

# Iterative Joint Video and Channel Decoding in a Trellis-Based Vector-Quantized Video Codec and Trellis-Coded Modulation Aided Wireless Videophone

R. G. Maunder, J. Kliewer, S. X. Ng, J. Wang, L-L. Yang and L. Hanzo

School of ECS, University of Southampton, SO17 1BJ, UK.

Email: {rm02r,jk03v,sxn,jw02r,lly,lh}@ecs.soton.ac.uk, http://www-mobile.ecs.soton.ac.uk

**Abstract** – We propose a Vector-Quantized (VQ) video codec that induces deliberately-introduced code constraints. The complete set of these code constraints are unambiguously described by a novel block-based modification of a symbol-based variable length coding trellis structure. This trellis is employed during VQ encoding to represent video information with a Minimum Mean Squared Error (MMSE) reconstruction. Additionally, the proposed trellis structure is employed in the video decoder to obtain *a posteriori* probability based MMSE estimates of the transmitted VQ tiles. Since the trellis structure unambiguously describes the code constraints induced by the video codec, its employment guarantees the automatic recovery of valid sets of VQ tiles. The proposed video codec is serially concatenated with a trellis-coded modulation scheme for transmission over a fast-fading Rayleigh fading channel. In the receiver, iterative-decoding convergence that eliminates all perceivable channel-induced video distortion is shown to be supported within 1.29 dB of the system’s 2 bit/s/Hz channel-capacity limit of 3.96 dB.

## I. INTRODUCTION

The limited validity of Shannon’s source and channel coding separation theorem [1] for practical video transmission systems over realistic wideband fading mobile radio channels [2] motivates the application of joint source and channel coding techniques. One such technique is offered by the extension of serially-concatenated iterative decoding [3] to exploit the residual redundancy offered by the code constraints induced by a video codec. These code constraints are either deliberately induced during the design of the source codec or are inherently present, owing to low-complexity source codec design, for example. In such systems, the inner and outer iterative-decoding stages are comprised by the channel and the source decoder, respectively. Extrinsic information is exchanged between these iterative-decoding stages, with each using a specific trellis-based description of the corresponding code constraints.

Video codecs typically have an algorithmically-complex design due to their exploitation of the psycho-visual imperfections of human vision [2]. Hence, it is typically difficult to describe the complete specification of the code constraints induced by a video codec in a trellis form. Additionally, a serially-concatenated iterative-decoding aided system that aims for exploiting the residual redundancy resulting from the code constraints of a typical video codec may not support iterative-decoding convergence that eliminates all perceivable channel-induced video distortion. Thus, in [4], a trellis-based intermediate decoder comprised the outer iterative-decoding stage of a serially-concatenated video transmission system. Following the iterative recovery of the video-encoded bits, they were forwarded to a stand-alone video decoder for video reconstruction. Since the code constraints induced by the video codec were not exploited during the iterative-decoding process,

The financial support of the EPSRC, Swindon UK, the EU under the auspices of the NEWCOM, NEXWAY and Phoenix projects and the Leverhulme Trust, London UK is gratefully acknowledged.

the recovered video-encoded bits may not constitute a legitimate representation of video information, resulting in reduced-quality video reconstruction.

Against this background, the concept of iterative joint video and channel decoding is investigated in this contribution, where the outer iterative-decoding stage is constituted by a specifically designed trellis-based video decoder, rather than the intermediate decoder of [4]. In order to demonstrate the benefits of this principle, we propose the VQ-TCM system in Section II. Here, the inner and outer iterative-decoding stages are provided respectively by a Trellis-Coded Modulation codec and a block-based video codec, which employs both Vector Quantization (VQ) and Variable Length Coding (VLC). As it will be described in Section III, a novel block-based modification of the VLC-based symbol-level trellis structure of [5] is employed to unambiguously describe the complete set of code constraints induced by the video codec. Using a similar philosophy to Trellis Coded Quantization (TCQ) [6], a novel employment of this trellis structure is invoked in the proposed video encoder to perform VQ, as it will be described in Section IV. Finally, a novel block-based modification of the The Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm [7] is employed by the proposed video decoder, which also operates on the basis of the proposed trellis structure, as it will be described in Section V. In Section VI, both low- and high-latency implementations of the proposed VQ-TCM system are introduced and their video reconstruction quality is assessed. Additionally, the proposed VQ-TCM system’s iterative-decoding convergence performance is investigated with the aid of an Extrinsic Information Transfer (EXIT) chart analysis [8]. Finally, the performance of the proposed joint video and channel coding approach is compared to that of the bench marker of [4]. Our conclusions are offered in Section VII.

## II. SYSTEM OVERVIEW

In this section, we introduce the proposed VQ-TCM system, which employs a joint video and channel coding philosophy, as alluded to in Section I. The video codec employs an algorithmically-simple design, having only frame differencing and VQ components, with only the latter imposing code constraints on the video-encoded bits. We commence by describing the operation of frame differencing and iterative decoding, with the specifics of VQ encoding and decoding being saved for Sections IV and V, respectively.

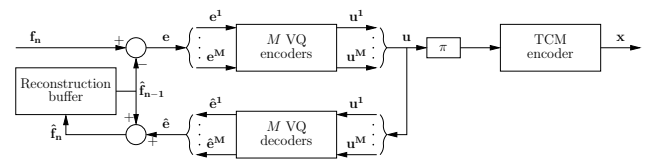


Fig. 1. The proposed VQ-TCM system’s transmitter.

In the proposed VQ-TCM system’s transmitter, shown in Figure 1, the Frame Difference (FD)  $\mathbf{e} = \mathbf{f}_n - \hat{\mathbf{f}}_{n-1}$  is vector quantized for the sake of conveying the video information to the receiver. Each FD  $\mathbf{e}$  is represented by a transmission frame  $\mathbf{u}$ , as will be detailed in Section IV. The transmission frame  $\mathbf{u}$  is interleaved ( $\pi$ ) and TCM-encoded, giving the channel-input symbols  $\mathbf{x}$ , as seen in Figure 1. As usual, a local decoder is employed to

obtain the VQ reconstruction  $\hat{\mathbf{e}}$  of the FD  $\mathbf{e}$ . This classic local-decoding technique is employed to obtain the local reconstruction of the current video frame  $\hat{\mathbf{f}}_{\mathbf{n}} = \hat{\mathbf{e}} + \hat{\mathbf{f}}_{\mathbf{n}-1}$ , which becomes the frame-differencing reference for the next video frame.

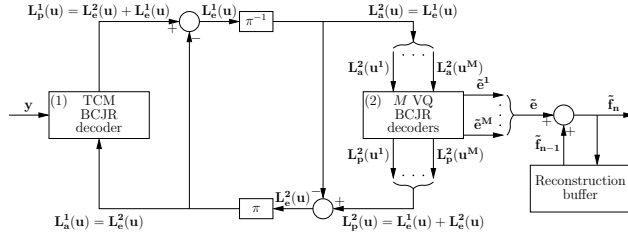


Fig. 2. The proposed VQ-TCM system's receiver.

A serially-concatenated iterative-decoding design is employed by the proposed VQ-TCM system's receiver, as shown in Figure 2. As in [4], the inner iterative-decoding stage is constituted by a TCM decoder. However, the outer iterative-decoding stage of the proposed VQ-TCM system's receiver is constituted by the VQ decoder, rather than an intermediate decoder as employed in [4]. Observe in Figure 2 that, during each decoding iteration, both iterative-decoding stages generate *a posteriori* Logarithmic Likelihood-Ratios (LLRs) [9]  $\mathbf{L}_{\mathbf{P}}(\mathbf{u})$  by applying their associated code constraints to the *a priori* LLRs  $\mathbf{L}_{\mathbf{a}}(\mathbf{u})$  supplied by the other iterative-decoding stage and, in the case of the inner iterative-decoding stage, by the received channel-output symbols  $\mathbf{y}$ . Following each iterative-decoding stage, the subtraction of  $\mathbf{L}_{\mathbf{a}}(\mathbf{u})$  from  $\mathbf{L}_{\mathbf{P}}(\mathbf{u})$  is necessary for the sake of generating the extrinsic LLRs  $\mathbf{L}_{\mathbf{e}}(\mathbf{u})$  (plus the information recovered from  $\mathbf{y}$  in the case of the inner iterative-decoding stage), which allows us to prevent the re-use of already-exploited information. It is the resultant extrinsic information that provides the *a priori* information for the next iterative-decoding stage. De-interleaving ( $\pi^{-1}$ ) is applied to  $\mathbf{L}_{\mathbf{e}}^1(\mathbf{u})$  in order to generate  $\mathbf{L}_{\mathbf{a}}^2(\mathbf{u})$ , whilst interleaving is applied to  $\mathbf{L}_{\mathbf{e}}^2(\mathbf{u})$  for the sake of producing  $\mathbf{L}_{\mathbf{a}}^1(\mathbf{u})$ . Iterative decoding continues until convergence is achieved, as it will be detailed in the context of the EXIT chart analysis of Section VI. Following this, a soft FD reconstruction  $\hat{\mathbf{e}}$  is provided, as it will be highlighted in Section V. Finally, the reconstruction of the current video frame is obtained as  $\hat{\mathbf{f}}_{\mathbf{n}} = \hat{\mathbf{e}} + \hat{\mathbf{f}}_{\mathbf{n}-1}$ .

As will be described in Sections IV and V respectively, VQ encoding and decoding is performed with the assistance of the novel trellis structure to be described in Section III-C. Owing to the specific nature of this trellis structure, its complexity is proportional to the square of the represented FD area. For this reason, each FD  $\mathbf{e}$  and the corresponding transmission frame  $\mathbf{u}$  is divided into  $M$  sub-frames, as will be described in Section III-A, with each encoded and decoded separately. These sub-frames are identified by the denotive superscript  $m$ , where  $m \in [1 \dots M]$  is the sub-frame index, and the decompositions into sub-frames, as well as their recombinations, are represented by braces in Figures 1 and 2.

### III. VQ-INDUCED CODE CONSTRAINTS

As stated in the previous section, each FD  $\mathbf{e}$  is represented by a single transmission frame  $\mathbf{u}$  in the proposed VQ-TCM system. This section outlines the format of these transmission frames and describes the VQ-induced code constraints. These code constraints must be adhered to during VQ encoding and they deliberately introduce residual redundancy, without unduly increasing the associated bitrate, which may be exploited during VQ decoding to assist the iterative-decoding process of the proposed VQ-TCM system's receiver, as described in Section II. These code constraints are induced by the specific decomposition of each FD  $\mathbf{e}$  into FD sub-frames  $\mathbf{e}^m$  and by the specific nature of the VQ codebook employed. Owing to the simple nature of the proposed video codec, the complete set of these code constraints may be represented using a novel trellis structure, which may be employed to assist VQ encoding and decoding. These issues are discussed

in the following sub-sections and are described with the aid of an example that is illustrated in Figures 3, 4 and 5.

#### III-A. Frame difference decomposition

As stated in Section II, the FD  $\mathbf{e}$  is decomposed into  $M$  FD sub-frames, with each FD sub-frame  $\mathbf{e}^m$  represented by a separate transmission sub-frame  $\mathbf{u}^m$ . This is explained with the aid of an example, which is provided in Figure 3.

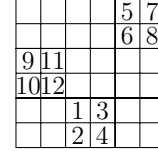


Fig. 3. Example selection of  $J = 12$  ( $8 \times 8$ )-pixel video blocks from a ( $48 \times 48$ )-pixel FD to provide one of  $M = 3$  FD sub-frames.

The VQ codec operates on a block-based philosophy, with the *video block* being defined as the smallest unit of video information considered in isolation and having dimensions of ( $8 \times 8$ ) pixels. As shown in Figure 3, each FD sub-frame  $\mathbf{e}^m$  comprises a unique combination of  $J$  ( $8 \times 8$ )-pixel video blocks in disjointed groups from the FD  $\mathbf{e}$ , where  $\mathbf{e}$  comprises ( $M \cdot J$ ) video blocks in total. A video-block index  $j \in [1 \dots J]$  is allocated to each of the  $J$  ( $8 \times 8$ )-pixel video blocks, denoted  $e_j^m$ , in the FD sub-frame  $\mathbf{e}^m$ , with consecutive indices being allocated within video-block groups.

The specifics of the FD decomposition are not detailed owing to space limitations. Suffice to say that each FD sub-frame has similar statistical properties and identical disjointed-group shapes and block-index allocations.

#### III-B. VQ codebook

Each FD sub-frame  $\mathbf{e}^m$  is represented by the reconstructed FD sub-frame  $\hat{\mathbf{e}}^m$  and the transmission sub-frame  $\mathbf{u}^m$  on the basis of VQ. Since, as stated in Section III-A, all FD sub-frames have similar statistical properties, identical designs and a single  $K$ -entry VQ codebook can be employed for the  $M$  separate VQ encoders and decoders. This  $K$ -entry VQ codebook comprises the VQ tiles and their VLC-based representations known to both the transmitter and receiver. An example of a  $K = 5$ -entry VQ codebook is provided in Figure 4. In the VQ tiles of this figure, dark pixels indicate negative frame-difference values, grey represents zero values and light pixels represent positive frame-difference values.

Index $k$	Video blocks $J_x^k \times J_y^k = J^k$	Bits $I^k$	VLC code $\mathbf{VLC}^k$
1	$2 \times 2 = 4$	1	0
2	$1 \times 2 = 2$	2	11
3	$1 \times 1 = 1$	3	101
4	$1 \times 1 = 1$	4	1001
5	$1 \times 1 = 1$	5	10001

VQ tiles:

Fig. 4. Example of a  $K = 5$ -entry VQ codebook.

The  $K$  VQ tiles of the VQ codebook have a variety of dimensions, all of which are multiples of the ( $8 \times 8$ )-pixel video block dimensions, as shown in Figure 4. This allows the adequate representation of both small areas of high video activity and large areas of low video activity. Each VQ codebook entry, having an index  $k \in [1 \dots K]$ , corresponds to a VQ tile  $\mathbf{VQ}^k$  having the dimensions of ( $J_x^k \times J_y^k$ ) ( $8 \times 8$ )-pixel video blocks. Additionally, the VQ codebook entry with index  $k$  has a VLC code  $\mathbf{VLC}^k$  of length  $I^k$  bits,

as shown in Figure 4.

Each reconstructed FD sub-frame  $\hat{\mathbf{e}}^m$  comprises  $J$  ( $8 \times 8$ )-pixel video blocks from the FD  $\mathbf{e}$ . These  $J$  video blocks are provided by an appropriately-selected tessellation of VQ tiles from the employed VQ codebook. Each employment of the VQ tile  $\mathbf{VQ}^k$  of the VQ codebook entry with index  $k$  provides the video blocks for a  $(J_x^k \times J_y^k)$ -block region of the reconstructed FD sub-frame  $\hat{\mathbf{e}}^m$ . A constraint is imposed to allow the formation of the novel trellis structure described in the next sub-section. Specifically, the  $J^k = J_x^k \cdot J_y^k$  video blocks of  $\hat{\mathbf{e}}^m$  provided by the employment of  $\mathbf{VQ}^k$  must have consecutive indices. This constraint therefore depends upon the employed decomposition of the FD into sub-frames, as described in Section III-A.

The transmission sub-frame  $\mathbf{u}^m$  comprises the concatenation of the VLC codes associated with the VQ tiles providing the  $J$  video blocks of the reconstructed FD sub-frame  $\hat{\mathbf{e}}^m$ . However, in the proposed video codec, each transmission sub-frame is constrained to having a length of  $I$  bits. Since the VQ codebook entry having index  $k$  represents  $J^k$  video blocks with  $I^k$  bits, the task of finding an appropriate  $J$ -block reconstructed FD sub-frame  $\hat{\mathbf{e}}^m$  and its  $I$ -bit transmission sub-frame  $\mathbf{u}^m$  is non-trivial. This motivates the use of a trellis structure.

### III-C. VQ trellis structure

A trellis structure, comprising a set of transitions between trellis states, is employed by the proposed VQ encoder and decoder to represent the complete set of VQ-induced code constraints. A novel block-based modification of the symbol-level VLC trellis structure described in [5] is employed. Here, the VLC symbol axis of [5] is replaced by a video block axis. In contrast to [5], transitions are permitted encompass a number of consecutive positions along this axis.

The proposed VQ trellis structure allows the consideration of every valid combination of VQ codebook entries that provide  $J$  video blocks for the reconstructed FD sub-frame  $\hat{\mathbf{e}}^m$  and  $I$  bits for the transmission sub-frame  $\mathbf{u}^m$ . The specific structure of the proposed VQ trellis therefore depends on the VQ codebook employed. This structure also depends on the employed decomposition into sub-frames owing to the constraint described in Section III-A. The VQ trellis structure example of Figure 5 contains all valid transitions, and highlights the associated trellis states, corresponding to the  $K = 5$ -entry VQ codebook of Figure 4 and the FD sub-frame of Figure 3 for  $J = 12$  and  $I = 17$ .

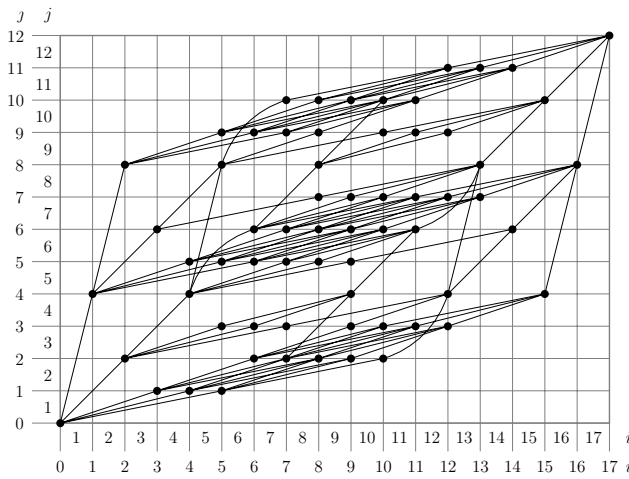


Fig. 5. Example of a VQ trellis structure for the VQ codebook provided in Figure 4 and the example FD sub-frame provided in Figure 3, where we have  $J = 12$  and  $I = 17$ .

In Figure 5, the video-block and bit indices,  $j \in [1 \dots J]$  and  $i \in [1 \dots I]$  respectively, occur between the corresponding trellis state indices,

namely  $j \in [0 \dots J]$  for the video blocks and  $i \in [0 \dots I]$  for the bits. Each trellis state  $S_{j,i}$  represents the encoding progress made immediately after the consideration of the first  $j$  video blocks in the  $J$ -block reconstructed FD sub-frame  $\hat{\mathbf{e}}^m$  and the first  $i$  bits in the  $I$ -bit transmission sub-frame  $\mathbf{u}^m$ , as illustrated in Figure 5. Each transition  $T$  represents the employment of the VQ codebook entry with index  $k^T$  at a specific point during the FD sub-frame encoding process. This point is identified by the state indices  $j^T$  and  $i^T$ , with the transition  $T$  embarking from the trellis state  $S_{j^T, i^T}$ . As stated in Section III-B, the employment of the VQ codebook entry with index  $k^T$  provides  $J^{k^T}$  ( $8 \times 8$ )-pixel video blocks for the reconstructed FD sub-frame and  $I^{k^T}$  bits for the transmission sub-frame. This results in arriving at the trellis state  $S_{j^T+J^{k^T}, i^T+I^{k^T}}$  immediately after the employment of the VQ codebook entry with index  $k^T$ .

The proposed VQ trellis structure contains all valid transition paths between the trellis states  $S_{0,0}$  and  $S_{J,I}$ , as shown in Figure 5, and therefore represents the complete set of VQ-induced code constraints.

## IV. VQ ENCODER

In the proposed VQ-TCM system's transmitter shown in Figure 1 and described in Section II, the VQ encoding process is separately applied to all of the  $M$  FD sub-frames. Each VQ encoder operates on the basis of the proposed VQ trellis structure outlined in Section III-C, which is reminiscent of TCQ [6]. Here, we seek the optimal tessellation of the VQ codebook's tiles, which have different dimensions, as described in Section III-B and are exemplified in Figure 4. The aim of this template-matching optimisation is to provide the Minimum Mean Squared Error (MMSE) reconstruction  $\hat{\mathbf{e}}^m$  of the FD sub-frame  $\mathbf{e}^m$ .

Since the proposed VQ trellis structure represents the complete set of VQ-induced code constraints, its employment represents the consideration of every possible valid FD sub-frame encoding, allowing the most appropriate one to be selected. This is achieved using the philosophy of Viterbi decoding [10], with a survivor path being selected at each trellis state  $S_{j,i}$ . This selection yields the  $i$ -bit encoding of the first  $j$  video blocks in  $\mathbf{e}^m$  that introduces the lowest possible distortion  $D(S_{j,i})$ .

As stated in Section III-C, a transition  $T$  in the proposed VQ trellis structure, which is associated with the VQ codebook index  $k^T$ , corresponds to employing the VQ tile  $\mathbf{VQ}^{k^T}$  to represent a total of  $J^{k^T}$  ( $8 \times 8$ )-pixel video blocks of the FD sub-frame  $\mathbf{e}^m$ . The distortion associated with the transition  $T$ , namely  $d(T)$ , is the sum of the squared difference between  $\mathbf{VQ}^{k^T}$  and the corresponding  $J^{k^T}$  video blocks of  $\mathbf{e}^m$ .

The survivor path at the trellis state  $S_{j,i}$  is deemed to be that associated with the specific merging transition  $T$  having the minimum cumulative distortion  $D(T) = d(T) + D(S_{j^T, i^T})$ , where we have  $D(S_{0,0}) = 0$ . Having determined the survivor path at the trellis state  $S_{J,I}$ , the optimum MMSE VQ encoding of the FD sub-frame  $\mathbf{e}^m$  has been found. The reconstructed FD sub-frame  $\hat{\mathbf{e}}^m$  is formed as the tessellation of the VQ tiles associated with the survivor path transitions and the transmission sub-frame  $\mathbf{u}^m$  is formed as the concatenation of the associated VLC codes, as exemplified in Figure 4.

## V. VQ BCJR DECODER

In the proposed VQ-TCM system's receiver seen in Figure 2, BCJR decoding is employed for all of the  $M$  FD sub-frames, as described in Section II. More specifically, each VQ BCJR decoder considers the *a priori* LLRs  $\mathbf{L}_a^2(\mathbf{u}^m)$ , which pertain to the bits of the transmission sub-frame  $\mathbf{u}^m$ , to generate the *a posteriori* LLRs  $\mathbf{L}_p^2(\mathbf{u}^m)$  and the reconstructed FD sub-frame  $\hat{\mathbf{e}}^m$ . This process employs the proposed VQ trellis to exploit the residual redundancy provided by the VQ-induced code constraints. Before describing the VQ BCJR decoder's operation, we provide some additional background on the subject.

In addition to proposing a symbol-level VLC trellis structure, [5] also pro-

vides a specific BCJR algorithm that gives soft symbol-level output. With the introduction of a normalisation factor, [11] extends this BCJR algorithm to additionally provide soft bit-level output, for the sake of facilitating iterative extrinsic information exchange. This is achieved following the calculation of the *a posteriori* transition probabilities by considering a cross-section of the trellis that is perpendicular to the bit index axis and positioned at the bit of interest. Next, the bisected transitions are divided into two sub-sets according to the bit value they represent at their bisection. Finally, the *a posteriori* probability of the bit of interest having a particular value is arranged to be proportional to the sum of the appropriate sub-set's *a posteriori* transition probabilities.

On the basis of the proposed VQ trellis structure of Section III-C, the BCJR technique of [11] is employed by each VQ BCJR decoder to provide *a posteriori* bit probabilities, which are converted to the *a posteriori* LLRs  $\mathbf{L}_p^2(\mathbf{u}^m)$  [9]. This method was additionally adapted to provide *a posteriori* video-block probabilities. During VQ encoding, each application of the VQ tile  $\mathbf{VQ}^k$  is employed to represent a set of  $J^k$  number of  $(8 \times 8)$ -pixel video blocks of the FD sub-frame  $\mathbf{e}^m$ . We seek the *a posteriori* probability that  $\mathbf{VQ}^k$  use employed by the VQ encoder for the representation of the video block  $e_j^m$  of  $\mathbf{e}^m$ . This probability is termed  $P(\hat{e}_j^m \in \mathbf{VQ}^k | \mathbf{y})$  and, in analogy to the *a posteriori* bit probabilities, is obtained by making a trellis cross-section perpendicular to the block axis and positioned at the video-block index  $j$ .

The *A Posteriori* Probability (APP)-based MMSE estimation of the representation of the video block  $e_j^m$  is obtained as

$$\tilde{e}_j^m = \sum_{k=1}^K P(\hat{e}_j^m \in \mathbf{VQ}^k | \mathbf{y}) \cdot \mathbf{VQ}^k,$$

where  $\mathbf{VQ}^k$ , with  $j' \in [1 \dots J^k]$ , is the particular  $(8 \times 8)$ -pixel video block of  $\mathbf{VQ}^k$  that may have been employed to represent  $e_j^m$ . Finally, the set of video blocks  $\tilde{e}_j^m$  are tessellated to provide  $\tilde{\mathbf{e}}^m$ . Since the proposed VQ trellis structure represents the complete set of VQ-induced code constraints, its employment during VQ decoding guarantees the recovery of valid, although not necessarily correct, sets of VQ tiles.

## VI. RESULTS

In this section, the achievable performance of the proposed VQ-TCM system is assessed using the 10 frame/s grey-scale head-and-shoulders  $(176 \times 144)$ -pixel Quarter Common Intermediate Format (QCIF) 'Lab' video sequence [2]. This video sequence exhibits a moderate level of video activity and was not included in the VQ training set, which comprised five 250-frame video sequences. Two VQ-TCM schemes were implemented using two different interleaver lengths. The first scheme imposed a low latency equal to the duration of a single video frame (0.1 s at 10 frame/s), which is suitable for real-time interactive video-telephony applications. The second scheme had a high latency of 50 video frames (5 s at 10 frame/s), which is suitable for non-realtime video streaming and wireless internet download applications. In the low-latency scheme, the length of each transmission frame  $\mathbf{u}$  and that of the interleaver  $\pi$  equals  $M \cdot J = 1485$  bits. In the high-latency scheme, 50 transmission frames  $\mathbf{u}$  are concatenated before interleaving, giving an interleaver length of  $50 \cdot M \cdot J = 74250$  bits. Both schemes have an encoded-video bitrate of 14.85 kbps. Having described the differences between the two schemes, we continue by discussing their common features. This section will be concluded by considering the performance of the different-latency schemes and comparing them to that of the benchmark scheme of [4].

In both schemes, each FD was divided into  $M = 33$  FD sub-frames. Each FD sub-frame comprised  $J = 12$   $(8 \times 8)$ -pixel video blocks and each transmission sub-frame consisted of  $I = 45$  bits. A  $K = 512$ -entry VQ codebook was employed, comprising the five VQ tiles shown in Figure 4 and an additional 507 single-video-block VQ tiles, which were generated using the Linde-Buzo-Gray (LBG) algorithm [12]. The probability of occurrence of all VQ tiles was evaluated and the VQ codebook entries were assigned

Reversible VLC (RVLC) codes [13] having a minimum free-distance of 2. The coding rate of the RVLC-represented VQ codec is  $R_{VQ} = 0.666$ , which is the ratio of its entropy to its actual average RVLC length. A terminated  $R_{TCM} = 3/4$ -rate TCM scheme employing a code memory of 6,  $M_{TCM} = 16$ QAM modulation and IQ-interleaving [14] was employed for transmission over an uncorrelated narrowband Rayleigh fading channel. The bandwidth efficiency of the proposed VQ-TCM system is given by  $\eta = R_{VQ} \times R_{TCM} \times \log_2(M_{TCM}) = 2.00$  bit/s/Hz, when ideal Nyquist filtering is assumed and the presence of the code termination symbols added by the TCM encoder is ignored. Note that at  $\eta = 2$  bit/s/Hz, the uncorrelated Rayleigh-fading channel-capacity limit for 16QAM is  $E_b/N_0 = 3.96$  dB [15], where  $E_b/N_0 = SNR/\eta$  is the Signal to Noise Ratio (SNR) per bit.

The EXIT characteristics of the TCM decoder operating at various  $E_b/N_0$  values are provided in Figure 6, along with the VQ decoder's inverted EXIT characteristic. Employing RVLC codes with a minimum free-distance of 2 in the VQ codebook was found to yield a VQ decoder EXIT characteristic that achieves unity extrinsic mutual information  $I_E^2$  for unity *a priori* mutual information  $I_A^2$ . Iterative-decoding convergence at unity mutual information, which is associated with the elimination of all perceivable channel-induced video distortion, is therefore supported. A 'tunnel' between the TCM and inverted VQ EXIT characteristics, giving iterative-decoding convergence at unity mutual information, can be seen for  $E_b/N_0 > 5.25$  dB in Figure 6, which is just 1.29 dB away from the proposed VQ-TCM system's 2 bit/s/Hz channel-capacity limit of 3.96 dB. The EXIT trajectories are provided for both the low- and high-latency schemes at  $E_b/N_0 = 6$  dB, with the latter closely matching the VQ and TCM EXIT characteristics and achieving the desired unity mutual information. By contrast, the low-latency scheme deviates from these EXIT characteristics and fails to converge to the desired unity mutual information limit at  $E_b/N_0 = 6$  dB. The improved performance of the high-latency scheme is explained by its employment of a longer interleaver length [9].

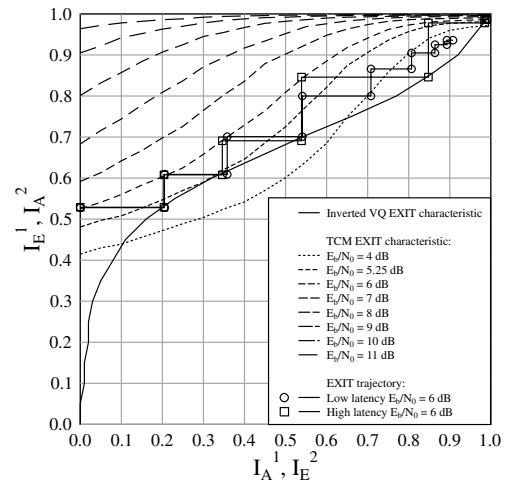


Fig. 6. The proposed VQ-TCM system's EXIT chart [8] for the 'Lab' video sequence and a 3/4-rate TCM scheme.

The video reconstruction quality of both of the different-delay VQ-TCM schemes may be assessed in Figure 7 for various number of decoding iterations and for a range of channel  $E_b/N_0$  values. The Peak Signal to Noise Ratio (PSNR) [2] is the objective video reconstruction quality metric employed. Substantial PSNR iteration gains can be observed for both schemes, with cumulative  $E_b/N_0$  gains of 4.34 dB for the low-latency and 5.61 dB for the high-latency schemes being achieved after eight decoding iterations at a PSNR of 29.5 dB. This value of PSNR is associated with an

aesthetically pleasing reconstruction which contains only a negligible level of channel-induced distortion. The high-latency scheme can be seen to outperform the low-latency scheme, regardless of the number of decoding iterations, as predicted by our EXIT chart analysis. In Figure 7, a PSNR value higher than 29.5 dB can be seen to be achieved by both the low- and high-latency schemes after eight decoding iterations for  $E_b/N_0 > 7.00$  dB and  $E_b/N_0 = 5.75$  dB, respectively.

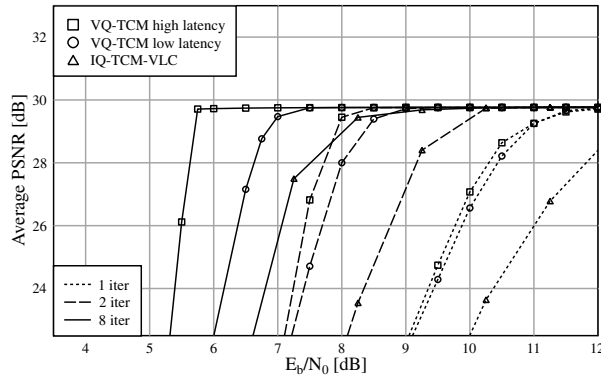


Fig. 7. PSNR performance of the proposed VQ-TCM system and the IQ-TCM-VLC benchmark for the ‘Lab’ video sequence when communicating over an uncorrelated narrowband Rayleigh fading channel.

The IQ-TCM-VLC system of [4] has a single-video-frame latency and a similar coding rate, bitrate and error-free reconstruction quality to those of the proposed VQ-TCM system. The PSNR performance of the IQ-TCM-VLC system of [4], which is also shown in Figure 7, was found to be improved by the low-latency VQ-TCM scheme, regardless of the number of decoding iterations. At a PSNR of 29.5 dB, this improvement was approximately 1.5 dB, again, regardless of the number of decoding iterations. The observed poorer performance of the IQ-TCM-VLC system [4] may be explained by its employment of an intermediate decoder. Since this intermediate decoder is unable to assess the validity of the recovered video-encoded bits, channel-induced-error masking may be required, when error-free recovery is not achieved. This may inflict significant channel-induced distortion upon the reconstructed video sequence.

## VII. CONCLUSIONS

In this contribution, the joint video and channel coding based VQ-TCM transmission system has been proposed. The outer iterative-decoding stage of this serially-concatenated video transmission system was constituted by the proposed trellis-based video decoder, rather than an intermediate decoder as was the case in [4]. By adopting an algorithmically-simple design it was possible to unambiguously describe the video-codec-induced code constraints by a novel trellis structure, which was employed to perform MMSE VQ encoding. A modified BCJR algorithm utilising the novel trellis structure was employed to obtain soft APP-based block-level outputs. These were employed to calculate a MMSE estimation of the video-encoded information. Since the novel trellis structure represented the complete set of VQ-induced code constraints, its employment guaranteed the recovery of valid sets of VQ tiles, eliminating the need for error masking during video reconstruction, which was required in [4]. In the proposed VQ-TCM system’s receiver, iterative-decoding convergence that eliminates all perceivable channel-induced video distortion was shown to be supported within 1.29 dB of the system’s 2 bit/s/Hz channel-capacity limit of 3.96 dB. Finally, the proposed VQ-TCM system was shown to outperform the benchmark system of [4], which employed an intermediate decoder, rather than the video decoder, to exchange extrinsic information with the inner iterative-decoding stage.

## VIII. REFERENCES

- [1] C. E. Shannon, “A Mathematical Theory of Communication,” *The Bell System Technical Journal*, vol. 27, pp. 379–656, July 1948.
- [2] L. Hanzo, P. J. Cherriman and J. Street, *Wireless Video Communications: Second to Third Generation Systems and Beyond*. New York, USA: IEEE Press, 2001.
- [3] S. Benedetto, D. Divsalar, G. Montorsi and F. Pollara, “Serial Concatenation of Interleaved Codes: Performance Analysis, Design and Iterative Decoding,” *IEEE Transactions on Information Theory*, vol. 44, pp. 909–926, May 1998.
- [4] S. X. Ng, R. G. Maunder, J. Wang, L-L. Yang and L. Hanzo, “Joint Iterative-Detection of Reversible Variable-Length Coded Constant Bit Rate Vector-Quantized Video and Coded Modulation,” in *European Signal Processing Conference (EUSIPCO)*, (Vienna, Austria), pp. 2231–2234, September 2004.
- [5] R. Bauer and J. Hagenauer, “Symbol-by-Symbol MAP Decoding of Variable Length Codes,” in *ITG Conference on Source and Channel Coding*, (Munich, Germany), pp. 111–116, January 2000.
- [6] M. W. Marcellin and T. R. Fischer, “Trellis Coded Quantization of Memoryless and Gauss-Markov Sources,” *IEEE Transactions on Communications*, vol. 38, pp. 82–93, January 1990.
- [7] L. R. Bahl, J. Cocke, F. Jelinek and J. Raviv, “Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate,” *IEEE Transactions on Information Theory*, vol. 20, pp. 284–287, March 1974.
- [8] S. ten Brink, “Convergence Behaviour of Iteratively Decoded Parallel Concatenated Codes,” *IEEE Transactions on Communications*, vol. 49, pp. 1727–1737, October 2001.
- [9] J. Hagenauer, E. Offer and L. Papke, “Iterative Decoding of Binary Block and Convolutional Codes,” *IEEE Transactions on Information Theory*, vol. 42, pp. 429–445, March 1996.
- [10] L. Hanzo, T. H. Liew and B. L. Yeap, *Turbo Coding, Turbo Equalisation and Space Time Coding for Transmission over Wireless Channels*. Chichester, UK: Wiley, 2002.
- [11] J. Kliewer and R. Thobaben, “Iterative Joint Source-Channel Decoding of Variable-Length Codes Using Residual Source Redundancy,” *IEEE Transactions on Wireless Communications*, vol. 4, May 2005.
- [12] Y. Linde, A. Buzo and R. Gray, “An Algorithm for Vector Quantizer Design,” *IEEE Transactions on Communications*, vol. 28, pp. 84–95, January 1980.
- [13] Y. Takishima, M. Wada and H. Murakami, “Reversible Variable Length Codes,” *IEEE Transactions on Communications*, vol. 43, pp. 158–162, Feb/Mar/Apr 1995.
- [14] S. X. Ng and L. Hanzo, “Space-Time IQ-Interleaved TCM and TTCM for AWGN and Rayleigh Fading Channels,” *IEE Electronics Letters*, vol. 38, pp. 1553–1555, November 2002.
- [15] L. Hanzo, S. X. Ng, T. Keller and W. Webb, *Quadrature Amplitude Modulation*. Chichester, UK: Wiley, 2004.