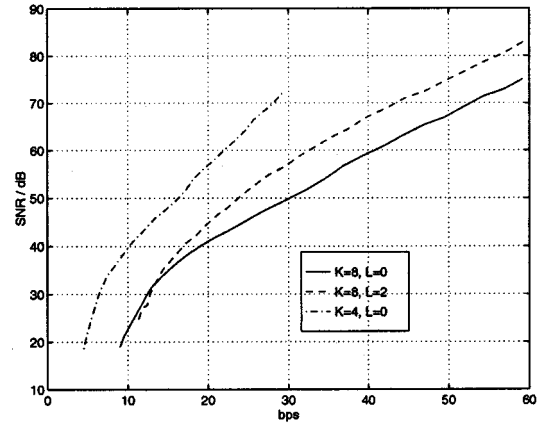

Fig. 11. Distortion-rate characteristic of oversampled noise-shaping subband coders with and without noise shaping.

this resolution enhancement comes at the cost of increased sample rate. It is therefore natural to ask how oversampled noise-shaping subband coders perform from a rate-distortion point of view, i.e., how the coding rate behaves in relation to the resolution enhancement. The following simulation results pertain to this problem. This is investigated by the following simulation study. We coded an audio signal, using a paraunitary, odd-stacked, cosine-modulated FB with $N = 64$ channels, filter length $L_h = 256$, and oversampling factors $K = 1, 2, 4, 8, 16, 32, 64$. Uniform quantizers with equal step sizes in all subbands were employed. The quantizer outputs were entropy-coded using a Huffman coder which jointly operates on all channel outputs, i.e., all subband signal samples $v_k[m]$ (for $k = 1, 2, \ldots, N$ and for the total range of m values) were collected and jointly Huffman coded. The optimum noise-shaping system was calculated under the assumption of uncorrelated white quantization noise with equal variance in all channels.

Fig. 11 shows the measured distortion-rate performance, i.e., the SNR as a function of the number of bits per sample (bps) required to encode the input signal. The distortion-rate performance obtained with $K = 8$ and noise-shaping system order $L = 2$ is seen to be better than that obtained with $K = 8$ and no noise shaping but poorer than that obtained with $K = 4$ and no noise shaping. We furthermore observed that the distortion rate performance of oversampled noise-shaping coders is poorer than that of critically sampled coders without noise shaping.

IV. OVERSAMPLED SIGNAL-PREDICTIVE SUBBAND CODERS

This section introduces an alternative method for noise reduction in oversampled FBs. This method is based on linear prediction of the FB's subband signals. The resulting oversampled signal predictive subband coders can be motivated by oversampled signal-predictive A/D converters [11], which will be briefly reviewed first.

A. Oversampled Signal-Predictive A/D Converters

In contrast to noise-shaping A/D converters which predict the inband quantization noise (see Section III-A), signal-predictive A/D converters predict the current sample of the signal to be
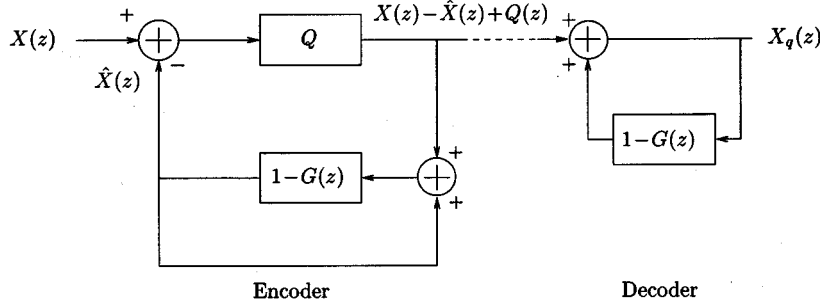
Fig. 12.   Signal-predictive A/D converter.

quantized. The signal-predictive coder (again modeled as an entirely discrete-time system) is depicted in Fig. 12. Here

$$G(z) = 1 - \sum_{l=1}^{L} g[l] z^{-l}$$

is the prediction error filter of order L. The predictor uses the past L noisy signal samples to estimate the current signal sample

$$\hat{x}[n] = \sum_{l=1}^{L} g[l](x[n-l] + q[n-l]).$$

The prediction error $x[n] - \hat{x}[n]$, which forms the input to the quantizer, is given by

$$X(z) - \hat{X}(z) = G(z)X(z) - [1 - G(z)]Q(z).$$

Choosing the filter $G(z)$ such that the prediction error is minimized leads to a reduced dynamic range over which the quantizer must operate. This allows to improve the effective quantizer resolution for a fixed number of quantization intervals. The decoder output is given by $X_q(z) = X(z) + Q(z)$, so that the overall reconstruction error is equal to the quantization error $E(z) = X_q(z) - X(z) = Q(z)$.

An oversampled signal-predictive coder exploits two types of redundancies: the "natural" redundancy which is inherent in the input signal whenever it has a nonflat power spectral density function, and the "synthetic" redundancy which is introduced by oversampling the analog signal, i.e., by expanding the input signal into a redundant signal set (time-shifted sinc functions, see Section II-B). Increasing the oversampling factor yields more synthetic redundancy and hence better prediction accuracy.

B. Signal Prediction in Oversampled FBs

Signal-predictive oversampled A/D converters exploit the redundancy inherent in the signal samples to estimate the current sample to be quantized. This principle will now be extended to oversampled PR FBs whose subband signals are a redundant representation of the input signal. The resulting oversampled signal-predictive subband coders extend critically sampled signal-predictive subband coders [14], [16], [15], [25].

1) The Signal-Predictive Subband Coder: Fig. 13 shows the structure of the oversampled signal-predictive subband coder.

The prediction error system is an $N \times N$ MIMO system given by

$$\mathbf{G}(z) = \mathbf{I}_N - \sum_{l=1}^{L} \mathbf{G}_l z^{-l} \qquad (31)$$

which results in a strictly causal feedback loop (prediction) system

$$\mathbf{I}_N - \mathbf{G}(z) = \sum_{l=1}^{L} \mathbf{G}_l z^{-l}.$$

The predictor uses the past L noisy subband signal vectors to estimate the current subband signal vector

$$\hat{\mathbf{v}}[m] = \sum_{l=1}^{L} \mathbf{G}_l(\mathbf{v}[m-l] + \mathbf{q}[m-l]).$$

This is a "noisy" vector prediction problem. For subband coding using high-resolution quantizers, the effect of quantization noise can be neglected and hence

$$\hat{\mathbf{v}}[m] \approx \sum_{l=1}^{L} \mathbf{G}_l \mathbf{v}[m-l].$$

However, here we are primarily interested in low-resolution quantization.

The prediction error $\boldsymbol{\epsilon}[m] \triangleq \mathbf{v}[m] - \hat{\mathbf{v}}[m]$ forms the input to the quantizer. It can be shown that

$$\boldsymbol{\epsilon}(z) = \mathbf{v}(z) - \hat{\mathbf{v}}(z) = \mathbf{G}(z)\mathbf{v}(z) - [\mathbf{I}_N - \mathbf{G}(z)]\mathbf{q}(z). \quad (32)$$

By choosing $\mathbf{G}(z)$ such that the dynamic range of the quantizer input vector $\boldsymbol{\epsilon}[m] = \mathbf{v}[m] - \hat{\mathbf{v}}[m]$ is reduced, it is possible to improve the effective quantizer resolution for a fixed number of quantization intervals.

With (32), it follows that the quantizer output is $\mathbf{a}(z) = \mathbf{G}(z)[\mathbf{v}(z) + \mathbf{q}(z)]$, which, in turn, implies (assuming existence of the inverse of $\mathbf{G}(z)$) that the decoder output is $\mathbf{x}_q(z) = \mathbf{R}(z)[\mathbf{v}(z) + \mathbf{q}(z)]$. Using a PR FB (i.e., $\mathbf{R}(z)\mathbf{E}(z) = \mathbf{I}_M$), we have $\mathbf{R}(z)\mathbf{v}(z) = \mathbf{R}(z)\mathbf{E}(z)\mathbf{x}(z) = \mathbf{x}(z)$ so that

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

This yields the following result that can be interpreted as an extension of the fundamental theorem of predictive quantization [40].
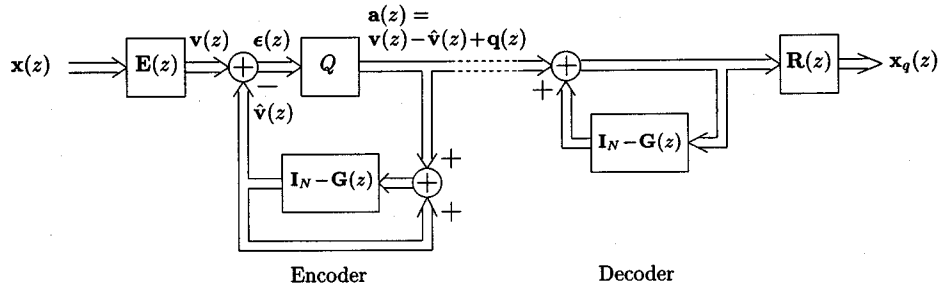
Fig. 13.   Oversampled signal-predictive subband coder (polyphase-domain representation).

Proposition 2: For an oversampled signal-predictive subband coder using a PR FB, the reconstruction error $\mathbf{e}(z) = \mathbf{x}_q(z) - \mathbf{x}(z)$ is given by

$$\mathbf{e}(z) = \mathbf{R}(z)\mathbf{q}(z).$$

Thus, the reconstruction error is the quantization noise filtered by the synthesis FB $\mathbf{R}(z)$. With our results from Section II-C.4, this leads to the important conclusion that the para-pseudo-inverse $\hat{\mathbf{R}}(z)$ minimizes the reconstruction error variance in the case of uncorrelated and white quantization noise since it suppresses the component of $\mathbf{q}(z)$ that lies in $\mathcal{R}^{\perp}$.

Just like an oversampled signal-predictive A/D converter, an oversampled signal-predictive subband coder exploits two types of redundancies: the natural redundancy that is inherent in the input signal and the synthetic redundancy that is introduced by the oversampled analysis FB, i.e., by expanding the input signal into a redundant set of functions (see Section II-C). An increase of the oversampling factor yields more synthetic redundancy in the subband signals and hence better prediction accuracy.

Since, in general, the matrices $\mathbf{G}_k$ are not diagonal, we are performing interchannel (cross-channel) prediction in addition to intrachannel prediction. Exploiting interchannel correlations (which are due to the overlap of the channel filters' transfer functions) may yield an important performance gain. In fact, it has previously been demonstrated [15] for a two-channel, critically sampled Haar FB that using information from the high-frequency band for prediction in the low-frequency band yields rate-distortion optimality. Critically sampled subband coders employing interchannel prediction have also been considered in [16].

The MIMO system $\mathbf{G}(z)$ is said to be *minimum phase* or *minimum delay* if all the roots of $\det \mathbf{G}(z) = 0$ lie inside the unit circle in the z-plane. This condition ensures that the inverse filter $\mathbf{G}^{-1}(z)$, and hence the feedback loop, will be stable [40]. In the noiseless case ($\mathbf{q}(z) = \mathbf{0}$), it is shown in [40] that $\mathbf{G}(z)$ is minimum phase if the process $\mathbf{v}[m]$ is stationary and nondeterministic. Although we do not have a proof of the minimum phase property of $\mathbf{G}(z)$ in the noisy case, we always observed stability of $\mathbf{G}^{-1}(z)$ in our simulation examples.

2) Calculation of the Optimum Prediction System: We now derive the optimum prediction system. In contrast to the case of noise-shaping subband coders, the input signal $x[n]$ will here be modeled as a random process that is assumed wide-sense stationary, zero-mean, real-valued, and uncorrelated with the quan-

tization noise process $\mathbf{q}[m]$. For simplicity, the analysis and synthesis filters are assumed real-valued as well. It will be convenient to introduce the "FB input vector"

$$\mathbf{x}[m] = [x[mM]\, x[mM+1] \ldots x[mM+M-1]]^T$$

with $M \times M$ correlation matrices

$$\mathbf{C}_x[l] = \mathcal{E}\{\mathbf{x}[m]\mathbf{x}^T[m-l]\}$$

and power spectral matrix

$$\mathbf{S}_x(z) = \sum_{l=-\infty}^{\infty} \mathbf{C}_x[l]z^{-l}.$$

Using $\mathbf{v}(z) = \mathbf{E}(z)\mathbf{x}(z)$, the power spectral matrix of $\mathbf{v}[m]$ is given by

$$\mathbf{S}_v(z) = \sum_{l=-\infty}^{\infty} \mathbf{C}_v[l]z^{-l} = \mathbf{E}(z)\mathbf{S}_x(z)\tilde{\mathbf{E}}(z)$$

where

$$\mathbf{C}_v[l] = \mathcal{E}\{\mathbf{v}[m]\mathbf{v}^T[m-l]\} = \sum_{i=-\infty}^{\infty} \mathbf{E}_i \sum_{j=-\infty}^{\infty} \mathbf{C}_x[j]\mathbf{E}_{i+j-l}^T$$

$$(33)$$

with $\mathbf{C}_v^T[-l] = \mathbf{C}_v[l]$. With (32) and using the fact that $x[n]$ (and hence also $\mathbf{v}[m]$) is uncorrelated with $\mathbf{q}[m]$, it follows that the power spectral matrix of the prediction error $\boldsymbol{\epsilon}[m] = \mathbf{v}[m] - \hat{\mathbf{v}}[m]$ is given by

$$\mathbf{S}_{\epsilon}(z) = \mathbf{G}(z)\mathbf{S}_v(z)\tilde{\mathbf{G}}(z) + [\mathbf{I}_N - \mathbf{G}(z)]\mathbf{S}_q(z)[\mathbf{I}_N - \tilde{\mathbf{G}}(z)].$$

Hence, the prediction error variance is obtained as

$$\sigma_{\epsilon}^2 = \frac{1}{N}\int_{-1/2}^{1/2} \mathrm{Tr}\Big\{\mathbf{G}(e^{j2\pi\theta})\mathbf{S}_v(e^{j2\pi\theta})\mathbf{G}^H(e^{j2\pi\theta})$$

$$+ [\mathbf{I}_N - \mathbf{G}(e^{j2\pi\theta})]\mathbf{S}_q(e^{j2\pi\theta})[\mathbf{I}_N - \mathbf{G}^H(e^{j2\pi\theta})]\Big\}\, d\theta. \quad (34)$$

Inserting (31) into (34) and using $\mathbf{C}_v^T[-l] = \mathbf{C}_v[l]$ and $\mathbf{C}_q^T[-l] = \mathbf{C}_q[l]$, we obtain further

$$\sigma_{\epsilon}^2 = \frac{1}{N}\,\mathrm{Tr}\Bigg\{\mathbf{C}_v[0] - \sum_{l=1}^{L}(\mathbf{C}_v^T[l]\mathbf{G}_l + \mathbf{C}_v[l]\mathbf{G}_l^T)$$

$$+ \sum_{m=1}^{L}\mathbf{G}_m \sum_{l=1}^{L}(\mathbf{C}_v[l-m] + \mathbf{C}_q[l-m])\mathbf{G}_l^T\Bigg\}.$$

$$(35)$$

In order to calculate the matrices $\mathbf{G}_l$ minimizing σ_{ϵ}^2, we set[9] $\frac{\partial \sigma_{\epsilon}^2}{\partial \mathbf{G}_i} = \mathbf{0}$ and use the matrix derivative rules from Sec-

---

[9]The optimum prediction system can equivalently be derived using the orthogonality principle.

tion III-B2. This yields the following block Toeplitz system of linear equations:

$$\sum_{l=1}^{L} (\mathbf{C}_v[l-i] + \mathbf{C}_q[l-i]) \mathbf{G}_l^T = \mathbf{C}_v^T[i],$$

$$\text{for } i = 1, 2, \ldots, L \quad (36)$$

or, equivalently,

$$\begin{bmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_{L-1} \\ \boldsymbol{\Gamma}_{-1} & \boldsymbol{\Gamma}_0 & \cdots & \boldsymbol{\Gamma}_{L-2} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\Gamma}_{-(L-1)} & \boldsymbol{\Gamma}_{-(L-2)} & \cdots & \boldsymbol{\Gamma}_0 \end{bmatrix} \begin{bmatrix} \mathbf{G}_1^T \\ \mathbf{G}_2^T \\ \vdots \\ \mathbf{G}_L^T \end{bmatrix} = \begin{bmatrix} \mathbf{C}_v^T[1] \\ \mathbf{C}_v^T[2] \\ \vdots \\ \mathbf{C}_v^T[L] \end{bmatrix},$$

$$\text{with } \boldsymbol{\Gamma}_l = \mathbf{C}_v[l] + \mathbf{C}_q[l]. \quad (37)$$

Using (36) in (35), the minimum prediction error variance is obtained as

$$\sigma_{\epsilon, \min}^2 = \frac{1}{N} \text{Tr} \left\{ \mathbf{C}_v[0] - \sum_{l=1}^{L} \mathbf{C}_v[l] \mathbf{G}_{l, \text{opt}}^T \right\} \quad (38)$$

where $\mathbf{G}_{l, \text{opt}}$ denotes the solution of (36) or (37).

In the noiseless case, (36) reduces to

$$\sum_{l=1}^{L} \mathbf{C}_v[l-i] \mathbf{G}_l^T = \mathbf{C}_v^T[i], \qquad \text{for } i = 1, 2, \ldots, L \quad (39)$$

which can be solved efficiently using the multichannel Levinson recursion [39]. Another important special case where this is possible is the noisy case with white (but possibly correlated) quantization noise, i.e., $\mathbf{C}_q[l] = \mathbf{C}_q[0]\delta[l]$. Here, (36) reduces to (39) with $\mathbf{C}_v[0]$ replaced by $\mathbf{C}_v[0] + \mathbf{C}_q[0]$.

We finally note that the above derivation can be extended to incorporate correlations between $\mathbf{v}[m]$ and $\mathbf{q}[m]$.

3) Constrained Optimum Prediction: Reducing the computational complexity of predictive subband coding is often important, especially for adaptive prediction. Hence, in analogy to noise shaping (see Section III-B3), we shall consider the case of no interchannel prediction ("intrachannel prediction") or interchannel prediction between neighboring channels only.

We shall first calculate the optimum intrachannel prediction system. Specializing (35) to diagonal $\mathbf{G}_l$, we obtain

$$\sigma_{\epsilon}^2 = \frac{1}{N} \sum_{i=0}^{N-1} \left[c_{i,i}^{(0)} - 2\sum_{l=1}^{L} c_{i,i}^{(l)} g_{i,i}^{(l)} + \sum_{m=1}^{L} g_{i,i}^{(m)} \sum_{l=1}^{L} \gamma_{i,i}^{(m-l)} g_{i,i}^{(l)} \right] \quad (40)$$

where $g_{i,j}^{(l)} = [\mathbf{G}_l]_{i,j}$, $c_{i,j}^{(l)} = [\mathbf{C}_v[l]]_{i,j}$, and $\gamma_{i,j}^{(l)} = [\boldsymbol{\Gamma}_l]_{i,j}$ with $\boldsymbol{\Gamma}_l = \mathbf{C}_v[l] + \mathbf{C}_q[l]$. Proceeding as in Section III-B3, we obtain the following $NL \times L$ Toeplitz systems of equations:

$$\sum_{m=1}^{L} \gamma_{i,i}^{(l-m)} g_{i,i}^{(m)} = c_{i,i}^{(l)},$$

$$l = 1, 2, \ldots, L, \; i = 0, 1, \ldots, N-1 \quad (41)$$

or, briefly, $\mathbf{A}_i \mathbf{g}_i = \mathbf{c}_i$ for $i = 0, 1, \ldots, N-1$, where $[\mathbf{A}_i]_{l,m} = \gamma_{i,i}^{(l-m)}$, $[\mathbf{g}_i]_l = g_{i,i}^{(l)}$, and $[\mathbf{c}_i]_l = c_{i,i}^{(l)}$. Inserting (41) into (40)

and using $\gamma_{i,i}^{(m-l)} = \gamma_{i,i}^{(l-m)}$, we obtain the minimum prediction error variance as

$$\sigma_{\epsilon, \min}^2 = \frac{1}{N} \sum_{i=0}^{N-1} \left[c_{i,i}^{(0)} - \sum_{l=1}^{L} c_{i,i}^{(l)} g_{i,i;\text{opt}}^{(l)} \right] \quad (42)$$

where $g_{i,i;\text{opt}}^{(l)}$ denotes the solution of (41).

We next calculate the optimum prediction system that exploits only neighboring channel dependencies. Proceeding similarly to the case of noise shaping, we obtain the block Toeplitz system of equations

$$\sum_{m=1}^{L} \mathbf{A}_{l-m} \mathbf{g}_m = \mathbf{c}_l, \qquad l = 1, 2, \ldots, L$$

with the $(3N-2)$-dimensional vectors $\mathbf{g}_l$ and $\mathbf{c}_l$ shown at the bottom of this page and the $(3N-2) \times (3N-2)$ block diagonal matrices $\mathbf{A}_l = \text{diag}\{\mathbf{A}_i^{(l)}\}_{i=0}^{N-1}$ where

$$\mathbf{A}_i^{(l)} = \begin{bmatrix} \gamma_{i-1,i-1}^{(l)} & \gamma_{i,i-1}^{(l)} & \gamma_{i+1,i-1}^{(l)} \\ \gamma_{i-1,i}^{(l)} & \gamma_{i,i}^{(l)} & \gamma_{i+1,i}^{(l)} \\ \gamma_{i-1,i+1}^{(l)} & \gamma_{i,i+1}^{(l)} & \gamma_{i+1,i+1}^{(l)} \end{bmatrix},$$

$$i = 1, 2, \ldots, N-2$$

and

$$\mathbf{A}_0^{(l)} = \begin{bmatrix} \gamma_{0,0}^{(l)} & \gamma_{1,0}^{(l)} \\ \gamma_{0,1}^{(l)} & \gamma_{1,1}^{(l)} \end{bmatrix}$$

$$\mathbf{A}_{N-1}^{(l)} = \begin{bmatrix} \gamma_{N-2,N-2}^{(l)} & \gamma_{N-1,N-2}^{(l)} \\ \gamma_{N-2,N-1}^{(l)} & \gamma_{N-1,N-1}^{(l)} \end{bmatrix}.$$

The minimum prediction error variance is here obtained as

$$\sigma_{\epsilon, \min}^2 = \frac{1}{N} \sum_{i=0}^{N-1} \left\{ c_{i,i}^{(0)} - \sum_{l=1}^{L} \left[c_{i-1,i}^{(l)} g_{i-1,i;\text{opt}}^{(l)} + c_{i,i}^{(l)} g_{i,i;\text{opt}}^{(l)} \right. \right.$$

$$\left. \left. + c_{i+1,i}^{(l)} g_{i+1,i;\text{opt}}^{(l)} \right] \right\}.$$

4) An Example: Let us reconsider the two-channel FB with oversampling factor $K = 2$ previously considered in Section III-B4. The input process is an AR-1 process defined by $x[n] = ax[n-1] + u[n]$ with correlation coefficient $a = 0.5$ and white driving noise $u[n]$ with variance 1. The autocorrelation function of $x[n]$ is $C_x[l] = \frac{4}{3}(\frac{1}{2})^{|l|}$ [37]. Inserting $\mathbf{E}_0 = \frac{1}{\sqrt{2}}[1 \; 1]^T$ and $\mathbf{E}_1 = \frac{1}{\sqrt{2}}[1 \; -1]^T$ into (33), we obtain

$$\mathbf{C}_v[0] = \begin{bmatrix} 2 & 0 \\ 0 & 2/3 \end{bmatrix} \qquad \mathbf{C}_v[1] = \begin{bmatrix} 3/2 & 1/2 \\ -1/2 & -1/6 \end{bmatrix}.$$

The quantization noise is assumed uncorrelated white with variance $\sigma_q^2 = 1$ in each channel, i.e., $\mathbf{C}_q[l] = \mathbf{I}_2\delta[l]$. Without prediction (i.e., $L = 0$ or $\boldsymbol{\epsilon}[m] = \mathbf{v}[m]$), the variance at the input

$$\mathbf{g}_l = \left[g_{0,0}^{(l)} g_{0,1}^{(l)} \left| g_{1,0}^{(l)} g_{1,1}^{(l)} g_{1,2}^{(l)} \right| g_{2,1}^{(l)} g_{2,2}^{(l)} g_{2,3}^{(l)} \right| \cdots \left| g_{N-2,N-3}^{(l)} g_{N-2,N-2}^{(l)} g_{N-2,N-1}^{(l)} \right| g_{N-1,N-2}^{(l)} g_{N-1,N-1}^{(l)} \right]^T$$

$$\mathbf{c}_l = \left[c_{0,0}^{(l)} c_{0,1}^{(l)} \left| c_{1,0}^{(l)} c_{1,1}^{(l)} c_{1,2}^{(l)} \right| c_{2,1}^{(l)} c_{2,2}^{(l)} c_{2,3}^{(l)} \right| \cdots \left| c_{N-2,N-3}^{(l)} c_{N-2,N-2}^{(l)} c_{N-2,N-1}^{(l)} \right| c_{N-1,N-2}^{(l)} c_{N-1,N-1}^{(l)} \right]^T$$

of the quantizer is obtained as $\sigma_\epsilon^2 = 4/3 \approx 1.33$. Next, we use a first-order prediction system $\mathbf{G}(z) = \mathbf{I}_2 - \mathbf{G}_1 z^{-1}$. From (37)

$$\mathbf{G}_{1,\text{opt}} = \mathbf{C}_v[1][\mathbf{C}_v[0] + \mathbf{I}_2]^{-T} = \begin{bmatrix} 1/2 & 3/10 \\ -1/6 & -1/10 \end{bmatrix}.$$

The resulting (minimum) prediction error variance is obtained from (38) as

$$\sigma_{\epsilon,\text{min}}^2 = \frac{1}{2} \text{Tr}\{\mathbf{C}_v[0] - \mathbf{C}_v[1]\mathbf{G}_{1,\text{opt}}^T\} = 5/6 \approx 0.83.$$

Let us compare this result with the optimum first-order *intra-channel* prediction system (see Section IV-B3). From (41) it follows that

$$g_{0,0}^{(1)} = \frac{c_{0,0}^{(1)}}{\gamma_{0,0}^{(0)}} = \frac{1}{2} \quad \text{and} \quad g_{1,1}^{(1)} = \frac{c_{1,1}^{(1)}}{\gamma_{1,1}^{(0)}} = -\frac{1}{10}$$

and hence

$$\mathbf{G}_{1,\text{opt}} = \begin{bmatrix} 1/2 & 0 \\ 0 & -1/10 \end{bmatrix}.$$

The corresponding prediction error variance is obtained from (42) as $\sigma_{\epsilon,\text{min}}^2 = 0.95$. It is larger than the error variance obtained with interchannel prediction but still smaller than that obtained without prediction.

5) Simulation Study 4: The next simulation example demonstrates that linear prediction is able to exploit the synthetic redundancy introduced by the oversampled analysis FB for improving prediction accuracy and hence enhancing resolution. The coder uses a paraunitary, odd-stacked, cosine-modulated FB [3], [8], [9] with $N = 16$ channels, normalized analysis filters of length $L_h = 64$, and various oversampling factors K. We evaluated the expression (38) for the prediction error variance $\sigma_{\epsilon,\text{min}}^2$ for a zero-mean white input process $x[n]$ (i.e., $\mathbf{C}_x[l] = \mathbf{I}_M \delta[l]$) and in the absence of quantization (noiseless prediction). Since the input is white, it contains no natural redundancy and hence all the prediction gain is due to synthetic redundancy. Fig. 14(a) shows $10 \log \sigma_{\epsilon,\text{min}}^2$ as a function of the predictor order L for different values of K. For increasing L, $\sigma_{\epsilon,\text{min}}^2$ is seen to decrease up to a certain point, after which it remains constant. There is no prediction gain for $K = 1$ since the function set corresponding to the FB is orthogonal. Note that there is a one-to-one correspondence between the prediction error variance and the overall prediction gain.

Fig. 14(b) shows the corresponding *measured* prediction error variance $10 \log \sigma_{\epsilon,\text{min}}^2$ obtained for an implemented coder. This result was obtained by averaging over five realizations (of length 1024) of the white input process. For prediction system order $L > 3$ (not shown), the performance of the implemented coder deteriorated significantly. This is probably due to the near-singularity of the block matrix in (37) for $L > 3$, which introduces numerical errors in the computation of the prediction system coefficient matrices. These numerical problems also explain the deviation between the computed and measured performance for $L = 3$.

6) Simulation Study 5: Our next simulation example demonstrates that oversampling combined with linear prediction is a powerful means to improve the effective resolution of a subband coder. We coded realizations of an AR-1 process
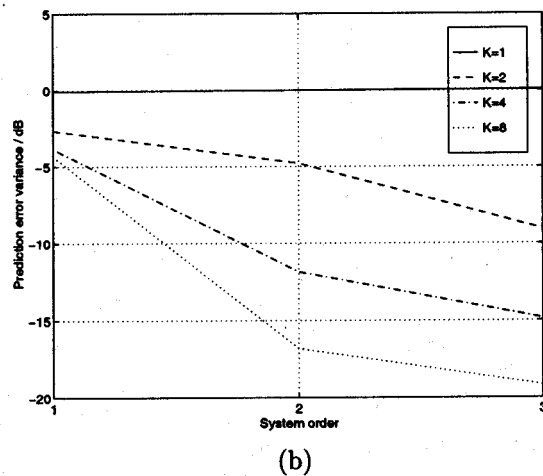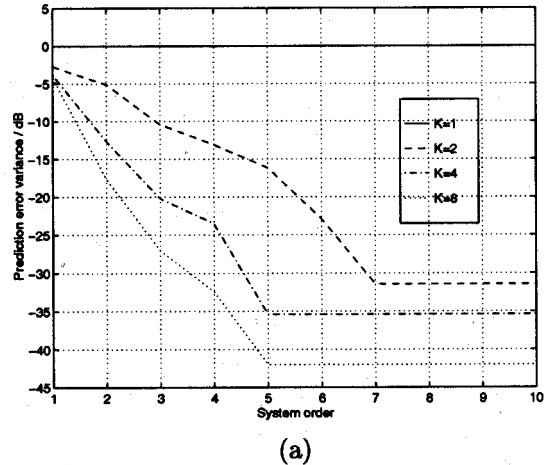

(a)

(b)

Fig. 14. Prediction error variance $10 \log \sigma_{\epsilon,\text{min}}^2$ for a white input signal and no quantization noise as a function of the predictor order L. (a) Computed according to (38). (b) Measured.

(length 1024) with correlation coefficient $a = 0.5$ using a paraunitary, 16-channel, critically sampled, odd-stacked, cosine-modulated FB and quantizers with 152 quantization intervals (8-bit quantizers) in each channel. The resulting SNR $= \frac{\|x\|^2}{\|x_q - x\|^2}$ was 32.49 dB. Next, we coded the same signal using an FB with oversampling factor $K = 4$ and a predictor with order $L = 10$ (designed under the assumption of uncorrelated and white quantization noise[10]). Here, quantizers with only 15 quantization intervals (4-bit quantizers) achieved an SNR of 32.51 dB. Hence, oversampling and prediction allowed us to save 4 bits of quantizer resolution in each of the 16 channels, of course at the cost of increased sample rate. For oversampling factor 8, quantizers with 15 quantization intervals (4-bit quantizers), and a predictor with order $L = 15$, we obtained an SNR of 50.48 dB. In order to achieve an SNR of 50.43 dB with a critically sampled subband coder without prediction, we had to use 1219 quantization intervals (11-bit quantizers). Hence, oversampling and prediction here saved 7 bits of quantizer resolution. Table II summarizes these results.

---

[10]We recall, however, that especially in the oversampled case the assumption of uncorrelated and white quantization noise is not realistic.
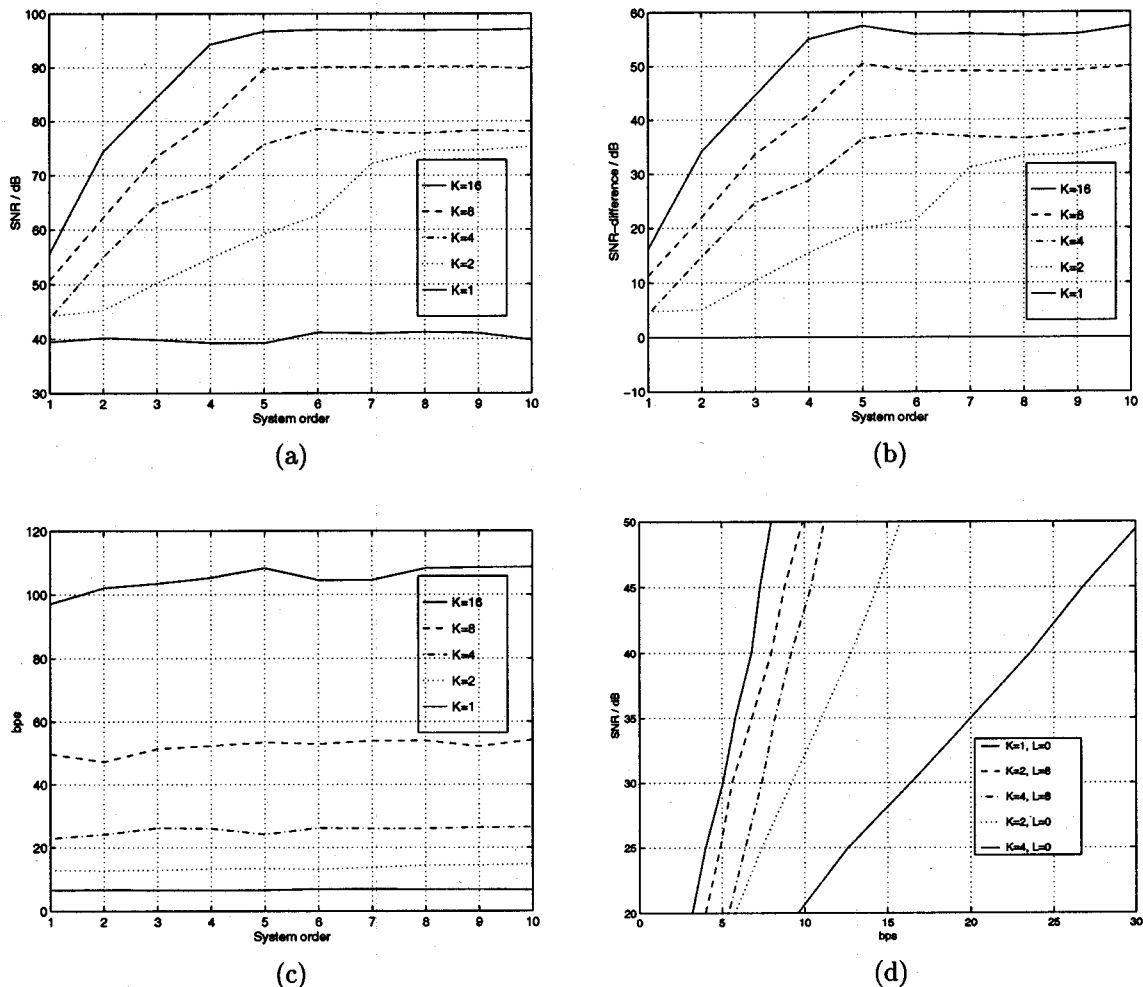
Fig. 15. Signal-predictive subband coder with 255 quantization intervals and various oversampling factors; simulation results for an AR-1 signal. (a) SNR as a function of the prediction system order L. (b) SNR differences with respect to $K = 1$. (c) bps as a function of the prediction system order L for different oversampling factors K. (d) Distortion-rate characteristic in comparison to alternative subband coders with various oversampling factors and predictor orders.

TABLE II
IMPROVING THE EFFECTIVE RESOLUTION OF A SUBBAND CODER BY
MEANS OF OVERSAMPLING AND PREDICTION (N_Q DENOTES THE NUMBER
OF QUANTIZATION INTERVALS REQUIRED)

K	L	SNR/dB	N_Q
1	0	32.49	152
4	10	32.51	15
1	0	50.43	1219
8	15	50.48	15

7) Simulation Study 6: We finally investigate the rate distortion and related properties of an implemented oversampled signal-predictive subband coder. As we observed in Section IV-B5, the variance of the quantizer input decreases for increasing oversampling factor K and for increasing prediction system order L. Therefore, for a fixed number of quantization intervals (which in this case was 255), we can reduce the quantization step size, thereby reducing the quantization error and, in turn, the overall reconstruction error $x_q[n] - x[n]$ (see Proposition 2). Fig. 15(a) shows the SNR, averaged over five realizations (of length 1024) of an AR-1 input signal with correlation coefficient $a = 0.5$, as a function of the

predictor order L for various oversampling factors K. The FB is as in Simulation Study 4. The predictor was designed for uncorrelated and white quantization noise with variance $\frac{\Delta^2}{12}$ in each channel, where Δ denotes the quantization stepsize used. In Fig. 15(b), the differences of the curves in Fig. 15(a) with respect to the $K = 1$ curve are depicted. One can observe that a predictive subband coder of order $L = 5$ and oversampling factor $K = 16$ leads to SNR improvements of more than 55 dB as compared to the critical case.

Fig. 15(c) shows the number of bps required by the predictive subband coder (with subsequent Huffman coding as in Section III-B7) as a function of the predictor order L for various oversampling factors K. We see that the number of bps increases slightly with L, which is due to the fact that prediction whitens the signal. Throughout this experiment, the number of quantization intervals was fixed to 255.

Finally, Fig. 15(d) shows the distortion-rate characteristic (SNR versus bps) of the signal-predictive subband coder, again with Huffman coding, for various oversampling factors K and predictor orders L. The distortion rate performance obtained with $K = 2$ and $L = 8$ (which in this case is the maximum

possible predictor order) is seen to be poorer than that of a critically sampled subband coder without prediction. Hence, whereas the proposed oversampled signal-predictive subband coders yield substantial noise reduction and allow the use of low-resolution quantizers, they cannot compete with critically sampled subband coders from a rate distortion point of view.

V. CONCLUSION

We have introduced two methods for noise reduction in oversampled filter banks. These methods are based on predictive quantization; they can be viewed as extensions of oversampled predictive A/D converters. We demonstrated that predictive quantization in oversampled FBs yields considerable quantization noise reduction at the cost of increased rate. The combination of oversampled filter banks with noise shaping or linear prediction improves the effective resolution of subband coders and is thus well suited for applications where—for technological or other reasons—quantizers with low resolution (even single bit) have to be used. Using low-resolution quantizers increases circuit speed and allows for lower circuit complexity.

Our simulation results furthermore suggested that, from a rate-distortion point of view, oversampled subband coders are inferior to critically sampled subband coders. However, it should be noted that from a perceptual point of view, oversampled subband coders have potential advantages over critically sampled coders. Finally, it is worthwhile to point out that the proposed methods are not limited to oversampled FBs but can be generalized to arbitrary frame expansions.

ACKNOWLEDGMENT

REFERENCES

[1] H. Bölcskei, F. Hlawatsch, and H. G. Feichtinger, "Frame-theoretic analysis of oversampled filter banks," *IEEE Trans. Signal Processing*, vol. 46, pp. 3256–3268, Dec. 1998.

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

[3] H. Bölcskei and F. Hlawatsch, "Oversampled cosine modulated filter banks with perfect reconstruction," *IEEE Trans. Circuits Syst. II (Special Issue on Multirate Systems, Filter Banks, Wavelets, and Applications)*, vol. 45, pp. 1057–1071, Aug. 1998.

[4] Z. Cvetković, "Oversampled modulated filter banks and tight Gabor frames in $l^2(Z)$," in *Proc. IEEE ICASSP-95*, Detroit, MI, May 1995, pp. 1456–1459.

[5] ——, "Overcomplete expansions for digital signal processing," Ph.D. dissertation, Univ. Calif., Berkeley, CA, Dec. 1995.

[6] H. Bölcskei, F. Hlawatsch, and H. G. Feichtinger, "Oversampled FIR and IIR DFT filter banks and Weyl–Heisenberg frames," in *Proc. IEEE ICASSP-96*, vol. 3, Atlanta, GA, May 1996, pp. 1391–1394.

[7] A. J. E. M. Janssen, "Density theorems for filter banks," Philips Res. Lab., Eindhoven, The Netherlands, Tech. Rep. 6858, Apr. 1995.

[8] H. Bölcskei and F. Hlawatsch, "Oversampled modulated filter banks," in *Gabor Analysis and Algorithms: Theory and Applications*, H. G. Feichtinger and T. Strohmer, Eds. Boston, MA: Birkhäuser, 1998, ch. 9, pp. 295–322.

[9] H. Bölcskei, "Oversampled filter banks and predictive subband coders," Ph.D. dissertation, Vienna Univ. Technol., Vienna, Austria, Nov. 1997.

[10] H. Bölcskei and F. Hlawatsch, "Oversampled filter banks: Optimal noise shaping, design freedom, and noise analysis," in *Proc. IEEE ICASSP-97*, vol. 3, Munich, Germany, Apr. 1997, pp. 2453–2456.

[11] S. K. Tewksbury and R. W. Hallock, "Oversampled, linear predictive and noise-shaping coder of order $N \geq 1$," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 436–447, July 1978.

[12] J. C. Candy and G. C. Temes, *Oversampling Delta-Sigma Data Converters*. New York: IEEE Press, 1993.

[13] R. M. Gray, "Oversampled sigma-delta modulation," *IEEE Trans. Commun.*, vol. 35, pp. 481–489, Apr. 1987.

[14] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[15] P. W. Wong, "Rate distortion efficiency of subband coding with crossband prediction," *IEEE Trans. Inform. Theory*, vol. 43, pp. 352–356, Jan. 1997.

[16] L. Vandendorpe and B. Maison, "Multiple-input/multiple-output prediction of subbands and image compression," in *Proc. COST 254 (Emerging Techniques for Communication Terminals)*, Toulouse, France, July 1997.

[17] R. Zamir and M. Feder, "Rate-distortion performance in coding bandlimited sources by sampling and dithered quantization," *IEEE Trans. Inform. Theory*, vol. 41, pp. 141–154, Jan. 1995.

[18] ——, "Information rates of pre/post-filtered dithered quantizers," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1340–1353, Sept. 1996.

[19] R. J. Duffin and A. C. Schaeffer, "A class of nonharmonic Fourier series," *Trans. Amer. Math. Soc.*, vol. 72, pp. 341–366, 1952.

[20] C. E. Heil and D. F. Walnut, "Continuous and discrete wavelet transforms," *SIAM Rev.*, vol. 31, pp. 628–666, Dec. 1989.

[21] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[22] A. J. Jerri, "The Shannon sampling theorem—Its various extensions and applications: A tutorial review," *Proc. IEEE*, vol. 65, pp. 1565–1596, 1977.

[23] R. J. Marks, *Introduction to Shannon Sampling Theory and Interpolation Theory*. New York: Springer-Verlag, 1991.

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

[25] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[26] T. Chen and P. P. Vaidyanathan, "Vector space framework for unification of one- and multidimensional filter bank theory," *IEEE Trans. Signal Processing*, vol. 42, pp. 2006–2021, Aug. 1994.

[27] M. Vetterli and C. Herley, "Wavelets and filter banks: Theory and design," *IEEE Trans. Signal Processing*, vol. 40, pp. 2207–2232, Sept. 1992.

[28] A. J. E. M. Janssen, "The duality condition for Weyl–Heisenberg frames," in *Gabor Analysis and Algorithms: Theory and Applications*, H. G. Feichtinger and T. Strohmer, Eds. Boston, MA: Birkhäuser, 1998, pp. 33–84.

[29] V. K. Goyal, M. Vetterli, and N. T. Thao, "Quantized overcomplete expansions in r^N: Analysis, synthesis and algorithms," *IEEE Trans. Inform. Theory*, vol. 44, pp. 16–31, Jan. 1998.

[30] N. J. Munch, "Noise reduction in tight Weyl–Heisenberg frames," *IEEE Trans. Inform. Theory*, vol. 38, pp. 608–616, Mar. 1992.

[31] N. T. Thao and M. Vetterli, "Deterministic analysis of oversampled A/D conversion and decoding improvement based on consistent estimates," *IEEE Trans. Signal Processing*, vol. 42, pp. 519–531, Mar. 1994.

[32] ——, "Lower bound on the mean squared error in oversampled quantization of periodic signals using vector quantization analysis," *IEEE Trans. Inform. Theory*, vol. 42, pp. 469–479, Mar. 1996.

[33] N. T. Thao, "Vector quantization analysis of sigma-delta modulation," *IEEE Trans. Signal Processing*, vol. 44, pp. 808–817, Apr. 1996.

[34] Z. Cvetković and M. Vetterli, "Overcomplete expansions and robustness," in *Proc. IEEE TFTS-96*, Paris, France, June 1996, pp. 325–328.

[35] N. Wiener and P. Masani, "The prediction theory of multivariate stochastic processes—I," *Acta. Math.*, vol. 98, pp. 111–150, 1957.

[36] ——, "The prediction theory of multivariate stochastic processes—II," *Acta Math.*, vol. 99, pp. 93–137, 1958.

[37] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

[38] A. Weinmann, *Uncertain Models and Robust Control*. Vienna, Austria: Springer, 1991.

[39] S. M. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[40] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.