Joint Iterative Decoding of Trellis-Based VQ and TCM

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

Abstract— A joint video and channel coded system employing an iteratively decoded serial concatenation of a Vector Quantization (VQ) based video codec and a Trellis-Coded Modulation (TCM) scheme is proposed. The video codec imposes VQ-induced code constraints, which may be completely described by a trellis structure, which is employed as the basis for optimal minimum mean-squared-error VQ-encoding and -decoding. In the latter case, the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm is employed to facilitate the iterative exchange of soft information between the VQ and TCM decoder. An error-free video reconstruction quality is supported using 16-level Quadrature Amplitude Modulation (16QAM) based TCM for transmission over Rayleigh-fading channels at a Signal-to-Noise Ratio (SNR) per bit of 5.25 dB. This value is within 1.29 dB of the Rayleigh channel's capacity at our system's effective bandwidth-efficiency of 2 bits/s/Hz. Owing to its ability to exploit the VQ-induced code constraints during iterative decoding, the joint video and channel coding approach is found to consistently outperform the Shannonian source and channel separation philosophy. This is achieved at the cost of a 1.6 times higher computational complexity. Finally, the convergence of the iterative decoder is investigated with the aid of a novel so-called EXtrinsic Information Transfer (EXIT) chart.

Index Terms— Video coding, vector quantization, joint source and channel coding, trellis codes, BCJR.

I. INTRODUCTION

SHANNON'S source and channel separation theorem [1] states that under certain idealized conditions, source and channel coding can be performed in isolation without any loss in performance. This motivated the designs of the Vector Quantization (VQ) based [2] video transmission system in [3] and the MPEG4-based [4] system of [5]. However, Shannon's findings only apply under a number of idealistic assumptions. These assumptions have limited validity for practical video transmission systems communicating over realistic mobile-radio channels [2]. This motivates the application of joint source and channel coding techniques in wireless video transmission, as exemplified by [6].

Let us now summarise the novelty of this paper in contrast to the state-of-the-art. In [6], joint source and channel coding techniques were applied at the bit-level to mitigate channel-induced distortion within an MPEG-4 [4] coded video sequence. More specifically, knowledge of the residual correlation of bits within the MPEG-4 bitstream was exploited to assist their recovery upon reception. In our approach, joint

source and channel coding techniques are applied at the Video Block (VB)-level within a specially designed novel proprietary video codec, which offers the advantage of directly assisting the recovery of the transmitted video information, rather than of its intermediate bit-based representation.

The design of the proposed video transmission system extends the serial concatenation and iterative decoding proposals of [7], in a manner similar to that of [8]. More explicitly, a novel trellis-decoded VQ-based video codec is serially concatenated with the separate In-phase, Quadrature-phase (IQ) interleaved TCM codec of [9]. In the receiver, VQ- and TCM-decoding are performed iteratively.

The video codec of the proposed VQ-TCM system is designed to maintain a simple VQ-based transmission frame structure. The rules governing the formation of the legitimate bit-sequences of a transmission frame are referred to as the *VQ-induced code constraints* in this contribution. The video codec was specifically designed to impose VQ-induced code constraints that may be completely described by a novel VQ trellis structure, reminiscent of the symbol-level trellis structure of [10]. This is in contrast to the video transmission systems of [3] and [5], in which the code constraints imposed by the video codec did not lend themselves to a trellis-based description.

As we will outline in Section III-D, the employment of the proposed VQ trellis structure represents the consideration of all legitimate transmission frame permutations. More explicitly, we perform optimal Minimum Mean-Squared-Error (MMSE) VQ encoding of the video sequence in a novel manner that is reminiscent of Trellis-Coded Quantisation (TCQ) [11]. Additionally, the employment of the proposed VQ trellis structure during VQ decoding guarantees the recovery of legitimate – although not necessarily error-free – video information. Hence, unlike in the video transmission systems of [3] and [5], the proposed video decoder is never forced to discard information. During VQ decoding, a novel modification of the Bahl-Cocke-Jelinek-Raviv (BCJR) [12] algorithm provided in [13] is performed on the basis of the proposed VQ trellis structure. This allows the exploitation of the VQ-induced code constraints to assist in joint iterative VQ- and TCM-decoding. Additionally, this allows the unconventional soft *A Posteriori* Probability (APP) based [14] MMSE reconstruction of the transmitted video frames.

This paper is organised as follows. In Section II, the proposed VQ-TCM system is introduced. Section III describes the VQ-TCM system's transmission frame structure and the VQ-induced code constraints. Additionally, the complete description of the VQ-induced code constraints by the proposed VQ trellis structure is discussed in Section III. The employment of the VQ trellis structure in VQ-encoding and -decoding is

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

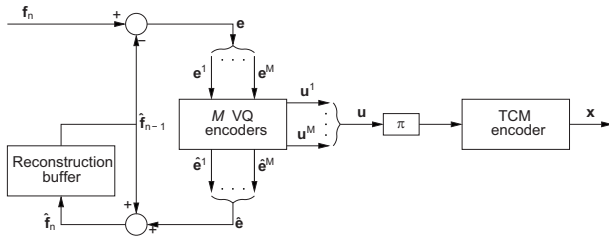

Fig. 2. The proposed VQ-TCM system's receiver.

described in Sections IV and V, respectively. In Section VI, both low- and high-latency implementations of the proposed VQ-TCM system are introduced and their video reconstruction quality is assessed. Finally, the performance of the proposed joint video and channel coding approach is compared to that of the video transmission systems of [3] and [5]. Our conclusions are offered in Section VII.

II. SYSTEM OVERVIEW

In this section we introduce the proposed VQ-TCM video transmission system, which employs a joint video and channel coding philosophy, as introduced in Section I. The video codec of this system achieves compression by employing low-complexity frame differencing and VQ of the resultant Frame Difference (FD) [2]. These are chosen in favour of the more advanced but higher complexity compression techniques, such as sub-pixel-accuracy MPEG4-style Motion Compensation (MC) or the Discrete Cosine-Transform (DCT) [2]. This ensures that a simple transmission frame structure is maintained, as we will detail in Section III. Rather than performing an excessive-complexity single VQ-encoding or -decoding operation for the entire video frame as usual, a decomposition into a number of smaller VQ operations is performed. This approach is associated with a significant computational complexity reduction, as will be shown in Section III-D. The proposed video codec is protected by the serially concatenated TCM codec of [9] and joint iterative decoding is employed in the receiver. These issues are further detailed in the following sub-sections with reference to the proposed VQ-TCM system's transmitter and receiver schematics seen in Figures 1 and 2, respectively.

A. Compression

In the VQ-TCM system's transmitter of Figure 1, the previous reconstructed video frame $\hat{\mathbf{f}}_{\mathbf{n-1}}$ is employed as a prediction for the current video frame $\mathbf{f_n}$. The FD $\mathbf{e} = \mathbf{f_n} - \hat{\mathbf{f}}_{\mathbf{n-1}}$ has a lower variance than the current video frame $\mathbf{f_n}$. Hence, the transmission of the FD $\mathbf{e}$ to convey the current video frame $\mathbf{f_n}$ provides compression.

As will be discussed in Section III, VQ is employed to represent the FD $\mathbf{e}$ in a compact bit-sequence form, namely the transmission frame $\mathbf{u}$. Hence, the VQ offers compression. This compression is termed 'lossy' because the VQ approximates the FD $\mathbf{e}$ by the reconstructed FD $\hat{\mathbf{e}}$, as shown in Figure 1. The reconstructed FD $\hat{\mathbf{e}}$ is employed to obtain the current reconstructed video frame $\hat{\mathbf{f}}_{\mathbf{n}} = \hat{\mathbf{e}} + \hat{\mathbf{f}}_{\mathbf{n-1}}$, which is stored for the sake of providing the prediction of the next video frame.

B. Vector quantization decomposition

In the proposed VQ-TCM system's transmitter and receiver, VQ-encoding and -decoding are performed, respectively. As will be detailed in Sections III-A and III-D, we decompose each FD $\mathbf{e}$ into M number of smaller sub-frames, which may be represented by VQs operating in parallel. This decomposition implies the corresponding decomposition of both the reconstructed FD $\hat{\mathbf{e}}$ and the transmission frame $\mathbf{u}$, as indicated by the braces { and } in Figure 1. In the transmitter, M number of separate VQ operations are employed. Each VQ encoder approximates the FD sub-frame $\mathbf{e^m}$ by the reconstructed FD sub-frame $\hat{\mathbf{e}}^m$, where $m \in [1 \ldots M]$ denotes the sub-frame index. Additionally, the transmission sub-frame $\mathbf{u^m}$ is generated, as will be detailed in Section IV. The bit-based transmission frame $\mathbf{u}$ is constituted by the concatenation of the set of these M number of transmission sub-frames.

C. Serial concatenation and iterative decoding

In the proposed VQ-TCM system, the video codec is serially concatenated with the TCM codec of [9]. In the transmitter, the current video frame $\mathbf{f_n}$ is conveyed by the reconstructed FD $\hat{\mathbf{e}}$, where the latter is generated using VQ encoding and is represented by the transmission frame $\mathbf{u}$, as discussed in Section II-A. The transmission frame $\mathbf{u}$ is interleaved in the block π of Figure 1 and TCM encoded to generate the channel's input symbols $\mathbf{x}$. These are transmitted over the channel and are received as the channel's output symbols $\mathbf{y}$, as shown in Figure 2. Finally, TCM- and VQ-decoding are employed to recover the reconstructed FD estimate $\tilde{\mathbf{e}}$. This allows the recovery of the current video frame estimate $\tilde{\mathbf{f}}_{\mathbf{n}} = \tilde{\mathbf{e}} + \tilde{\mathbf{f}}_{\mathbf{n-1}}$, as shown in Figure 2.

In the receiver, Soft-In Soft-Out (SISO) TCM- and VQ-decoding are performed iteratively, as shown in Figure 2. Soft information, represented in the form of Logarithmic Likelihood-Ratios (LLRs) [14], is exchanged between the TCM and VQ iterative decoding stages for the sake of assisting each other's operation. With each successive decoding iteration, the reliability of this soft information increases, until iterative decoding convergence is achieved. In Figure 2, $L(\cdot)$ denotes the LLRs of the specified bits, where the superscript (1) denotes TCM decoding and (2) represents VQ decoding. Additionally, a subscript denotes the role of the LLRs, with a, p and e indicating *a priori*, *a posteriori* and extrinsic information, respectively.

During each decoding iteration, *a priori* LLRs $L_a(\mathbf{u})$ are provided for each iterative decoding stage, as shown in Figure 2. Naturally, in the case of TCM decoding, the channel's output symbols $\mathbf{y}$ are also exploited. Note that the *a priori* LLRs $L_a(\mathbf{u})$ are obtained from the most recent operation of
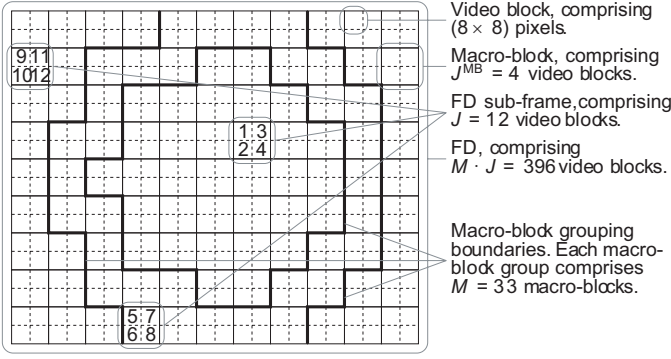
Fig. 3. Example of selecting $J = 12$ (8×8)-pixel VBs from a (176×144)-pixel FD to provide one of the $M = 33$ FD sub-frames.
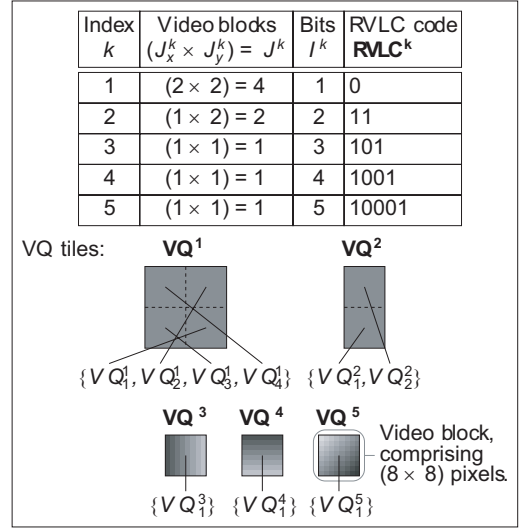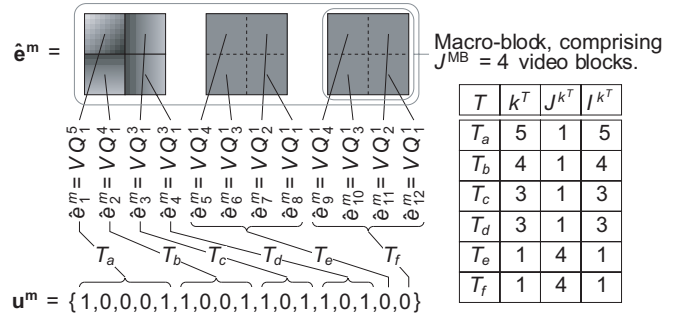

Fig. 4. Example of a $K = 5$-entry VQ codebook.

Fig. 5. Example of a reconstructed FD sub-frame, comprising $J = 12$ VBs selected as exemplified in the FD decomposition of Figure 3, and the corresponding $I = 17$-bit transmission sub-frame. These comprise the entries k^T, where $T \in [T_a \dots T_f]$, from the VQ codebook example of Figure 4.

the other decoding stage, as will be highlighted below. In the case of the first decoding iteration no previous VQ decoding has been performed. In this case, the *a priori* LLRs $L_a^1(\mathbf{u})$ provided for TCM decoding are all zero, corresponding to a probability of 0.5 for both '0' and '1'. Each iterative decoding stage applies the BCJR [12] algorithm, as described in [9] and Section V for TCM- and VQ-decoding, respectively. The result is the generation of the *a posteriori* LLRs $L_p(\mathbf{u})$, as shown in Figure 2.

During iterative decoding, it is necessary to prevent the re-use of already-exploited information, since this would limit the attainable iteration gain [15]. This is achieved following each decoding stage by the subtraction of $L_a(\mathbf{u})$ from $L_p(\mathbf{u})$, as shown in Figure 2. Following VQ decoding, we arrive at the extrinsic LLRs $L_e^2(\mathbf{u})$. In the case of TCM decoding, the LLRs $L_e^1(\mathbf{u})$ additionally contain information extracted from the channel's output symbols $\mathbf{y}$. It is these sets of LLRs that provide the *a priori* LLRs for the next iteration of the other decoding stage. De-interleaving, indicated by the block π^{-1} in Figure 2, is applied to $L_e^1(\mathbf{u})$ in order to generate $L_a^2(\mathbf{u})$. Similarly, interleaving is applied to $L_e^2(\mathbf{u})$ for the sake of providing $L_a^1(\mathbf{u})$, as shown in Figure 2. These interleaving and de-interleaving operations are necessary for the sake of mitigating the correlation of consecutive LLRs, before forwarding them to the next iterative decoding stage [14]. As always, the interleaver's ability to provide this desirable statistical independence is related to its length.

As stated in Section II-B, M number of separate VQ decoding processes are employed in the proposed VQ-TCM system's receiver. Similarly to the decomposition of the bit-based transmission frame $\mathbf{u}$, the *a priori* LLRs $L_a^2(\mathbf{u})$ are decomposed into M number of sub-frames, as shown in Figure 2. This decomposition is accompanied by the corresponding decomposition of the *a posteriori* LLRs $L_p^2(\mathbf{u})$ and the reconstructed FD estimate $\tilde{\mathbf{e}}$. Each VQ decoder is provided with the *a priori* LLR sub-frame $L_a^2(\mathbf{u^m})$ and generates the *a posteriori* LLR sub-frame $L_p^2(\mathbf{u^m})$, where $m \in [1 \dots M]$. Additionally, each of the M number of VQ decoding process recovers the reconstructed FD sub-frame estimate $\tilde{\mathbf{e}}^\mathbf{m}$, as will be detailed in Section V.

III. TRANSMISSION FRAME STRUCTURE

As stated in Section II, each video frame $\mathbf{f_n}$ is conveyed between the proposed VQ-TCM system's transmitter and receiver by means of a single bit-based transmission frame

$\mathbf{u}$. This comprises the concatenation of M number of VQ-based transmission sub-frames. Again, the formation of legitimate bit-sequences within these transmission sub-frames is governed by simple VQ-induced code constraints, which are imposed by the decomposition of FDs into sub-frames and by the specific nature of the VQ codebook to be outlined in Section III-B. These VQ-induced code constraints may be described by a novel VQ trellis structure. These issues are discussed in the following sub-sections and are described with the aid of an example that continues throughout Figures 3, 4, 5 and 6. It should be noted that the generalization of this example is straightforward.

A. Frame difference decomposition

As stated in Section II-B, the FD $\mathbf{e}$, the reconstructed FD $\hat{\mathbf{e}}$ and the reconstructed FD estimate $\tilde{\mathbf{e}}$ are decomposed into M number of sub-frames, as detailed in the example of Figure 3.

For the sake of implementational simplicity, the FD decomposition is designed to yield FD sub-frames with statistical properties that are similar to each other. This allows the allocation of an equal number of bits, namely I, to each transmission sub-frame $\mathbf{u^m} = \{u_i^m\}_{i=1}^I$, where $u_i^m \in \{0, 1\}$ and $i \in [1 \dots I]$ is the bit index. Additionally, the same codebook may be employed for the M number of VQ-encoding and -decoding processes.
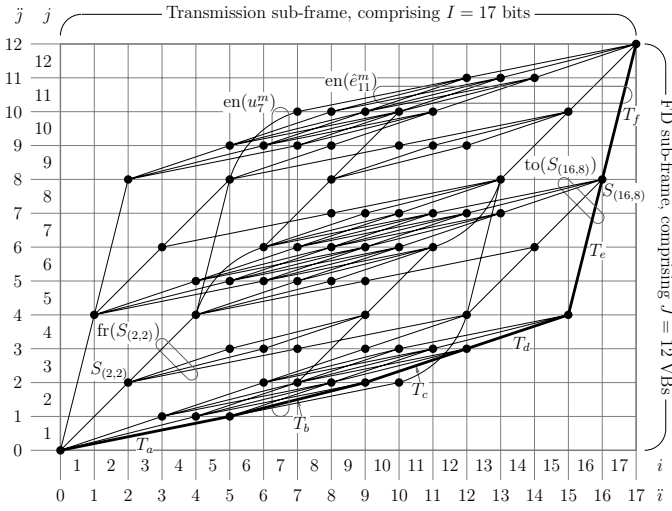
Fig. 6.   Example of a VQ trellis structure for the FD decomposition example of Figure 3 and for the $K = 5$-entry VQ codebook example of Figure 4, where we have $J = 12$ and $I = 17$. Trellis states occur between the consideration of specific bits and VBs, hence bit state indices $\ddot{i}$ and VB state indices $\ddot{j}$ occur between the bit indices i and the VB indices j, respectively.

The proposed video codec operates on a block-based philosophy. In this contribution, a *video block* (VB) has the dimensions of (8×8) pixels and is defined as the smallest unit of video information that is considered in isolation. Each of the M number of FD sub-frames $\mathbf{e^m}$ comprises a unique combination of J number of (8×8)-pixel VBs of the FD $\mathbf{e}$. Hence, the FD $\mathbf{e}$ comprises $M \cdot J$ number of VBs in total. In the example of Figure 3, each of the $M = 33$ FD sub-frames in the (176×144)-pixel FD comprises $J = 12$ of the $M \cdot J = 396$ (8×8)-pixel VBs, which are shown with dashed boundaries.

We proceed by decomposing the FD $\mathbf{e}$ into a set of perfectly tiling *macro-blocks* (MBs). Each MB has the dimensions of $(J_x^{\mathrm{MB}} \times J_y^{\mathrm{MB}}) = J^{\mathrm{MB}}$ number of VBs, where $J = 12$ is an integer multiple of J^{MB}. In the example of Figure 3, MBs are shown with solid boundaries and have the dimensions of $(J_x^{\mathrm{MB}} \times J_y^{\mathrm{MB}}) = (2 \times 2)$ VBs, giving $J^{\mathrm{MB}} = 4$. The MBs are then assigned to $J/J^{\mathrm{MB}} = 3$ different groups on the basis of the distance between their centre and the FD centre, with each group comprising $M = 33$ MBs. This results in $J/J^{\mathrm{MB}} = 3$ quasi-concentric MB groups, as indicated by the thick boundaries in Figure 3. A pseudo-random selection of one MB from each of the $J/J^{\mathrm{MB}} = 3$ groups is performed for each FD sub-frame $\mathbf{e^m}$. It is the $J = 12$ (8×8)-pixel VBs identified by this pseudo-random MB selection that constitute the FD sub-frame $\mathbf{e^m}$. We note that a pre-determined fixed seed is employed for the sake of allowing identical pseudo-random MB selections to be made independently by both the video encoder and decoder.

Each of the $J = 12$ (8×8)-pixel VBs e_j^m, that constitute the FD sub-frame $\mathbf{e^m} = \{e_j^m\}_{j=1}^J$, is allocated a VB index $j \in [1 \dots J]$. These $J = 12$ VB indices are allocated *amongst* the $J/J^{\mathrm{MB}} = 3$ pseudo-randomly selected MBs in a quasi-radial ordering. In this way, the $J^{\mathrm{MB}} = 4$ lowest-valued VB indices are allocated to the MB nearest to the FD centre, as seen in Figure 3. Similarly, the MB nearest to the FD perimeter is assigned the $J^{\mathrm{MB}} = 4$ highest-valued VB indices. The allocation of VB indices *within* MBs should be made with

specific consideration of the VQ codebook employed, as will be detailed in Section III-C. Again, Figure 3 exemplifies the indices allocated to the $J = 12$ VBs of a FD sub-frame.

As mentioned, the FD sub-frames have statistical properties that are similar to each other, however it should be noted that each of the $J/J^{\mathrm{MB}} = 3$ constituent MBs is likely to exhibit different statistical properties. For example, the MB that is allocated the lowest-valued VB indices in each FD sub-frame can be expected to exhibit a high level of video activity, since this MB is located near the centre of the FD. By contrast, a low level of video activity can be expected for the MB that is assigned the highest-valued VB indices.

The reconstructed FD sub-frame $\hat{\mathbf{e}}^{\mathbf{m}} = \{\hat{e}_j^m\}_{j=1}^J$ employs the same VB selection and indexing as the corresponding FD sub-frame $\mathbf{e^m} = \{e_j^m\}_{j=1}^J$. The same applies to the reconstructed FD sub-frame estimate $\tilde{\mathbf{e}}^{\mathbf{m}} = \{\tilde{e}_j^m\}_{j=1}^J$.

In contrast to the FD decomposition described above, an alternative approach could comprise a selection of $J = 12$ *adjacent* VBs for each FD sub-frame. Although this alternative would permit the exploitation of correlation between adjacent VBs, each FD sub-frame would have different statistical properties. As a result, the corresponding transmission sub-frames would require different individual bit-allocations on a demand-basis. Additionally, each transmission sub-frame's different bit-allocation would require a separate VQ-encoder and -decoder design. Finally, the transmission of side information would be required for the sake of signalling this allocation between the transmitter and the receiver. Hence, this alternative approach was discarded for the sake of maintaining simplicity.

B. VQ codebook

As stated in Section II, each FD sub-frame $\mathbf{e^m}$ is approximated by the reconstructed FD sub-frame $\hat{\mathbf{e}}^{\mathbf{m}}$ and represented by the bit-based transmission sub-frame $\mathbf{u^m}$ on the basis of the same VQ codebook. This comprises K number of VQ tiles addressed by their Reversible Variable-Length Coding (RVLC) based [16] codebook index. These are known to both the video encoder and decoder. 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 FD values and light pixels represent positive FD values.

As exemplified in Figure 4, the K number of VQ tiles in the VQ codebook may have different dimensions, which must be multiples of the (8×8)-pixel VB dimensions. This allows the adequate representation of both small areas of high video activity and large areas of low video activity. In this way, coding efficiency is maintained. The VQ codebook entry with index $k \in [1 \dots K]$ is associated with the VQ tile $\mathbf{VQ^k}$. This has the dimensions of $(J_x^k \times J_y^k) = J^k$ number of (8×8)-pixel VBs, as exemplified in Figure 4. However, a VQ tile's dimensions must not exceed the MB dimensions of $(J_x^{\mathrm{MB}} \times J_y^{\mathrm{MB}}) = (2 \times 2)$ VBs, as defined in Section III-A. Each VB $VQ_{j^k}^k$, from the set of J^k VBs that constitute the VQ tile $\mathbf{VQ^k} = \{VQ_{j^k}^k\}_{j^k=1}^{J^k}$, is allocated an index $j^k \in [1 \dots J^k]$. Note that these indices may be arbitrarily allocated. To emphasise this point, a random allocation of VB indices is employed in Figure 4, as exemplified by the VQ tiles $\mathbf{VQ^1}$ and $\mathbf{VQ^2}$.

Additionally, the VQ codebook entry index k is represented by the RVLC code $\mathbf{RVLC^k}$, as exemplified in Figure 4. Each RVLC code $\mathbf{RVLC^k} = \{RVLC_{i^k}^k\}_{i^k=1}^{I^k}$ comprises I^k number of bits having values of $RVLC_{i^k}^k \in \{0, 1\}$, where $i^k \in [1 \ldots I^k]$ is the bit index. The employment of RVLC codes instead of their more efficient, higher coding-rate alternatives, such as Huffman codes [17], will be justified in Section III-C.

The VQ codebook should be designed by considering the statistical properties of the FD sub-frames. This can be achieved by employing the Linde-Buzo-Gray algorithm [18] to design the VQ tile set and a Huffman coding based algorithm [16] to design the RVLC code set.

Recall from Section III-A that all FD sub-frames exhibit similar statistical properties, but different statistical properties are exhibited by each of the $J/J^{\mathrm{MB}} = 3$ MBs that constitute the FD sub-frames. In an alternative approach, a separate VQ codebook could be designed and employed to model the statistical properties of each of these MBs. This alternative approach would allow the achievement of a higher coding efficiency. However, it was discarded for the sake of simplicity.

C. VQ-induced code constraints

Let us now elaborate on how the employment of VQ imposes code constraints that govern the formation of legitimate bit-sequences within the transmission sub-frames. In Figure 5 an example of a VQ-based $J = 12$-block reconstructed FD sub-frame $\hat{\mathbf{e}}^m$ is provided. Here, we employed the FD decomposition example of Figure 3 and the $K = 5$-entry VQ codebook example of Figure 4. The corresponding $I = 17$-bit transmission sub-frame $\mathbf{u}^m$ is subject to the VQ-induced code constraints outlined below.

As stated in Section III-A, each reconstructed FD sub-frame $\hat{\mathbf{e}}^m = \{\hat{e}_j^m\}_{j=1}^J$ comprises $J = 12$ (8×8)-pixel VBs. These $J = 12$ VBs constitute $J/J^{\mathrm{MB}} = 3$ MBs, as shown in Figure 5. Each of these MBs comprises an appropriate tessellation of VQ tiles from the K-entry VQ codebook. The tiles $\mathbf{VQ^k} = \{VQ_{j^k}^k\}_{j^k=1}^{J^k}$ may cover regions of $(J_x^k \times J_y^k) = J^k$ number of (8×8)-pixel VBs. Note that, in Figure 5 each of the entries k^T, where $T \in [T_a \ldots T_f]$, invoked from the VQ codebook example of Figure 4 provides J^{k^T} number of (8×8)-pixel VBs for the reconstructed FD sub-frame $\hat{\mathbf{e}}^m$.

A specific constraint is imposed that restricts the positioning of the tile $\mathbf{VQ^k}$ in legitimate tessellations. Specifically, the J^k number of VBs in $\hat{\mathbf{e}}^m = \{\hat{e}_j^m\}_{j=1}^J$ that are represented by the tile $\mathbf{VQ^k}$ must have consecutive VB indices j. Note that this is true in all cases in Figure 5. This VQ tile positioning constraint is imposed to allow the formation of the novel trellis structure to be described in Section III-D. In order that the video-degradation imposed by this constraint is minimised, the VQ tiles and the allocation of VB indices within the MBs should be jointly designed, as stated in Section III-A. In view of this, the presence of the vertical VQ tile $\mathbf{VQ^2}$ in the VQ codebook example of Figure 4 motivates the vertical allocation of consecutive VB indices within each MB, as seen in Figure 3.

The employment of the VQ tile $\mathbf{VQ^k}$ is expressed using J^k number of mappings of the form $\hat{e}_j^m = VQ_{j^k}^k$, as exemplified in Figure 5. With reference to the examples of Figures 3 and 4, we note that the resultant values of $j \in [1 \ldots J]$ and

$j^k \in [1 \ldots J^k]$ in such mappings are dependent on the specific positioning of the VQ tile.

Each of the VQ tiles $\mathbf{VQ^k}$ that comprise the reconstructed FD sub-frame $\hat{\mathbf{e}}^m$ is associated with an RVLC code $\mathbf{RVLC^k}$, where k is the VQ codebook entry index. These RVLC codes are concatenated in the order of the VB indices j that are associated with the employment of the corresponding VQ tiles in the reconstructed FD sub-frame $\hat{\mathbf{e}}^m = \{\hat{e}_j^m\}_{j=1}^J$. This is exemplified in Figure 5, in which each of the entries k^T, where $T \in [T_a \ldots T_f]$, invoked from the VQ codebook example of Figure 4 is represented by I^{k^T} number of bits. In the proposed video codec this concatenation is constrained to having a total length of $I = 17$ bits, since it constitutes the $I = 17$-bit transmission sub-frame $\mathbf{u}^m$.

It follows from the above discussions that each $I = 17$-bit transmission sub-frame must represent a legitimate tessellation of VQ tiles, comprising $J = 12$ VBs in total. These VQ-induced code constraints represent redundancy within the bit-based transmission sub-frames. Note that the degree of this redundancy is dependent on the distance properties of the RVLC code set employed. As will be outlined in Section VI, the employment of RVLC codes rather than Huffman codes is justified, because their additional redundancy is desirable.

D. VQ trellis structure

As described in Section III-C, the formation of legitimate bit-sequences within each of the bit-based transmission sub-frames is governed by the VQ-induced code constraints. During VQ-encoding and -decoding we consider only legitimate $I = 17$-bit transmission sub-frame permutations. Since satisfying this constraint is non-trivial, the employment of a trellis structure is proposed to describe the complete set of VQ-induced code constraints. These VQ-induced code constraints depend on the FD decomposition and the VQ codebook employed, as described in Section III-C. Hence, the design of the proposed VQ trellis structure also depends on these aspects of the proposed video codec's operation. The VQ trellis structure example of Figure 6 describes the complete set of VQ-induced code constraints that correspond to the FD decomposition example of Figure 3 and to the $K = 5$-entry VQ codebook example of Figure 4. In this example, each reconstructed FD sub-frame $\hat{\mathbf{e}}^m$ comprises $J = 12$ VBs and each transmission sub-frame $\mathbf{u}^m$ comprises $I = 17$ bits.

The proposed VQ trellis structure comprises a set of transitions between trellis states, as exemplified in Figure 6. A novel block-based modification of the Variable Length Coding (VLC) symbol-level trellis structure described in [10] is employed. Whilst the bit index axis of [10] is retained in the proposed VQ trellis structure, the VLC symbol index axis of [10] is replaced by a VB index axis. In contrast to [10], transitions are permitted to skip a number of consecutive indices along this axis.

During a VQ-encoding or -decoding operation, there are a number of instances when it is possible to employ each VQ codebook entry, as will be elaborated on below. The proposed VQ trellis structure represents each of these legitimate possibilities with a transition, as exemplified in Figure 6. In this way, the proposed VQ trellis structure describes the complete set of VQ-induced code constraints. Each transition

T represents a possible employment of the VQ codebook entry having the index $k^T \in [1 \ldots K]$. This is associated with a unique combination of (a) the positioning of the VQ tile $\mathbf{VQ^{k^T}}$ within the reconstructed FD sub-frame $\mathbf{\hat{e}^m}$ and (b) the corresponding positioning of the associated RVLC code $\mathbf{RVLC^{k^T}}$ within the transmission sub-frame $\mathbf{u^m}$. Note that for reasons to be discussed below, (a) is subject to the legitimate VQ tile positioning constraint that was stated in Section III-C.

Each trellis state $S_{(\ddot{i},\ddot{j})}$ in the proposed VQ trellis structure represents the progress made at a particular point during the VQ-encoding or -decoding operation. This point occurs immediately after the consideration of the first $\ddot{j}$ number of VBs in the reconstructed FD sub-frame $\mathbf{\hat{e}^m}$ and the first $\ddot{i}$ number of bits in the transmission sub-frame $\mathbf{u^m}$. Here, $\ddot{j} \in [0 \ldots J]$ denotes a VB state index. These occur between the VB indices $j \in [1 \ldots J]$ that were introduced in Section III-A. Likewise, bit state indices $\ddot{i} \in [0 \ldots I]$ occur between the bit indices $i \in [1 \ldots I]$, as shown in Figure 6.

Each transition T represents the employment of the VQ codebook entry with index $k^T \in [1 \ldots K]$ immediately after reaching a particular point during the VQ-encoding or -decoding operation. This point is identified by the state indices $\ddot{i}^T \in [0 \ldots I]$ and $\ddot{j}^T \in [0 \ldots J]$. Hence, the transition T emerges from the trellis state $S_{(\ddot{i}^T,\ddot{j}^T)}$. The VQ tile $\mathbf{VQ^{k^T}}$ provides J^{k^T} number of VBs for the reconstructed FD sub-frame $\mathbf{\hat{e}^m}$, as stated in Section III-C. Additionally, the RVLC code $\mathbf{RVLC^{k^T}}$ provides I^{k^T} number of bits for the transmission sub-frame $\mathbf{u^m}$. Hence, the transition T merges into the trellis state $S_{(\ddot{i}^T+J^{k^T},\ddot{j}^T+I^{k^T})}$. Note that the VQ tile positioning constraint described in Section III-C is necessary to ensure that each transition is continuous with respect to the VB index axis. Also note that a particular employment of a VQ codebook entry is only possible if the associated transition T contributes to a legitimate transition path between the trellis states $S_{(0,0)}$ and $S_{(I,J)}$. This condition is satisfied if at least one transition path exists between the trellis states $S_{(0,0)}$ and $S_{(\ddot{i}^T,\ddot{j}^T)}$ and between the trellis states $S_{(\ddot{i}^T+I^{k^T},\ddot{j}^T+J^{k^T})}$ and $S_{(I,J)}$.

Note that the reconstructed FD sub-frame and transmission sub-frame examples of Figure 5 correspond to the bold trellis path in Figure 6. Here, each transition $T \in [T_a \ldots T_f]$ corresponds to the similarly labelled employment of a VQ codebook entry in Figure 5. With reference to Figure 5, observe that each transition $T \in [T_a \ldots T_f]$ encompasses I^{k^T} number of bit indices and J^{k^T} number of VQ indices. Whilst the trellis path considered comprises six transitions, it should be noted that the trellis structure of Figure 6 additionally contains trellis paths comprising seven transitions. In general, each transition path in the proposed VQ trellis structure comprises a varying number of transitions.

Note furthermore that a single VQ trellis structure considering $J \cdot M = 396$ VBs and $I \cdot M = 561$ bits would contain more transitions than the combination of $M = 33$ VQ trellis structures, each considering $J = 12$ VBs and $I = 17$ bits. This justifies the decomposition of the VQ-encoding and -decoding operations of an entire FD into $M = 33$ less complex trellis-based VQ operations, as described in Section II-B. This decomposition is associated with a reduced grade of VQ

encoding freedom and hence a reduced video reconstruction quality. However, this slight video degradation is insignificant compared to the resultant computational complexity reduction.

For the benefit of the following sections, we now introduce some trellis-transition set notation, which is exemplified in Figure 6. The set of all transitions that encompasses (en) the VB $j \in [1 \ldots J]$ of the reconstructed FD sub-frame $\mathbf{\hat{e}^m}$ is termed $\mathrm{en}(\hat{e}_j^m)$. Furthermore, $\mathrm{en}(\hat{e}_j^m = VQ_{j^k}^k)$ is the specific sub-set of $\mathrm{en}(\hat{e}_j^m)$ that maps the VB $j^k \in [1 \ldots J^k]$ of the VQ tile $\mathbf{VQ^k}$ onto $\hat{e}_j^m$. Note that some of these sub-sets may be empty. This is a consequence of the VQ tile positioning constraint described in Section III-C. The set of all transitions that encompasses the bit $i \in [1 \ldots I]$ of the transmission sub-frame $\mathbf{u^m}$ is termed $\mathrm{en}(u_i^m)$. Additionally, $\mathrm{en}(u_i^m = b)$ is the specific sub-set of $\mathrm{en}(u_i^m)$ that maps the bit value $b \in \{0, 1\}$ onto u_i^m. Note that for a particular transition T, we have $RVLC_{i-\ddot{i}^T}^{k^T} = b$ if $T \in \mathrm{en}(u_i^m = b)$. Finally, the set of transitions emerging from (fr) the trellis state $S_{(\ddot{i},\ddot{j})}$ is denoted as $\mathrm{fr}(S_{(\ddot{i},\ddot{j})})$, whilst the set merging into (to) that trellis state is represented as $\mathrm{to}(S_{(\ddot{i},\ddot{j})})$.

IV. VQ ENCODING

In the proposed VQ-TCM system's transmitter of Figure 1, VQ encoding is performed separately for each of the M number of FD sub-frames. Each VQ encoder operates on the basis of the proposed VQ trellis structure described in Section III-D and exemplified in Figure 6. Since it describes the complete set of VQ-induced code constraints, the employment of the proposed VQ trellis structure represents the consideration of every legitimate FD sub-frame encoding. This allows us to find the MMSE approximation of the FD sub-frame $\mathbf{e^m}$. The result is the optimal reconstructed FD sub-frame $\mathbf{\hat{e}^m}$ and the corresponding bit-based transmission sub-frame $\mathbf{u^m}$.

We quantize the video sequence in a novel manner, which is reminiscent of TCQ [11], but considers the tessellation of potentially differently sized VQ tiles. The philosophy of Viterbi decoding [15] is employed, with a survivor path being selected at each trellis state $S_{(\ddot{i},\ddot{j})}$. This selection yields the $\ddot{i}$-bit encoding of the first $\ddot{j}$ number of VBs in the FD sub-frame $\mathbf{e^m}$ that introduces the lowest possible cumulative video distortion $D(S_{(\ddot{i},\ddot{j})})$.

As stated in Section III-D, each transition T in the proposed VQ trellis structure represents the employment of the VQ codebook entry with index k^T during VQ encoding. This corresponds to employing the VQ tile $\mathbf{VQ^{k^T}}$ to represent a total of J^{k^T} number of (8×8)-pixel VBs of the FD sub-frame $\mathbf{e^m}$. The distortion $d(T)$ associated with the transition T is the sum of the squared difference between $\mathbf{VQ^{k^T}}$ and the corresponding J^{k^T} number of VBs of $\mathbf{e^m}$.

The survivor path at the trellis state $S_{(\ddot{i},\ddot{j})}$ is deemed to be that associated with the specific merging transition $T \in \mathrm{to}(S_{(\ddot{i},\ddot{j})})$ having the minimum cumulative video distortion. This is calculated as $D(T) = d(T) + D(S_{(\ddot{i}^T,\ddot{j}^T)})$, where $D(S_{(0,0)}) = 0$. Having determined the survivor path at the trellis state $S_{(I,J)}$, the MMSE VQ encoding of the FD sub-frame $\mathbf{e^m}$ has been found. The reconstructed FD sub-frame $\mathbf{\hat{e}^m}$ is formed as the tessellation of the VQ tiles associated with the survivor path transitions. Additionally, the I-bit transmission sub-frame $\mathbf{u^m}$ is formed as the concatenation of the associated RVLC codes, as described in Section III-C.

V. VQ DECODING

In the proposed VQ-TCM system's receiver of Figure 2, VQ decoding is performed separately for each of the M number of I-bit transmission sub-frames. Each of the M number of VQ decoders operates on the basis of the proposed VQ trellis structure. The recovery of legitimate – although not necessarily error-free – video information is therefore guaranteed, since the proposed VQ trellis structure describes the complete set of VQ-induced code constraints. As stated in Section III-C, these VQ-induced code constraints impose redundancy within each transmission sub-frame. This redundancy may be exploited to assist during VQ decoding by invoking the BCJR [12] algorithm on the basis of the proposed VQ trellis structure. Additionally, residual redundancy is exhibited by the transmission sub-frame $\mathbf{u^m}$ owing to the simplicity of the proposed video codec. This may be exploited by the BCJR algorithm with consideration of the statistical properties of FD sub-frames, as will be detailed below.

Each VQ decoder is provided with the *a priori* LLR sub-frame $L_a^2(\mathbf{u^m})$ and generates the *a posteriori* LLR sub-frame $L_p^2(\mathbf{u^m})$, as stated in Section II-C. Additionally, each VQ decoder recovers the optimal MMSE-based reconstructed FD sub-frame estimate $\mathbf{\tilde{e}^m}$. This is achieved using soft APP-based reconstruction, as follows.

A novel block-based modification of the BCJR algorithm of [13] is invoked in the proposed VQ trellis structure. This obtains an APP for each transition, giving the probability that this specific transition was in the survivor path during the corresponding VQ encoding operation, as described in Section IV. These APPs are calculated as

$$P(T|\mathbf{y}) = \frac{1}{C_1} \cdot \alpha(S_{(\tilde{i}^T, \tilde{j}^T)}) \cdot \gamma(T) \cdot \beta(S_{(\tilde{i}^T + I^{k^T}, \tilde{j}^T + J^{k^T})}). \quad (1)$$

The specific value of the normalisation factor $C_1 = p(\mathbf{y})$ in (1) may be ignored in this application, since we are only concerned with the ratios of *a posteriori* transition probabilities, as we will show below. The terms $\alpha(S_{(\tilde{i}^T, \tilde{j}^T)})$, $\gamma(T)$ and $\beta(S_{(\tilde{i}^T + I^{k^T}, \tilde{j}^T + J^{k^T})})$ consider the probability of the requisite trellis activity before, during and after the occurrence of the transition T, respectively.

Specifically, the term $\gamma(T)$ in (1) is calculated as

$$\gamma(T) = \frac{P(k^T)}{C_2(S_{(\tilde{i}^T, \tilde{j}^T)})} \cdot \prod_{i^k=1}^{I^{k^T}} P(u_{\tilde{i}^T + i^k}^m = RVLC_{i^k}^{k^T}|y), \quad (2)$$

where $P(u_i^m = b|y)$ is the *a priori* probability that the transmission sub-frame bit u_i^m has a value of $b \in \{0, 1\}$. This is obtained from the *a priori* LLR $L_a^2(u_i^m) = \ln(\frac{P(u_i^m=0|y)}{P(u_i^m=1|y)})$ [14]. Furthermore, $P(k)$ in (1) is the probability of occurrence for the VQ codebook entry with index k. This is obtained based on knowledge of the statistical properties of the FD sub-frames. As stated in Section III-A, different statistical properties are associated with each of the J/J^{MB} number of MBs constituting a FD sub-frame. Hence, different values of $P(k)$ are employed, depending on the specific MB that T is constituent of. Finally, the normalisation factor $C_2(S_{(\tilde{i}, \tilde{j})}) = \sum_{T \in \text{fr}(S_{(\tilde{i}, \tilde{j})})} P(k^T)$ was proposed in [13] for the sake of ensuring that (2) represents a true probability.

The BCJR algorithm's forward recursion emerging from the trellis state $S_{(0,0)}$ is employed to obtain the values of

$$\alpha(S_{(\tilde{i}, \tilde{j})}) = \sum_{T \in \text{to}(S_{(\tilde{i}, \tilde{j})})} \gamma(T) \cdot \alpha(S_{(\tilde{i}^T, \tilde{j}^T)}), \quad (3)$$

where $\alpha(S_{(0,0)}) = 1$. Similarly, a backward recursion from the trellis state $S_{(I,J)}$ is employed to obtain the values of

$$\beta(S_{(\tilde{i}, \tilde{j})}) = \sum_{T \in \text{fr}(S_{(\tilde{i}, \tilde{j})})} \gamma(T) \cdot \beta(S_{(\tilde{i}^T + I^{k^T}, \tilde{j}^T + J^{k^T})}), \quad (4)$$

where $\beta(S_{(I,J)}) = 1$.

Having determined the *a posteriori* transition probabilities, the method of [13] is employed for obtaining bit-based soft outputs. This is facilitated by the employment of the normalisation factor $C_2(S_{(\tilde{i}^T, \tilde{j}^T)})$ during the calculation of $\gamma(T)$, as described above. For each of the I number of bits in the transmission sub-frame $\mathbf{u^m} = \{u_i^m\}_{i=1}^I$, we consider a cross-section of the trellis structure, which is perpendicular to the bit index axis at the particular index i. The associated *a posteriori* LLR is calculated as

$$L_p^2(u_i^m) = \ln \left(\frac{\sum_{T \in \text{en}(u_i^m=0)} P(T|\mathbf{y})}{\sum_{T \in \text{en}(u_i^m=1)} P(T|\mathbf{y})} \right). \quad (5)$$

The recovery of the reconstructed FD sub-frame estimate $\mathbf{\tilde{e}^m}$ is performed on an individual block-by-block basis. A soft APP-based output is obtained for each of the J number of (8×8)-pixel VBs in $\mathbf{\tilde{e}^m}$. Again, a novel modification of the method of [13] is employed for obtaining these block-based soft outputs. For each of the J number of VBs in the reconstructed FD sub-frame $\mathbf{\hat{e}^m} = \{\hat{e}_j^m\}_{j=1}^J$, we consider a cross-section of the trellis structure. In analogy to the generation of the previously mentioned bit-based soft outputs, this cross-section is now perpendicular to the VB index axis at the particular index j. The APP of the VB $\hat{e}_j^m$ being provided by a particular one of the J^k number of VBs in the VQ tile $\mathbf{VQ^k} = \{VQ_{j^k}^k\}_{j^k=1}^{J^k}$ is calculated as

$$P(\hat{e}_j^m = VQ_{j^k}^k|\mathbf{y}) = \frac{\sum_{T \in \text{en}(\hat{e}_j^m = VQ_{j^k}^k)} P(T|\mathbf{y})}{\sum_{T \in \text{en}(\hat{e}_j^m)} P(T|\mathbf{y})}. \quad (6)$$

Note that some of the sets $\text{en}(\hat{e}_j^m = VQ_{j^k}^k)$ may be empty, as described in Section III-D. In this case, the corresponding APP is zero-valued.

Each of the J number of (8×8)-pixel VBs that constitute the reconstructed FD sub-frame estimate $\mathbf{\tilde{e}^m} = \{\tilde{e}_j^m\}_{j=1}^J$ is obtained using optimal MMSE estimation according to

$$\tilde{e}_j^m = \sum_{k=1}^K \sum_{j^k=1}^{J^k} P(\hat{e}_j^m = VQ_{j^k}^k|\mathbf{y}) \cdot VQ_{j^k}^k. \quad (7)$$

The Log-MAP algorithm [19] is employed to reduce the computational complexity of VQ decoding. Specifically, the above mentioned calculations are performed in the logarithmic domain by employing the Jacobian approximation with an eight-entry table-lookup correction factor [15]. This reduces the number of multiplications required by the BCJR algorithm. Additionally, the T-BCJR algorithm [20] is employed to prune insignificant transitions from the proposed VQ trellis structure,
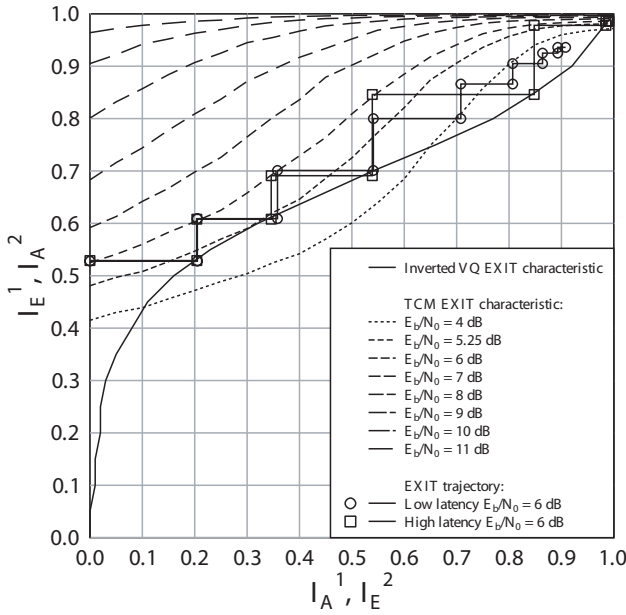
Fig. 7. The proposed VQ-TCM system's EXIT chart for the 'Lab' video sequence and a $3/4$-rate TCM scheme.

where the transition paths passing through the trellis state $S_{(\check{i},\check{j})}$ are pruned during the forward recursion if

$$\frac{\alpha(S_{(\check{i},\check{j})})}{\sum_{\check{i}'=0}^{I} \alpha(S_{(\check{i}',\check{j})})} < 0.001. \tag{8}$$

In the next section we consider the performance of the proposed VQ-TCM system, where the above-mentioned complexity-reduction methods were observed to impose no significant performance degradation.

VI. RESULTS

In this section we assess the performance of the proposed VQ-TCM system. We transmitted 250 video frames of the 'Lab' video-sequence [2]. This 10 frames/s grey-scale head-and-shoulders (176×144)-pixel Quarter Common Intermediate Format (QCIF) video sequence exhibits a moderate level of video activity.

The proposed video codec was designed to achieve an attractive trade-off between the conflicting design requirements associated with bitrate, video reconstruction quality and computational complexity. The 396-block FDs were decomposed into $M = 33$ FD sub-frames, each comprising $J = 12$ (8×8)-pixel VBs. This was performed exactly as exemplified in Figure 3. However, in contrast to our simplified example of Figure 6 using $I = 17$ bits per sub-frame, each transmission sub-frame comprised $I = 45$ bits in the system implemented. A $K = 512$-entry VQ codebook was employed. This comprised the five VQ tiles shown in Figure 4 and an additional 507 single-VB VQ tiles. The corresponding RVLC codes were designed to have a minimum free distance of two [16]. The coding rate of the proposed video codec is defined here as the ratio of its entropy to its average RVLC code length. This may be estimated as the area beneath the inverted VQ decoding EXIT characteristic [14] provided in Figure 7, which will be discussed in more detail below. This gives a coding rate of $R_{VQ} \approx 0.666$.

Two VQ-TCM schemes associated with different latencies were employed. The first scheme imposed a low latency equal to the duration of a single video frame, namely 0.1 s at 10 frames/s. This is suitable for realtime interactive video-telephony applications. In this scheme the length of each transmission frame $\mathbf{u}$ and that of the interleaver π equals $M \cdot I = 1485$ bits. The second VQ-TCM scheme had a high latency of 50 video frames, i.e. 5 s at 10 frames/s. This is suitable for non-realtime video streaming and wireless-Internet download applications. Here, 50 transmission frames $\mathbf{u}$ are concatenated before interleaving, giving an interleaver length of $50 \cdot M \cdot I = 74250$ bits. Note that both schemes have a video encoded bitrate of 14.85 kbps.

The same TCM codec was employed in both VQ-TCM schemes. This terminated $R_{TCM} = 3/4$-rate TCM codec employed a coding memory of 6, $M_{TCM} = 16$-level QAM modulation and IQ-interleaving [9] 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 without a Nyquist excess bandwidth. More explicitly, we assume ideal Nyquist filtering and ignore the code termination symbols added by the TCM encoder. Note that at $\eta = 2$ bit/s/Hz the uncorrelated Rayleigh-fading channel's capacity limit for 16QAM is $E_b/N_0 = 3.96$ dB [21], where $E_b/N_0 = SNR/\eta$ is the Signal to Noise Ratio (SNR) per bit.

Again, in Figure 7 we provide the EXIT characteristics [22] for TCM decoding at various E_b/N_0 values. These can be seen to achieve unity extrinsic mutual information I_E^1 for unity a priori mutual information I_A^1. Additionally, Figure 7 provides the inverted VQ decoding EXIT characteristic. Similarly to TCM decoding, VQ decoding achieves unity extrinsic mutual information I_E^2 for unity a priori mutual information I_A^2. This was found to be the benefit of the employment of an RVLC code set having a minimum free distance of two.

Additionally, Figure 7 provides EXIT trajectories [22] for both the low- and high-latency VQ-TCM schemes at $E_b/N_0 = 6$ dB. The low-latency trajectory can be seen to deviate from the EXIT characteristics and fails to converge to the desired unity mutual information. By contrast, the high-latency trajectory closely matches these EXIT characteristics and converges to unity mutual information. The improved performance of the high-latency scheme is a benefit of its longer interleaver [14], justified by the reasons noted in Section II-C. Note that iterative decoding convergence to unity mutual information is associated with the achievement of an infinitesimally low video degradation. The proposed VQ-TCM system supports this for $E_b/N_0 \geq 5.25$ dB, as shown in Figure 7. With the advent of a sufficiently long interleaver and of a sufficiently high number of decoding iterations, infinitesimally low video degradation could be achieved at $E_b/N_0 = 5.25$ dB. This is just 1.29 dB from the proposed VQ-TCM system's channel capacity limit of 3.96 dB.

In Figure 8, the video reconstruction quality of both the low- and high-latency VQ-TCM schemes is assessed after a number of decoding iterations and for a range of channel E_b/N_0 values. We employ the Peak-Signal to Noise Ratio (PSNR) [2] as the objective video reconstruction quality metric. In this application, a PSNR of 29.5 dB is associated with an aesthetically pleasing video reconstruction quality.
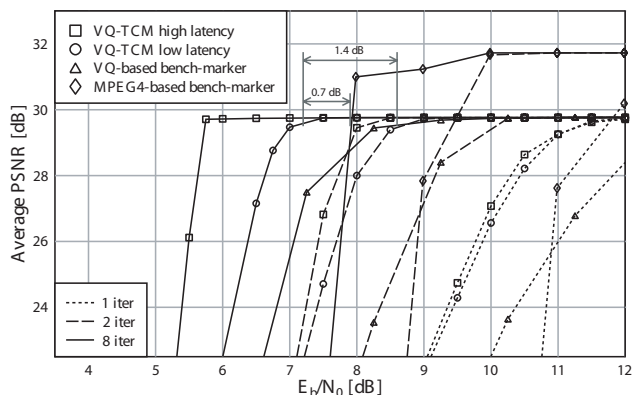
Fig. 8. PSNR performance of the proposed VQ-TCM system as well as of the VQ- and MPEG4-based bench-markers for the 'Lab' video sequence, when communicating over an uncorrelated narrowband Rayleigh fading channel.

The proposed VQ-TCM system exhibited substantial iteration gains, as shown in Figure 8. After eight decoding iterations an E_b/N_0 iteration gain of 4.34 dB was achieved by the low-latency scheme at a PSNR of 29.5 dB. In the case of the high-latency scheme, the corresponding iteration gain was 5.61 dB. The high-latency scheme was seen to outperform the low-latency scheme, regardless of the number of decoding iterations. This is in agreement with the findings of the EXIT chart analysis of Figure 7, as discussed above. For the low-latency scheme, a PSNR higher than 29.5 dB was achieved after eight decoding iterations for $E_b/N_0 > 7.00$ dB, as shown in Figure 8. This is 3.04 dB from the proposed VQ-TCM system's channel capacity limit of 3.96 dB. In the case of the high-latency scheme, this is achieved for $E_b/N_0 > 5.75$ dB, which is just 1.79 dB from the channel's capacity limit.

We now compare the performance of the proposed VQ-TCM system with two bench-markers. Similarly to the proposed VQ-TCM system, the IQ-TCM-VLC system of [3] is also VQ-based and is hence referred to as the *VQ-based bench-marker* here. In contrast, the IQ-TCM-RVLC system of [5] is MPEG4-based and is referred to here as the *MPEG4-based bench-marker*. Similarly to the proposed VQ-TCM system, the bench-markers employ both RVLC and TCM. However, they adopt Shannon's source and channel coding separation philosophy [1], with video decoding being performed independently of iterative RVLC- and TCM-decoding. Since the bench-markers have the same latency, coding rate and bitrate as the low-latency VQ-TCM scheme, their direct comparison is justified.

In Figure 8, the low-latency VQ-TCM scheme can be seen to outperform the VQ-based bench-marker, regardless of the number of decoding iterations. Although the MPEG4-based bench-marker achieves a slightly higher error-free video reconstruction quality than the proposed VQ-TCM system, it is outperformed by the scheme advocated at low values of E_b/N_0. At a PSNR of 29.5 dB, the proposed VQ-TCM system offers a consistent improvement over both the VQ- and MPEG4-based bench-markers, which amounts to about 1.4 dB and 0.7 dB, respectively. This is a benefit of the iterative exploitation of the video codec's code constraints, as outlined in Section V. Additionally, this approach guarantees the recovery of legitimate video information. Hence, the proposed VQ-TCM system is never forced to discard any video information.

By contrast, both the bench-markers of [3] and [5] drop video information, when the iterative decoding process does not recover legitimate video information, reducing the attainable video reconstruction quality.

The improved performance offered by the proposed VQ-TCM system is achieved at the cost of an increased computational complexity. A simple metric of computational complexity considers the number of BCJR trellis transitions encountered during iterative decoding. Note that this ignores the complexity contribution imposed by trellis pruning and video reconstruction. Although an equal number of BCJR transitions are employed during TCM decoding in the proposed VQ-TCM system and in the bench-markers, the complexity of the respective VQ- and RVLC-decoding operations are different. More explicitly, in the proposed VQ-TCM system, VQ decoding employs a channel-condition-dependent number of BCJR transitions, which is typically equal to the number of trellis transitions encountered during TCM decoding. By contrast, RVLC decoding has approximately a quarter of the number of trellis transitions, in both bench-markers. Hence, the proposed VQ-TCM system has a computational complexity that may be deemed approximately 1.6 times that of the bench-markers.

VII. CONCLUSIONS

In this contribution we have considered the concept of joint video and channel coding. The proposed video codec imposed VQ-induced code constraints, which were described by a novel VQ trellis structure. This VQ trellis structure was the basis of the proposed MMSE VQ-encoding and -decoding processes. In the latter case, a novel BCJR algorithm was employed to facilitate the iterative exchange of soft information with the serially concatenated TCM decoder. The proposed VQ-TCM system was shown to support error-free video reconstruction at 1.29 dB from the Rayleigh-channel's capacity limit of 3.96 dB.

The joint video and channel coding approach of the proposed VQ-TCM system was shown to consistently outperform two powerful bench-markers, both employing the Shannonian source and channel separation philosophy [1]. However, this performance improvement was found to accrue at the cost of a 1.6 times increase in computational complexity. Our further research will consider the application of the techniques introduced here to standard video codecs.

REFERENCES

[1] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, no. 3, 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. X. Ng *et al.*, "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, Sept. 2004, pp. 2231–2234.

[4] *Information Technology – Coding of Audio-Visual Objects – Part 2: Visual*, ISO/IEC 14496-2:2004 Std.

[5] S. X. Ng *et al.*, "A Turbo-Detection Aided Serially Concatenated MPEG-4/TCM Videophone Transceiver," in *IEEE Vehic. Technol. Conf. (VTC)*, Los Angeles, USA, Sept. 2004.

[6] Q. Chen and K. P. Subbalakshmi, "Joint source-channel decoding for MPEG-4 video transmission over wireless channels," *IEEE J. Select. Areas Commun.*, vol. 21, no. 10, pp. 1780–1789, 2003.

[7] S. Benedetto *et al.*, "Serial Concatenation of Interleaved Codes: Performance Analysis, Design and Iterative Decoding," *IEEE Trans. Info. Theory*, vol. 44, no. 3, pp. 909–926, May 1998.

[8] J. Hagenauer and N. Görtz, "The Turbo Principal in Joint Source-Channel Coding," in *Proc. IEEE Info. Theory Wksp.*, Paris, France, Mar. 2003, pp. 275–278.

[9] 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, Nov. 2002.

[10] R. Bauer and J. Hagenauer, "Symbol-by-Symbol MAP Decoding of Variable Length Codes," in *ITG Conference on Source and Channel Coding*, Munich, Germany, Jan. 2000, pp. 111–116.

[11] M. W. Marcellin and T. R. Fischer, "Trellis Coded Quantization of Memoryless and Gauss-Markov Sources," *IEEE Trans. Commun.*, vol. 38, no. 1, pp. 82–93, Jan. 1990.

[12] L. R. Bahl *et al.*, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, Mar. 1974.

[13] J. Kliewer and R. Thobaben, "Iterative Joint Source-Channel Decoding of Variable-Length Codes Using Residual Source Redundancy," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, May 2005.

[14] J. Hagenauer, E. Offer, and L. Papke, "Iterative Decoding of Binary Block and Convolutional Codes," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 429–445, Mar. 1996.

[15] 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.

[16] Y. Takishima, M. Wada, and H. Murakami, "Reversible Variable Length Codes," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 158–162, Feb./Mar./Apr. 1995.

[17] D. A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes," in *Proc. IRE*, vol. 40, no. 9, pp. 1098–1101, 1951.

[18] Y. Linde, A. Buzo, and R. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–95, Jan. 1980.

[19] P. Robertson, E. Villebrun, and P. Hoeher, "A Comparison of Optimal and Sub-Optimal MAP Decoding Algorithms Operating in the Log Domain," in *IEEE Int'l. Conf. Commun.*, Seattle, USA, June 1995, pp. 1009–1013.

[20] V. Franz and J. B. Anderson, "Concatenated Decoding with a Reduced-Search BCJR Algorithm," *IEEE J. Select. Areas Commun.*, vol. 16, no. 2, pp. 186–195, Feb. 1998.

[21] L. Hanzo *et al.*, *Quadrature Amplitude Modulation*. Chichester, UK: Wiley, 2004.

[22] S. ten Brink, "Convergence Behaviour of Iteratively Decoded Parallel Concatenated Codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

Robert G. Maunder has studied with the School of Electronics and Computer Science, University of Southampton, UK, since October 2000. He was awarded a first class honours BEng in Electronic Engineering in July 2003, shortly before beginning his current PhD studies in the Communications Research Group at the same university. His research interests include video coding, joint source/channel coding and iterative decoding.

Jörg Kliewer (S'97–M'99–SM'04) received the Dipl.-Ing. degree in Electrical Engineering from Hamburg University of Technology, Hamburg, Germany, in 1993 and the Dr.-Ing. degree (Ph.D.) in Electrical Engineering from the University of Kiel, Kiel, Germany, in 1998, respectively. From 1993 to 1998, he was a Research Assistant at the Institute for Circuits and Systems Theory at the University of Kiel, Germany. In 1996, he spent five months at the Image Processing Lab, University of California, Santa Barbara, CA, as a visiting researcher. Since 1999, he is with the Faculty of Engineering, University of Kiel, Germany, as a Senior Researcher and Lecturer. In 2004 he was visiting the University of Southampton, U.K., for one year, and since 2005 he is with the University of Notre Dame, IN, as a Visiting Assistant Professor. He was the recipient of a both Leverhulme fellowship award and a German Research Foundation fellowship award in 2003 and 2004, respectively. His current research interests include joint source and channel coding, error correcting codes, network information theory, and network coding

Soon Xin Ng (S'99-M'03) received the B.Eng. degree (First class) in electronics engineering and the Ph.D. degree in wireless communications from the University of Southampton, Southampton, U.K., in 1999 and 2002, respectively. From 2003 to 2006, he was a postdoctoral research fellow working on collaborative European research projects known as SCOUT, NEWCOM and PHOENIX. Since August 2006, he has been a lecturer in wireless communications at the University of Southampton. His research interests are mainly in adaptive coded modulation, channel coding, space-time coding, joint source and channel coding, OFDM and MIMO. He has published numerous papers and coauthored a book in this field.

Jin Wang (S'03) received the B.E. degree from University of Science and Technology of China (USTC), Hefei, China in 1999 and the M.E. degree in video signal processing from Graduate School of Chinese Academy of Sciences (GSCAS), Beijing, China, in 2002.

He is currently pursuing his PhD degree with the Communications Research Group at the School of ECS, University of Southampton, Southampton, UK. His research interests include video coding, channel coding, joint source/channel coding, and iterative detection and decoding for digital communication systems.

Dr. Lie-Liang Yang received his B.Eng. degree in communication engineering from Shanghai TieDao University, Shanghai, China in 1988, and his M.Eng, Ph.D. degrees in communications and electronics from Northern Jiaotong University, Beijing, China in 1991 and 1997, respectively. From June 1997 to December 1997 he was a visiting scientist of the Institute of Radio Engineering and Electronics, Academy of Sciences of the Czech Republic. Since December 1997, he has been with the Communications Research Group, School of Electronics and Computer Science, University of Southampton, U.K, where he was first a Postdoctoral Research Fellow (Dec. 1997 - Aug. 2002), then a Lecturer (Sept. 2002 - Feb. 2006), and currently holds the academic post of Readership.

Dr. Yang's research has covered a wide range of areas in telecommunications, which include error control coding, modulation and demodulation, spread-spectrum communications and multiuser detection, synchronization, space-time processing, adaptive wireless systems, as well as wideband, broadband and ultrawide-band code-division multiple-access (CDMA). He has published over 130 papers in journals and conference proceedings, coauthored one book and published several book chapters. He was awarded the Royal Society Sino-British Fellowship in 1997 and the EPSRC Research Fellowship in 1998. Dr. Yang is currently an associate editor for both the *Journal of Communications and Networks* (JCN) and the *Journal of Communications* (JCM).

Lajos Hanzo , Fellow of the Royal Academy of Engineering, received his first-class degree in electronics in 1976 and his doctorate in 1983. In 2004 he was awarded the Doctor of Sciences (DSc) degree by the University of Southampton, UK. During his career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the Department of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has co-authored 14 books, totalling 10 000 pages on mobile radio communications, published in excess of 650 research papers, has acted as TPC Chair of numerous major IEE and IEEE conferences, presented various keynote lectures and has been awarded a number of distinctions. Currently he heads an academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European IST Programme and the Mobile Virtual Centre of Excellence (VCE), UK. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. Lajos is also an IEEE Distinguished Lecturer of both the Communications as well as the Vehicular Technology Society, a Fellow of both the IEEE and the IEE. He is an editorial board member of the Proceedings of the IEEE and a Governer of the IEEE VT Society. For further information on research in progress and associated publications, please refer to http://www-mobile.ecs.soton.ac.uk