Abstract—This paper studies low-delay Wyner–Ziv coding, i.e., lossy source coding with side information at the decoder, with emphasis on the extreme of zero delay. To achieve zero delay, a scalar quantizer is followed by scalar coding of quantization indices. In the fixed-length coding scenario, under high-resolution assumptions and appropriately defined decorrelation constraints, the optimal quantization level density is conjectured to be periodic. This conjecture, which is provable when the correlation is high, allows for a precise analysis of the rate-distortion tradeoff. The performance of variable-length coding with periodic quantization is also characterized. The results are then incorporated in predictive Wyner–Ziv coding for Gaussian sources with memory, and optimal prediction filters are numerically designed so as to strike a balance between maximally exploiting both temporal and spatial correlation and limiting the propagation of distortion due to occasional decoding errors. Finally, the zero-delay schemes are also employed in transform coding with small block lengths, where the Gaussian source and side information are transformed separately with the premise that corresponding transform coefficient pairs exhibit good spatial correlation and minimal temporal correlation. For the specific source-side information pairs studied, it is shown that transform coding, even with a small block-length, outperforms predictive coding. Performance of both predictive and transform coding are also compared with the asymptotic rate-distortion bounds.

Index Terms—High-resolution quantization, low-delay, predictive coding, transform coding, Wyner–Ziv coding.

I. INTRODUCTION

The Wyner–Ziv coding scenario, where a source is to be transmitted in a lossy fashion to a receiver with side information unknown to the sender [1], has received considerable attention over the past decade. Among the most notable contributions is the work of Zamir et al. [2], [3], where the authors proposed a structured algebraic binning scheme based on a pair of nested linear/lattice codes for binary symmetric/quadratic Gaussian sources. It is shown in [3] that the Wyner–Ziv rate-distortion (RD) function in the quadratic Gaussian case is asymptotically achievable as the dimensionality goes to infinity. Motivated by this, recent research focused on nested lattice codes with fixed-length coding. However, high-dimensional lattice codes are difficult to implement in practice. One approach to mitigate that difficulty is to use trellis-based codes [4]. Another approach is to use low-dimensional lattices followed by ideal Slepian–Wolf coding [5]. Liu et al. [6] showed that if ideal Slepian–Wolf coding is assumed for coding of low-dimensional quantization indices, nested lattice codes come close to Wyner–Ziv RD function for binary symmetric and jointly Gaussian source-side information pairs. Similar approaches were taken in [7]–[9]. However, in order to achieve the theoretical Slepian–Wolf rates, one has to utilize capacity achieving channel codes (e.g., LDPC, Turbo, etc.), which require very large block lengths. Though this approach is acceptable for applications such as video coding, it is problematic for sensor network applications, which tend to be delay sensitive.

In this work, with the motivation of delay-sensitive sensor networks, we study low-delay coding with emphasis on zero delay. One approach to achieve low (zero) delay is to use joint source-channel coding, i.e., directly map short source blocks (samples) into channel blocks, generally in a nonlinear fashion [10]–[12]. Similar mappings were also explored in the scenario where a relay channel creates the side information at the receiver [13], [14]. Here, we take a separate source-channel coding approach and tackle the source coding problem only, with the assumption of a noiseless channel. The benefit of this separation is, as usual, that all that must be known during the design of the source coder is the allowed rate of transmission over the channel (i.e., not the channel statistics).

A. Zero-Delay Source Coding

Our zero-delay source coder performs both the quantization and the noiseless coding of the quantization indexes (i.e., binning) in a scalar fashion. It should be noted that the subsequent channel coder may introduce significant delay if one strives to approach the capacity of the channel while keeping the probability of transmission error very low. However, if zero or low delay is strictly imposed on the overall system (as we assume), capacity-achieving codes are out of the question. In that case, it would be best to use channel codes of low rate and focus on achieving a very low probability of error.1

1Although channel codes are out of the scope of this paper, it is worth pointing out that even simple repetition codes might possibly be utilized in this context. For example, if sensor measurements are taken every 250 milliseconds and compressed with 5 bits/sample, simply repeating every source bit 51 times in the channel code would produce only about 1 kbits to transmit, and yet, if the channel randomly flips 10% (or 5%) of these bits, the resultant probability that a source bit is erroneously decoded becomes as low as \( \approx 2 \cdot 10^{-13} \) (or \( \approx 10^{-20} \)).
We employ both nonuniform and uniform quantization followed by fixed- and variable-length coding of indices, respectively. In the regime of high resolution fixed-length coding, we conjecture that the optimal quantization level density is periodic under the requirement that for every interval of a certain width $2\Delta$, the number of quantization levels is upper bounded by a given number $W$. We prove the conjecture in two extreme situations: i) extremely high correlation and ii) extremely low correlation. Based on the conjecture, the optimal distributed quantization level density in one period can be computed by solving the usual (nondistributed) quantization level density problem for a new source. The overall RD tradeoff is then obtained by considering all $(\Delta, W)$ pairs. For variable-length coding, we look only at the case where the level density is periodic, as the analysis becomes extremely difficult otherwise.

The periodic characteristic of optimal quantization was previously postulated in [15], although the authors did not consider the high resolution case. They transformed the original pdf of the source, by periodizing and truncating. They performed Lloyd quantization on the new pdf and then “periodize” the quantizer without proof of optimality of periodicity. Servetto [16] also explored periodic nonuniform quantization. However, due to the erroneous analysis of nonuniform quantization in [16], the resultant periodic quantizer was not optimal.

B. Zero- and Low-Delay Coding for Sources With Memory

Most existing work on distributed source coding focused on memoryless sources. To exploit the temporal correlation typically found in many sources, we extend our results to sources with memory by applying prediction filters to both the source and the side information in order to improve efficiency. Previously, [17] exploited memory in source coding by applying predictive lattice quantization and Wyner–Ziv DPCM. The author used the prediction filter at the decoder to treat memory as “side information.” Generalizing the technique of asymptotic closed-loop predictive quantization to distributed source coding, [18] proposed an iterative algorithm whereby binning is included in the loop. In contrast, we place binning after the prediction loop and minimize the combination of the resultant “overload” distortion and the granular distortion obtained prior to binning. After a high-resolution analysis, we arrive at the same conclusion reached immaturely in [19], i.e., that optimal prediction in the Wyner–Ziv regime is fundamentally different from that in nondistributed coding in both fixed-length and variable-length scenarios. The difference stems from two factors: i) the prediction filters must strike the optimal balance between reducing temporal redundancy and keeping spatial redundancy to maximize coding efficiency and ii) at the same time, the prediction filter of the source must suppress the propagation of decoding errors to prevent catastrophic distortion. Unfortunately, as was also shown in [20] for zero-delay Wyner–Ziv coding of memoryless sources, our analysis shows that at very high rates, predictive Wyner–Ziv coding has no advantage over its nondistributed counterpart. Fortunately, simulations agree with the high-resolution analysis even at moderate (i.e., practically useful) rates where there is still significant advantage over nondistributed coding.

We then turn to transform coding, where we transform both the source and the side information blocks (of relatively short length) and treat each corresponding transform coefficient pair as a separate scalar Wyner–Ziv coding problem. The justification for this approach is that if the source and the side information are highly correlated spatially, then the resultant coefficient pairs will be almost uncorrelated temporally. In fact, if the joint structure of the source and the side information is such that both are noisy observations of a stationary Gaussian source, then this decorrelation is exact. More specifically, if we apply the (ordinary) Karhunen-Loève transform (KLT) designed for the underlying source to both the observed source and the side information, the corresponding transform coefficient pairs will be independent. It should be noted that in this case, although the resultant KLT coincides with the “conditional KLT” as introduced by Gastpar et al. [21], the subsequent bit allocation and coding are different because in [21], transform coefficients from a large number of consecutive blocks are assumed to be jointly encoded in an asymptotically RD-optimal manner. Our approach is also different from that in [22], [23], where the authors studied high-rate scalar quantization of transform coefficients, because ideal Slepian–Wolf coding was assumed for the quantization indices, which still implies very large block lengths.

Our numerical results show that the gap between transform distributed and nondistributed coding is larger than that between predictive distributed and nondistributed coding. Interestingly, when the underlying source is a first-order Gauss–Markov process, distributed transform coding even with a small block length performs better than first-order distributed predictive coding. This performance gap is more prominent in fixed-length coding than in variable-length coding. In contrast, for nondistributed coding, the performance of transform coding could catch up with that of predictive coding only after a block length of approximately 50, because first-order prediction of such a source is already very successful in eliminating the temporal redundancy.

The rest of this paper is organized as follows. We begin in the next section with optimal fixed-length scalar quantization, and continue with optimal periodic variable-length quantization in Section III. We then incorporate the results in predictive and transform coding in Sections IV and V, respectively. In Appendix A, we briefly discuss the asymptotic Wyner–Ziv rate-distortion function for a certain class of stationary Gaussian sources. Finally, the paper is concluded in Section VI.

II. High-Resolution Fixed-Length WZ Scalar Quantization

We assume that the source-side information pair $(X, Y) \in \mathbb{R} \times \mathbb{R}$ has a probability density function (pdf) that satisfies $p_{XY}(x, y) > 0$ everywhere. Further, let the conditional pdf $p_{X|Y}(x|y)$ be a smooth function peaking around $E[X|Y = y]$. A fixed-length code consists of an encoder-decoder pair

$$
\phi : \mathbb{R} \rightarrow \{0, 1\}^R
$$

$$
\psi : \{0, 1\}^R \times \mathbb{R} \rightarrow \mathbb{R}
$$

(1)
Therefore, if the reconstruction levels 
formly for all the bins. Further, to ensure 
are at most

$$D = \mathbb{E}[d(X, \psi(X(Y)))].$$

A. Encoder–Decoder Scheme

As shown in Fig. 1, we use encoders of the form \( \psi = I_{SW} \circ Q \), where:

1) the quantizer \( Q : \mathbb{R} \to \mathbb{Z} \) is a nearest neighbor partition with reconstruction levels \( \hat{x}_j \); we use the notation \( X = Q(X) \), and sometimes abuse the notation by using \( Q^{-1}(j) \) instead of \( \hat{x}_j \);

2) the Slepian–Wolf (SW) mapping \( I_{SW} : \mathbb{Z} \to \{0, \ldots, W - 1\} \) maps quantizer indexes to transmission indexes (i.e., “bins”). The corresponding rate becomes \( R = \log_2 W \).

We denote by \( \hat{X} \) the decoder output, i.e., \( \hat{X} = \psi(Y) \).

Given the side information \( Y = y \) and \( i = \phi(Y) \), the optimal decoder would compute

$$\psi_i(i, y) = \mathbb{E}[X \mid Y = y, \phi(Y) = i].$$

However, it would be difficult to optimize the end-to-end distortion with respect to the quantizer (and later on, with respect to the prediction filters and transforms) if we use (2) and make no further assumptions about the structure of \( I_{SW} \). To that end, we use the simple binning scheme

$$I_{SW}(j) = j \mod W$$

and the decoding function

$$\psi(i, y) = \arg \min_{\hat{x}_j : I_{SW}(j) = i} [\hat{x}_j - E[X \mid Y = y]].$$

Indeed, when the quantizer is of high resolution, it is true that

$$\phi^{-1}(i) \triangleq \{x : I_{SW}(Q(x)) = i\} \approx \{\hat{x}_j : I_{SW}(j) = i\}.$$ 

Therefore, if the reconstruction levels \( \hat{x}_j \) with \( I_{SW}(j) = i \) are sufficiently far apart, (3) approximately behaves as the maximum a posteriori estimator for \( X \), thus becoming a competitive alternative to the mean-square estimator (2). The special structure of \( I_{SW} \) above enforces maximum such separation uniformly for all the bins. Further, to ensure \( \hat{X} = \hat{X} \) with high probability, we enforce that for some large enough \( \Delta > 0 \), there are at most \( W \) reconstruction levels in any interval of length \( 2\Delta \) on the real axis. Based on (3), \( \hat{X} = \hat{x}_j \) is then correctly decoded when

$$[\hat{x}_j - E[X \mid Y = y]] < \Delta$$

or since \( x \approx \hat{x}_j \), approximately when

$$|x - E[X \mid Y = y]| < \Delta.$$

When \((X, Y)\) is a pair of zero-mean jointly Gaussian random variables whose covariance matrix is

$$\begin{bmatrix} \sigma_X^2 & \rho \sigma_X \sigma_Y \\ \rho \sigma_X \sigma_Y & \sigma_Y^2 \end{bmatrix}$$

the conditional pdf \( p_{X \mid Y}(x \mid y) \) is also Gaussian with mean

$$E[X \mid Y = y] = \frac{\rho \sigma_X}{\sigma_Y} y$$

and variance

$$\text{Var}[X \mid Y = y] = \sigma_X^2 (1 - \rho^2).$$

It can be seen from (4), (5), and (6) that to achieve a high probability of correct decoding of \( \hat{X} \), one needs to set

$$\Delta = \beta \sigma_X \sqrt{1 - \rho^2}$$

with some appropriate large \( \beta > 0 \), resulting in

$$\Pr[\hat{X} = \hat{X}] = \text{erf} \left( \frac{\beta}{\sqrt{2}} \right).$$

We refer to \( \beta \) as the loading factor of the quantizer, as it plays the same role as that of a quantizer in nondistributed source coding.

B. Distortion Analysis

As is well known (see, for example, [26]), the quantizer design problem at high resolution for a given rate \( R \) can be formulated in terms of the density of quantization levels \( \lambda(x) \). Since the number of levels can be infinite, let this be the “raw” density corresponding to the actual numbers of levels.\(^2\) The density must satisfy

$$\int_{x-\Delta}^{x+\Delta} \lambda(x') \, dx' \leq W, \quad \forall x \in \mathbb{R}.$$ (7)

Letting \( Z = X - \hat{X} \), the distortion incurred on \( X \) is then given by

$$D = \mathbb{E}[(\hat{X} - X)^2] = \mathbb{E}[(\hat{X} - \hat{X} + \hat{X} - X)^2] = \mathbb{E}[Z^2] + \mathbb{E}[(\hat{X} - \hat{X})^2] + 2\mathbb{E}[(\hat{X} - \hat{X})Z]$$

where

$$D_g = \mathbb{E}[Z^2] \approx \int_{-\infty}^{\infty} \frac{p(X)}{12\lambda(x)^2} \, dx$$

\(^2\)That is, \( \lambda(x) = \lim_{N \to \infty} \lambda(x) \) instead of the more commonly used \( \lambda(x) = \lim_{N \to \infty} (N(x)/N) \), where \( N(x) \, dx \) is the number of quantization levels in the interval \((x, x + dx)\).
is the granular distortion correspond to correct decoding of \( \hat{X} \), and
\[
D_o = E[(\hat{X} - \bar{X})^2]
\]
is the overload distortion resulting from a possible decoding error due to binning. The cross-term \( 2E[\hat{X} - \bar{X} \mid Z] \) is negligible compared to \( D_o \), because
\[
\begin{align*}
|E[(\hat{X} - \bar{X})Z]| & \leq \|E[\hat{X} - \bar{X}] \mid Z\| \cdot \|E[\hat{X} - \bar{X} \mid Z] \| \cdot \Pr[\hat{X} - \bar{X} \neq 0] \\
& \leq E[(\hat{X} - \bar{X})^2 \mid \hat{X} - \bar{X} \neq 0] \Pr[\hat{X} - \bar{X} \neq 0] \\
& = E[(\hat{X} - \bar{X})^2].
\end{align*}
\]
That, in turn, is because for large \( \Delta \), the decoding error \( \hat{X} - \bar{X} \) is much larger than the high-resolution quantization error \( Z \) in magnitude. Therefore,
\[
D \approx D_g + D_o. \tag{9}
\]

1) Analysis of the Granular Distortion \( D_g \): Even in the simplified framework of (9), choosing the optimal \( \lambda(x) \) minimizing \( D_g + D_o \) is a tedious task. Instead, we strive to choose \( \lambda(x) \) so as to minimize \( D_o \) only. To that end, we start with a conjecture.

Conjecture 1: The optimal \( \lambda^*(x) \) which minimizes (8) subject to the constraint (7) is periodic with period \( 2\Delta \). In both [15] and [16], the authors also assumed a periodic quantizer. When \( \lambda(x) \) is periodic, we have
\[
D_g = \sum_{k \in \mathbb{Z}} \int_{-\Delta}^{\Delta+2k\Delta} \frac{p_X(x)}{12\lambda(x)^2} dx
\]
using which we can prove the following lemma:

Lemma 1: The optimal periodic quantizer with period \( 2\Delta \) is given by
\[
\lambda^*(x) = \frac{W}{C_{X,\Delta}} \left( \sum_{k \in \mathbb{Z}} p_X(x + 2k\Delta) \right)^{1/3}
\]
with
\[
C_{X,\Delta} = \int_{-\Delta}^{\Delta} \left( \sum_{k \in \mathbb{Z}} p_X(x + 2k\Delta) \right)^{1/3} dx
\]
which, using \( W = 2^R \), results in
\[
D_g(R) = \frac{(C_{X,\Delta})^3}{12} 2^{-2R}. \tag{11}
\]

Proof: Define
\[
q_X(x) = \sum_{k \in \mathbb{Z}} p_X(x + 2k\Delta)
\]
and a corresponding random variable \( \hat{X} \) such that
\[
p_X(x) = \begin{cases} q_X(x), & -\Delta \leq x \leq \Delta \\ 0, & \text{otherwise} \end{cases}
\]
The granular distortion can then be rewritten as
\[
D_g = \int_{-\Delta}^{\Delta} \frac{q_X(x)}{12\lambda(x)^2} dx = \int_{-\Delta}^{\Delta} \frac{p_X(x)}{12\lambda(x)^2} dx.
\]
It then follows from [26, Sec. 6.3] and the periodicity of the quantizer that
\[
\lambda^*(x) = W - \frac{q_X(x)^{1/3}}{\int_{-\Delta}^{\Delta} q_X(x')^{1/3} dx'}
\]
where the extra \( W \) appears because \( \lambda^*(x) \) is the raw density, and must therefore integrate to \( W \) over the interval \([ -\Delta, \Delta ] \).

For general (i.e., possibly nonperiodic) \( \lambda(x) \), one can observe using well-known techniques in calculus of variations that the necessary and sufficient conditions for any \( \lambda(x) \) to minimize (8) subject to the constraint (7) is given by
\[
\int_{-\infty}^{\infty} \frac{p_X(x)}{(\lambda(x) + \epsilon \mu(x))^2} dx \geq \int_{-\infty}^{\infty} \frac{p_X(x)}{\lambda(x)^2} dx
\]
for arbitrary \( \epsilon \geq 0 \) and \( \mu(x) \) satisfying
\[
\int_{x-\Delta}^{x+\Delta} (\lambda(x') + \epsilon \mu(x')) dx' \leq W, \quad \forall x \in \mathbb{R}.
\]
This translates to
\[
d \frac{d}{dx} \int_{-\infty}^{\infty} \frac{p_X(x)}{(\lambda(x) + \epsilon \mu(x))^2} dx \bigg|_{\epsilon = 0} \geq 0
\]
or, taking the derivative,
\[
\int_{-\infty}^{\infty} \frac{p_X(x)}{\lambda(x)^3} \mu(x) dx \leq 0 \tag{12}
\]
for any \( \mu(x) \) satisfying
\[
\int_{x-\Delta}^{x+\Delta} \mu(x') dx' \leq 0, \quad \forall x \in \mathbb{R}. \tag{13}
\]
Now, substituting (10) into (12), the necessary and sufficient condition for \( \lambda^*(x) \) to indeed be the optimal density function becomes
\[
\int_{-\infty}^{\infty} \frac{p_X(x)}{q_X(x)} \mu(x) dx \leq 0 \tag{14}
\]
for any \( \mu(x) \) satisfying (13). It is easy to observe that if equality in (13) holds, which implies \( \mu(x) \) is periodic with period \( 2\Delta \), then equality in (14) would automatically be satisfied.
Although it proved difficult to show (14), we could verify it in the following two extreme situations.

1) When \( \Delta \rightarrow \infty \), we are effectively performing nondistributed coding. If (13) holds, (14) would naturally be true, because

\[
\lim_{\Delta \to \infty} \frac{p_X(x)}{q_X(x)} = 1
\]

uniformly for all \( x \).

2) When \( \Delta \rightarrow 0 \), (13) implies that \( \mu(x) \leq 0 \) for all \( x \in \mathbb{R} \). Hence, (14) is automatically satisfied. The significance of this extreme case is that it corresponds to \( \rho \rightarrow 1 \).

2) Analysis of the Overload Distortion \( D_o \): Defining

\[ S = \hat{X} - E[X | Y] \]

the overload distortion can be written as

\[ D_o \approx 4\Delta^2 \sum_{k \in \mathbb{Z}, k \neq 0} k^2 p_T[(2k-1)\Delta \leq S \leq (2k+1)\Delta] \]  

(15)

which follows from \( \hat{X} - \hat{X} = -Q_{2\Delta}(S) \), where \( Q_{2\Delta} \) is the uniform quantizer rounding to the nearest integer multiple of \( 2\Delta \). This is due to the periodic property of quantizer as we discussed above. See Fig. 2 for an illustration.

3) Comparison With Related Work: In [16], Servetto defined a scaling factor \( s \), which plays the same role as \( \Delta \) here, and argued that for Gaussians it should satisfy

\[
\lim_{\rho \to 1} s(\rho) = 0 \quad \text{and} \quad \lim_{\rho \to 1} \frac{s(\rho)}{\sigma_X \sqrt{1 - \rho^2}} = \infty. \]  

(16)

Our method is different in several aspects. First, we control \( \Delta \) using \( \beta \) for any \( \rho \), whereas [16] fixes \( s(\rho) \). Second, (17) implies that \( s(\rho) \) shrinks with a rate less than that of \( \sigma_X |Y| \), whereas in our case, the shrinking rate is controlled by the optimum \( \beta \).

Finally, [16] focuses on \( \rho \rightarrow 1 \), while we target all values of \( \rho \). Nevertheless, in the special case of \( \rho \rightarrow 1 \), both granular and overload distortion becomes more tractable analytically in our framework as well. In particular, as \( \rho \rightarrow 1 \), it is clear that \( \Delta \rightarrow 0 \), in which case \( p_X(x) \) can be approximated as the uniform distribution. That, in turn, implies that \( \lambda^*(x) \) is constant, i.e., the optimal quantizer is uniform, and the corresponding granular distortion can be found as

\[ D_g = \frac{\Delta^2}{3} 2^{-2R}. \]  

Assuming a large enough \( \beta \), \( D_o \) can also be approximated as

\[ D_o = E[(\hat{X} - \hat{X})^2] \approx (2\Delta)^2 \left[ 1 - \text{erf} \left( \frac{\beta}{\sqrt{2}} \right) \right]. \]

Adding the two distortion values together, we obtain

\[ D = 4\Delta^2 (1 - \rho^2)^2 \left[ \frac{2 - 2R}{12} + 1 - \text{erf} \left( \frac{\beta}{\sqrt{2}} \right) \right]. \]  

(18)

The behavior of (18) as \( \beta \) varies is shown in Fig. 3. The optimal \( \beta \) values shown on Fig. 3 coincide with what we observed experimentally for very large \( \rho \), thus validating (18).

Turning back to general \( \rho \), we see that our \( \lambda^*(x) \) differs from the derivation in [16], which, although periodic, was given by

\[ \lambda(x) = W \frac{p_X(x)0^{1/\beta}}{\int_{-\Delta}^{\Delta} p_X(x)0^{1/\beta} dx}. \]  

(19)

in the interval \( [-\Delta, \Delta] \). In Fig. 4, we compare the performance of our scheme with that in [16]. First of all, as can be seen from the figure, not only is our scheme better, but the performance gap between the two schemes widens as \( R \) increases. The second observation is that the gap between the Wyner–Ziv RD curve and the achieved performance of either scheme diverges as \( R \rightarrow \infty \).

Note that this contrasts with the claim in [16] that the latter gap converges to \( G_1 \) as \( \rho \rightarrow 1 \), where \( G_1 \) is a constant satisfying \( (1/12) \leq G_1 \leq (1/2) \). This discrepancy, as well as the performance gap between the two schemes, is probably because (19) was obtained by erroneously applying the distortion expression derived for uniform quantization in [16, Sec. III.B] to nonuniform quantization in [16, Sec. III.C].

III. HIGH RESOLUTION VARIABLE-LENGTH WZ SCALAR QUANTIZATION

For variable length coding, on top of the fixed-length encoder \( I_{SW} \circ Q \), with \( Q: \mathbb{R} \rightarrow \mathcal{Z} \) and \( I_{SW}(j) = j \) \#\( W \), we also employ a prefix-free mapping \( \tilde{ia}_1: [0, 1, \ldots, W-1] \rightarrow \{0, 1\}^* \). At the receiver, the variable length codeword is first decoded.

Fig. 2. A decoding error example. The correct quantized value \( \hat{X} \) is missed because the estimate of \( X \) is closer to \( \hat{X} - 2\Delta \). The overload error is \( -Q_{2\Delta}(S) \approx -2\Delta \). The triangles represent all the reconstruction levels \( \hat{X}_j \) with \( I_{SW}(j) = i \), where \( i \) is the transmission index.

Fig. 3. Behavior of the total distortion as \( \beta \) varies when \( \rho \rightarrow 1 \). Optimal values of \( \beta \) for various \( R \) are also shown.
Since the rate with variable-length coding is a complicated function of the quantizer level density, minimizing the granular distortion for a fixed rate seems to be extremely difficult. Instead, motivated by Conjecture 1, we only look at the case where the level density is periodic. A by-product of this is that the overload distortion is still given by (15). The optimal granular distortion is derived in the next lemma.

Lemma 2: The optimal periodic level density for entropy coded WZ quantization is uniform and the resulting granular distortion is

\[
D_g(R) = \frac{c^2 h(X)}{12} 2^{-2R} 
\]

where

\[
h(\hat{X}) = -\int_{-\Delta}^{\Delta} p_X(x) \log p_X(x) dx.
\]

Proof: The proof follows the same lines as in [26, Sec. 9.9] after observing that, as in the fixed-length case, designing a periodic WZ quantizer for \((X, Y)\) is no different than designing a nondistributed quantizer for \(\hat{X}\). This observation leads to

\[
D \approx \frac{1}{12W^2} \int_{-\Delta}^{\Delta} \frac{p_X(x)}{\lambda(x)^2} dx = \frac{1}{12} E \left[ \frac{1}{(W \lambda(X))^2} \right]
\]

and

\[
H(\hat{X}) \approx h(\hat{X}) - E \left[ \log \frac{1}{W \lambda(X)} \right]
\]

where \(\hat{\lambda}(x) = (\lambda(x)/W)\).

IV. HIGH-RESOLUTION PREDICTIVE WZ CODING OF GAUSSIAN SOURCES

A. Fixed-Length Predictive Coding

Consider two jointly Gaussian processes \(X(n)\) and \(Y(n)\). We strive for optimal exploitation of time and space correlation by performing scalar Wyner–Ziv coding of \(X(n)\) after passing both \(X(n)\) and \(Y(n)\) through appropriate first-order “prediction” filters \(A(z) = 1 - az^{-1}\) and \(B(z) = 1 - bz^{-1}\), respectively, as shown in Fig. 5. The filters need not perform prediction in the classical sense of minimizing the variance of the prediction error. Instead, they need to strike the optimal balance between low time correlation and high space correlation. We denote the filter outputs by \(E_X(n)\) and \(E_Y(n)\).³

The conditional pdf \(P_{E_X(n) | E_Y(n)}(e_X | e_Y)\) is Gaussian with mean

\[
E[E_X(n) | E_Y(n) = e_Y] = \frac{\rho_{E_X E_Y} \sigma_{E_X}}{\sigma_{E_Y}} e_Y 
\]

and variance

\[
\text{VAR}[E_X(n) | E_Y(n) = e_Y] = \sigma_{E_X}^2 \left( 1 - \rho_{E_X E_Y}^2 \right) 
\]

³In the high-resolution quantization regime, the open-loop predictor \(A(z)\) is a good approximation to the actual closed-loop predictor.
the overload distortion, we could define prediction and reconstruction filters. That is because an encoder—even though we perform closed-loop prediction with matching

Similar to CHEN AND TUNCEL: LOW-DELAY PREDICTION- AND TRANSFORM-BASED WYNER–ZIV CODING 659

We adopt the same periodic quantization regime as described as before.

S(n) \triangleq \mathbb{E}[X(n) - \hat{X}(n)] - \mathbb{E}[E_X(n)E_Y(n)] = \mathbb{E}[X(n) - \hat{X}(n)] - \frac{\rho E_x E_y}{\sigma_{E_x} \sigma_{E_y}} E_Y(n)

\approx E_X(n) - \frac{\rho E_x E_y \sigma_{E_x}}{\sigma_{E_Y}} E_Y(n), \quad (23)

We adopt the same periodic quantization regime as described in the previous section, and optimize over \(a, b,\) and \(\beta.\) Unfortunately, unlike in nondistributed predictive source coding, we do not enjoy the relation

\[ \mathbb{E}[(X(n) - \hat{X}(n))^2] = \mathbb{E}[(E_X(n) - \hat{E}_X(n))^2] \]

even though we perform closed-loop prediction with matching prediction and reconstruction filters. That is because an encoder-decoder mismatch occurs whenever \(|S(n)| > \Delta.\) On the other hand, we can redraw the block diagram of our algorithm as in Fig. 6, from which it can be seen that

\[ \hat{X}(n) = [(X(n) + Z(n)) * a(n) - \hat{S}(n)] * a^{-1}(n) = X(n) + Z(n) - \hat{S}(n) \]

\[ (24) \]

where \(\star\) denotes convolution, \(a(n)\) and \(a^{-1}(n)\) are respectively the impulse responses of the prediction filter \(A(z) = 1 - a z^{-1}\) and the reconstruction filter \(A^{-1}(z) = 1/(1 - a z^{-1})\), and

\[ \hat{S}(n) = \hat{S}(n) * a^{-1}(n). \]

However, (24) together with the high resolution assumption implies

\[ D = \mathbb{E}[(\hat{X}(n) - X(n))^2] \approx \mathbb{E}[Z(n)^2] + \mathbb{E}[\hat{S}(n)^2] = D_g + D_o. \]

where, the cross-term \(\mathbb{E}[Z(n)\hat{S}(n)]\) vanishes as before.

It is a tedious task to choose the optimal \(\lambda(e_X)\) minimizing \(D_g + D_o.\) Instead, we choose \(\lambda(e_X)\) so as to minimize \(D_g\) only, and as discussed above, it becomes

\[ \lambda(e_X) = \frac{W}{C_{E_x, \Delta}} \left( \sum_{k \in \mathbb{Z}} p_{E_X}(e_X + 2k\Delta) \right)^{1/3} \]

for any \(e_X \in \mathbb{R},\) where

\[ C_{E_x, \Delta} \triangleq \int_{-\Delta}^{\Delta} \left( \sum_{k \in \mathbb{Z}} p_{E_X}(e_X + 2k\Delta) \right) \, de_X. \]

The corresponding minimum granular distortion is then given by

\[ D_g \approx \frac{(C_{E_x, \Delta})^3}{12} 2^{-2R}. \]
We have from (25) that
\[ E[\hat{S}(n)^2] = R_{SS}(0) = \sum_{\tau = -\infty}^{\infty} \frac{|a|^2 R_{SS}(\tau)}{1 - a^2} \]
\[(28)\]
where
\[ R_{SS}(\tau) = 4\Delta^2 \sum_{i=\infty}^{\infty} \sum_{j=\infty}^{\infty} ij \times \Pr[\hat{S}(n) = 2i\Delta, \hat{S}(n - \tau) = 2j\Delta] \]
which reduces to the unfiltered overload distortion in (15) when \( \tau = 0 \), as it should. To evaluate
\[ \Pr[\hat{S}(n) = 2i\Delta, \hat{S}(n - \tau) = 2j\Delta] = \int_{(2i+1)\Delta}^{(2i+1)\Delta} \int_{(2j+1)\Delta}^{(2j+1)\Delta} p_S(s_1, s_2) ds_1 ds_2 \]
we use the approximation (23) which makes \( S(n) \) a Gaussian process with zero-mean and autocorrelation
\[ R_{SS}(\tau) = R_{E}\sigma_{E}\sigma_{S} + \frac{2\sigma_{S}^2}{\sigma_{E}^{2}} R_{EY} (\tau) \]
\[-\frac{\sigma_{E}^{2}}{\sigma_{S}^2} \] \[ R_{EY} (\tau) + R_{EY} (\tau) \].

It can be deduced from (28) that if \( |a| \approx 1 \), (i) even occasional decoding errors will propagate for a long time, and (ii) these errors will be amplified immensely by a factor \( 1 / (1 - a^2) \). Conversely, if \( a \approx 0 \), then \( R_{SS}(0) \approx R_{SS}(0) \), and overload distortion will be given by (15), i.e., binning errors will be forgotten quickly and will not be amplified.

The choice of \( a \), \( b \), and \( \beta \) affect both \( D_{\text{y}} \) and \( R_{SS}(\tau) \) in so complex a manner that analytical optimization is an almost hopeless task. We instead use experimental analysis and perform a brute-force search in the space \( -1 < a, b < 1 \) and \( \beta \geq 2.5 \). In our experiment, we use the following model. Let \( T(n) \) be a first-order Gauss-Markov process. That is,
\[ T(n) = \rho T(n - 1) + W(n) \]
where \( W(n) \) is i.i.d. zero-mean Gaussian, \( W(n) \perp T(n - 1) \), and \( \sigma_{S}^{2} = 1 - \rho^2 \) so that \( \sigma_{S}^{2} = 1 \). Also assume that \( T(0) \) has zero mean and unit variance so that the process is stationary.

Let the source \( X(n) \) observed at the encoder and the side information \( Y(n) \) observed at the decoder be noisy observations of \( T(n) \), i.e.,
\[ X(n) = T(n) + U(n) \] \[ Y(n) = T(n) + V(n) \]

where \( U(n) \) and \( V(n) \) are i.i.d. zero-mean Gaussians independent of each other.

Fig. 7(a)–(c) shows the resultant high-resolution rate-distortion performance for three sample cases in which the space correlation is high enough. Also shown on the same figures are the high-resolution rate-distortion performances when i) the naive choice of filter parameters \( a = \rho / (1 + \sigma_{S}^{2}) \) and \( b = \rho / (1 + \sigma_{S}^{2}) \) is used, and ii) the side information at the decoder is ignored. Optimal \( \beta \) increases without bound as we climb up the rate-distortion curve. Thus, as \( R \to \infty \), the scalar Wyner–Ziv coder reduces to a scalar nondistributed coder.

\footnote{This choice minimizes \( \sigma_{S}^{2} \) and \( \sigma_{S}^{2} \) simultaneously, and corresponds to the optimal first-order prediction filters in nondistributed source coding.}
explaining why all three curves meet at high rates, and also why the optimal choice of \( a \) approaches the naive one. Since the high-resolution assumption is accurate only at high rates, one may wonders whether any of our analysis is useful. To address this, we implemented actual quantizers corresponding to \( \lambda(e_X) \) and simulated the coding of sequences of length \( 1000 \, \text{samples} \). Fortunately, as shown also on Fig. 7(a)–(c), the simulated performance is not only always better than that of nondistributed source coding, but also catches the theoretical performance at moderate rates where there is still room for Wyner–Ziv coding gain.

On the low rate side, since optimal \( \beta \) is also low, the decoding error probability becomes high, resulting in the need to quickly forget the errors as discussed above. This explains the very low values \( a \) assumes at low rates.

Finally, Fig. 7(a)–(c) also shows the corresponding asymptotically optimal rate-distortion performances, i.e., when arbitrarily large block-codes are allowed. We defer the detailed computation of the asymptotic rate-distortion curve to Appendix A.

### B. Variable-Length Predictive Coding

We use the same coding technique as in Fig. 6. Therefore, the overload distortion \( D_o \) is the same as in fixed-length predictive coding. However, the granular distortion \( E(Z^2(n)) \) has to be reformulated. Following (20),

\[
D_g = E(Z^2(n)) = \frac{\epsilon_X^2}{12}2^{-2R}
\]

where

\[
h(\bar{E}_X) = -\int_{-\Delta}^{\Delta} p_{E_X}(e_X) \log p_{E_X}(e_X) \, de_X
\]  

(31)

and

\[
p_{E_X}(e_X) = \begin{cases} 
\frac{e_X^2}{\Delta^2} & \text{for } -\Delta \leq e_X \leq \Delta \\
0 & \text{otherwise} 
\end{cases}
\]

Fig. 8(a)–(c) shows the resultant high-resolution rate-distortion performance for the same sample cases as in fixed-length predictive coding. The behavior of optimal \( a \) and \( \beta \) is similar to what we observed in fixed-length coding. Also similar are i) how the gap between optimal distributed and nondistributed coding performances vanish as the rate increases, and ii) how the simulated performance is superior to nondistributed coding even in low rates.

We also compare the performance of fixed- and variable-length scenarios in Fig. 9 for the same three cases. In the low rate regime, the performance of fixed- and variable-length distributed schemes is almost identical. This is expected, as in the low rates, \( \beta \) (and thus \( \Delta \)) tends to be low, which results in an almost uniform \( p_{E_X} \). On the other hand, at high rates, although there is a significant gap between the two, both converge to their nondistributed counterparts as \( \beta \) becomes very large. It is interesting to observe that even at medium rates, where there is potential for variable-length distributed coding to have a significant gain over both fixed-length distributed coding and
variable-length distributed coding is approximately the same as the upper envelope of those of fixed-length distributed and variable-length nondistributed coding schemes.

V. HIGH RESOLUTION TRANSFORM WZ CODING OF GAUSSIAN SOURCES

Our transform coding model is as shown in Fig. 10. In this section, we exploit the structure of \( X(n) \) and \( Y(n) \) as given in (29) and (30) with arbitrary stationary Gaussian \( T(n) \) to design optimal transform matrices \( \mathbf{A} \) and \( \mathbf{B} \). More specifically, the \( k \times k \) covariance matrices \( \mathbf{C}_{XX}, \mathbf{C}_{YY}, \) and \( \mathbf{C}_{XY} \) can be written as

\[
\mathbf{C}_{XX} = \mathbf{C}_{TT} + \sigma_t^2 \mathbf{I} \\
\mathbf{C}_{YY} = \mathbf{C}_{TT} + \sigma_v^2 \mathbf{I} \\
\mathbf{C}_{XY} = \mathbf{C}_{TT}.
\]

In the transform domain, the corresponding covariance matrices become

\[
\mathbf{C}_{XX'} = \mathbf{A} \mathbf{C}_{TT} \mathbf{A}^{-1} + \sigma_t^2 \mathbf{I} \\
\mathbf{C}_{YY'} = \mathbf{B} \mathbf{C}_{TT} \mathbf{B}^{-1} + \sigma_v^2 \mathbf{I} \\
\mathbf{C}_{XY'} = \mathbf{A} \mathbf{C}_{TT} \mathbf{B}^{-1}.
\]

Thus, setting both \( \mathbf{A} \) and \( \mathbf{B} \) to the Karhunen–Loève transform (KLT) for the vector source \( T(1), T(2), \ldots, T(k) \), one can make all three matrices diagonal. This, in turn, implies that the transform coefficient pairs \( (X'_1, Y'_1), (X'_2, Y'_2), \ldots, (X'_k, Y'_k) \) are independent of each other, and therefore, should be coded separately after proper allocation of the available bits.

It should be noted that in this case, although our transform coincides with the “conditional KLT” for \( X(n) \) given \( Y(n) \) as introduced by Gastpar et al. [21], the subsequent bit allocation and coding is different because in [21], transform coefficients from a large number of consecutive blocks are assumed to be jointly encoded in an asymptotically RD-optimal manner.

The average distortion on \( X(1), \ldots, X(k) \), or equivalently on \( X'_1, \ldots, X'_k \), can be written as

\[
D(b_1, b_2, \ldots, b_k) = \frac{1}{k} \sum_{i=1}^{k} D_i(b_i)
\]

where \( b_1, \ldots, b_k \) are the number of bits allocated to component \( i \) so that

\[
\sum_{i=1}^{k} b_i = kR
\]

and \( D_i(b_i) \) is the distortion of the \( i \)th transform coefficient, which, as before, comprises of granular and overload components:

\[
D_i(b_i) = D_{g,i}(b_i) + D_{o,i}.
\]

Depending on whether fixed- or variable-length coding is used, \( D_{g,i}(b_i) \) can be calculated as in (11) or (20), respectively.
using $X'_i$ in place of $X$. Similarly, $D_{D_i}$ is given as in (15) using

$$S'_i = X'_i = \frac{\rho X'_i Y'_i \sigma X'_i}{\sigma Y'_i}$$

in place of $S$. Of course, for each integer $b_k$, the parameter $\Delta$ should be chosen so as to minimize $D_{D_i}(b_k)$.

We convexify $D_{D_i}(b_k)$ so that $D(b_1, b_2, \ldots, b_k)$ becomes convex as well, in which case the bit allocation problem

Minimize $\sum_{i=1}^{k} D_i(b_i)$

subject to $\sum_{i=1}^{k} b_i = kR$  \hspace{1cm} (32)

can alternatively be tackled by minimizing the Lagrangian

$$L = \sum_{i=1}^{k} \left[D_i(b_i) + \lambda b_i \right] \triangleq \sum_{i=1}^{k} L_i$$

for every $\lambda > 0$. It is clear that the minimum is achieved by independently minimizing each $L_i$. That, in turn, implies that as we initialize $\lambda$ as a sufficiently large number and gradually decrease it, we can successfully track the optimal $\{b_i\}_{i=1}^{k}$ by the following simple algorithm.

1) Initialize with $b_i = 0$ for all $i$.
2) Calculate $\delta_i = D_i(b_i) - D_i(b_i + 1)$.
3) Find $i^* = \arg \max_{i} b_i$, and increment $b_{i^*}$.
4) Repeat 2) and 3) until $\sum_{i=1}^{k} b_i = kR$.

This algorithm is essentially the same as that given in [26, Sec. 16.5], where the lower envelope of $\sum_{i=1}^{k} [D_i(b_i) + \lambda b_i]$ for all possible bit assignments was treated as a piecewise linear monotonically increasing function of $\lambda$, and singular values of $\lambda$ where two line segments meet are visited until the desired number of total bits is reached. In our algorithm, each maximum $\delta_i$ corresponds to the next singular $\lambda$ to be visited.

For the same Gauss–Markov $T(n)$ as in predictive coding, with $\rho = 0.99$ and $\sigma_{Y_i}^2 = \sigma_{Y_i}^2 = 0.005$, we compare the performance of distributed transform coding to those of predictive coding as well as nondistributed coding in Fig. 11. Fig. 11(a) and (b) shows how the performance of transform coding (both distributed and nondistributed) evolves with increasing block-length for fixed- and variable-length coding, respectively. Fig. 11(c) and (d) compares the performance of transform coding to that of predictive coding, again for fixed- and variable-length coding, respectively. Finally, Fig. 11(e) compares the performance of fixed- and variable-length predictive and transform coding.

First of all, we observe that the gap between distributed and nondistributed transform coding is larger than that between distributed and nondistributed predictive coding. Secondly, distributed transform coding yields better results than distributed predictive coding, contrary to the nondistributed case. This can be explained by the fact that the source $X(n)$ is nearly first-order Markov, and first-order prediction is thus very powerful in nondistributed coding. In fact, we had to increase the block-length to 50 for transform coding to catch up with predictive coding. On the other hand, due to the fact that significant temporal correlation remains after prediction in the distributed case, the performance of predictive coding falls below that of transform coding.

VI. CONCLUSION

We studied zero-delay, i.e., scalar, lossy source coding with side information at the decoder. Instead of imposing a periodic structure on the scalar quantizer directly, we addressed optimal quantization under the constraint that within every interval of
Fig. 11. Transform coding results. (a) and (b) compare the performances with various block-lengths for fixed- and variable-length coding, respectively. (c) and (d) compare the performance of transform coding (with block-length 10) and predictive coding (with order 1) for fixed- and variable-length coding, respectively. Finally, (e) compares fixed- and variable-length coding for both transform and predictive coding.
size $2\Delta$, there are at most $W$ cells. We conjectured that optimal quantizers under this regime are periodic, and proved this conjecture for the important extreme case of a very high correlation coefficient between the source and side information.

When we incorporated our zero-delay coding results into predictive coding for Gaussian sources with memory, we observed that optimal prediction in the Wyner–Ziv regime is fundamentally different from that in nondistributed coding. That is because the prediction filters must not only jointly exploit the temporal and spatial redundancies, but the prediction filter of the source must also suppress the propagation of occasional decoding errors. At low rates, the optimal action turns out to be to allow more decoding errors but prevent their propagation aggressively using a very low prediction coefficient for the source. As the rate increases, it is more advantageous to allow less decoding errors and an increased prediction coefficient. We also employed our scalar codes in transform coding with small block lengths (thereby achieving a low delay), where the source and side information are transformed separately. For the specific source-side information pairs studied, we showed that transform coding, even with a small block-length, outperforms predictive coding, despite the first-order Markov-like structure of the signals.

To keep our results in perspective, we also derived the asymptotic rate-distortion function for the special structure of sources we studied. Because we do not use the ideal Slepian–Wolf coding assumption anywhere in our results, the gap between the asymptotic rate-distortion function and performance of our schemes is more significant than usual.

**APPENDIX A**

**THE ASYMPTOTIC RATE-DISTORTION TRADEOFF**

**Theorem 1:** The Wyner–Ziv rate-distortion function for the pair $(X(n), Y(n))$ described in (29) and (30) for any stationary Gaussian $T(n)$ is parameterized as

$$R_\theta(k) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \max \left[ 0, \log \frac{A(e^{j\omega})}{\theta} \right] d\omega$$

(33)

and

$$D_\theta(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min\left[ \theta, A(e^{j\omega}) \right] d\omega$$

(34)

where

$$A(e^{j\omega}) = \Phi_{XX}(e^{j\omega}) - \frac{\Phi_{XY}(e^{j\omega})^2}{\Phi_{YY}(e^{j\omega})}$$

and $\Phi_{XX}(e^{j\omega})$, $\Phi_{YY}(e^{j\omega})$, and $\Phi_{XY}(e^{j\omega})$ are the discrete-time Fourier transforms (DTFT) of $R_{XX}(\tau)$, $R_{YY}(\tau)$, and $R_{XY}(\tau)$, respectively.

**Proof:** For the given source model, following the same logic as in Section V, we can deduce that by applying the same de-correlating transformation $A$ to both sources, i.e., obtain $X' = AX$ and $Y' = AY$ with $X = [X(1)\ X(2)\ \ldots\ X(n)]^T$ and $Y = [Y(1)\ Y(2)\ \ldots\ Y(n)]^T$, all three covariance matrices $C_{XX}$, $C_{YY}$, and $C_{XY}$ are diagonalized simultaneously. This, in turn, implies that transform coefficient pairs $(X'_k, Y'_k)$ are independent of each other and thus can be coded separately with optimal rate

$$R_\theta(k) = \max \left[ 0, \frac{1}{2} \log \frac{\sigma_{X_k}^2}{\theta} \right]$$

(35)

and distortion

$$D_\theta(k) = \min \left[ \theta, \sigma_{X_k}^2 \right]$$

(36)

for some $\theta > 0$, where the formulas (35) and (36) follow by generalizing the optimal (i.e., water-filling) bit allocation discussed in [25] to Wyner–Ziv rate-distortion function [1]. The relevant quantity $\sigma_{X_k}^2(1 - \rho_{X_k}^2)$ can be expressed as

$$\sigma_{X_k}^2(1 - \rho_{X_k}^2) = \lambda_{XX}^k \left[ \frac{1}{\lambda_{XX}^k} - \frac{(\lambda_{XX}^k)^2}{\lambda_{YY}^k} \right]$$

(37)

where $\lambda_{XX}^k$, $\lambda_{YY}^k$, and $\lambda_{XY}^k$ are the $k$th eigenvalues of $C_{XX}$, $C_{YY}$, and $C_{XY}$, respectively. We observe that $\lambda^k$ is the $k$th eigenvalue of $C = \Delta C_{XX} - C_{XY}^2 C_{YY}^{-1}$.

This follows because $A$ also diagonalizes $C$.

Now, [27, Theorem 4.2] shows that for an infinite Toeplitz matrix $T$ with entries $t_k$ on the top row, and for any function $F$ that is continuous on the range of the eigenvalues $\lambda_k^T$ of $T$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} F(\lambda_k^T) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\Phi(\omega))d\omega$$

where $\Phi(\omega)$ is the DTFT of $t_k$. Further, [27, Theorem 5.2(c)] and [27, Theorem 5.3(b)] state that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} F(\lambda_k^{T_1}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F\left[ \frac{1}{\Phi(\omega)} \right]d\omega$$

and

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} F(\lambda_k^{T_2}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F[\Phi_1(\omega)\Phi_2(\omega)]d\omega$$

respectively, where $\Phi_i(\omega)$, $i = 1, 2$, are the DTFT of the top row of $T_i$. Choosing $F(\lambda^k) = \max[0, (1/2) \log(\lambda^k/\theta)]$ and $F(\lambda^k) = \min[\theta, \lambda^k]$ as in (35) and (36), respectively, yield (33) and (34).

**REFERENCES**


Xuechen Chen received the B.S. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2007. She is currently working towards the Ph.D. degree in the Department of Electrical Engineering, University of California, Riverside.

Her research is mainly on low delay Wyner–Ziv coding and joint source channel coding.

Ertan Tuncel (S’99–M’04) received the Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, in 2002.

In 2003, he joined the Department of Electrical Engineering, University of California, Riverside, where he is currently an Associate Professor. His research interests include rate-distortion theory, multiterminal source coding, joint source-channel coding, zero-error information theory, and content-based retrieval in high dimensional databases.

In 2007, Dr. Tuncel received the National Science Foundation CAREER Award.