

Turbo-Detected Unequal Protection MPEG-4 Wireless Video Telephony using Trellis Coded Modulation and Space-Time Trellis Coding

S. X. Ng, J. Y. Chung and L. Hanzo

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

Tel: +44-23-8059 3125, Fax: +44-23-8059 4508

Email: {sxn,jyc00r,lh}@ecs.soton.ac.uk, http://www-mobile.ecs.soton.ac.uk

Abstract – A jointly optimised turbo transceiver capable of providing unequal error protection is proposed for MPEG-4 aided wireless video telephony. The transceiver advocated consists of Space-Time Trellis Coding (STTC), Trellis Coded Modulation (TCM) and two different-rate Non-Systematic Convolutional codes (NSCs). A benchmark scheme combining STTC and NSC is used for comparison with the proposed scheme. The video performance of the both schemes is evaluated when communicating over uncorrelated Rayleigh fading channels. It was found that the proposed scheme requires about two dBs lower transmit power than the benchmarker scheme in the context of the MPEG-4 videophone transceiver, when aiming for an effective throughput of 2 bits/symbol at a similar decoding complexity.

1. MOTIVATION AND BACKGROUND

The MPEG-4 standard [1, 2] offers a standardised framework for a whole range of multimedia applications. Examples include tele-shopping, interactive TV, internet games, or mobile video communication. MPEG-4 integrates different types of multimedia services by the introduction of a so-called object-based approach for the description and coding of multimedia contents. The key functionalities of MPEG-4 include independent coding of objects in a video frame, the ability to interactively embed these video objects into a scene shown on the screen, the transmission of 3D scene descriptions, quality versus bitrate based temporal and spatial scalability and improved error resilience [3].

As the MPEG-4 standard targets a broader range of different applications and bitrates than the previously defined hybrid video coding standards such as MPEG-1, 2 or H.263, it employs a higher variety of different algorithms and coding modes. In the MPEG-4 coding algorithm a scene consists of one or more audio-visual objects potentially generated from multiple sources.

The MPEG-4 algorithm employed for encoding natural video scenes is the classic block-based hybrid coding scheme [4],

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

which is known from the well-established MPEG-1, 2 or H.263 codecs. However, these codecs were further developed in order to allow the encoding of arbitrarily shaped video objects. For employment in error-prone environments, error resilient encoding features were introduced by several parts of the MPEG-4 standards. This renders the MPEG-4 coding standard particularly suitable for wireless video telephony.

In this study the MPEG-4 video codec was incorporated in a sophisticated unequal-protection turbo transceiver using joint coding and modulation as inner coding, twin-class convolutional outer coding as well as space time coding based spatial diversity. Specifically, Trellis Coded Modulation (TCM) [5, 6] constitutes a bandwidth-efficient joint channel coding and modulation scheme, which was originally designed for transmission over Additive White Gaussian Noise (AWGN) channels. In an effort to mitigate the effects of Rayleigh fading channels and render them Gaussian-like for the sake of supporting the operation of the TCM scheme, Space-Time Trellis Coding (STTC) [7] employing multiple transmit and receive antennas was invoked for providing spatial diversity gain. Furthermore, maximal minimum distance Non-Systematic Convolutional codes (NSCs) [8, p. 331] having two different code-rates were used for providing unequal video protection.

Again, in this contribution, a novel unequal-protection joint source-coding, channel-coding, modulation and spatial diversity aided turbo-transceiver is proposed, which consists of a STTC, a TCM and two different-rate NSCs. This STTC-TCM-2NSC scheme is proposed for MPEG-4 video telephony. We will demonstrate that significant iteration gains are attained with the aid of the proposed turbo transceiver. The paper is structured as follows. In Section 2 we describe the proposed system's architecture and highlight the interactions of its constituent elements. We elaborate further by characterising the achievable system performance in Section 3 and conclude with a range of system design guideline in Section 4.

2. THE TURBO TRANSCEIVER

The block diagram of the serially concatenated STTC-TCM-2NSC turbo scheme using a STTC, a TCM and two RSCs as its constituent codes is depicted in Figure 1. The MPEG-4 codec operated at $R_f=30$ frames per second using the $(176 \times$

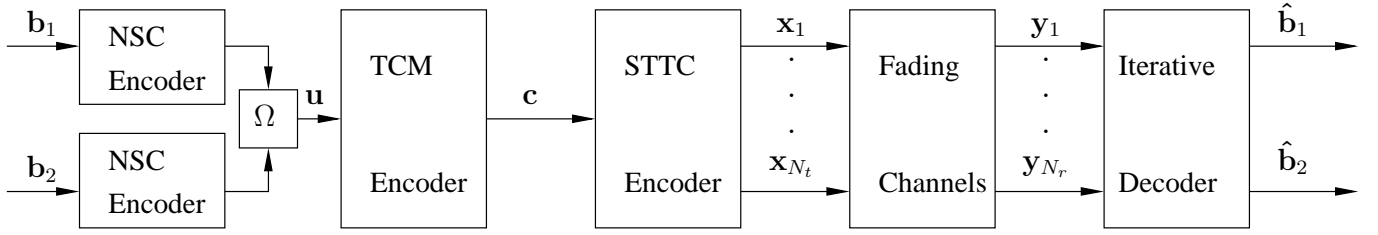


Figure 1: Block diagram of the serially concatenated STTC-TCM-2NSC scheme. The notations \mathbf{b}_i , $\hat{\mathbf{b}}_i$, \mathbf{u} , \mathbf{c} , \mathbf{x}_j and \mathbf{y}_k denote the vectors of the class- i video bits, the estimates of the class- i video bits, the resultant encoded bits of both NSC encoders, the TCM coded symbols, the STTC coded symbols for transmitter j and the received symbols at receiver k , respectively. Furthermore, Ω is a parallel-to-serial converter, while N_t and N_r denote the number of transmitters and receivers, respectively. The symbol-based channel interleaver between the STTC and TCM schemes as well as the two bit-based interleavers at the output of NSC encoders are not shown for simplicity. The iterative decoder seen at the right is detailed in Figure 2.

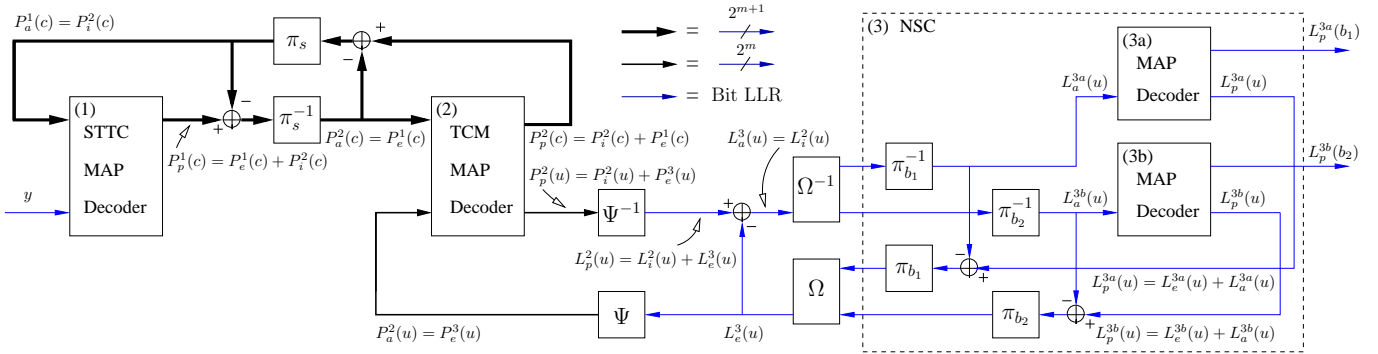


Figure 2: Block diagram of the STTC-TCM-2NSC turbo detection scheme seen at the right of Figure 1. The notations $\pi_{(s,b_i)}$ and $\pi_{(s,b_i)}^{-1}$ denote the interleaver and deinterleaver, while the subscript s denotes the symbol-based interleaver of TCM and the subscript b_i denotes the bit-based interleaver for class- i NSC. Furthermore, Ψ and Ψ^{-1} denote LLR-to-symbol probability and symbol probability-to-LLR conversion, while Ω and Ω^{-1} denote the parallel-to-serial and serial-to-parallel converter, respectively. The notation m denotes the number of information bits per TCM coded symbol. The thickness of the connecting lines indicates the number of non-binary symbol probabilities spanning from a single LLR per bit to 2^m and 2^{m+1} probabilities.

144)-pixel Quarter Common Intermediate Format Miss America video sequence, encoded at a near-constant bitrate of $R_b=69$ kbps. Hence, we have $R_b/R_f = 2300$ bits per video frame. We partition the video bits into two unequal protection classes. Specifically, class-1 and class-2 consist of 25% (which is 575 bits) and 75% (which is 1725 bits) of the total number of video bits. The more sensitive video bits constituted mainly by the MPEG-4 framing and synchronisation bits are in class-1 and they are protected by a stronger binary NSC having a coding rate of $R_1 = k_1/n_1 = 1/2$ and a code memory of $L_1 = 3$. The less sensitive video bits predominantly signalling the MPEG-4 Discrete Cosine Transform (DCT) coefficients and motion vectors are in class-2 and they are protected by a weaker non-binary NSC having a coding rate of $R_2 = k_2/n_2 = 3/4$ and a code memory of $L_2 = 3$. Note that the number of MPEG-4 framing and synchronisation bits is only about 10% of the total video bits. Hence, about 25%-10%=15% of class-1 bits are constituted by the video bits signalling the most sensitive MPEG-4 DCT coefficients. We invoke code termination bits in both NSCs and hence the number of coded bits emerging from

the $R_1 = 1/2$ binary NSC encoder is $(575 + k_1 L_1)/R_1 = 1156$ bits, while that generated by the $R_2 = 3/4$ non-binary NSC encoder is $(1725 + k_2 L_2)/R_1 = 2312$ bits. The class-1 and class-2 NSC coded bit sequences are interleaved by two separate bit interleavers of length 1156 and 2312 bits, respectively.

The two interleaved bit sequences are then concatenated in the parallel-to-serial converter block, denoted by Ω in Figure 1 to form a bit sequence of $1156 + 2312 = 3468$ bits. This bit sequence is then fed to the TCM encoder having a coding rate of $R_3 = k_3/n_3 = 3/4$ and a code memory of $L_4 = 3$. We employ code termination also in the TCM scheme and hence at the TCM encoder's output we have $(3468 + k_3 L_3)/R_3 = 4636$ bits or $4636/4=1159$ TCM symbols. The TCM symbol sequence is then symbol-interleaved and fed to the STTC encoder. We invoke a 16-state STTC scheme having a code memory of $L_4 = 4$ and $N_t = 2$ transmit antennas, employing $M = 16$ -level Quadrature Amplitude Modulation (16QAM). We terminate the STTC code by a 4-bit 16QAM symbol, since we have $N_t = 2$. Therefore, at each transmit antenna we have

1159 + 1 = 1160 16QAM symbols or $4 \times 1160 = 4640$ bits in a transmission frame. The overall coding rate is given by $R = 2300/4640 \approx 0.5$ and the effective throughput of the system is $\log_2(M)R \approx 2.0$ Bits Per Symbol (BPS).

At the receiver, we employ $N_r = 2$ receive antennas and the received signals are fed to the iterative decoders for the sake of estimating the video bit sequences in both class-1 and class-2, as seen in Figure 1. The STTC-TCM-2NSC scheme's turbo decoder structure is illustrated in Figure 2, where there are four constituent decoders, each labelled with a round-bracketed index. Symbol-based and bit-based MAP algorithms [6] operating in the logarithmic-domain are employed by the TCM as well as the rate $R_2 = 3/4$ NSC decoders and by the $R_1 = 1/2$ NSC decoder, respectively. The notations $P(\cdot)$ and $L(\cdot)$ in Figure 2 denote the logarithmic-domain symbol probabilities and the Logarithmic-Likelihood Ratio (LLR) of the bit probabilities, respectively. The notations c , u and b_i in the round brackets (\cdot) in Figure 2 denote TCM coded symbols, TCM information symbols and the class- i video bits, respectively. The specific nature of the probabilities and LLRs is represented by the subscripts a , p , e and i , which denote *a priori*, *a posteriori*, *extrinsic* and *intrinsic* information, respectively. The probabilities and LLRs associated with one of the four constituent decoders having a label of $\{1, 2, 3a, 3b\}$ are differentiated by the identical superscripts of $\{1, 2, 3a, 3b\}$. Note that the superscript 3 is used for representing the two NSC decoders of $3a$ and $3b$.

As we can observe from Figure 2, the STTC decoder of block 1 benefits from the *a priori* information provided by the TCM decoder of block 2, namely $P_a^1(c) = P_i^2(c)$ regarding the 2^{m+1} -ary TCM coded symbols, where m is the number of information bits per TCM coded symbol. More specifically, $P_i^2(c)$ is referred to as the intrinsic probability of the 2^{m+1} -ary TCM coded symbols, because it contains the inseparable extrinsic information provided by the TCM decoder itself as well as the *a priori* information regarding the uncoded 2^m -ary TCM input information symbols emerging from the NSC decoders of block 3, namely $P_a^2(u) = P_e^3(u)$. Hence, the STTC decoder indirectly also benefits from the *a priori* information $P_a^2(u) = P_e^3(u)$ provided by the NSC decoders of block 3, potentially enhanced by the TCM decoder of block 2. Similarly, the intrinsic probability of $P_i^2(u)$ provided by the TCM decoder for the sake of the NSC decoders' benefit consists of the inseparable extrinsic information generated by the TCM decoder itself as well as of the systematic information of the STTC decoder, namely $P_a^2(c) = P_e^1(c)$. Note that after the symbol probability-to-LLR conversion, $P_i^2(u)$ becomes $L_i^2(u)$. Therefore, the NSC decoders of block 3 benefit directly from the *a priori* information provided by the TCM decoder of block 2, namely from $L_a^3(u) = L_i^2(u)$ as well as indirectly from the *a priori* information provided by the STTC decoder of block 1, namely from $P_a^2(c) = P_e^1(c)$. On the other hand, the TCM decoder benefits directly from the STTC and NSC decoders through the *a priori* information of $P_a^2(c) = P_e^1(c)$

and $P_a^2(u) = P_e^3(u)$, respectively, as it is shown in Figure 2.

2.1. The Turbo Benchmarker

In order to evaluate the proposed scheme, we created a powerful benchmarker scheme by replacing the TCM and NSC encoders of Figure 1 by a single NSC codec having a coding rate of $R_0 = k_0/n_0 = 1/2$ and a code memory of $L_0 = 6$. We will refer to this benchmarker scheme as the STTC-NSC arrangement. All video bits are equally protected in the benchmarker scheme by a single NSC encoder and a STTC encoder. A bit-based channel interleaver is inserted between the NSC encoder and STTC encoder. Taking into account the bits required for code termination, the number of output bits of the NSC encoder is $(2300 + k_0L_0)/R_0 = 4612$, which corresponds to 1153 16QAM symbols. Again, a 16-state STTC scheme having $N_t = 2$ transmit antennas is employed. After code termination, we have $1153 + 1 = 1154$ 16QAM symbols or $4(1154) = 4616$ bits in a transmission frame at each transmit antenna. The overall coding rate is given by $R = 2300/4616 \approx 0.5$ and the effective throughput is $\log_2(16)R \approx 2.0$ BPS, both of which are identical to the corresponding values of the proposed scheme. The decoder structure of the STTC-NSC benchmarker scheme is similar to that of Figure 2, replacing the TCM and NSC decoders of block 2 and 3 by a single NSC decoder.

2.2. Complexity

Let us define a single decoding iteration as a combination of a STTC decoding, a TCM decoding, a class-1 NSC decoding and a class-2 NSC decoding step for the proposed STTC-TCM-2NSC scheme. Similarly, a decoding iteration of the STTC-NSC benchmarker scheme is comprised of a STTC decoding and a NSC decoding step. We will quantify the decoding complexity of the proposed STTC-TCM-2NSC scheme and that of the benchmarker scheme using the number of decoding trellis states. The total number of decoding trellis states per iteration of the proposed scheme employing 2 NSC decoders having a code memory of $L_1 = L_2 = 3$, TCM having $L_3 = 3$ and STTC having $L_4 = 4$, is $S = 2^{L_1} + 2^{L_2} + 2^{L_3} + 2^{L_4} = 40$. By contrast, the total number of decoding trellis states per iteration for the benchmarker scheme having a code memory of $L_0 = 6$ and STTC having $L_4 = 4$, is given by $S = 2^{L_0} + 2^{L_4} = 80$. Therefore, the complexity of the proposed STTC-TCM-2NSC scheme having two iterations is equivalent to that of the benchmarker scheme having a single iteration, which corresponds to 80 decoding states. Similarly, the complexity of the STTC-TCM-2NSC scheme having four iterations is equivalent to that of the benchmarker scheme having two iterations, which corresponds to 160 decoding states.

3. SIMULATION RESULTS

In this section we evaluate the performance of the proposed MPEG-4 based video telephone schemes using both the Bit Error Ratio (BER) and the average video Peak Signal to Noise Ratio (PSNR) [4]. Figures 3 and 4 depict the BER versus Signal to Noise Ratio (SNR) per bit, namely E_b/N_0 , performance of the proposed 16QAM-based STTC-TCM-2NSC scheme and that of the STTC-NSC benchmarker scheme, respectively, when communicating over uncorrelated Rayleigh fading channels. Furthermore, Figures 5 and 6 portray the PSNR versus E_b/N_0 performance of the proposed 16QAM-based STTC-TCM-2NSC scheme and that of the STTC-NSC benchmarker scheme, respectively, for transmissions over uncorrelated Rayleigh fading channels. Again, the overall throughput of both systems is 2 BPS.

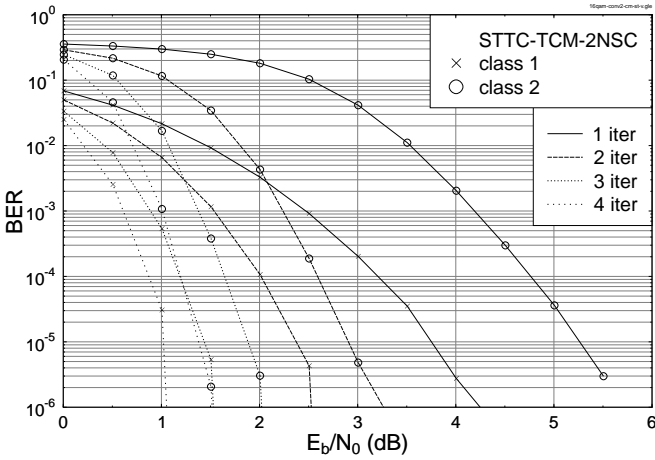


Figure 3: BER versus E_b/N_0 performance of the proposed 16QAM-based STTC-TCM-2NSC assisted MPEG-4 scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was 2 BPS.

It can be seen from Figure 3 that the class-1 video bits have a lower BER compared to the class-2 video bits. Specifically, at a BER of 10^{-4} , the class-1 bits can tolerate a 1.5 dB lower E_b/N_0 value at the first decoding iteration. Viewing this unequal-protection scheme from a different perspective, the class-1 bits benefit from more than an order of magnitude lower BER at a given SNR, than the class-2 bits. However, the E_b/N_0 difference of the two classes reduces to 0.5 dB at the fourth iteration. This is due to the different iteration gains attained by the two different bit classes. Explicitly, at a BER of 10^{-4} the iteration gains of the class-1 bits and class-2 bits are approximately 2.4 dB and 3.5 dB, respectively, when the number of iterations is increased from one to four. It is demonstrated in Figure 3 that the class-1 and class-2 bits required an extremely low E_b/N_0 of approximately 0.8 dB and 1.3 dB at BER= 10^{-4} in order to attain a throughput of 2 BPS. By contrast, the STTC-NSC benchmarker scheme required an E_b/N_0 of approximately 4.4 dB and 3.3 dB at the first and second it-

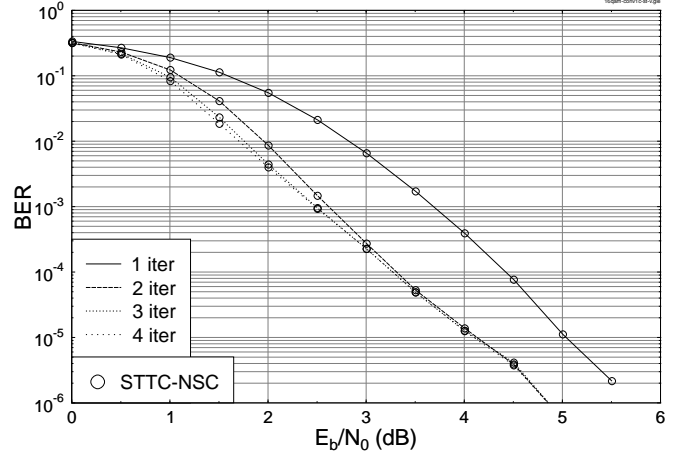


Figure 4: BER versus E_b/N_0 performance of the 16QAM-based STTC-NSC assisted MPEG-4 benchmarker scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was 2 BPS.

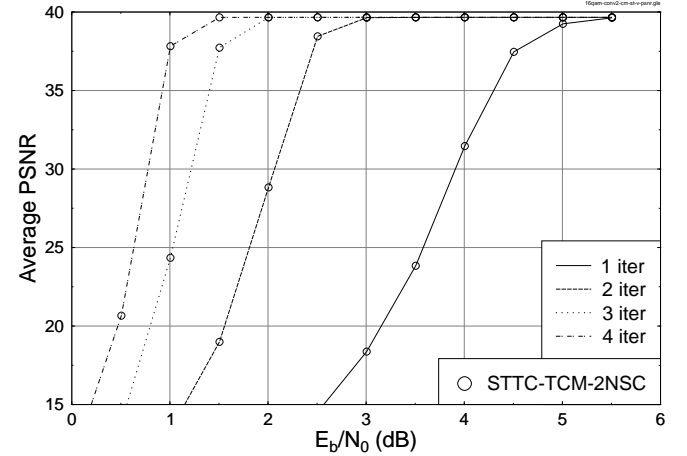


Figure 5: Average PSNR versus E_b/N_0 performance of the proposed 16QAM-based STTC-TCM-2NSC assisted MPEG-4 scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was 2 BPS.

eration, respectively. Further marginal gains of about 0.1 dB were attained, when the number of iterations was increased to three or four.

Let us now consider the PSNR versus E_b/N_0 performance of the systems characterised in Figures 5 and 6. It is evidenced in Figure 5 that similar to our observations made in the context of the achievable BER results, an approximately 3.5 dB of iteration gain was attained by the proposed STTC-TCM-2NSC scheme, when the number iterations was increased from one to four at a PSNR of 37.5 dB. Again, the complexity of STTC-TCM-2NSC having four iterations corresponds to 160 trellis states, which is similar to that of the STTC-NSC scheme having two iterations. As shown in Figure 6, at PSNR=37.5 dB the STTC-NSC benchmarker scheme having two iterations re-

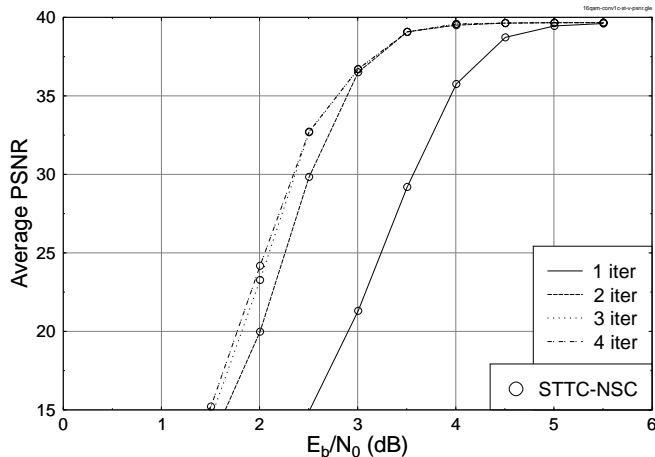


Figure 6: Average PSNR versus E_b/N_0 performance of the 16QAM-based STTC-NSC assisted MPEG-4 benchmarker scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was **2 BPS**.

quired an E_b/N_0 value of 3.2 dB, which is about 2.2 dB higher than that required by the similar-complexity STTC-TCM-2NSC arrangement having four iterations, as shown in Figure 5. Observe in Figure 6 that the STTC-NSC scheme only attains a significant iteration gain during the second iteration. Further increasing the number of iterations results in a higher decoding complexity, but attains no noteworthy iteration gain. Hence, the best possible attainable performance is PSNR=37.5 dB at $E_b/N_0=3.2$ dB. Note that if we reduce the code memory of the NSC constituent code of the STTC-NSC benchmarker scheme from $L_0=6$ to 3, the best possible performance becomes poorer. If we increased L_0 from 6 to 7 (or higher), the decoding complexity increased significantly, while the attainable best possible performance is only marginally increased. Hence, the STTC-NSC scheme having $L_0=6$ constitutes a good benchmarker scheme in terms of its performance versus complexity tradeoffs.

4. CONCLUSIONS

In conclusion, a jointly optimised source-coding, outer channel-coding, inner coded modulation and spatial diversity aided turbo transceiver was proposed for MPEG-4 wireless video telephony. With the aid of two different-rate NSCs the video bits were protected differently according to their sensitivity. The employment of TCM improved the bandwidth efficiency of the system and by utilising STTC spatial diversity was attained. The performance of the proposed STTC-TCM-2NSC scheme was enhanced with the advent of an efficient iterative decoding structure. It was shown in Figure 5 that the proposed STTC-TCM-2NSC scheme required $E_b/N_0=1$ dB in order to attain a PSNR of 37.5 dB. At the cost of a similar complexity, the STTC-TCM-2NSC scheme is approximately 2.2 dB more efficient in terms of the required E_b/N_0 than the STTC-NSC

benchmarker scheme.

5. REFERENCES

- [1] ISO/IEC JTC1/SC29/WG11, "Information Technology - Generic coding of Audio-visual Objects.," in *Part 2: Visual. Draft ISO/IEC 14496-2 (MPEG-4), version 1, ISO/IEC*, (Geneva), 1998.
- [2] ISO/IEC JTC1/SC29/WG11 W2502 in *ISO/IEC 14496-2. Final Draft International Standard. Part 2: Visual*, (Atlantic City), 1998.
- [3] B. Haskell, A. Puri, and L. Rabiner, "Image and Video Coding-Emerging Standards and beyond," vol. 8, pp. 814–837, November 1998.
- [4] L. Hanzo, P.J. Cherriman and J. Street, *Wireless Video Communications: Second to Third Generation Systems and Beyond*. NJ, USA : IEEE Press., 2001.
- [5] G. Ungerböck, "Channel Coding with Multilevel/Phase Signals," *IEEE Transactions on Information Theory*, vol. 28, pp. 55–67, January 1982.
- [6] L. Hanzo, T. H. Liew and B. L. Yeap, *Turbo Coding, Turbo Equalisation and Space Time Coding for Transmission over Wireless channels*. New York, USA: John Wiley IEEE Press, 2002.
- [7] V. Tarokh, N. Seshadri and A. R. Calderbank, "Space-time Codes for High Rate Wireless Communication: Performance analysis and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, March 1998.
- [8] S. Lin and D. J. Costello, Jr, *Error Control Coding: Fundamentals and Applications*. Inc. Englewood Cliffs, New Jersey 07632: Prentice-Hall, 1983.