Layered Wyner–Ziv Video Coding

Qian Xu, *Student Member, IEEE*, and Zixiang Xiong, *Senior Member, IEEE*

***Abstract*—Following recent theoretical works on successive Wyner–Ziv coding (WZC), we propose a practical *layered* Wyner–Ziv video coder using the DCT, nested scalar quantization, and irregular LDPC code based Slepian–Wolf coding (or lossless source coding with side information at the decoder). Our main novelty is to use the base layer of a standard scalable video coder (e.g., MPEG-4/H.26L FGS or H.263+) as the decoder side information and perform layered WZC for quality enhancement. Similar to FGS coding, there is no performance difference between layered and monolithic WZC when the enhancement bitstream is generated in our proposed coder. Using an H.26L coded version as the base layer, experiments indicate that WZC gives slightly worse performance than FGS coding when the channel (for both the base and enhancement layers) is noiseless. However, when the channel is noisy, extensive simulations of video transmission over wireless networks conforming to the CDMA2000 1X standard show that H.26L base layer coding plus Wyner–Ziv enhancement layer coding are more robust against channel errors than H.26L FGS coding. These results demonstrate that layered Wyner–Ziv video coding is a promising new technique for video streaming over wireless networks.**

***Index Terms*—Distributed source coding, layered coding, MPEG-4/H.26L FGS coding, Slepian–Wolf coding (SWC), wireless video, Wyner–Ziv coding (WZC).**

I. Introduction

THE growing popularity of video streaming over heterogeneous networks has prompted the need for scalable and robust video compression. Video coding standards such as MPEG-4 [1] and H.264 [2] perform well under perfect network conditions, but do not support rate scalability or provide error robustness over packet losses. The MPEG-4 fine granularity scalability (FGS) [3] streaming profile has been introduced to address the heterogeneity issue with client bandwidths. However, the MPEG-4 FGS base layer bitstream is very sensitive to packet losses. This is due to the DPCM paradigm that underlies standard video coding, where packet losses often cause the encoder and the decoder to lose synch, resulting in error drifting/propagation with severe degradation of the video quality.

FGS coding, thus, makes the assumption that the base layer can be delivered error-free with strong forward error correction (FEC) coding in practical streaming applications. However, there is always a probability of error with FEC. Thus, the problem of error drifting is unavoidable in practice with FGS coding due to the strong coupling between the base and enhancement layers. To alleviate the inherent problem with FGS coding and, hence, increase the base layer's error robustness, we are motivated to seek scalable video coding techniques based on distributed source coding (DSC) principles that independently generate the base layer and enhancement layers.

DSC refers to compression of two or more correlated sources that do not communicate with each other. The main issue with DSC is to achieve the same coding efficiency as with joint (e.g., DPCM) encoding. For lossless compression of two discrete correlated sources, Slepian and Wolf [4] showed the surprising result that there is no loss of coding efficiency with separate encoding when compared to joint encoding as long as joint decoding is performed. For the more general case of lossy coding with side information at the decoder, Wyner and Ziv [5] showed that it generally suffers rate loss when compared to lossy coding of the source with the side information available at both the encoder and the decoder. However, one special case of the Wyner–Ziv problem is when the source X and side information Y are zero-mean and stationary Gaussian memoryless sources and the distortion metric is MSE. The minimum bit rate needed to encode X for a given distortion when Y is available only at the decoder is equal to the rate when Y is known at both sides. In other words, there is no rate loss for this quadratic Gaussian case in Wyner–Ziv coding (WZC).

To approach the Wyner–Ziv rate-distortion (R–D) function established in [5], information-theoretic approaches were presented in [6] and several practical coding schemes for ideal jointly Gaussian sources have been proposed (see the two tutorial papers [7], [8] and references therein).

Several groups have recently explored video compression based on DSC principles. One approach targets emerging applications (e.g., "uplink" video communications from handheld devices) that demand low encoding complexity – a scenario that is the opposite of video broadcast, for which standard coders with heavy encoding are designed. Puri and Ramchandran proposed a coder in [9] that attempts to swap the encoder-decoder complexity of standard coders. Their encoder consists of the DCT, uniform quantization and trellis coding for Slepian–Wolf compression, while their decoder performs heavy-duty motion estimation. Girod *et al.* [10] also investigated distributed video coding using a relatively low-complexity turbo code based Slepian–Wolf encoder. Whereas both coders perform better than independent intraframe (e.g., motion JPEG) coding (with the lowest encoding complexity), they suffer substantial R–D penalty when compared to standard MPEG-4 and H.264 coding (with high encoding complexity, mainly due to motion estimation). Thus, there is still a large gap between what WZC or DSC theory promises (to the extent of no performance loss in certain special cases when compared to joint encoding) and

Manuscript received July 13, 2004; revised June 11, 2006. This work was supported by the National Science Foundation under Grant CCF-0430720. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Aria Nosratinia.

The authors are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: qianxu@ece.tamu.edu; zx@ece.tamu.edu).

Color versions of Figs. 4 and 8–10 are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2006.884925

what practical low-complexity distributed video coders can achieve.

Another approach is to de-emphasize low-complexity encoding while focusing on error robust Wyner–Ziv video coding. For example, Sehgal *et al.* [11] discussed how coset-based Wyner–Ziv video coding can be used to alleviate the problem of prediction mismatch in DPCM-based standard video coders. Their coder is "state free" in the sense that the decoder does not have to maintain the same states as the encoder. Girod *et al.* [10] presented a robust video transmission system by using WZC to generate parity bits for protecting an MPEG encoded bitstream of the same video; however, since their Wyner–Ziv coder outputs parity bits (as in systematic channel coding), this scheme is better categorized as systematic source/channel coding [12]—with an MPEG systematic part plus a Wyner–Ziv parity part.

In this paper, we present a novel *layered* video coder based on standard video coding and successive WZC [13], [14]. Treating a standard coded video as the base layer and decoder side information, our layered Wyner–Ziv encoder consists of DCT, nested scalar quantization (NSQ), and irregular low-density parity-check (LDPC) code based Slepian–Wolf coding (SWC) [15]. The DCT is applied as an approximation to the conditional Karhunen–Loeve transform (cKLT) [16], which makes the components of the transformed block conditionally independent given the decoder side information. Nested scalar quantization is a binning scheme that facilitates layered bit plane coding of the bin indices while reducing the bit rate. SWC plays the role of conditional entropy coding (with side information at the decoder) for further compression. Our Wyner–Ziv decoder performs joint decoding by combining the base layer and the Wyner–Ziv bitstream for enhanced video quality

We aim to generate a layered Wyner–Ziv bitstream from the *original video sequence* such that it is still decodable with commensurate qualities at rates corresponding to layer boundaries. Thus, our proposed layered WZC scheme is very much like MPEG-4/H.26L FGS [3], [17] and H.263+ coding [18] in "spirit" in terms of having an embedded enhancement layer with good R–D performance. However, the key difference is that the enhancement layer is generated "blindly" without using the base layer in WZC. This alleviates the problem of error drifting/propagation associated with encoder-decoder mismatch in standard DPCM-based coders because, as we shall see later in Section VI, a corrupted base layer (with errors within a certain range) can still be combined with the enhancement layer for Wyner–Ziv decoding. This inherent error-resilience in the base layer is the main advantage of our proposed Wyner–Ziv coder over standard FGS coding.

Our layered WZC scheme has the attractive feature that encoding is done only once but decoding allowed at many lower bit rates with commensurate qualities, i.e., there is no performance difference between layered and monolithic WZC for the enhancement layer. This is because our work is underpinned by recent theoretical results [13] and [14] that extend successive refinability of Gaussian sources from classic source coding to WZC and because our design is based on scalar quantization and bit plane coding. While the code design in [14] assumes ideal Gaussian sources with MSE distortion, results here are the first reported on practical layered WZC of video that *do not* suffer performance loss due to layering in WZC. The conference version of this paper appeared in [19]. Other groups' works on scalable Wyner–Ziv video coding are [20]–[22].

Relying on the H.26L coded version as the base layer,[1] our Wyner–Ziv video coder is capable of achieving roughly the same R–D performance as the H.26L FGS coder in [17]. For example, using irregular LDPC codes of lengths in the order of 8×10^4 bits for SWC, the former performs 0.3 dB worse in average PSNR than the latter at high rate. The performance of our layered Wyner–Ziv coder only degrades slightly when the LDPC code lengths are decreased by a factor of four to reduce latency.

Besides the thrust of applying the theory of successive WZC to practical video compression and showing the competitiveness of layered Wyner–Ziv video coding with standard FGS coding, this paper also aims to highlight error robustness of the base layer due to Wyner–Ziv enhancement layer coding via simulations, where a fixed amount of macroblock loss is introduced to the H.26L coded base layer. After combining the corrupted base layer with enhancement layers generated from WZC instead of FGS coding, we observe an improvement of 0.6–2.71 dB in video quality measured in average PSNR. This means that in video streaming applications, error-free delivery of the base layer is less critical with our layered Wyner–Ziv video coder than with standard scalable coders (e.g., MPEG-4/H.26L FGS).

Finally, we use a wireless channel simulator [23] from Qualcomm Inc. that conforms to the CDMA2000 1X standard to test the robustness of our layered Wyner–Ziv coder under wireless environments *for both the base and enhancement layers*. Extensive simulations show that layered WZC is more robust against channel errors than H.26L FGS coding, offering 0.3–1.5 dB gain in average PSNR. In addition, the advantage of layered WZC is more pronounced for high motion sequences, when error drifting with H.26L FGS becomes more severe. Our results clearly demonstrate that layered Wyner–Ziv video coding is a promising new technique for video streaming over wireless networks.

The rest of the paper is organized as follows. Section II gives the theoretical background on WZC. Section III covers Wyner–Ziv code design for jointly Gaussian sources. Section IV puts forth our layered Wyner–Ziv video coding framework. Section V presents our video compression results, while Section VI focuses on error robustness of our layered Wyner–Ziv video coder. Section VII concludes the paper and points to future research directions.

II. Theoretical Background on WZC

Consider $\{(X_i, Y_i)\}_{i=1}^{\infty}$ as a sequence of independent drawings of a pair of correlated discrete random variables X and Y. The problem of separate lossless encoding and joint decoding of X and Y was first considered by Slepian and Wolf [4], who gave the achievable rate region as

$$R_X \geq H(X|Y),\ R_Y \geq H(Y|X),\ R_X + R_Y \geq H(X,Y).$$

[1]H.26L refers to the video codec of the now well-known H.264 standard [2], which is the result of the combined efforts of ITU and ISO MPEG.

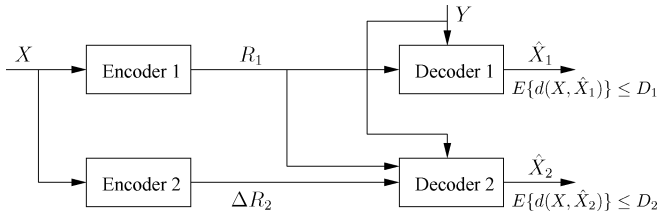


Fig. 1. Two-stage successive refinement with identical side information at the decoders.

This surprising result indicates that the joint entropy $H(X,Y)$ is still achievable. Hence, there is no performance loss with SWC when compared to joint encoding, but the caveat is that SWC is only lossless asymptotically with respect to the code length.

Lossless source coding with side information at the decoder is a special case of the SWC problem. Assume Y is encoded with $H(Y)$ bits so that it can be perfectly decoded at the decoder, then according to (1), SWC of X boils down to compressing it to the rate limit $H(X|Y)$. We henceforth limit ourselves to equating SWC with lossless source coding with side information in this paper.

WZC [5] generalizes the setup of SWC in that coding of X is with respect to a fidelity criterion rather than lossless. In addition, the source X could be either discrete or continuous. The work of [5] examines the question of how many bits are needed to encode the source X under the constraint that the average distortion between X and decoded version $\hat{X}$ satisfies $E\{d(X,\hat{X})\} \leq D$, assuming that the side information Y (discrete or continuous) is available only at the decoder. Denote $R^*_{WZ}(D)$ as the achievable lower bound of the bit rate for an expected distortion D for WZC, and $R^*_{X|Y}(D)$ as the R–D function of coding X with side information Y available also at the encoder.

In general, there is a rate loss associated with WZC, that is $R^*_{WZ}(D) \geq R^*_{X|Y}(D)$. However, $R^*_{WZ}(D) = R^*_{X|Y}(D)$ when X and Y are zero-mean and jointly Gaussian and the distortion measure is MSE [5].[2] We restrict ourselves to this jointly Gaussian case in WZC because there is no rate loss and it is of special interest in practice, where many image and video sources can be modeled as jointly Gaussian after mean subtraction.

A. Successive WZC

A successive refinement code for the Wyner–Ziv problem consists of multistage encoders and decoders where each decoder uses all the information generated from decoders of its earlier stages [13]. Fig. 1 depicts a special case of two-stage successive coding for the Wyner–Ziv problem with the side information at each stage being the same.

Let Y be the side information available to the decoder at both the coarse stage and the refinement stage, and the corresponding coding rates (distortions) are $R_1(D_1)$ and $R_2(D_2)$, respectively. A source X is said to be *successively refinable* from D_1 to D_2 ($D_1 > D_2$) with side information Y if

$$R_1 = R^*_{WZ}(D_1) \quad \text{and} \quad R_1 + \Delta R_1 = R^*_{WZ}(D_2). \tag{1}$$

[2]Recently, Pradhan *et al.* [24] extended this no rate loss result to the more general case with $X = Y + Z$, where Y and Z are independent, and only Z is i.i.d. Gaussian (Y can follow arbitrary distribution).

The notion of successive coding can be naturally extended to any finite number of stages [13]. Consider the case when the side information fed into the K decoders at each level is the same, the source X is *multistage successively refinable* with side information Y if

$$R_1 = R^*_{X|Y}(D_1) \quad \text{and} \quad R_i + \Delta R_i = R^*_{X|Y}(D_{i+1}) \tag{2}$$

for $i = 1, 2, \ldots k-1$.

Necessary and sufficient conditions for successive refinability are given in [13] and the jointly Gaussian source (with MSE measure) shown to be multistage successively refinable in the Wyner–Ziv setting. Extending the successive refinement result of [13] on jointly Gaussian sources, Cheng and Xiong [14] proved that the jointly Gaussian condition can be relaxed to the case that only the difference $X - Y$ between the source X and the side information Y is Gaussian and independent of the side information Y, i.e., the more general class of sources without rate loss in WZC defined in [24] is also successively refinable.

III. Wyner–Ziv Code Design for Jointly Gaussian Sources

Wyner–Ziv code design involves both source coding (quantization) and channel coding [8]. The simplest Wyner–Ziv coder involves a 1-D nested lattice/uniform quantizer [6] with a coarse coset code nested in a fine coset code, i.e., the coarse code is a subcode of the fine code. The fine code does source coding while each coarse coset code performs channel coding. This coset coding scheme amounts to binning, which refers to dividing the space of all possible outcomes of a source into disjoint subsets (or bins). To encode, X is first quantized by the fine source code, resulting in quantization errors, then only the index of the bin that the quantized X belongs to is coded to save the rate. Using this coded index, the decoder finds in the bin (or coset code) the codeword closest to the side information Y as the best estimate of X. There is quantization error due to source coding and binning loss due to channel coding.

Usually, there is still correlation remaining in the quantized version of X and the side information Y, and SWC can be employed to exploit this correlation to reduce rate. Thus, SWC is an integral part of WZC, much like the way entropy coding is used in classic lossy source coding to achieve further compression after quantization.

The connection between SWC and channel coding was first made in [25] and an explicit syndrome-based binning scheme was outlined there based on parity-check codes. The idea is to partition the 2^n possible source inputs (assuming binary source X with block coding of length n) into 2^{n-k} bins, each with 2^k elements, and index them with syndromes of a binary (n,k) parity-check code. The bin with zero syndrome corresponds to all valid codewords of the channel code, and the rest are different shifted versions of it. This way, the distance property of the channel code is preserved among elements in each bin. Encoding only involves multiplying the $(n-k) \times n$ parity check matrix of the channel code with the n-bit input sequence and outputting the $n-k$ syndrome bits that index the bin to which the input sequence belongs. The resulting compression ratio is

$n : n-k$. The decoder takes the bin index and finds the element closest to the side information in the bin as the best estimate of the input sequence. This syndrome-based binning scheme can approach the rate limit $H(X|Y) = 1 - r$ of SWC if a near-capacity parity-check code with rate $r = k/n$ is designed for the channel that characterizes the correlation between X and Y. The first practical design that follows this scheme using LDPC codes [26] was reported in [15], showing performance very close to the Slepian–Wolf limit $H(X|Y)$.

For WZC of jointly Gaussian sources, the limit-approaching Slepian–Wolf code design in [15] can be combined with strong source codes (e.g., TCQ [27]) to approach the theoretical limit. This forms the base of the Slepian–Wolf coded quantization paradigm [8] for WZC that generalizes entropy-coded quantization for classic source coding. The role of SWC is to approach the rate limit $H(Q(X)|Y)$, where $Q(X)$ is the quantized version of the input Gaussian source X. This requires a simple extension of the syndrome-based binning scheme [25] from SWC of binary sources to M-ary sources with multilevel LDPC codes.

According to [8], the performance gap of high-rate Slepian–Wolf coded quantization to the Wyner–Ziv distortion-rate (D–R) function is exactly the same as that of high-rate classic source coding to the D–R function. A practical layered Wyner–Ziv code design based on NSQ and multilevel LDPC codes for SWC was presented in [14], yielding results that are 2.9 to 1.65 dB away from the Wyner–Ziv D–R function for rates ranging from 0.48 to 6.0 b/s.

IV. Layered Wyner–Ziv Video Coding

Successive or scalable image/video coding made popular by EZW [28] and 3-D SPIHT [29] is attractive in practical applications such as networked multimedia. By producing a video stream that can be decoded at more than one quality levels, scalable video coding achieves graceful quality degradation as the available bandwidth for data transmission decreases. This is very desirable in video streaming applications.

When decoder side information is available, it is also important and rewarding to explore successive Wyner–Ziv video coding in practice. Although practical code designs in [8] and [14] for WZC of jointly Gaussian sources perform close to the theoretical limit, it is not straightforward to apply these designs directly to video sources. The first issue in Wyner–Ziv video coding is the identification of the decoder side information. Our novelty is to use a standard decoded low-quality video as the side information, which is highly correlated with the original video source. In addition, there are several other issues involved in Wyner–Ziv video coding.

1) Transform Design: Unlike i.i.d. Gaussian sources, the neighboring pixels in a video frame are highly correlated with each other. In standard video coding, the DCT has been widely used to decorrelate the image pixels to facilitate compression. For Wyner–Ziv video coding, ideally the cKLT [16] should be applied to both the video source and the side information to make the former conditionally independent given the latter before performing WZC. But the cKLT is signal-dependent, in practice a signal-independent approximation has to be used.

2) Correlation Modeling: In WZC of jointly Gaussian sources, the joint statistics of the sources is assumed to be known *a priori*. In Wyner–Ziv video coding, the source correlation depends on the video quality of the side information. In practice it has to be estimated via correlation modeling, which is a critical step as it directly determines the performance limit of WZC.

3) Quantization: To approach the Wyner–Ziv limit, strong quantizers such as TCQ have to be employed in conjunction with limit-approaching Slepian–Wolf codes. However, TCQ does not facilitate successive refinement (although it can lead to progressive coding). Thus, we are confined to NSQ that allows bit plane coding for successive refinement. Fortunately, the performance loss of using NSQ for WZC of jointly Gaussian sources is only 1.53 dB at high rate (assuming ideal SWC) [8].

4) Slepian–Wolf Code Design and Rate Control: Capacity-achieving channel codes such as LDPC codes [26] have to be used to approach the Slepian–Wolf limit. However, to achieve high performance with these advanced channel codes requires long block length, which is not a problem with jointly Gaussian sources but introduces long delay in video coding. In addition, the code rate for SWC and convergence at the Slepian–Wolf decoder heavily rely on the correlation between the source and the side information. Our main contribution here lies in the design of efficient multilevel LDPC codes to realize layered Wyner–Ziv video coding via SWC of successive bit planes after NSQ, starting from the most significant bit plane.

A. Proposed Code Design

We now present our layered Wyner–Ziv video coder using LDPC code based SWC. Treating a standard H.26L decoded video as the base layer (and side information), a layered Wyner–Ziv bitstream of the original video sequence is generated to enhance the base layer such that it is still decodable with commensurate qualities at rates corresponding to layer boundaries. Denote the current frame of the original video as $\mathbf{x}$, which is encoded with H.26L to obtain the base layer (or side information) $\mathbf{y}$. Fig. 2 depicts the block diagram of our layered Wyner–Ziv coder, whose encoder consists of three components: the DCT, NSQ, and SWC based on irregular LDPC codes.

NSQ introduces both a binning loss, which should be kept small with strong channel coding, and a quantization loss that should be optimally traded off with rate in source coding. In addition, there is still correlation between the quantized version (bit planes in the middle) of the source $\mathbf{X}$ and the side information $\mathbf{Y}$, and SWC can be employed to exploit this correlation and achieve further compression.We use the DCT as an approximation to the cKLT [16], which makes the coefficients of the transformed block of the original video $\mathbf{x}$ conditionally independent given the same transformed block of the side information $\mathbf{y}$. NSQ is a binning scheme that assigns the input DCT coefficients $\mathbf{X}$ to cosets and outputs only the coset indices. The DCT coefficients are split into several bit planes with their binary representations. The upper significant bit planes of the DCT coefficients are skipped in NSQ since they are highly correlated to those in the side information. There will be a significant loss in quality if the side information cannot be used to correctly recover these bit planes at the joint Wyner–Ziv decoder. The lower significant bit planes are less important and, hence, quantized to zero by NSQ to save rate. Therefore, both the upper and lower significant bit

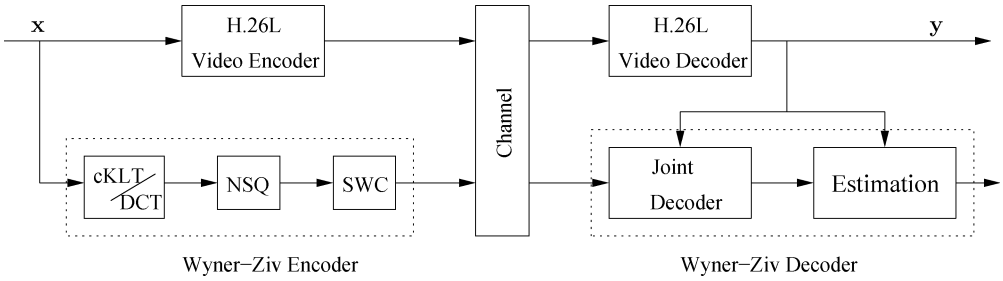


Fig. 2. Block diagram of the proposed layered Wyner–Ziv video coder.

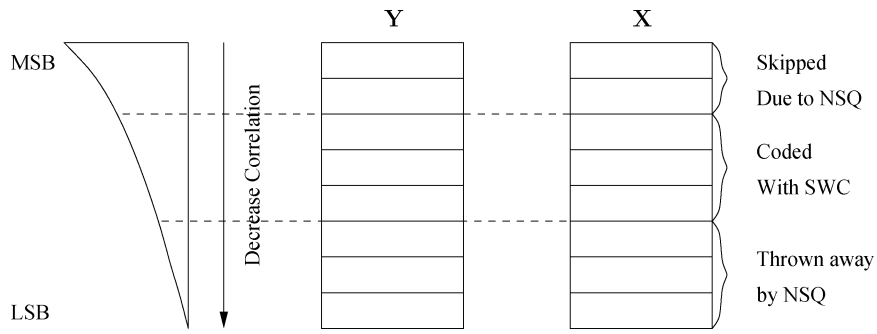


Fig. 3. NSQ throws away both the upper significant bit planes (with nesting) and the lower significant bit planes (with quantization). The number of thrown away bit planes depends on the nesting ratio N (quantization stepsize q).

planes are thrown away in NSQ depending on the nesting ratio N and quantization stepsize q, respectively, and only those in between are coded (see Fig. 3). We employ multilevel LDPC codes for SWC in the third component of the encoder and output one layer of compressed bitstream for each bit plane after NSQ. In doing so, we note that the correlation decreases as we move from the most significant bit (MSB) to the least significant bit (LSB). Thus, higher rate LDPC codes are designed for higher bit planes to achieve more compression; while lower rate LDPC codes are given to lower bit planes for less compression. Furthermore, although theoretically the overall rate required is the same with different orders of bit plane coding, to facilitate layered coding, the order of encoding proceeds from the MSB to the LSB after NSQ. In the following, we will explain each component in details.

1) Transform: For WZC of $\mathbf{x}$, we first apply the cKLT (approximated by the DCT) to every 4×4 block of $\mathbf{x}$ so that the components of the transformed block $\mathbf{X} = \mathbf{T}\mathbf{x}$ are conditionally independent given the side information $\mathbf{y}$, which is also transformed into $\mathbf{Y} = \mathbf{T}\mathbf{y}$. Each frequency component of $\mathbf{Y}$ (denoted by Y) acts as the side information for the corresponding component of $\mathbf{X}$ (denoted by X). We assume that X and Y are jointly Gaussian with $X = Y + Z$, where Z is zero-mean Gaussian and independent of Y (although DCT coefficients of images/video are better modeled as Laplacian distributed [30]).

2) Quantization: The next step is NSQ (see Fig. 4), which consists of a coarse coset channel code with minimum distance $d_{\min} = Nq$ nested in a fine uniform scalar quantizer with stepsize q. To encode, X is first quantized by the fine source code (uniform quantizer), resulting the "good" distortion, which is the average quantization error of $q^2/12$ at high rate. However, only the index B ($0 \le B \le N-1$) of the coset in the coarse channel code that the quantized X belongs to is coded to save rate. Using the decoded coset index B, the decoder finds in the coset the codeword closest to the side information Y as the best estimate of X. Due to the coset channel code employed in nesting process, the Wyner–Ziv decoder suffers a small probability of "bad" distortion that is inversely proportional to $d_{\min} = Nq$. It is desirable to choose a small quantization stepsize q to minimize the "good" distortion due to source coding. On the other hand, $d_{\min}$ should be maximized to minimize the "bad" distortion due to channel coding. Thus, for a fixed N, there exists an optimal q that minimizes the total distortion, which is the sum of the "good" distortion and the "bad" distortion.

3) SWC: Due to the correlation between X and Y, there still remains correlation between the coset index B and the side information Y. SWC can be used to compress B to the rate of $R = H(B|Y)$. Express B in its binary representation as $B = B_0B_1\ldots B_{m-1}$, where B_0 is the MSB, B_{m-1} is the LSB, and $m = \lceil \log_2 N \rceil$. We employ multilevel LDPC codes to compress $B_0B_1\ldots B_{m-1}$ based on the syndrome-based approach [15], [25]. The rate of the LDPC code for B_i ($0 \le i \le m-1$) depends on the conditional entropy $H(B_i|B_0,\ldots,B_{i-1},Y)$[14], which denotes the minimum rate needed for lossless recovery of B_i given $B_0\ldots B_{i-1}$ and Y at the decoder.

We initially assume ideal SWC in the sense that the rate $R = H(B|Y)$ can be achieved. Then for each fixed N (number of cosets in the channel code), we vary the uniform quantization step size q to generate a set of R–D points (R, D) and pick the optimal q^* corresponding to the point with the steepest R–D slope from the zero-rate point in WZC. Note that the distortion for the zero-rate point is just $\|X - Y\|^2$, which is the average

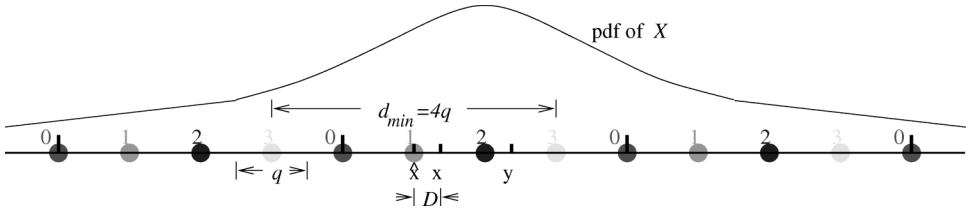


Fig. 4. NSQ with nesting ratio $N = 4$.

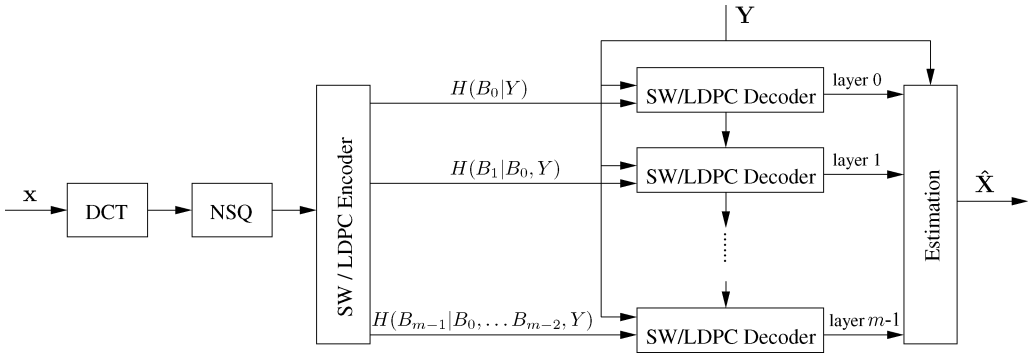


Fig. 5. Bit-plane-based multistage SWC using multilevel LDPC codes for layered WZC after the DCT and NSQ.

distortion of base layer coding due to H.26L. After identifying the optimal R–D points for different N, the lower convex hull of these points forms the operational R–D curve of WZC. Due to the fact that quadratic Gaussian sources are successively refinable [13], [14], the same operational R–D curve should be traversed[3] by starting with a large N (with its corresponding q^*) first and then sequentially dropping bit planes of B. In other words, by setting different low bit plane levels of B to zero, the resulting R–D points after Wyner–Ziv decoding should all lie on the operational R–D curve. Our simulations verify this property of successive refinement and justify our approach of coding B_i into the ith layer with rate $H(B_i|B_0,\ldots,B_{i-1},Y)$ (see Fig. 5). By the chain rule

$$H(B|Y) = H(B_0|Y) + H(B_1|B_0,Y) + \cdots + H(B_{m-1}|B_0,\ldots B_{m-2},Y).$$

So, layered coding suffers no rate loss when compared with monolithic coding.

In our irregular LDPC code designs, the code degree distribution polynomials $\lambda(x)$ and $\rho(x)$ of the LDPC codes are optimized using density evolution based on the Gaussian approximation [31]. The bipartite graph for the irregular LDPC code is then randomly constructed based on the optimized code degree distribution polynomials $\lambda(x)$ and $\rho(x)$. To compress bit plane B_i, only the corresponding syndrome determined by the sparse parity check matrix of the irregular LDPC code is transmitted to the decoder. At the decoder, each additional bitstream/syndrome layer is combined with previously decoded bit planes to decode a new bit plane before joint estimation of the output video. Let $\hat{B}_i$ represent the reconstruction of B_i. The message-passing algorithm [32] is used for iterative LDPC decoding, in which the received syndrome bits correspond to the check nodes on the bipartite graph, the side information and the previously decoded bit planes provide *a priori* information about the probability of the current bit being "1" or "0," i.e.,

$$\mathrm{LLR} = \log \frac{p(B_i = 0|\hat{B}_0,\ldots,\hat{B}_{i-1},Y)}{p(B_i = 1|\hat{B}_0,\ldots,\hat{B}_{i-1},Y)}.$$

After decoding B_0 to $\hat{B}_0$, both $\hat{B}_0$ and Y will be fed into the decoder for decoding of B_1. Since the allocated bit rate for coding B_1 is $H(B_1|B_0,Y)$, B_1 can be correctly decoded as long as $\hat{B}_0 = B_0$. By multistage decoding, B_i can be correctly recovered with the help of Y and the previously decoded bit planes $B_0, B_1, \ldots, B_{i-1}$, which are already available at the decoder. The more syndrome layers the decoder receives (or the higher the bit rate), the more bit planes of B will be recovered to better reconstruct X. Therefore, successive WZC provides the flexibility to accommodate a wide range of bit rates. Progressive decoding is desirable for applications where only a coarse description of the source suffices at the first stage with low bit rate, and fine details are needed at some later stage with higher bit rate.

We perform optimal estimation at the joint decoder. The decoded coset index $\hat{B}_0\hat{B}_1\ldots\hat{B}_i$ specifies the uncertainty region of X. The side information essentially supplies the conditional pdf of X given Y, which is a Gaussian with mean Y and variance proportional to the correlation between Y and X. The optimal estimate of X is computed as the conditional centroid $\hat{X} = E(X|\hat{B}_0\hat{B}_1\ldots\hat{B}_i, Y)$. Finally, the inverse DCT is applied to $\hat{\mathbf{X}}$ to obtain $\hat{\mathbf{x}}$ in the pixel domain.

V. Compression Results

A. Successive Refinement

Due to the approximation of the cKLT by the DCT and the Gaussian assumption of X and Y in our practical Wyner–Ziv

[3]Here, we assume that X and Y are jointly Gaussian, even though this is only approximately true.

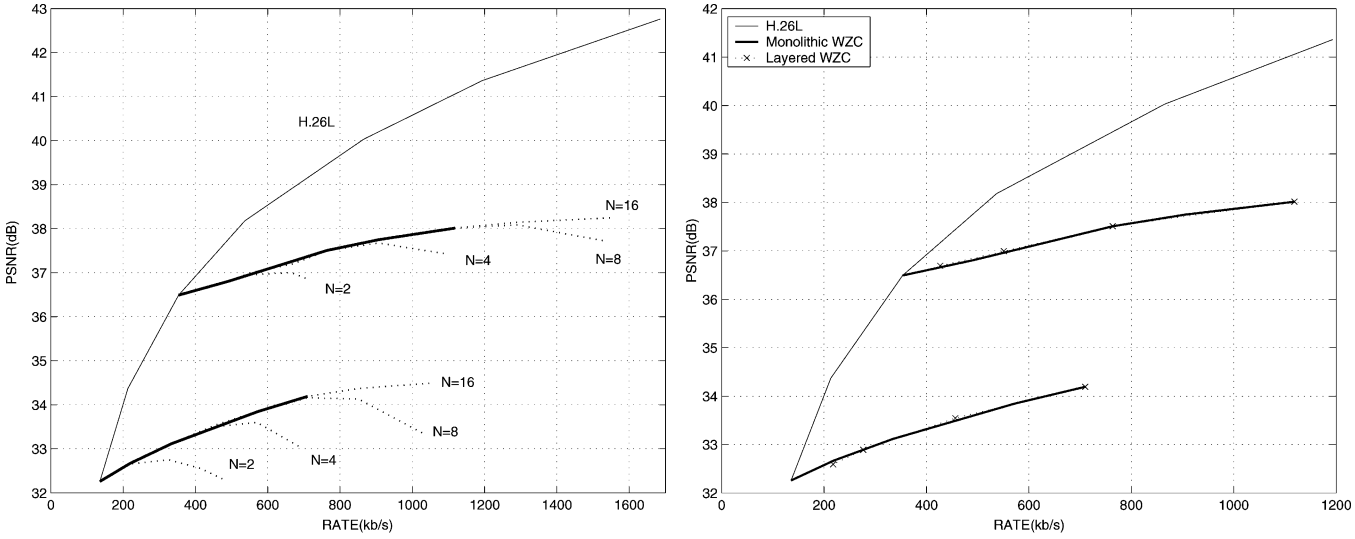


Fig. 6. Illustration of successive refinement in our layered Wyner–Ziv video coder, assuming ideal SWC. (a) The operational rate-PSNR function of WZC is formed as the upper concave hull of different rate-PSNR points. (b) There is almost no performance loss between monolithic WZC and layered WZC.

video coder, experiments are carried out on the CIF Foreman sequence to verify the validity of our practice and illustrate successive refinement by assuming ideal SWC, that is the rate $R = H(B|Y)$.

The input video is first encoded with H.26L to obtain base layer (and the side information). Then the proposed Wyner–Ziv video coding scheme which consists of the DCT, NSQ and *ideal* bit plane based SWC [with rate $R = H(B|Y)$] is used to generate a Wyner–Ziv bitstream to enhance the base layer. The rate-PSNR performance for four different values of $N \in \{2, 4, 8, 16\}$ with different q's for each N, starting from two different "zero-rate" points for WZC from the H.26L PSNR-rate curve, is plotted in Fig. 6(a). After that, the operational rate-PSNR function of WZC is formed as the upper concave hull of different rate-PSNR points. Then starting at a high rate point on the operational rate-PSNR function in Fig. 6(a) (e.g., with $N = 16$ and its corresponding q^*), we perform layered coding by dropping more and more lower bit planes of the coset index B to achieve lower rates. Fig. 6(b) shows a good match between the performance of monolithic WZC and that of layered WZC.

B. Layered Coding

We implement SWC based on irregular LDPC codes and investigate the layered WZC performance for Football (352×240) and CIF Foreman and Mother_daughter, since these sequences represent different amount of motion. Standard H.26L encoded video is treated as side information at the decoder. One hundred frames are compressed with a frame rate of 30 Hz. For each of these sequences, the first frame is coded as I frame, and all the subsequent frames as P frames by H.26L. Different quantization stepsizes are used in the H.26L coder to generate different "zero-rate" points for WZC. After DCT of the original video, different transform coefficients are encoded independently, and we only code the first three DCT coefficients.

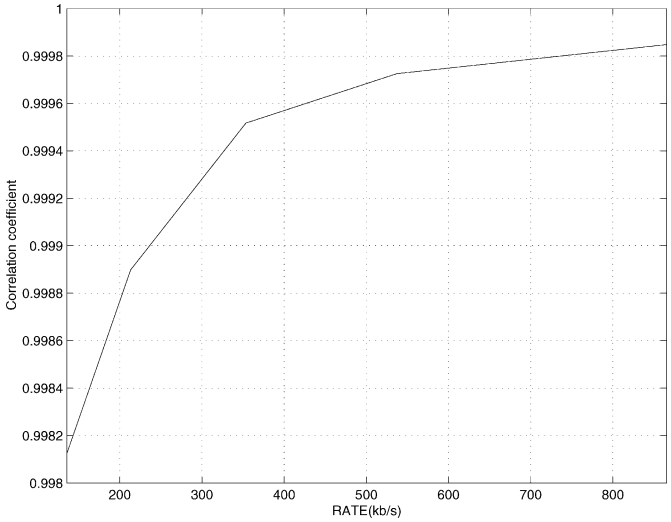


Fig. 7. Estimated correlation coefficient between the DC component of the original video and that of the side information for CIF Foreman. The horizontal axis represents the rate for the H.26L base layer.

1) Correlation Modeling: We assume that $X = Y + Z$ in the DCT domain, where the side information $Y \sim N(0, \sigma_Y^2)$ and the quantization noise $Z \sim N(0, \sigma_Z^2)$ due to H.26L coding are independent. We estimate σ_Z^2 separately for different DCT coefficients by computing the MSE between X and Y. Fig. 7 shows our empirically estimated correlation coefficient $\sqrt{1 - \sigma_Z^2/\sigma_X^2}$ between the DC component of the original video X and that of the side information Y for Foreman.

2) LDPC Code Design for SWC: Recall that we only apply NSQ to the first three DCT coefficients and compress them with SWC. For each transform coefficient, we use a four-bit nested scalar quantizer to generate four bit planes. The 12 bit planes are then encoded by 12 different LDPC codes (of different rates). The LDPC code rate for the ith bit plane B_i is maximized to approach the conditional entropy $H(B_i|B_0, \ldots, B_{i-1}, Y)$,

TABLE I
CONDITIONAL ENTROPY, LDPC CODE RATE, AND THE CORRESPONDING DEGREE DISTRIBUTION POLYNOMIALS $\lambda(x)$ AND $\rho(x)$ FOR EACH BIT PLANE AFTER NSQ OF (A) THE DC COEFFICIENT, (B) THE FIRST AC COEFFICIENT, AND (C) THE SECOND AC COEFFICIENT AFTER THE DCT

Bit plane	Conditional entropy	LDPC code rate	Degree polynomials	
			$\lambda(x)$	$\rho(x)$
0	0.03433	0.94	$0.1827x + 0.2609x^2 + 0.0805x^3 + 0.3954x^8 + 0.0806x^9$	$0.5x^{65} + 0.5x^{66}$
1	0.03884	0.94	-	-
2	0.10157	0.85	$0.2137x + 0.2482x^2 + 0.0795x^3 + 0.0695x^7 + 0.3889x^8$	$0.5x^{24} + 0.5x^{25}$
3	0.19593	0.73	$0.2103x + 0.2062x^2 + 0.0615x^4 + 0.1667x^5 + 0.3002x^{13} + 0.0551x^{14}$	$0.5x^{14} + 0.5x^{15}$

(a)

Bit plane	Conditional entropy	LDPC code rate	Degree polynomials	
			$\lambda(x)$	$\rho(x)$
0	0.00178	0.99	$0.2530x + 0.3067x^2 + 0.4403x^3$	$0.5x^{194} + 0.5x^{195}$
1	0.03768	0.94	$0.1827x + 0.2609x^2 + 0.0805x^3 + 0.3954x^8 + 0.0806x^9$	$0.5x^{65} + 0.5x^{66}$
2	0.02532	0.96	$0.1703x + 0.2714x^2 + 0.0136x^3 + 0.1046x^4 + 0.4400x^9$	$0.5x^{99} + 0.5x^{100}$
3	0.16754	0.77	$0.2313x + 0.2201x^2 + 0.1092x^3 + 0.3477x^8 + 0.0917x^9$	$0.5x^{15} + 0.5x^{16}$

(b)

Bit plane	Conditional entropy	LDPC code rate	Degree polynomials	
			$\lambda(x)$	$\rho(x)$
0	0.02785	0.95	$0.1779x + 0.2605x^2 + 0.0843x^3 + 0.3452x^8 + 0.1322x^9$	$0.5x^{79} + 0.5x^{80}$
1	0.00488	0.98	$0.2695x + 0.3885x^2 + 0.3420x^3$	$0.5x^{134} + 0.5x^{135}$
2	0.03240	0.94	$0.1827x + 0.2609x^2 + 0.0805x^3 + 0.3954x^8 + 0.0806x^9$	$0.5x^{65} + 0.5x^{66}$
3	0.44291	0.49	$0.2330x + 0.1589x^2 + 0.0361x^3 + 0.0785x^5 + 0.1672x^6 + 0.0709x^{19} + 0.2554x^{20}$	$0.5x^7 + 0.5x^8$

(c)

meaning LDPC code design has to be tailored to the specific sequence or GOF. We note that this code rate only depends on the joint statistics between B and Y (or more specifically between B_i and $\{B_0, \ldots, B_{i-1}, Y\}$) – the reason why we can encode X without using the actual value of Y.

The degree distribution polynomials of the LDPC codes are optimized using the Gaussian approximation [31] and the bipartite graphs of them are generated randomly. As an example, for CIF Foreman with 530 Kbps for the base layer, the optimal quantization stepsize for the NSQ is $q = 32.0$ with nesting ratio $N = 16$. For the first GOF, the conditional entropy (or rate limit of SWC), LDPC code rate[4] and the corresponding degree distribution polynomials $\lambda(x)$ and $\rho(x)$ for each bit plane after NSQ of the DC component and the first two AC components of the DCT coefficients are listed in Table I, which indicates more loss at lower bit planes due to practical LDPC coding when the conditional entropy is higher.

3) Coding Performance: Starting with the largest $N = 16$ and its corresponding optimal q^*, we quantize X into B and sequentially decode B_0, B_1, B_2, and B_3. When the LDPC code lengths are in the order of 8×10^4 bits, we group 20 frames together in WZC. One hundred iterations are used for LDPC iterative decoding to achieve the bit error probability of 5×10^{-5}. The same pseudorandom seed is used at both the encoder and the decoder such that the same codebooks are used. The joint decoder performs optimal estimation based on the side information Y and the decoded coset index. Compared to ideal SWC with $R = H(B|Y)$, the loss due to practical LDPC coding is 0.05 b/s. Layered WZC results in terms of rate-PSNR performance is shown in Fig. 8. We see that as more bit planes are decoded, the video quality improves. We also observe that the performance loss due to WZC rather than FGS coding decreases as the bit rate for H.26L base layer (or "zero-rate" for WZC) increases, with a maximum PSNR loss of 0.3 dB at the same rate. This is partially because the correlation between X and Y is higher when the base layer is coded at higher rate with better quality (see also Table I).

To cut the latency introduced by SWC, we reduce the LDPC code lengths to the order of 2×10^4 bits (while using the same degree profiles as before). This way only five frames are grouped together in WZC. The rate loss in SWC is increased slightly from 0.05 to 0.09 b/s, which is translated into a maximum PSNR loss of 0.5 dB. As seen from Fig. 8, the PSNR performance loss for Foreman is small when the code length for SWC is scaled down from 8×10^4 to 2×10^4 bits.

VI. ERROR ROBUSTNESS

A. Against Simulated Errors in the Base Layer

Our layered Wyner–Ziv video coding framework is very similar to FGS coding [3], [17] in the sense that both schemes treat the standard coded video as the base layer and generate an embedded bitstream as the enhancement layer. However, the key difference is that instead of coding the difference between the original video and the base layer reconstruction as with FGS, the enhancement layer is generated "blindly" without knowing the base layer in Wyner–Ziv video coding. Therefore, the stringent

[4]The actual compression rate is one minus the LDPC code rate.

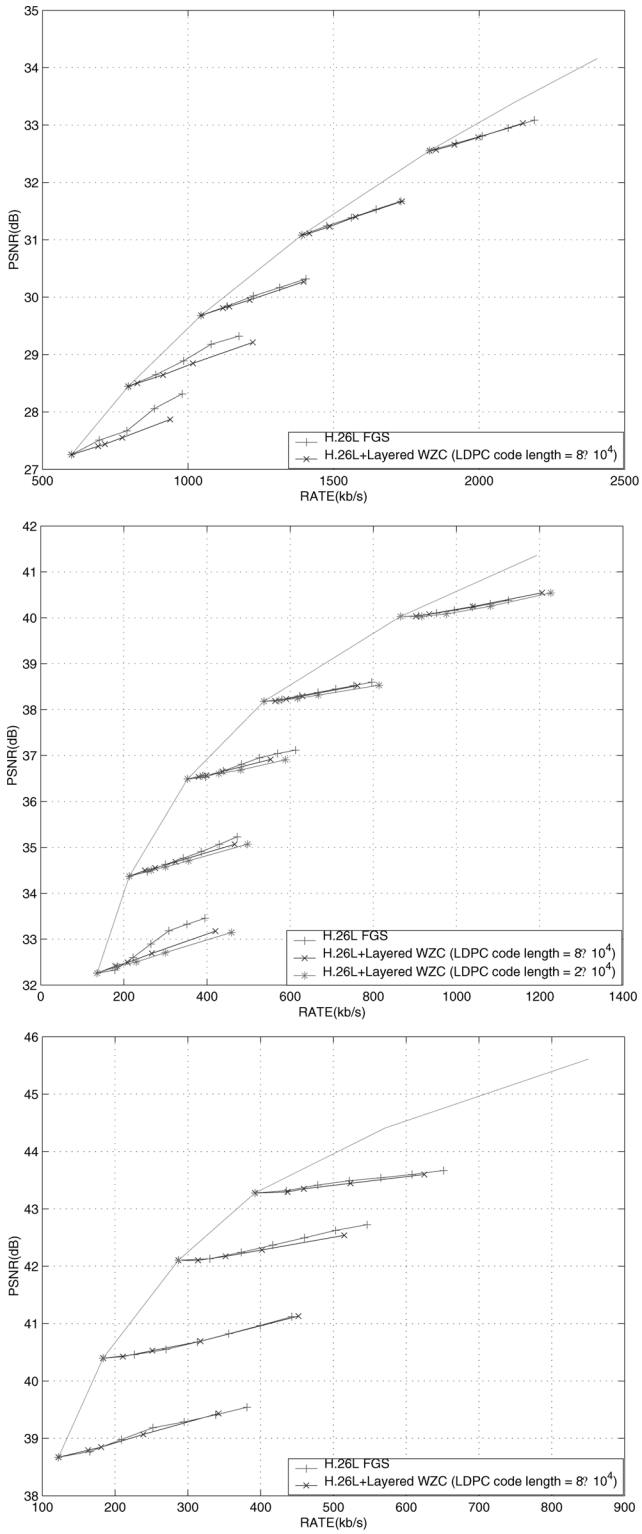


Fig. 8. Layered WZC of (top) Football, (middle) CIF Foreman, and (bottom) Mother_daughter, starting from different "zero-rate" points. The sum of the rates for H.26L coding and WZC is shown in the horizontal axis.

requirement of FGS coding that the base layer is always available losslessly at the decoder/receiver can be loosened somewhat because an error-concealed version of the base layer can still be used in the joint Wyner–Ziv decoder. That the latter statement is true can be easily seen from the NSQ depicted in Fig. 4, since any side information $y \in (\hat{x} - d_{\min}/2, \hat{x} + d_{\min}/2)$ will result in the same decoded $\hat{x}$.

In our experiments, the same video sequence is compressed by both our Wyner–Ziv video coder and the H.26L FGS coder [17] at a frame rate of 30 Hz. The bit rate for the base layer is the same, so is for the enhancement layers. Every 15 frames start with one I frame, followed by 14 P frames. We introduce the same amount of macroblock loss to the base layer for both coders and compare their error robustness.

For the Football sequence, the base layer is encoded at 1450 Kb/s and the bit rate for the enhancement layer of both WZC and FGS coding is 200 Kb/s. Then 1% macroblock loss in the base layer is simulated with simple error concealment [33] performed during decoding of the base layer. The PSNRs of the first 15 frames are shown in Fig. 9. The performance of Wyner–Ziv video coding is 2.71 dB better on average than H.26L FGS coding. This is because the basic assumption of error-free delivery of the base layer in FGS coding is no longer valid in this setup while the error-concealed version of the base layer can still be used as side information in Wyner–Ziv decoding.

Results from similar experiments with CIF Foreman and Mother_daughter are also given in Fig. 9. For Foreman, the base layer is encoded at 190 Kb/s and the bit rate for the enhancement layer of both WZC and FGS coding is 60 Kb/s. 5% macroblock loss is introduced to the base layer. The performance of Wyner–Ziv video coding is 1.93 dB better on average than H.26L FGS coding. For Mother_daughter, the base layer is encoded at 146 Kb/s and the bit rate for the enhancement layer of both WZC and FGS coding is 108 Kb/s. 5% macroblock loss is introduced to the base layer. The performance of Wyner–Ziv video coding is 0.6 dB better on average than H.26L FGS coding.

B. Against Errors From a Qualcomm Wireless Channel Simulator

To test error robustness of the *base and enhancement layer* bitstream of our Wyner–Ziv coder, a wireless channel simulator [23] is obtained from Qualcomm Inc. This simulator adds packet errors to streams of real-time transport protocol (RTP) packets transmitted over wireless networks conforming to the CDMA2000 1X standard. It assumes the use of a dedicated radio channel for the RTP packet stream under a given maximum transmission rate. Furthermore, protocol data unit (PDU) losses are introduced in the radio link control (RLC) layer. Each RTP packet is fragmented into equal-size PDUs and it is considered successfully received by the decoder only when all its PDUs are received and the arriving time of its last PDU is still within the maximum end-to-end delay. The Qualcomm simulator also provides FEC emulation with Reed-Solomon (RS) code, which is assumed to have ideal error correction capability. In RS coding, source symbols are encoded into RS parity symbols to provide protection for PDU losses.

In our experiments, each video sequence is compressed into slice RTP packets by both our Wyner–Ziv video coder and the H.26L FGS coder [17] at a frame rate of 30 Hz. The GOF size is now set to 20 with the structure of $\mathrm{IP}\cdots\mathrm{P}$. The bit rate for the base layer is the same, so is for the enhancement layers. Both the base and enhancement layers are protected with RS parity

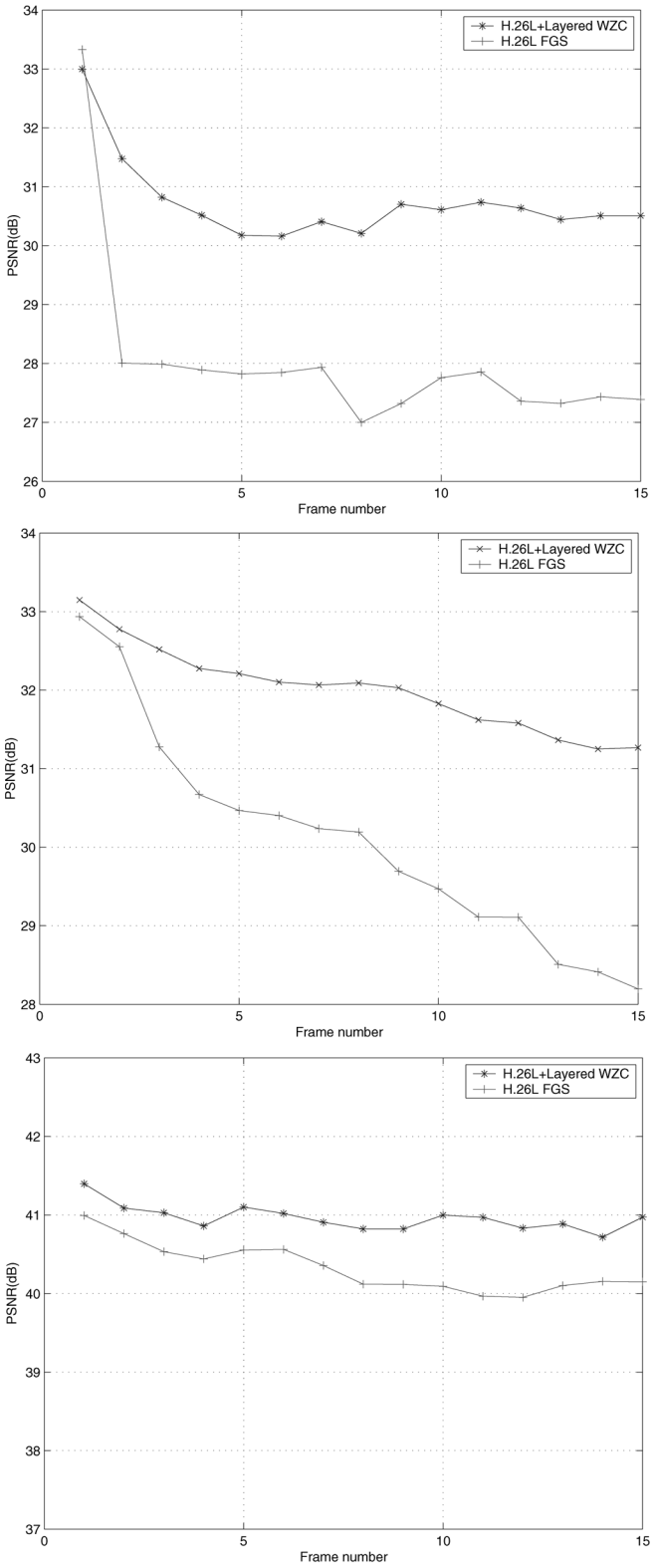


Fig. 9. Compared to FGS coding, Wyner–Ziv video coding offers substantial improvement in decoded video quality when the base layer (or decoder side information) suffers 1% macroblock loss for (top) Football, 5% macroblock loss for (middle) Foreman, and 5% macroblock loss for (bottom) Mother_daughter .

packets, whose rate is specified by the overhead percentage as an input parameter to the channel simulator. We set the FEC overhead percentage to 25% so that 20% of the overall bit rate is used for RS-based FEC. The resulting RTP packet streams are transmitted over CDMA2000 1X wireless networks simulated with the Qualcomm simulator. PDU losses are introduced to the RLC layer for both coders for error robustness comparisons.

For Football, the H.26L base layer is encoded at 1470 Kb/s and the bit rate for the enhancement layer of both WZC and FGS coding is 400 Kb/s. Thus, the total transmission rate is $(1470 + 400) \times (1 + 25\%) = 2337$ Kb/s. PDU loss rates of up to 10% are simulated across the *entire* packet stream on the RLC layer.

Despite 20% FEC, owing to the stringent latency constraint and packet fragmentation during transmission, there are still *residual* RTP packet losses at the decoder. For example, the residual RTP packet loss rates at 2%, 4%, 6%, 8%, and 10% PDU loss rate are 0.15%, 0.75%, 1.76%, 3.30%, and 5.19%, respectively. Simple error concealment is performed during decoding of the base layer. As for the enhancement layer bitstream, the *whole* layer with the first packet error, together with all subsequent layers, are discarded in the Wyner–Ziv video decoder since Slepian–Wolf decoding cannot proceed with corrupted syndromes in any layer. On the other hand, decoding the embedded FGS bitstream only stops at the first detected lost packet for the H.26L FGS coder.

For each PDU loss rate, 200 video transmissions are simulated, and Fig. 10 depicts the average PSNR performance versus PDU loss rate for the two coders. When the base layer is perfectly reconstructed (e.g., no error occurs in the base layer in 21% of our simulated transmissions when the PDU loss rate is 6%), the H.26L FGS coder performs slightly better than the Wyner–Ziv video coder (as seen in the coding performance of Section V-B3). However, our Wyner–Ziv coder outperforms H.26L FGS coding when there are packet losses in the base layer; and the higher the rate of the uncorrupted enhancement layer, the larger the performance gap between these two coders. The average PSNR performance gain of layered WZC increases with the PDU loss rate first, reaching 1.5 dB when the PDU loss rate is 6%. Because the amount of FEC is fixed (at 20%), the performance gain of layered WZC slightly decreases as the PDU loss rate (hence, the residual RTP packet loss rate) goes further up to make the enhancement layer more corrupted.

The decoded tenth frames of Football by H.26L FGS and layered Wyner–Ziv video coding (from the seventh simulated transmission) are shown in Fig. 11. It is easy to see that the decoded frame in Fig. 11(b) has higher visual quality than that in Fig. 11(a).

Similar simulations are also run on the CIF Foreman and Mother_daughter sequences, and results included in Fig. 10. For Foreman, the H.26L base layer is encoded at 305 Kb/s and the bit rate for the enhancement layer of both WZC and FGS coding is 255 Kb/s. The total transmission rate is 700 Kb/s. With 20% FEC, the residual RTP packet loss rates at 2%, 4%, 6%, 8%, and 10% PDU loss rate are 0.33%, 0.95%, 2.20%, 3.84%, and 5.66%, respectively. Again, 200 video transmissions are simulated for each PDU loss rate. The average PSNR performance gains of layered WZC over H.26L FGS are 0.67, 0.9, and 0.77 dB when the PDU loss rate is 6%, 8%, and 10%, respectively.

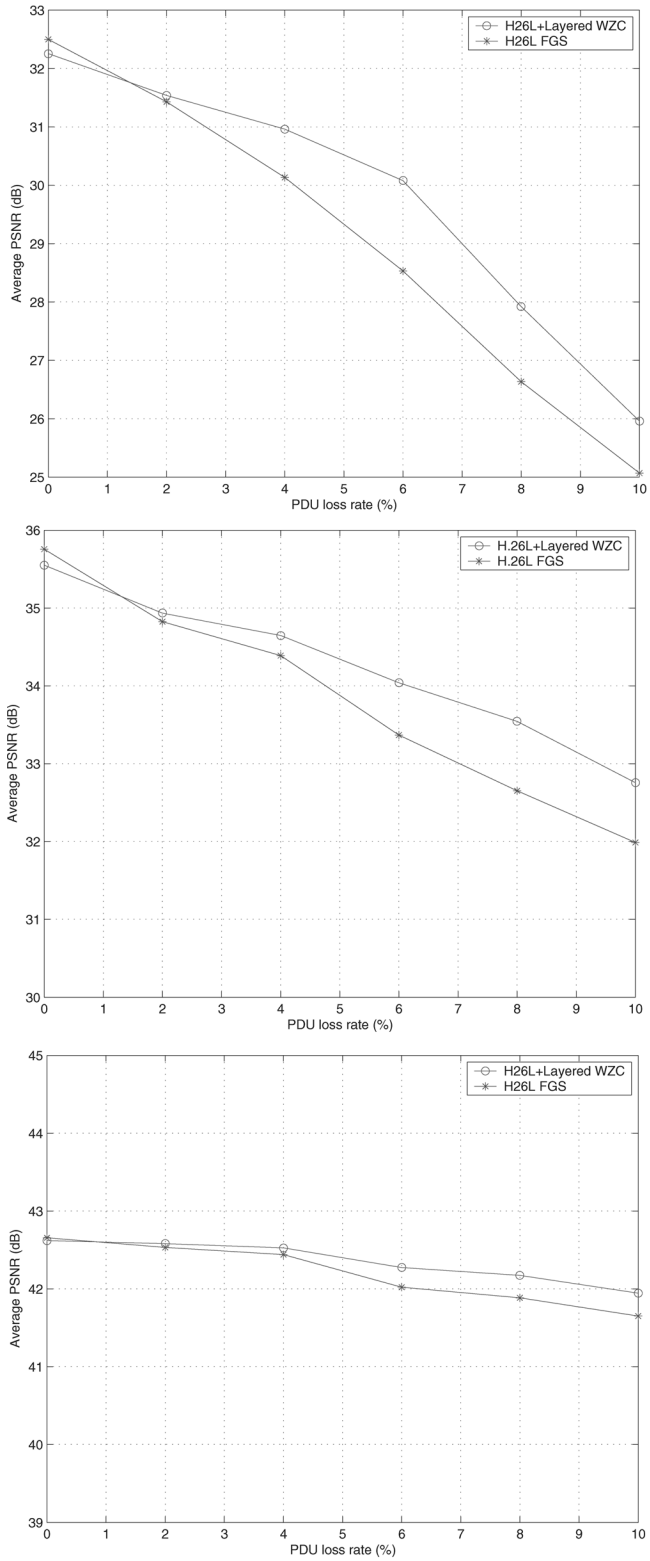
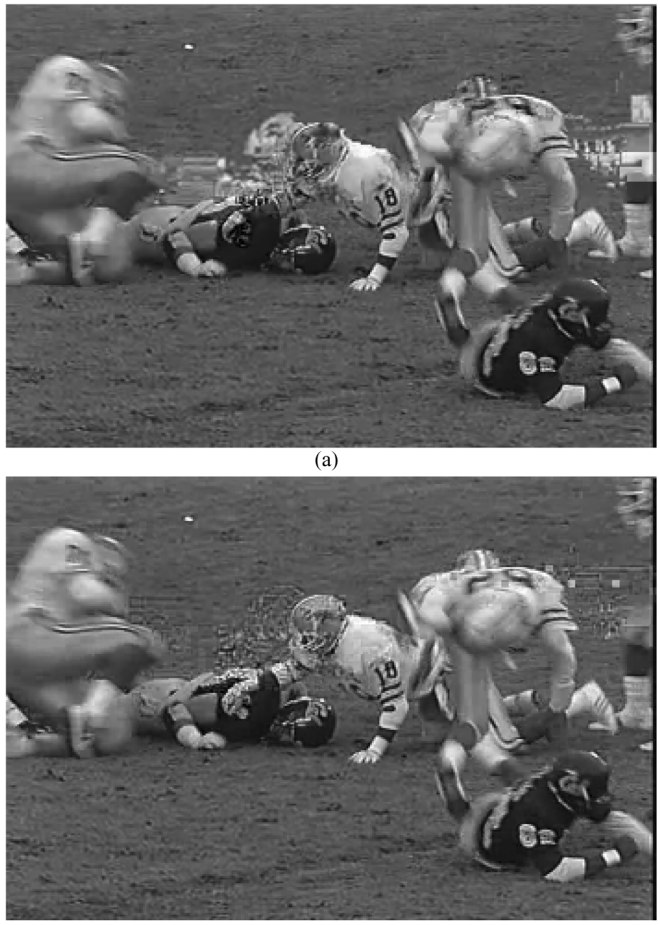


Fig. 10. Comparison of the Wyner–Ziv video coder and H.26L FGS coder when both are protected with RS-based FEC codes and transmitted over a simulated CDMA2000 1X channel for (top) Football, CIF (middle) Foreman, and (bottom) Mother_daughter.



(a)

(b)

Fig. 11. Error robustness performance of Wyner–Ziv video coding compared with H.26L FGS for Football when both the base layer and enhancement layer bitstreams are protected with 20% RS-based FEC and transmitted over a simulated CDMA2000 1X channel with 6% PDU loss rate. The tenth decoded frame by (a) H.26L FGS and (b) Wyner–Ziv video coding in the seventh simulated transmission.

For Mother_daughter, the base layer is encoded at 257 Kb/s and the bit rate for the enhancement layer is 143 Kb/s. The total transmission rate is 500 Kb/s. With 20% FEC, the residual RTP packet loss rates at 2%, 4%, 6%, 8%, and 10% PDU loss rate are 0.47%, 1.13%, 2.49%, 4.03%, and 5.68%, respectively. After 200 simulated video transmissions (at each RTP packet loss rate), the average PSNR performance gains of layered WZC over H.26L FGS are 0.25, 0.29, and 0.3 dB when the PDU loss rate is 6%, 8%, and 10%, respectively.

From both Figs. 9 and 10, we see that layered WZC is more error robust than H.26L FGS in video streaming applications. In addition, we see that more performance gain in terms of average PSNR is obtained for high motion sequences like Football. This is because the error-drifting problem becomes worse for high motion sequences in standard DPCM-based video coding; on the other hand, the distributed nature of WZC makes it a promising and viable technique for alleviating the effect of error drifting associated with standard coders.

VII. Conclusion and Future Directions

We have proposed a practical layered Wyner–Ziv video coding system using the DCT, NSQ, and irregular LDPC code based SWC. The low-complexity DCT is used as an

approximation to the cKLT. NSQ is the simplest nested quantization scheme that corresponds to quantization in classic source coding. LDPC code based SWC exploits the correlation between the quantized version of the source and the side information, and can be viewed as the counterpart of entropy coding in classic source coding. Our layered video coding system achieves scalability as the layered Wyner–Ziv bitstream enhances the standard base layer bitstream in such a way that it is still decodable with commensurate qualities at rates corresponding to layer boundaries. Simulation results demonstrate that layered WZC is more error robust than H.26L FGS coding in video streaming applications.

Although our results are encouraging, much remains to be done to make Wyner–Ziv video coding more viable in practice. First, more accurate statistical modeling will improve the quantizer and the SWC design. Second, LDPC code design can be further improved by using density evolution without the Gaussian approximation for different source distributions. In addition, it is desirable to improve the performance of LDPC codes with shorter block length to reduce the time delay in video coding. Our current implementation of layered WZC only allows decoding at layer boundaries—decoding at the middle of a layer (bit plane) will suffer a high performance loss as unavailable bits in the layer have to be treated as erasures. The counterpart of scalable source codes (e.g., arithmetic codes) or progressively decodable channel codes are needed for *scalable* SWC to achieve fine-grained scalability. But scalable SWC remains to be a challenging problem.

We have relied on extensive simulations when testing error robustness of our layered Wyner–Ziv video coder, which is only designed under the assumption of *noiseless* channels. When the channel for the Wyner–Ziv enhancement bitstream is noisy, we have a problem of distributed joint source-channel coding [34]. On the other hand, if the channel for the base layer bitstream is not perfect, the side information might be absent at the Wyner–Ziv decoder; although theoretical results for optimal encoding in this case appeared more than 20 years ago [35], we have not seen any corresponding practical code design even for ideal sources yet.

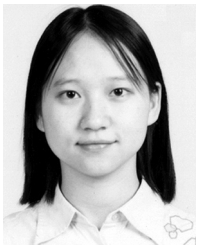
ACKNOWLEDGMENT

The authors would like to thank Profs. K. Ramchandran and J. Wolf and Dr. H. Garudadri at Qualcomm, Inc., for pointers to the CDMA channel simulator used in their simulations. They would also like to thank the anonymous reviewers and Prof. V. Stanković for their constructive comments, which have significantly improved the quality of this paper.

REFERENCES

[1] MPEG-4 Video VM 1999, ISO/IEC JTC 1/SC29/WG11 N2687, 13.0.
[2] T. Wiegand, G. Sullivan, G. Bjintegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
[3] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 301–317, Mar. 2001.
[4] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.
[5] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, pp. 1–10, Jan. 1976.
[6] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.
[7] S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Process. Mag.*, vol. 19, no. 3, pp. 51–60, Mar. 2002.
[8] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Process. Mag.*, vol. 21, no. 9, pp. 80–94, Sep. 2004.
[9] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," presented at the Allerton Conf. Communication, Control and Computing, Monticello, IL, Oct. 2002.
[10] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2005.
[11] A. Sehgal, A. Jagmohan, and N. Ahuja, "Wyner–Ziv coding of video: Applications to error resilience," *IEEE Trans. Multimedia*, vol. 6, no. 4, pp. 249–258, Apr. 2004.
[12] S. Shamai, S. Verdu, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 564–579, Mar. 1998.
[13] Y. Sterinberg and N. Merhav, "On successive refinement for the Wyner–Ziv problem," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1636–1654, Aug. 2004.
[14] S. Cheng and Z. Xiong, "Successive refinement for the Wyner–Ziv problem and layered code design," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3269–3281, Aug. 2005.
[15] A. Liveris, Z. Xiong, and C. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Commun. Lett.*, vol. 6, no. 10, pp. 440–442, Oct. 2002.
[16] M. Gastpar, P. Dragotti, and M. Vetterli, "The distributed, partial, and conditional Karhunen–Loeve transforms," presented at the DCC, Snowbird, UT, Mar. 2003.
[17] Y. He, R. Yan, F. Wu, and S. Li, "H.26L-based fine granularity scalable video coding," presented at the ISO/IEC MPEG 58th Meeting, M7788, Pattaya, Thailand, Dec. 2001.
[18] G. Cote, B. Erol, M. Gallant, and F. Kossentini, "H.263+: Video coding at low bit rates," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 6, pp. 849–866, Nov. 1998.
[19] Q. Xu and Z. Xiong, "Layered Wyner–Ziv video coding," presented at the VCIP, San Jose, CA, Jan. 2004.
[20] A. Sehgal, A. Jagmohan, and N. Ahuja, "Scalable Wyner–Ziv coding using Wyner–Ziv codes," presented at the PCS, San Francisco, CA, Dec. 2004.
[21] H. Wang and A. Ortega, "WZS: Wyner–Ziv scalable predictive video coding," presented at the PCS, San Francisco, CA, Dec. 2004.
[22] M. Tagliasacchi, A. Majumdar, and K. Ramchandran, "A distributed-source-coding based robust spatio-temporal scalable video codec," presented at the PCS, San Francisco, CA, Dec. 2004.
[23] H. Garudadri, H. Chung, N. Srinivasamurthy, and P. Sagetong, "3GPP2 simulation methodology for multimedia services," presented at the WPMC, La Jolla, CA, Sep. 2006.
[24] S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1181–1203, May 2003.
[25] A. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 1, pp. 2–10, Jan. 1974.
[26] D. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 399–431, Mar. 1999.
[27] M. Marcellin and T. Fischer, "Trellis coded quantization of memoryless and Gaussian-Markov sources," *IEEE Trans. Commun.*, vol. 38, no. 1, pp. 82–93, Jan. 1990.
[28] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3445–3463, Dec. 1993.
[29] B.-J. Kim, Z. Xiong, and W. Pearlman, "Very low bit-rate embedded video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1365–1374, Dec. 2000.
[30] D. Reininger and J. Gibson, "Distribution of the two-dimensional DCT coefficients," *IEEE Trans. Commun.*, vol. 31, no. 6, pp. 835–839, Jun. 1983.
[31] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[32] S. Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 657–670, Feb. 2001.
[33] Y. Wang and Q. Zhu, "Error control and concealment for video communications: A review," *Proc. IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.
[34] Q. Xu, V. Stanković, and Z. Xiong, "Distributed source-channel coding of video using Raptor codes," *IEEE J. Sel. Areas Commun.*, to be published.
[35] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 727–734, Nov. 1985.



Qian Xu (S'06) received the B.S. degree in computer science from the University of Science and Technology of China in 2002 and the M.S. degree in electrical engineering in 2004 from Texas A&M University, College Station, where she is currently pursuing the Ph.D. degree. Her research interests include distributed video coding, video compression, pattern recognition, and genomic signal processing.

Zixiang Xiong (S'91–M'96–SM'02) received the Ph.D. degree in electrical engineering in 1996 from the University of Illinois at Urbana-Champaign, Urbana.

From 1997 to 1999, he was with the University of Hawaii at Manoa. Since 1999, he has been with the Department of Electrical and Computer Engineering, Texas A&M University (TAMU), College Station, where he is an Associate Professor. He spent the summers of 1998 and 1999 at Microsoft Research, Redmond, WA, and the summers of 2000 and 2001 at Microsoft Research, Beijing, China. His current research interests are network information theory, code designs and applications, genomic signal processing, and networked multimedia.

Dr. Xiong received a National Science Foundation Career Award in 1999, an Army Research Office Young Investigator Award in 2000, and an Office of Naval Research Young Investigator Award in 2001. He also received faculty fellow awards in 2001, 2002, and 2003 from TAMU. He served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (1999 to 2005), the IEEE TRANSACTIONS ON IMAGE PROCESSING (2002 to 2005), and the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2002–2006). He is currently an Associate Editor for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS and a member of the multimedia signal processing technical committee of the IEEE Signal Processing Society. He is the Publications Chair of GENSIPS'06 and ICASSP'07 and the Technical Program Committee Co-Chair of ITW'07.