Inspired by Witsenhausen and Wyner's 1980 (now expired) patent on "interframe coder for video signals", this paper presents a Witsenhausen-Wyner video codec, where the previously decoded video frame is used at the decoder as side information for joint decoding. Specifically, we replace the predictive Inter coding in H.264/AVC by the syndrome-based coding scheme of Wistenhausen and Wyner, while keeping the Intra and Skip modes unchanged. Within Wistenhausen-Wyner coding mode, we optimize the decisions between syndrome coding and entropy coding among different DCT bands and among different bit-planes within each DCT coefficient. We employ forward motion estimation at the encoder and send the motion vectors to help generate the side information at the decoder, since our focus is not on low-complexity encoding. We also examine the tradeoff between the motion vector resolution and coding performance. Extensive simulations of video transmission over wireless networks show that Witsenhausen-Wyner video coding is much more robust against channel errors than H.264/AVC. The price paid for enhanced error-resilience with Witsenhausen-Wyner coding is a small loss in compression efficiency.

Index Terms—distributed video coding, error-resilience, syndrome coding

1. INTRODUCTION

After two decades of research on video compression, we now have a series of international standards such as MPEG2/H.262[1], H.263[2] and H.264/MPEG4 AVC[3], which are widely used in applications [4][5] for example satellite TV, DVDs, video telephony and so on. As portable devices such as camera phones and digital video cameras are penetrating deeper in people’s lives, wireless video applications are becoming more and more popular. The challenge in these applications is to deliver satisfactory video quality when the compressed video is transported over wireless channels, which are unstable and noisy. To meet this challenge, the key issue is error-resilient video coding. The well-known video coding standards such as MPEG2/H.262, H.263 and H.264/MPEG4 AVC all adopt the paradigm of joint-encoding and joint-decoding, which forcefully removes the redundancy in video signal and meanwhile makes the compressed bit-stream vulnerable to the channel noises. The errors introduced to the current frame will propagate to the following frames until the intra-coded one. Among three coding modes in H.264/AVC, namely Inter coding, Intra coding and Skip, Inter coding is chiefly responsible for exploiting the temporal correlation, but it is most sensitive to the channel errors and facilitates error-drifting. Intra coding achieves the best error robustness owing to independence on the previous frame, but its inferior spatial prediction lowers the coding efficiency. Skip mode is less fragile than Inter mode, since its reconstructions are directly derived from the previous frame and thus irrelevant of the channel errors in the bit-stream of current frame.

A natural question to ask is whether we can have a coding scheme that offers the middle ground - with coding efficiency approaching that of Inter coding while being error resilient (as Intra coding). The distributed source coding (DSC), grounded on the paradigm of separate-encoding and joint-decoding, offers an alternative solution to error-robust video compression for networked multimedia. Slepian and Wolf [6] laid the theoretical foundation of DSC by showing that the rate of near-lossless source coding with decoder side information only is the same as that of joint encoding (with side information at both the encoder and decoder). The rate-distortion function for lossy source coding with side information only available at the decoder, was established in [7]. Because the source signal is encoded independently of the side information and the channel decoding (e.g., nearest neighbor decoding) techniques are employed at the decoder, the distributed coding has the advantages of increased error resilience over joint encoding.
Following these theoretical works, some distributed video coding (DVC) technologies have been proposed [8-16] etc. Girod et al. proposed a DVC scheme with a low complexity encoder, which outputs parity bits to achieve compression, while the decoder counts on additional bits from the encoder through a feedback channel to adjust the bit rate [8]. The delay caused by the feedback requests prevents it from being used in real-time applications. Since the side information is generated with extrapolation or interpolation in [8], the size of GOP (i.e., group of pictures) is restricted and more intra-coded frames are required. A DSC-based video codec, which targets flexible allocation of complexity between encoder and decoder or error robustness, was presented in [9] [10] [11]. In it, the transform coefficients and bit-planes of nested scalar quantization (NSQ) indices are encoded with a designated coding method (i.e., entropy or syndrome coding) without considering the correlation statistics of each frequency band or bit-plane. Moreover, when motion estimation is performed at the encoder, the depth of NSQ used for each individual DCT coefficient depends on the specific residual between the original coefficient and its side information, additional overhead indicating the depth information for each coefficient must be inserted into the bit-stream. If the decoder utilizes Cyclic Redundancy Check (CRC) to estimate the motion vectors and generate the side information, the coding performance suffers from fake side information blocks chosen by fixed-length CRC checksums that have finite error correction capability. Besides the DVC technologies mentioned before, there are also some works concentrating on side information generation at the decoder [12-14], transform [15] or channel coding [16].

The idea of video coding based on DSC principles was first proposed by Witsenhausen and Wyner in their 1980 US patent on an “inter-frame coder for video signals” [17]. Despite the flurry of recent research activities on DVC, the number of references to Witsenhausen and Wyner’s patent on Google scholar is only 18 (at the time of this writing) and the patent itself expired without actually being implemented. We first take a closer look at the patent. Assume the video signal source is binary, and the current frame X and the previous frame Y are correlated after motion alignment, e.g., with i.i.d X and Y satisfying the relationship X = Y + D where D represents the residual between X and Y. Each frame is divided into n-bit blocks. At the heart of Witsenhausen and Wyner’s encoder is a syndrome former that outputs S(X) = TX, which is the syndrome of X, where X is a block of the current frame and T is an (n-k)xn parity-check matrix of an error correction code (ECC) for the binary channel with X and Y being its input and output. Since only the syndrome S(X) is transmitted to the decoder, the encoder achieves a compression ratio of n/(n-k). After receiving the transmitted syndrome S(X), the Witsenhausen-Wyner video decoder first computes the syndrome difference S(X)-S(Y) = T(X-Y), since it already has the previous frame Y, before applying hard-decision decoding on S(X)-S(Y) to recover D = X-Y noiselessly by picking D as the leader of the ECC coset indexed by S(X)-S(Y). Finally, D is added to Y to form the reconstructed X. For general M-ary video sources, the patent suggests the use of M-ary ECC or the above binary syndrome-based scheme for the most significant bit-plane in conjunction with conventional coding schemes for the remaining bit-planes. In section 2, we will further investigate how the patent is related to recent papers (e.g., [9][18] and related works).

Following Witsenhausen and Wyner’s patent, we propose a novel Witsenhausen-Wyner Video Coding (WWVC) scheme, in which the correlation in video sequence is exploited at four levels. First, the mode decision according to temporal correlation in H.264/AVC is directly applied to each block of current frame. In place of the predictive Inter coding, WWVC takes the charge of exploiting temporal correlation among video frames, thanks to its Inter-like coding efficiency and insensitivity to channel errors. The Intra and Skip modes are kept the same as in H.264/AVC. Second, within WWVC mode, the blocks having similar correlation are clustered together and the coding parameters for each class are specified. Third, the mode decision is performed over DCT domain. The coefficients that have much higher correlation with side information than the statistical correlation are fed into syndrome coding, while others are just entropy coded. Fourth, as for the syndrome-coded coefficients, the bit-planes from NSQ indices are also classified into syndrome-coding mode and entropy-coding mode based on the correlation of each bit-plane. In a word, in our proposed WWVC scheme, both the coding parameters and the coding modes are specified according to the correlation statistics, which is different from other DVC works.

We evaluate the proposed WWVC scheme both in noisy and noiseless channels through a wireless channel simulator of RTP/IP [19] over 3GPP [20] networks from Qualcomm Inc. The coding efficiency and error resilience of WWVC are compared with that of H.264/AVC and H.264/AVC-IntraSkip (i.e. a modified H.264/AVC only allowing Intra and Skip modes by making the Inter mode disabled, which can significantly enhance the error resilience of H.264/AVC). Extensive simulations show that, WWVC performs better than both H.264/AVC and H.264/AVC-IntraSkip when the bit-stream is delivered in noisy channel. The improved error resilience of WWVC scheme is achieved at a small loss of coding efficiency compared to H.264/AVC, when the transmission is noiseless.

The rest of the paper is arranged as follows. In Section 2, the relationship between Witsenhausen-Wyner coding and Slepian-Wolf coding is illustrated. The proposed WWVC scheme is described in Section 3. Section 4 gives the simulation results of WWVC in both noiseless and noisy channels, as well as comparisons with those of H.264/AVC. Finally, we conclude the whole paper in Section 5.

2. WITSENHAUSEN AND WYNER’S PATENT
In this section, we will take a fresh look at how Witsenhausen and Wyner's patent relates to Slepian-Wolf coding (SWC). The syndrome-based encoding step in Witsenhausen and Wyner's patent is the same as that in SWC [18]. The first decoding step in the patent, forming \( TD = T(X-Y) \), could easily be used in SWC [18] and decode for \( D \) instead of \( X \) directly, as done in SWC [18]. The “noisy” vector in this case would be an all-zero vector instead of \( Y \). The result of the decoding algorithm should then be added to \( Y \) to form the most likely \( X \), similarly to the patent. We will take a generalized instance of binary sources to further illustrate the equivalence.

Assume the source \( X \) and the side information \( Y \) are equiprobable binary triplets with \( X, Y \in \{0,1\} \) and they differ at most in one position. The Slepian-Wolf encoder partitions the set of all possible outcomes of \( X \) into four cosets \( C_{00}, C_{01}, C_{10}, C_{11} \) with \( C_{00} = \{000,111\}, C_{01} = \{001,110\}, C_{10} = \{010,101\} \) and \( C_{11} = \{011,100\} \), in order that the two elements in one coset have Hamming distance \( d_h = 3 \). The encoder only needs to send two syndrome bits for the index \( S \) of the coset \( C \) that \( X \) belongs to regardless of the side information \( Y \). The joint decoder picks in the coset \( C \) the \( X \) with \( d_h \leq 1 \). Hence, as for SWC, it is possible to send \( H(X|Y) = 2 \text{bits} \) instead of \( H(X) = 3 \text{bits} \) for \( X \) and decode it without loss at the joint decoder, where \( H(X) \) denotes the self entropy of \( X \) and \( H(X|Y) \) the conditional entropy of \( X \) given \( Y \). For example, \( X = [001] \) and \( Y = [101] \). \( X \) belongs to \( C_{01} \) and \( S(X) = [01] \) is sent. At the receiver, the decoder chooses the symbol \( [001] \) that is most correlated with the side information \( [101] \) from the coset specified by the syndrome \( S(X) \).

In Witsenhausen and Wyner's patent, the encoder is the same as that in SWC. At the decoder, the side information \( Y = [101] \) is first fed into syndrome former to get the syndrome \( S(Y) = [10] \), since \( Y \) is included in the coset \( C_{01} \). Then, the decoder calculates \( S(D) = S(X) - S(Y) = [11] \). The coset leader \( [100] \) is selected from \( C_{11} \) as \( D \). Finally, \( D \) is added to \( Y \) and get \( X = Y + D = [001] \), which is identical to the result of SWC discussed above.

Nevertheless, the patent differs from SWC [18] in terms of the decoding algorithm. Hard-input decoding is used in the patent, whereas soft-input iterative message-passing decoding is used in [18]. This could be because at the time the patent was written syndromes were mainly used for decoding of block codes (hard-input) while soft-input decoding was only considered possible for convolutional codes through the Viterbi algorithm. The hard-decision Witsenhausen-Wyner decoder is optimum under the binary symmetric channel (BSC) model, meaning that in the setup of [18] it will give the same performance as the message-passing algorithm. The reason is that, in BSC model, the soft-input probabilities to the decoder only have two levels (more likely and less likely) and the decoding with the minimum Hamming weight is thus the maximum-likelihood approach. Of course, this will not be the case when the correlation model is not the BSC in which the soft-input to the decoder has more levels, e.g., different bits have different cross-over probabilities or there is an additive white Gaussian noise (AWGN) channel model. Since Witsenhausen and Wyner first conceived the idea of video compression based on DSC principles and in some cases the decoder in their patent is optimal and acquires the same result as state-of-the-art Slepian-Wolf decoders, we advocate the use of Witsenhausen-Wyner coding when the previous video frame is adopted to generate the side information at the Slepian-Wolf/Wyner-Ziv decoder [8][9]. We think of this as a tangible way of giving due credits to Witsenhausen and Wyner.

3. THE PROPOSED WITSENHAUSEN-WYNER VIDEO CODER

We propose a Witsenhausen-Wyner video coder with the motivation of robust video coding for unstable video transmission. The proposed Witsenhausen-Wyner encoder and decoder are illustrated in Fig.1 and Fig.3 respectively. It can be seen that the main difference from H.264/AVC coding lies in that the WWVC is substituted for the Inter coding, so the encoder does not need to calculate the residual between the original video input and the side information. The encoder only enlists the side information in the mode decision among WWVC, Intra, and Skip, and the classification of the WWVC-mode blocks based on the mutual correlation. The side information is generated from the previously reconstructed frame with motion compensation at the Witsenhausen-Wyner encoder. The resolution of motion estimation is adaptively selected from full-pixel, half-pixel, and quarter-pixel, based on the available computational resources. The complexity of interpolating the previous frame and motion search can be saved at the expense of the quality of prediction. The resulting motion vectors are entropy encoded and sent to produce the side information at the decoder. The Intra and Skip modes in H.264/AVC coding are maintained in the proposed scheme, since they are inherently resilient against the channel noises.

3.1. The proposed Witsenhausen-Wyner video encoder

The Witsenhausen-Wyner video encoder is developed with sufficient consideration of correlation in video signal at four levels. First, the video blocks are classified into Intra, WWVC, and Skip modes based on the correlation between the original block to be encoded and the motion compensated side information. Second, the WWVC blocks with similar correlation are further clustered. The coding parameters of each class are achieved with the correlation. Third, based on the distribution of correlation in transform domain, the coding method is selected between syndrome coding and entropy coding for each DCT frequency band. Fourth, the coding mode decision is also made for each bit-plane of the NSQ indices considering the variation of correlation from the most significant bit-plane to the least significant one.
As illustrated in Fig. 2, the blocks of WWVC mode go through DCT, scalar quantization, NSQ, LDPC based syndrome coding and entropy coding in turn. All the components are introduced as follows.

**Transform:** In order to remove spatial redundancy, distributed coding is carried out in transform domain. The block-wise DCT is applied to both the original frame and its side information. In each video frame, the transform coefficients belonging to the same frequency band are grouped together. Denote $X_i$ and $Y_i$ ($0 \leq i < L$) as the coefficients in the $i^{th}$ frequency band of the original and the side information frames. There are $L$ frequency bands in total which are ordered in zig-zag manner starting from the DC (zero-frequency) one. In Fig.2, the column corresponding to each band is filled with a specific pattern.

**Coding mode decision:** The mode decision from H.264/AVC is used in WWVC with the only change of relabeling Inter as WWVC. Furthermore, the blocks of WWVC mode are classified according to how they correlate with their side information. The correlation is calculated in terms of the mean squared error (MSE) between the transform coefficients of the WWVC block and their corresponding side information coefficients. For each class and each DCT band, a set of coding parameters including the depth of NSQ denoted as $N_{ij}$ and the weighting factors involved in the joint reconstruction of the decoder is determined from the correlation statistics, where $i$ and $j$ represent the DCT index and the class index respectively. The class index needs to be transmitted to the decoder.

**Scalar quantization:** After DCT, all the coefficients are scalar quantized with uniform quantization parameter $q$, resulting in the quantized symbols as

$$X'_i = \text{int}(X_i / q),$$

where int(.) represents the function that returns the closest integer of a float value. It seems that the scalar quantization removes $\text{int}(\log_2(q))$ least significant bit-planes, which are marked in black in Fig.2.

**Mode decision at DCT-band level:** In joint encoding such as H.264/AVC, the temporal redundancy that the original frame has with reference to the side information is mainly removed by calculating the residual signal between them. Then the entropy coding is further used to eliminate the statistical redundancy in residual. However, in WWVC, the WWVC-mode blocks are coded independently of the side information, so the encoder needs to select the coding method from syndrome coding and entropy coding in order that the temporal correlation and the statistical one can be exploited as far as possible. Since the scalar quantization dominates the distortion, the optimization mainly targets bit-rate saving.

Since the correlation tends to decrease from low-frequency band to high-frequency one, it is necessary to determine how to code each DCT band individually. We analyze the statistics in terms of the self entropy $H(X'_i)$ and the conditional entropy $H(X'_i | Y_i)$ for each DCT band. If the conditional entropy gets close to the self one, the band had better be entropy coded. Otherwise, syndrome coding turns out to be more efficient. As shown in Fig.2, $M$ low-frequency bands are classified into syndrome coding (SC) mode and other $(L-M)$ bands belong to entropy coding (EC) mode. In noisy channel, a smaller $M$ is meant to be adopted.

**Nested scalar quantization:** Given a SC-mode coefficient $X'_i$, that is in the $i^{th}$ DCT band and taken from a block of
class $j$, a coarse channel code with the minimum distance $d_{\min}=2^{N_{W}}/q$ is applied, resulting in the NSQ index

$$B_{i,j} = X_{i,j}^{*} \mod (2^{N_{W}}) .$$

(2)

It seems that several most significant bit-planes are excluded from the delivered bit-stream and will be estimated from the side information at the decoder. When

$$|Y_{i,j} - X_{i,j}^{*}| < \frac{d_{\min}}{2}$$

(3)

the decoder is able to find out the correct $X_{i,j}$ by choosing the codeword closest to the side information coefficient $Y_{i,j}$ in the coset identified with the index $B_{i,j}$. However, when

$$|Y_{i,j} - X_{i,j}^{*}| \geq \frac{d_{\min}}{2}$$

(4)

the decoder is likely to select an incorrect codeword. Thus, besides the distortion introduced by the fine scalar quantization, there exists the distortion caused by coarse channel code. Since the scalar quantization parameter $q$ is set as in H.264/AVC coding, the NSQ parameter $N_{i,j}$ is determined based on the mutual correlation between the source coefficients and the side information so that the distortion caused by coarse channel code can be minimized. For each class and each DCT band,

$$N_{i,j} = \text{int}(\log_{2}(\frac{\gamma \times \sigma_{i,j}}{q})) ,$$

(5)

where $\sigma_{i,j}$ denotes the root MSE between $X_{i,j}$ and $Y_{i,j}$. The more correlated $X_{i,j}$ and $Y_{i,j}$ are, the more information is expected to be inferred from $Y_{i,j}$ about $X_{i,j}$ at the decoder, leading to a smaller $N_{i,j}$. The weighting factor $\gamma$ in (5) roughly describes the temporal correlation of video sequence and can be tuned based on the motion speed of video sequence.

**Mode decision at bit-plane level:** As for $M$ low frequency coefficients of SC-mode, the resulting NSQ indices $B_{i,j}$ are expressed with $N_{i,j}$ bit-planes in a top-down manner. In order to increase the code length of syndrome coding and the coding efficiency accordingly, the NSQ indices from different DCT bands and the blocks of different classes are grouped together. Given a bit-plane $E_{i}$, depending on the difference between the self-entropy $H(E_{i})$ and the conditional entropy $H(E_{i}|Y_{i},E_{i-1},...,E_{i-1})$, which are the minimum bit rates for entropy coding and syndrome coding respectively, a decision between SC-mode and EC-mode is made. The subscripts of $Y$ are omitted here, since there is no need to distinguish the DCT band and the class. For each bit-plane, a larger difference between two entropy values implies the syndrome coding is likely to save more bit rate over the entropy coding. As shown in Fig.2, $K$ most significant bit-planes denoted as $E_{0}E_{1}...E_{K-1}$, are anticipated to be syndrome coded. The remaining bit-planes are entropy coded directly.

[Fig.3 The proposed Witsenhausen-Wyner video decoder.]

**Syndrome encoding:** As for the SC-mode bit-planes, owing to the complexity and memory requirements of $M$-ary ECC, we decide to adopt the multilevel LDPC codes to compress $E_{0}E_{1}...E_{K-1}$ with the syndrome-based approach [18]. Each bit-plane in SC-mode is individually encoded with one LDPC code. Only the syndrome bits are transmitted to the decoder (to achieve compression). The rate of syndrome bits for $E_{i}(0 \leq i \leq K-1)$ depends on the conditional entropy $H(E_{i}|Y_{i},E_{i-1},...,E_{i-1})$, which denotes the minimum rate needed for lossless recovery of $E_{i}$ given previously decoded bit-planes $E_{0}E_{1}...E_{i-1}$ and the side information $Y$ at the decoder [21].

In our irregular LDPC code designs, the code degree distribution polynomials are optimized using density evolution [22] under Gaussian approximation. The bipartite graph for the irregular LDPC code, which determines the sparse parity check matrix, is then randomly constructed based on the optimized code degree distribution polynomials. **Entropy encoding:** Due to weakened correlation with the side information, $(L-M)$ high frequency coefficients of EC-mode and $(N_{i,j}-K)$ EC-mode bit-planes extracted from SC-mode coefficients are compressed with context-adaptive arithmetic coding (CABAC) in H.264/AVC coding.

**3.2. The proposed Witsenhausen-Wyner video decoder**

Fig.3 illustrates the proposed Witsenhausen-Wyner video decoder. At the receiver, Intra and Skip blocks are decoded as in H.264/AVC. The previously reconstructed frame is used to generate the side information frame together with the received motion vectors. After the same DCT as at the encoder, the coefficients of each band $Y_{i}(0 \leq i < L)$ are produced. Besides, indicated by the received class indices, the coding parameters of each block are known to the decoder.

**Entropy decoding:** The entropy coded bit-stream is decompressed directly to restore the EC-mode coefficients and bit-planes. The $(N_{i,j}-K)$ least significant bit-planes are reproduced and used to restore the NSQ indices $B_{i,j}$ together with the syndrome decoded $K$ most significant bit-planes.
The weighting factor, used in [18]. Since we adopt multilevel LDPC code instead in general optimal, we thus employ a soft-decision decoder decision decoder in Witsenhausen and Wyner's patent is not quantization indices, where \( p(.|.) \) denotes the conditional probability function and \( E_i \) denotes the decoded \( E_i \).

**Syndrome decoding:** As discussed in Section 2, the hard-decision decoder in Witsenhausen and Wyner’s patent is not in general optimal, we thus employ a soft-decision decoder in which the received syndrome bits correspond to the check nodes on the bipartite graph. The log-likelihood ratio (LLR), which describes the \( a \) priori probabilities about the candidate decoding output can be calculated with the side information and the previously decoded bit-planes as

\[
LLR = \log \frac{p(E_i = 0 | E_{i+1}, ..., E_n, Y)}{p(E_i = 1 | E_{i+1}, ..., E_n, Y)}
\]

(6)

where \( p(.) \) denotes the conditional probability function and \( E_i \) denotes the decoded \( E_i \).

**Dequantization and inverse transform:** The optimal reconstructed coefficient \( X_{ij}^* \) is estimated based on the quantization indices \( X_{ij} \) and the side information \( Y_{ij} \) as

\[
X_{ij}^* = \frac{\omega_{ij}^2 \times (X_{ij} \times q) + \omega_{ij}^p \times Y_{ij}}{\omega_{ij}^2 + \omega_{ij}^p}.
\]

(7)

The weighting factor \( \omega_{ij}^p \) is just the \( \sigma_{ij}^2 \), and \( \omega_{ij}^q \) is calculated from the MSE between \( X_{ij} \) and \( (X_{ij} \times q) \). Finally, an inverse DCT is performed over the reconstructed coefficients to obtain the video output.

**4. EXPERIMENTAL RESULTS**

The proposed WWVC scheme aims at achieving the error resilience when the compressed video bit-stream is transmitted over noisy wireless networks. We evaluate the performance of the WWVC scheme under both noiseless and noisy channel conditions, and compare WWVC with H.264/AVC. Since the resolution of the motion estimation is selected according to the available computational resources of the encoder in WWVC, we generate the results of WWVC and H.264/AVC with full-pixel, half-pixel and quarter-pixel based motion estimation. The curves are labeled as H.264/AVC(WWVC)-FullPel, H.264/AVC(WWVC)-HalfPel, and H.264/AVC(WWVC)-QuarterPel in this paper. Besides, H.264/AVC-IntraSkip coding, which disables Inter mode and only allows Intra and Skip modes, is also added to the simulation as a straightforward method to enhance the error resilience of H.264/AVC. When the bit-stream is sent over noiseless channel, the coding efficiency is given in terms of PSNR versus bit-rate. When the channel is noisy, the average PSNR values of the reconstructions achieved at the decoder with different packet loss rates are adopted to assess the error robustness of various schemes.

**4.1. The coding efficiency in noiseless channels**

In this subsection, we evaluate the coding efficiency of WWVC in noiseless transmission. All the bit-streams are generated with some common settings. For simplicity but without losing generality, 16x16, 16x8, 8x16 block-wise motion estimation and 8x8 block-wise DCT are used. CABAC is utilized to do the entropy coding in H.264/AVC scheme and H.264/AVC-IntraSkip. The EC-mode coefficients of the WWVC-mode blocks in WWVC scheme are also coded with CABAC. One coding slice contains one frame. Data partition is enabled in all the schemes. The bit-stream is divided into header part, intra part and inter (or WWVC) part. Each of them is coded and delivered individually. The first fifteen frames from Football (720x486, 15frames per second) are used in the simulation.

<table>
<thead>
<tr>
<th>Table 1 The classification of the WWVC-mode blocks.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 0</td>
</tr>
<tr>
<td>{0,70.5}</td>
</tr>
<tr>
<td>Class 4</td>
</tr>
<tr>
<td>{425, 687.5}</td>
</tr>
</tbody>
</table>

![Fig. 4 The coding efficiency of the proposed WWVC, H.264/AVC coding and H.264/AVC-IntraSkip coding schemes.](image-url)

The WWVC-mode blocks of Football are classified into eight classes. The thresholds for classification in terms of MSE between the transform coefficients of WWVC blocks and those of their corresponding side information are listed in Table 1. Thirty-six low-frequency coefficients are included in SC-mode. Three most significant bit-planes are fed into the irregular LDPC based syndrome coding while other bit-planes are entropy coded. Hence, three LDPC codes are needed. The degree distribution polynomials of the LDPC codes are optimized with the approximation of AWGN channel. We adopt the LDPC codes given in [24]. The coding efficiency in terms of PSNR versus bit-rate is shown in Fig. 4. With the same motion estimation, the curve of WWVC is up to 0.5dB lower than that of H.264/AVC.
Besides, H.264/AVC-IntraSkip scheme gives the worst coding efficiency.

4.2. The error robustness in noisy channels

We adopt a wireless channel simulator for RTP/IP [19] over 3GPP [20] from Qualcomm Inc. to evaluate the error robustness of WWVC, H.264/AVC and H.264/AVC-IntraSkip schemes. The simulator transmits the real-time transport protocol (RTP) stream with a radio channel. The transmission rate is 64kbps in our simulation. Each RTP packet is fragmented into equal-size protocol data units (PDU). In the simulation each PDU contains 640bytes. Channel errors are randomly introduced to the PDUs. If all the PDUs belonging to one packet are received and the arriving time of its last PDU is still within the maximum end-to-end delay, the packet is considered successfully received by the decoder. The Qualcomm simulator also provides FEC simulation with Reed-Solomon (RS) code.

Fig. 5 The error robustness of the proposed WWVC, H.264/AVC coding and H.264-IntraSkip coding schemes.

As mentioned before, all generated bit-streams are partitioned into header, intra and inter (or WWVC) three parts. In all simulations, the header part is assumed to be error free. When the packet loss randomly happens to the other two parts, a simple error concealment which extracts the reconstruction from the previously decoded frame or neighboring reconstructed blocks is utilized. Besides, ideal error detection is assumed. When large errors in previous frame are found out, LLR in Equation (6) is set to be zero. It means that when the previous frame is not that reliable, no prior knowledge from that is used.

For fair comparison, we use the identical bit rate to generate bit-streams with different coding schemes. The bit-rate for Football is 3780kbps. We add 25% FEC overhead to each bit-stream, i.e., 20% of the overall transmission rate is used for RS-based FEC. Thus, the overall transmission rate for Football is 4725kbps. For each PDU loss rate, 100 runs are simulated.

Fig. 5 depicts the average PSNR versus PDU loss rate achieved with three video coding schemes. It can be found, in each figure, the curves corresponding to WWVC and H.264/AVC counterpart have a crossing point. WWVC gets more advantageous than H.264/AVC as the channel gets worse. No matter whether the channel is noisy or noiseless, H.264/AVC-IntraSkip is interior to WWVC.

5. CONCLUSIONS

Inspired by Witsenhausen and Wyner's 1980 patent, we propose a WWVC scheme in this paper, which is mainly devoted to robust video transmission in dynamic and diverse networks. In the proposed scheme, the correlation of video signal is exploited at four levels. First, the video blocks to be coded are classified into Intra, WWVC, Skip modes according to the correlation between the original block and the side information that is achieved with the mode decision algorithm in H.264/AVC. Particularly, DSC-based WWVC replaces the predictive Inter coding in H.264/AVC. Second, within the WWVC mode, the blocks having similar correlation are grouped together and the coding parameters for each class are specified according to the correlation. Third, DCT domain of the current frame has uneven distribution of temporal correlation with the side information. Relying on whether the temporal correlation or the statistical one has more potential to save bit rate, a good judgement between syndrome coding and entropy coding is made for each DCT band at the encoder. Fourth, in order to exploit the temporal and statistical correlation of each bit-plane taken from the syndrome coded coefficients as far as possible, the mode decision between syndrome coding and entropy coding is made for each bit-plane.

In this paper, the simulation results of WWVC in terms of coding efficiency and error resilience in various networks are given. The proposed WWVC scheme achieves better performance than both H.264/AVC and H.264/AVC-IntraSkip in noisy networks. The coding efficiency of
WWVC in noiseless channel suffers a limited loss (up to 0.5dB), compared to H.264/AVC.

Besides, in WWVC, the resolution of motion estimation is chosen from full-pixel, half-pixel and quarter-pixel according to the computational resources at the encoder. The motion vectors are delivered to generate the side information at the decoder. The tradeoff between the performance and the complexity of motion estimation is also examined. The performance of both H.264/AVC and WWVC declines as the complexity is reduced.

6. REFERENCES