Turbo Detection of Space-time Trellis-Coded Constant Bit Rate Vector-Quantised Videophone System using Reversible Variable-Length Codes, Convolutional Codes and Turbo Codes

B. L. Yeap, R. G. Maunder, S. X. Ng, and L. Hanzo

School of ECS, University of Southampton, SO17 1BJ, UK.
Tel: +44-23-8059 3125, Fax: +44-23-8059 4508
Email: {bly,rm02r,sxn,lh}@ecs.soton.ac.uk, http://www-mobile.ecs.soton.ac.uk

Abstract

In this treatise we characterise the achievable performance of a proprietary video transmission system, which employs a Constant Bit Rate (CBR) video codec that is concatenated with one of of three error correction codecs, namely a Reversible Variable-Length Code (RVLC), a Convolutional Code (CC) or a convolutional-based Turbo Code (TC). In our investigations, the CBR video codec was invoked in conjunction with Space-Time Trellis Coding (STTC) designed for transmission over a dispersive Rayleigh fading channel. At the receiver, the channel equaliser, the STTC decoder and the RVLC, CC or TC decoder, as appropriate, employ the Max-Log Maximum A-Posteriori (MAP) algorithm and their operations are performed in an iterative 'turbo-detection' fashion. The systems were designed for maintaining similar error-free video reconstruction qualities, which were found to be subjectively pleasing at a Peak Signal to Noise Ratio (PSNR) of 30.6 dB, at a similar decoding complexity per decoding iteration. These design criteria were achieved by employing differing transmission rates, with the CC- and TC-based systems having a 22% higher bandwidth requirement. The results demonstrated that the TC-, RVLC- and CC-based systems achieve acceptable subjective reconstructed video quality associated with an average PSNR in excess of 30 dB for E_b/N_0 values above 4.6 dB, 6.4 dB and 7.7 dB, respectively. The design choice between the TC- and RVLC-based systems constitutes a trade-off between the increased error resilience of the TC-based scheme and the reduced bandwidth requirement of the RVLC-based scheme.

1. INTRODUCTION

In this contribution, we analyse and characterise the achievable performance of a proprietary Constant Bit Rate (CBR) video codec in the context of Space-Time Trellis Coded (STTC) systems [1]. The motivation for employing a CBR video codec is that of system design convenience, since CBR systems, such as the second-generation GSM system, exhibit a low complexity as a consequence of requiring no statistical multiplexing. Furthermore, the proposed CBR systems exhibit a low latency equal to the duration of each video frame, which was 100ms, facilitating lip-synchronisation. Additionally, the CBR design philosophy allows video frame synchronisation to be readily re-established in the presence of transmission errors. For the sake of comparison, we have invoked three different error-resilience codecs at the output of the CBR videophone scheme, namely a Reversible Variable-Length Code (RVLC) [2], a rate $R_{cc} = \frac{3}{4}$ constraint length $K = 5$ Convolutional Code (CC) and an $R_{tc} = \frac{3}{4}$ $K = 4$ convolutional-based Turbo Code (TC) [3]. These values of K were chosen for the sake of having similar complexities per decoding iteration amongst the three systems. Specifically, symmetrical RVLCs were utilised in the RVLC-protected scheme, since they may be designed for a free distance of two and they also represent a good compromise between coding rate and error resilience. We also considered CCs since they are known to yield good performance despite having short coding block lengths, while TCs, are capable of approaching the Shannonian limit.

In order to improve the achievable videophone link quality and to approach the capacity of the CBR videophone scheme, we propose systems that amalgamate STTC with the stated RVLC, TC and CC codecs. A STTC scheme has been invoked for providing additional diversity gains as well as for overcoming the limited capacity offered by hostile wireless channels [4]. STTC relies on the joint design of channel coding, modulation, transmit diversity and on the associated optional receiver diversity schemes. Rather than performing the channel equalisation, STTC decoding and consecutive source decoding operations separately, the proposed system performs these operations iteratively, exchanging soft information amongst them and yielding a performance improvement. This philosophy is based on that of turbo equalisation [5], which was first employed in a convolutional encoded BPSK system transmitting over dispersive channels. In order to reduce the complexity of the turbo detector, we have employed the In-phase/Quadrature-phase (I/Q) turbo detector proposed in [4, 6].

The outline of this contribution is as follows. In Section 2 an overview of the system and its corresponding parameters is presented. This is followed a discussion of the associated performance results in Section 3. Finally we conclude our discourse in Section 4.

The financial support of both the EPSRC, Swindon UK and the EU under the auspices of the Phoenix project is gratefully acknowledged.

2. SYSTEM OVERVIEW

We commence by describing the proprietary CBR video codec and the RVLC codec employed in the first of the three candidate videophone system, illustrated in Figure 1. The coding philosophy of this Vector-Quantised (VQ) scheme was described in [7, Chapter 13], which was designed for the encoding of (176×144)-pixel greyscale Quarter Common Intermediate Format (QCIF) head-and-shoulder video sequences.

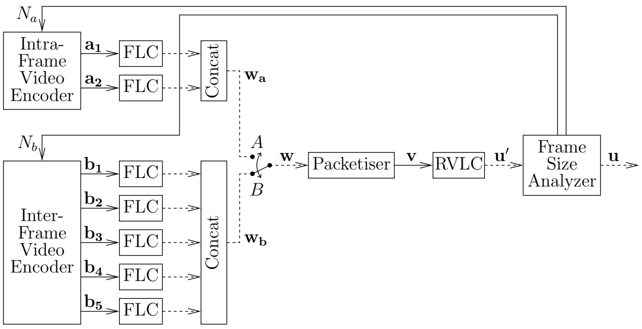

Figure 1: The proprietary CBR video encoder employing RVLC encoding. Solid lines indicate a transfer of symbols, while dashed lines indicate a transfer of bit strings.

In this system, each CBR coded video frame consists of 2048 bits, which completely fill five transmission frames, as will be described below. In Section 3, it is shown that this bit allocation permits the reconstruction of the 'Lab' video sequence [7] with a pleasing subjective reconstruction quality in the error-free case. At the 10 frames/sec video frame scanning rate employed, the resultant transmission bitrate is 20.48 kbps.

The proprietary video encoder [7, Chapter 13] incorporates both an intra- and an inter-frame encoder, as shown in Figure 1. **Inter-frame encoding** achieves a high level of compression by encoding the Motion Compensated Error Residual (MCER) between the current and previous frames. Partial Forced Update (PFU) [7, Section 12.8], Motion Compensation (MC) [7, Section 12.2] and Vector Quantisation (VQ) [7, Section 13.3] are employed. The Inter-frame encoding process operates on (8×8)-pixel blocks. There are 396 of these blocks in each (176×144)-pixel QCIF frame. The PFU, MC and VQ operations are applied only to a limited number of so-called *active blocks* in each frame. There are 22 PFU- and 63 MC-active blocks in each video frame, whilst the number of VQ-active blocks is variable and controlled by the 'Frame Size Analyzer' feedback parameter N_b, as shown in Figure 1. The specific selection of PFU-active blocks in each video frame is predetermined, while the specific choice of the MC- and VQ-active blocks is performed on a frame-by-frame basis using gain-cost control [7, Section 12.6].

The average luminances of PFU-active blocks are quantized and conveyed to the remote decoding using the PFU symbol set $\mathbf{a_1}$. They are used for mitigating the channel induced video degradation in the reconstruction of the previous video frame, which would otherwise persist during the reconstruction of the current inter-frame coded video frame. The motion vector symbols $\mathbf{b_2}$ are conveyed to the decoder for the set of MC-active blocks. Similarly, the VQ codebook entry symbols $\mathbf{b_3}$ are generated for the set of VQ-active blocks for the sake of representing the corresponding MCERs. The positions of PFU-active blocks are predetermined and known to the decoder. By contrast, the positions of each MC- and VQ-active block must be optimised and conveyed explicitly on a video frame-by-frame basis. A structured ordering of the MC- and VQ-active block indices is employed for the sake of signalling the active block positions $\mathbf{b_4}$ to the decoder. The structured ordering employed allows the detection of errors within the received active block indices. The value of N_b is conveyed using the symbols $\mathbf{b_5}$ of Figure 1.

However, inter-frame encoding is unable to represent the first frame of a video sequence, since there is no available previous frame to allow the generation of the MCER in this case. For this reason, **intra-frame encoding** must be employed for representing the first video frame of a sequence in isolation. The video frame is divided into a grid of $(N_a \times N_a)$ perfectly tiling blocks, where N_a is a feedback parameter, as shown in Figure 1. The average luminances of the blocks are quantised and conveyed using the intra-frame coding symbol set, $\mathbf{a_1}$, [7, Section 12.5]. The specific value of N_a is conveyed to the decoder using the symbols $\mathbf{a_2}$ of Figure 1.

During intra- and inter-frame encoding, two $[\mathbf{a_1} \dots \mathbf{a_2}]$ and five $[\mathbf{b_1} \dots \mathbf{b_5}]$ symbol sets are generated, respectively. As described above, the symbol set sizes are governed by the values of the parameters N_a and N_b, respectively. These values are provided by the system's 'Frame Size Analyzer' and bitrate control feedback mechanism, as shown in Figure 1. The employment of the above mechanism allows the generation of CBR encoded video frames $\mathbf{u}$, which efficiently use the 2048 bits that are available for their representation.

Bit-string representations of the symbols $[\mathbf{a_1} \dots \mathbf{a_2}]$ and $[\mathbf{b_1} \dots \mathbf{b_5}]$ are generated by applying Fixed Length Codes (FLC's). The allocation of FLCs to particular symbol values is based on their probability of occurrence. These bit strings are concatenated for the sake of forming the bit-strings $\mathbf{w_a}$ and $\mathbf{w_b}$, respectively, as seen in Figure 1.

A selection between $\mathbf{w_a}$ and $\mathbf{w_b}$ is made depending on whether the current video frame is to be intra- or inter-frame encoded. The selected concatenated bit string, $\mathbf{w}$, is packetised using a 4-bit packetiser, in order to form a set of 4-bit video symbols $\mathbf{v}$, as portrayed in Figure 1.

Due to the allocation of FLCs based on their probabilities of occurrence, these symbols have a reduced entropy, allowing an RVLC having a high coding rate to be designed. The RVLC employed consists of 16 possible codewords {00, 101, 0110, 1111, 010, 01110, 011110, 0111110, 1001, 11011, 100001, 110011, 10001, 1100011, 1000001, 1110111}. These were selected based on the probability of occurrence of the 16 possible values of video symbols in $\mathbf{v}$, as observed for the training video sequence set, and to give a free distance of two.

The average length of the bit string $\mathbf{w}$, following the adjustment of the feedback parameters N_a and N_b to give $\mathbf{u}$ with a length of 2048 bits for the 'Lab' video sequence, was found to be 1897 coded video bits. The corresponding average RVLC rate is therefore $R_{vlc} = \frac{1897}{2048} = 0.926$.

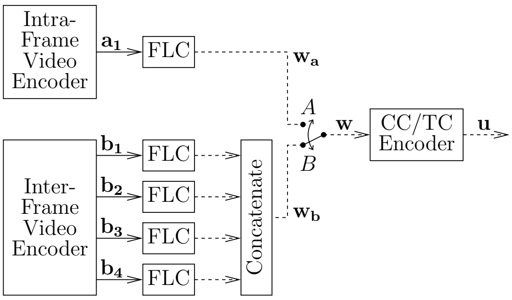

Figure 2: The proprietary CBR video encoder, which is concatenated with a convolutional codec or a convolutional-based turbo codec. Solid lines indicate a transfer of symbols, while dashed lines indicate a transfer of bit strings.

CC	TC
G_0: 1 1 0	G_0: 1 0 0 0 0 0
G_1: 1 0 1	G_1: 0 0 1 0 0 0
1 = transmitted bit	
0 = non transmitted bit	

Table 1: Puncturing pattern applied to the coded bits of the $R_{cc} = \frac{1}{2}$ CC and $R_{tc} = \frac{1}{3}$ TC in order to obtain a code rate of $\frac{3}{4}$ for the CC and TC, respectively.

In the second and third of the three candidate CBR videophone schemes, CC or TC codecs are employed instead of the RVLC codec for the sake of investigating whether they are capable of providing an improved error resilience. Unlike the RVLC, these codecs employ a fixed coding rate of $R_{cc} = R_{tc} = \frac{3}{4}$. In order to obtain $R_{cc} = \frac{3}{4}$ convolutional codes, we have employed the puncturing pattern used by Yasuda *et al.* [8]. For the sake of arriving at $R_{tc} = \frac{3}{4}$ TC, we have applied regular puncturing to the turbo codes. This puncturing pattern has been determined experimentally. For procedures on designing optimum high rate turbo codes via puncturing, the interested reader is referred to [9]. These puncturing patterns are summarised in Table 1.

Since in these systems a fixed coding rate of $\frac{3}{4}$ is employed, there is no uncertainty about the length of $\mathbf{w}$ required to give $\mathbf{u}$ having the desired length. Therefore, fixed values of N_a and N_b are employed. For this reason, the 'Frame Size Analyzer' and feedback mechanisms of Figure 1 are unnecessary in the specific scheme used here. This simplified video encoder is depicted in Figure 2. Note that since N_a and N_b are fixed, there is no need to convey their specific values to the decoder and hence the symbol sets $\mathbf{a_2}$ and $\mathbf{b_5}$ are not generated by the video encoder used. Each encoded video frame $\mathbf{w}$ is represented by 1876 bits, giving a similar reconstruction quality to that of the RVLC-protected scheme. The $\frac{3}{4}$-rate code is invoked for generating $\mathbf{u}$ having a length of 2502 bits.

For the RVLC protected system, $\mathbf{u}$ is interleaved by a 2048-bit channel interleaver π_c and subsequently directed to a $T_x = 2$-transmitter, $n = 4$-state and $M = 4$-PSK STTC encoder. For the sake of employing the same interleaver length and allowing a fair comparison to be made, the 2502-bit strings $\mathbf{u}$ of the CC- and TC-based schemes are concatenated and divided into 2048-bit partitions, before interleaving.

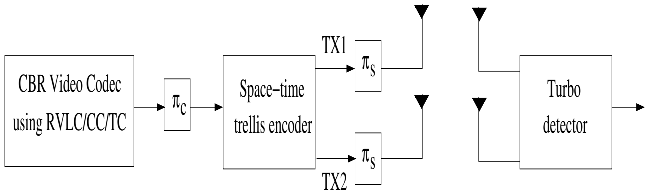

Figure 3: Schematic of the RVLC/CC/TC-coded and concatenated space-time trellis coded CBR video system employing turbo detection at the receiver. The notations π_c and π_s represent the channel interleaver and space-time trellis coding interleaver, respectively.

At the output of the STTC encoder, the encoded symbols are interleaved by a 1025-symbol random STTC interleaver represented by π_s in Figure 3. Note that random interleavers are utilised for both the bit-based channel interleaver and for the symbol-based STTC interleaver. As mentioned before, the same interleaving rule is used for all transmit antennas in order to preserve the rank property of the STTCs [1]. The 1025 STTC-deinterleaved symbols are mapped to five transmission frames each, having 205 data symbols.

Additionally, a dispersive two-path Rayleigh fading channel having equal symbol-spaced tap weights was used. The Rayleigh fading statistics obeyed a normalised Doppler frequency of 10^{-5}, where the fading magnitude and phase was kept constant for the duration of a transmission burst. At the receiver, the channel impulse response was assumed to be perfectly estimated.

As shown in Figure 3, $R_x = 2$ receive antennas are employed at the receiver. Turbo detection is used, whereby the channel equalisation, STTC decoding and RVLC, CC or TC decoding operations are performed iteratively. All three receiver components employed the Max-Log MAP [10, 11] algorithm. Note that for the TC-based scheme only a single inner turbo decoding iteration was performed before passing the soft information back to the outer turbo loop. The Signal-to-Noise Ratio (SNR) of the investigated system is defined as the ratio of the average received power to the noise power per receiver antenna. Since the long-term average received power is equal to the long-term average transmitted power, the SNR is $E[||s^k(t)||^2]/\sigma^2$, where σ^2 is the variance of the noise component $n^i(t)$. At the receiver the 4-bit RVLC symbol estimates $\hat{\mathbf{v}}$ are provided at the output of the RVLC decoder. These RVLC symbol estimates are then de-packetised and intra- or inter-frame video decoded, as appropriate. Having described the system's architecture, let us now analyse the performance of the RVLC-, CC- and TC-based schemes.

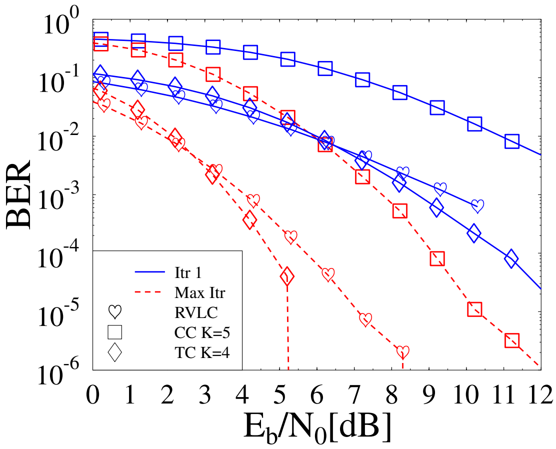

Figure 4: BER performance of the RVLC-, CC- and TC-coded STTC CBR videophone systems communicating over a dispersive two-path equal-power and symbol-spaced wideband Rayleigh fading channel, when employing turbo detection at the receiver. The term 'Max Itr' is used for indicating the number of turbo detection iterations that have to be performed such that no substantial further iteration gains were achieved using additional iterations. In the case of the RVLC, CC and TC schemes, this was three, four and four, respectively.

3. RESULTS AND DISCUSSION

Having described the turbo-detected videophone transceiver, let us now consider its performance. This was characterised using 20 repetitions of the 100-frame Lab video sequence's transmission in each simulation. The 'head-and-shoulders' Lab video sequence exhibits a moderate level of motion activity and it was not used in the video codec's training set.

As will be shown later, the number of turbo detection iterations that had to be performed such that no substantial further iteration gains were achieved using additional iterations, for the RVLC, CC and TC schemes was three, four and four respectively. The Bit Error Ratio (BER) performance after these numbers of decoding iterations was compared to that after a single decoding iteration, as shown in Figure 4. The latter scenario is equivalent to performing channel equalisation, STTC decoding and RVLC decoding separately.

The E_b/N_0 measure is employed in Figure 4 to compensate for the differing system throughputs and is defined as:

$$\frac{E_b}{N_0} = 10\log\left[\frac{1}{R \times \log_2(M)}\right] + \text{SNR}, \qquad (1)$$

where $R = R_{vlc},\ R_{cc}$ and R_{tc} for the RVLC, CC and TC protected systems, respectively. Additionally, for 4PSK, we have $M = 4$.

After the above-mentioned number of turbo detection iterations, similar gains of approximately 5.5 dB, 6 dB and 5 dB were achieved by performing iterative decoding at a BER of 10^{-3} for the RVLC-, CC- and TC-based systems, respectively.

At a BER of 10^{-3} and after the stated number of decoding iterations, the TC system (requiring an E_b/N_0 of 3.6 dB) outperforms the RVLC system (requiring an E_b/N_0 of 4 dB) by 0.4 dB, which in turn outperforms the CC system (requiring an E_b/N_0 of 7.8 dB) by 3.8 dB. Although the TC system using four decoding iterations outperforms the RVLC system employing three decoding iterations, this is achieved at a 33% higher complexity and a 22% higher bandwidth requirement. This is because $\mathbf{u}$ has a length of 2502 bits compared to the RVLC system's vector length of 2048 bits.

Figure 5 shows the variation of the 'Lab' sequence's objective reconstruction quality as a function of E_b/N_0 after various numbers of turbo detection iterations. The measure of objective video reconstruction quality was the average Peak Signal to Noise Ratio (PSNR) [7, Equation 12.8]. Each average PSNR value was calculated across the 20 repetitions of frames 26 to 100 of the Lab sequence's reconstruction. These frames were generated during the video codec's steady-state response following the codec's initial 'warm-up' phase, during which the relatively poor reconstruction quality of the intra-frame coded first video frame is gradually improved.

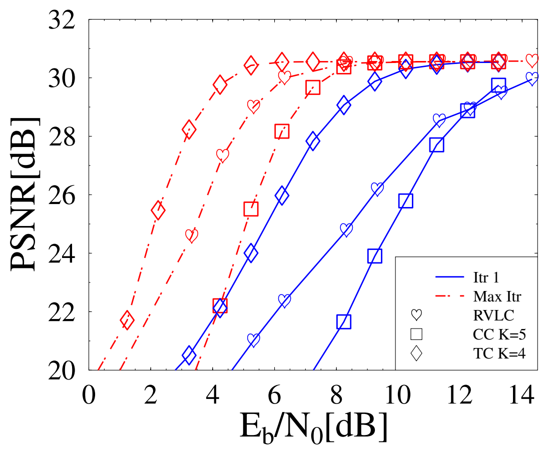

Figure 5: Variation of the Lab sequence's objective video reconstruction quality versus E_b/N_0 after various numbers of turbo detection iterations.

For high values of E_b/N_0, perfect error-free reconstructions of the 'Lab' sequence were obtained. These were associated with a PSNR of 30.6 dB and had a pleasing subjective reconstruction quality. With the subjective analysis of the reconstructed video sequences, acceptable video quality associated with no annoying virtual artifacts was obtained for PSNR values in excess of 30 dB. After the above-mentioned number of turbo detection iterations, gains of approximately 8 dB, 6 dB and 5 dB were achieved by performing iterative decoding at a PSNR of 30 dB for the RVLC-, CC- and TC-based systems, respectively.

At a PSNR of 30 dB and after the stated number of decoding iterations, the TC system (requiring an E_b/N_0 of 4.6 dB) outperforms the RVLC system (requiring an E_b/N_0 of 6.4 dB) by 1.8 dB, which outperforms the CC system (requiring an E_b/N_0 of 7.7 dB) by 1.3 dB. Recall that, although the TC system using four decoding iterations outperforms the RVLC system employing three decoding iterations, this is achieved at a higher complexity and bandwidth requirement. For lower E_b/N_0 values the subjective video reconstruction quality was seen to rapidly deteriorate.

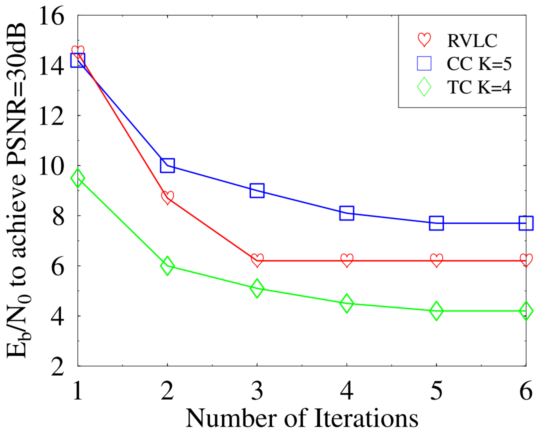

Figure 6: The E_b/N_0 ratio required at PSNR= 30 dB by the different complexity turbo detectors characterised in terms of the number of decoding iterations.

In Figure 6 we compare the relative performance of the three considered schemes at equal system complexities. Figure 6 depicts the E_b/N_0 value required by the RVLC-, TC- and CC-based videophone systems, in order to achieve a PSNR of 30 dB versus the complexity incurred, which is expressed in terms of the number of decoding iterations employed. As was asserted above, Figure 6 shows that only insignificant gains are achieved after more than three decoding iterations for the RVLC-based scheme and after four decoding iterations for the CC- and TC-based schemes. For all complexities, it is shown that the TC-based scheme outperforms the RVLC-based scheme, with E_b/N_0 gains of 2 dB being typical. Similarly, the RVLC-based scheme is seen to outperform the CC-based scheme at all complexities with similar gains being typical.

4. CONCLUSIONS

We have characterised the achievable performance of a STTC-based CBR videophone system employing $R = 0.926$ RVLC, $R_{cc} = \frac{3}{4}$ and $K = 5$ CC as well as an $R_{tc} = \frac{3}{4}$ and $K = 4$ TC. These systems were designed to give similar error-free video reconstruction qualities, which were found to be subjectively pleasing with a PSNR of 30.6 dB, and to have similar decoding complexities per decoding iteration. These design criteria were achieved by employing differing transmission throughputs, with the CC- and TC-based systems having a 22% higher bandwidth requirement.

It was observed that the RVLC-based scheme was capable of achieving its full potential after just three decoding iterations, therefore having a lower overall complexity than either of the CC- or TC-based schemes, which required four decoding iterations. These latter schemes were shown to have a 33% higher complexity than the RVLC-based scheme, if their full potential is to be realised. The adoption of iterative turbo decoding techniques yielded significant performance gains of over 5 dB in all systems.

The TC-based scheme was shown to outperform the RVLC-based scheme at all values of E_b/N_0 and at all complexities, with gains of 2 dB being typical. Likewise, the RVLC-based scheme was shown to outperform the CC-based scheme, with similar gains being typical. In situations where error-resilience is paramount, the TC-based system is preferred. However, as stated above, this system is associated with a high complexity and bandwidth requirement. In cases where these must be minimised, the use of the RVLC-based system is justified.

5. REFERENCES

[1] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, March 1998.

[2] V. Buttigieg and P. G. Farrell, "Variable-Length Error-Correcting Codes," *IEE Proceedings of Communications*, vol. 147, pp. 211–215, August 2000.

[3] C. Berrou and A. Glavieux, "Near optimum error-correcting coding and decoding: Turbo codes," *IEEE Transactions on Communications*, vol. 44, pp. 1261–1271, October 1996.

[4] L. Hanzo, T. H. Liew, and B. L. Yeap, *Turbo Coding, Turbo Equalisation and Space-Time Coding for Transmission over Fading Channels*. John-Wiley IEEE Press, 2002.

[5] C. Douillard, A. Picart, M. Jézéquel, P. Didier, C. Berrou, and A. Glavieux, "Iterative Correction of Intersymbol Interference: Turbo-Equalization," *European Transactions on Communications*, vol. 6, pp. 507–511, September-October 1995.

[6] B. L. Yeap, T. H. Liew, and L. Hanzo, "Reduced Complexity Turbo Equalization for Space-time Trellis Coded Systems," in *Proceedings of the IEEE Vehicular Technology Conference Spring*, (Birmingham, Alabama), May 6 - 9 2002.

[7] L. Hanzo, P. Cherriman, and J. Streit, *Wireless Video Communications: Second to Third Generation Systems and Beyond*. IEEE Press, 2001.

[8] Y. Yasuda, K. Kashiki, and Y. Hirata, "High-rate Punctured Convolutional Codes for Soft Decision Viterbi Decoding," *IEEE Transactions on Communications*, vol. COM-32, pp. 315–319, March 1984.

[9] Ömer. F. Açikel and W. E. Ryan, "Punctured turbo-codes for BPSK/QPSK channels," *IEEE Transactions on Communications*, vol. 47, pp. 1315–1323, September 1999.

[10] W. Koch and A. Baier, "Optimum and Sub-Optimum Detection of Coded Data Disturbed by Time-Varying Inter-Symbol Interference," in *Proceedings of the IEEE Global Telecommunications Conference 1990*, (San Diego, United States), pp. 1679–1684, 2-5 December 1990.

[11] J. A. Erfanian, S. Pasupathy, and G. Gulak, "Reduced Complexity Symbol Detectors with Parallel Structures for ISI Channels," *IEEE Transactions on Communications*, vol. 42, pp. 1661–1671, February/March/April 1994.