University of Southampton
Faculty of Engineering and Applied Sciences
School of Electronics and Computer Science

Wireless Speech and Audio Communications

by

Noor Shamsiah Othman
B.Eng., M.Sc.

A doctoral thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy at the University of Southampton

June 2008

SUPERVISORS:
Professor Lajos Hanzo
M.Sc., Ph.D, FREng, DSc, FIEEE, FIET
Chair of Telecommunications

Dr Soon Xin Ng (Michael)
BEng, PhD, MIEE, MIEEE

School of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ

This thesis is dedicated to:

my husband Mohd Fadzil

my little Khadijah and Muhammad

with all my love.

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND APPLIED SCIENCE
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

Wireless Speech and Audio Communications

by

Noor Shamsiah Othman

The limited applicability of Shannon's separation theorem in practical speech/audio systems motivates the employment of joint source and channel coding techniques. Thus, considerable efforts have been invested in designing various types of joint source and channel coding schemes. This thesis discusses two different types of Joint Source and Channel Coding (JSCC) schemes, namely Unequal Error Protection (UEP) aided turbo transceivers as well as Iterative Source and Channel Decoding (ISCD) exploiting the residual redundancy inherent in the source encoded parameters.

More specifically, in Chapter 2, two different UEP JSCC philosophies were designed for wireless audio and speech transmissions, namely a turbo-detected UEP scheme using twin-class convolutional codes and another turbo detector using more sophisticated Irregular Convolutional Codes (IRCC). In our investigations, the MPEG-4 Advanced Audio Coding (AAC), the MPEG-4 Transform-Domain Weighted Interleaved Vector Quantization (TwinVQ) and the Adaptive MultiRate WideBand (AMR-WB) audio/speech codecs were incorporated in the sophisticated UEP turbo transceiver, which consisted of a three-stage serially concatenated scheme constituted by Space-Time Trellis Coding (STTC), Trellis Coded Modulation (TCM) and two different-rate Non-Systematic Convolutional codes (NSCs) used for UEP. Explicitly, both the twin-class UEP turbo transceiver assisted MPEG-4 TwinVQ and the AMR-WB audio/speech schemes outperformed their corresponding single-class audio/speech benchmarkers by approximately 0.5 dB, in terms of the required E_b/N_0, when communicating over uncorrelated Rayleigh fading channels. By contrast, when employing the MPEG-4 AAC audio codec and protecting the class-1 audio bits using a 2/3-rate NSC code, a more substantial E_b/N_0 gain of about 2 dB was achieved. As a further design alternative, we also proposed a turbo transceiver employing IRCCs for the sake of providing UEP for the AMR-WB speech codec. The resultant UEP schemes exhibited a better performance when compared to the corresponding Equal Error Protection (EEP) benchmark schemes, since the former protected the audio/speech bits according to their sensitivity. The proposed UEP aided system using IRCCs exhibits

an E_b/N_0 gain of about 0.4 dB over the EEP system employing regular convolutional codes, when communicating over AWGN channels, at the point of tolerating a SegSNR degradation of 1 dB.

In Chapter 3, a novel system that invokes jointly optimised ISCD for enhancing the error resilience of the AMR-WB speech codec was proposed and investigated. The resultant AMR-WB coded speech signal is protected by a Recursive Systematic Convolutional (RSC) code and transmitted using a non-coherently detected Multiple-Input Multiple-Output (MIMO) Differential Space-Time Spreading (DSTS) scheme. To further enhance the attainable system performance and to maximise the coding advantage of the proposed transmission scheme, the system is also combined with multi-dimensional Sphere Packing (SP) modulation. The AMR-WB speech decoder was further developed for the sake of accepting the *a priori* information passed to it from the channel decoder as extrinsic information, where the residual redundancy inherent in the AMR-WB encoded parameters was exploited.

Moreover, the convergence behaviour of the proposed scheme was evaluated with the aid of both Three-Dimensional (3D) and Two-Dimenstional (2D) EXtrinsic Information Transfer (EXIT) charts. The proposed scheme benefitted from the exploitation of the residual redundancy inherent in the AMR-WB encoded parameters, where an approximately 0.5 dB E_b/N_0 gain was achieved in comparison to its corresponding hard speech decoding based counterpart. At the point of tolerating a SegSNR degradation of 1 dB, the advocated scheme exhibited an E_b/N_0 gain of about 1.0 dB in comparison to the benchmark scheme carrying out joint channel decoding and DSTS aided SP-demodulation in conjunction with separate AMR-WB decoding, when communicating over narrowband temporally correlated Rayleigh fading channels.

In Chapter 4, two jointly optimized ISCD schemes invoking the soft-output AMR-WB speech codec using DSTS assisted SP modulation were proposed. More specifically, the soft-bit assisted iterative AMR-WB decoder's convergence characteristics were further enhanced by using Over-Complete source-Mapping (OCM), as well as a recursive precoder. EXIT charts were used to analyse the convergence behaviour of the proposed turbo transceivers using the soft-bit assisted AMR-WB decoder.

Explicitly, the OCM aided AMR-WB MIMO transceiver exhibits an E_b/N_0 gain of about 3.0 dB in comparison to the benchmark scheme also using ISCD as well as DSTS aided SP-demodulation, but dispensing with the OCM scheme, when communicating over narrowband temporally correlated Rayleigh fading channels. Finally, the precoded soft-bit AMR-WB MIMO transceiver exhibits an E_b/N_0 gain of about 1.5 dB in comparison to the benchmark scheme dispensing with the precoder, when communicating over narrowband temporally correlated Rayleigh fading channels.

DECLARATION OF AUTHORSHIP

I, **Noor Shamsiah Othman**,
declare that the thesis entitled

Wireless Speech and Audio Communications

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:*(see list of publications)*

Signed:.....
Date:.....

Acknowledgements

I am very grateful to my supervisors Prof. Lajos Hanzo and Dr. Soon Xin Ng (Michael) for their generous guidance, kindness and patience. This research would not have progressed this far without them. All the help, advice, ideas and encouragement are greatly appreciated.

I would also like to thank all my colleagues in the Communications Group of University of Southampton, both past and present, for all their supports and helps throughout this project. Special thanks to Dr. Soon Xin Ng, Dr. Jin Wang, Mohammed El-Hajjar, Dr. Anh Pham Quang, Dr. Osamah Alamri, Dr. Jörg Kliewer and Nasruminallah for their technical support and collaborative work. The assistance provided by Denise Harvey and Rebecca Earl, the Communication Group secretaries, is gratefully acknowledged.

The financial support of University Tenaga Nasional Malaysia as well as that of the School of Electronics and Computer Science in the University of Southampton is gratefully acknowledged.

A special thanks also goes to all my friends in Southampton and elsewhere, for their constant kind wishes. There are two people that I would like to thank especially, Dr Rosmila Senik and Dr Azizah Mainal for their friendship and constant encouragement.

I would like to express my appreciation to my parents, my parents in law, my sisters and my brothers for their unconditional love and encouragement. My gratitude goes to my dearest husband Mohd Fadzil and my daughter Khadijah for their patience, sacrifices and encouragement, which motivates me to be steadfast and never bend to difficulties. Last but not least, my sincere gratitude to our newborn Muhammad. I owe my every achievement to all of them.

List of Publications

JOURNAL PAPERS:

1. **N. S. Othman**, M. El-Hajjar, O. Alamri and L. Hanzo "Iterative AMR-WB Source and Channel-Decoding Using Differential Space-Time Spreading Assisted Sphere Packing Modulation", to appear in IEEE Transactions on Vehicular Technology.

2. A. Q. Pham, L. -L. Yang, **N. S. Othman** and L. Hanzo "EXIT-Chart Optimized Block Codes for Wireless Video Telephony", to appear in IEEE Transactions on Circuits and Systems for Video Technology.

CONFERENCE PAPERS:

3. **N. S. Othman**, S. X. Ng and L. Hanzo, "Turbo-Detected Unequal Protection MPEG-4 Audio Transceiver Using Convolutional Codes, Trellis Coded Modulation and Space-Time Trellis Coding", IEEE 61st Vehicular Technology Conference, Stockholm, Sweden, 30 May-1 June 2005, pp. 1600 -1604.

4. **N. S. Othman**, S. X. Ng and L. Hanzo, "Turbo-Detected Unequal Protection Audio and Speech Transceivers Using Serially Concantenated Convolutional Codes, Trellis Coded Modulation and Space-Time Trellis Coding", IEEE 62nd Vehicular Technology Conference, Texas, USA, 25-28 September 2005, pp. 1044-1048.

5. J. Wang, **N. S. Othman**, J. Kliewer, L-L. Yang and L. Hanzo "Turbo Detected Unequal Error Protection General Configuration Irregular Convolutional Codes Designed for the Wideband Advanced Multirate Speech Codec", IEEE 62nd Vehicular Technology Conference, Texas, USA, 25-28 September 2005, pp. 1044-1048.

6. **N. S. Othman**, M. El-Hajjar, O. Alamri and L. Hanzo "Soft-Bit Assisted Iterative AMR-WB Source-Decoding and Turbo-Detection of Channel-Coded Differential Space-Time Spreading Using Sphere Packing Modulation", IEEE 65th Vehicular Technology Conference, Dublin, Ireland, 22-25 April 2007, pp. 2010-2014.

7. **N. S. Othman**, M. El-Hajjar, A. Q. Pham, O. Alamri, S. X. Ng and L. Hanzo "Over-Complete Source-Mapping Aided AMR-WB MIMO Transceiver Using Three-Stage Iterative Detection", IEEE International Conference on Communication, Beijing, China, 19-23 May 2008, pp. 751-755.

8. Nasruminallah, M. El-Hajjar, **N. S. Othman**, A. Q. Pham, O. Alamri and L. Hanzo "Over-Complete Mapping Aided, Soft-Bit Assisted Iterative Unequal Error

Protection H.264 Joint Source and Channel Decoding", to appear in Proceedings of IEEE 68th Vehicular Technology Conference, Calgary, Alberta, Canada, 21-24 September 2008.

9. **N. S. Othman**, M. El-Hajjar, O. Alamri, S. X. Ng and L. Hanzo "Three-Stage Iterative Detection of Precoded Soft-Bit AMR-WB for Speech MIMO Transceiver", to be submitted to IEEE 69th Vehicular Technology Conference, 2009.

BOOK CONTRIBUTIONS:

1. J. Wang, **N. S. Othman**, J. Kliewer, L-L. Yang and L. Hanzo "Section 9.6: Turbo-detected Unequal Error Protection Irregular Convolutional Coded AMR-WB Transceiver", L. Hanzo, C. Somerville and J. Woodard, "Voice and Audio Compression for Wireless Communications, 2nd Edition", Chichester, UK: John Wiley-Sons Inc., 2007, pp. 442-453.

2. **N. S. Othman**, S. X. Ng and L. Hanzo "Section 10.6: Turbo-detected Space-time Trellis Coded MPEG-4 Audio Transceivers", L. Hanzo, C. Somerville and J. Woodard, "Voice and Audio Compression for Wireless Communications, 2nd Edition", Chichester, UK: John Wiley-Sons Inc., 2007, pp. 516-524.

3. **N. S. Othman**, S. X. Ng and L. Hanzo "Section 10.7: Turbo-detected Space-time Trellis Coded MPEG-4 versus AMR-WB Speech Transceivers", L. Hanzo, C. Somerville and J. Woodard, "Voice and Audio Compression for Wireless Communications, 2nd Edition", Chichester, UK: John Wiley-Sons Inc., 2007, pp. 525-534.

Contents

Chapter 1

Introduction

Language is mankind's most important means of communication and speech is its primary medium. In the most general form, speech communication is the process of verbally transmitting an idea from one human being to another. Speech has remained the most desirable medium of communication between humans. However, analogue transmission of speech in telecommunication networks cannot exploit the benefits of Forward Error Correction (FEC) coding. By contrast, the digital transmission of speech is more robust against channel impairments, as a benefit of FEC coding, achieving lower production costs, consistent quality, security and a high spectral efficiency in the systems that exploit it.

The growth of research activities in the field of voice and audio coding is driven by an increasing demand for higher efficiency in the utilization of the allocated bandwidth in digital communication systems, in particular in wireless terrestrial and satellite communications. However, recently, most research activities have been focussed on wideband speech coding, in order to meet the demand for high-quality speech transmissions in multimedia applications, such as video-conferencing [1].

The diverse design factors of speech and audio coding for mobile communications are usually weighted differently according to the specific target application. Some of the factors to be taken into consideration include the bit rate, perceptual speech and audio quality, computational complexity, delay and error resilience. Since the transmission bandwidth constitutes a limited and valuable resource in digital communications, the compression of source signals is essential in order to increase the achievable bandwidth efficiency. The ultimate aim of source compression is to remove the redundancy from the source, with the aim of reducing the number of bits required for representing the source signal. On

the other hand, most modern audio codecs focus on removing perceptually irrelevant fine detail from the source signal, in an effort to achieve high compression. Specifically, the acoustic masking effect of the human ear is exploited [2] as in the MPEG-4 Advanced Audio Coding (AAC) and the MPEG-4 Transform-Domain Weighted Interleaved Vector Quantization (TwinVQ) schemes [3] which have found a wide range of communication-based and storage-based applications. However, striking an attractive trade-off between the bit-rate reduction and the reconstructed speech and audio quality, is essential.

In the past, the speech coding standardization activities dedicated to both wireline and wireless applications have been dominated by the Code Excited Linear Prediction (CELP)-based codecs. CELP speech coding was originally proposed by Schroeder and Atal in 1985 [4], which provided significantly better quality than existing low bit-rate algorithms at the time, which was achieved at an increased computational complexity. Its high computational complexity is associated with the search for the best codebook entry, where each candidate excitation sequence would undergo the analysis-by-synthesis process, by passing through the synthesis filters and then, would be evaluated as to how close the reconstructed speech would be to the original speech. An algorithm proposed in an effort to reduce the computational complexity of standard CELP coders is known as Algebraic Code Excited Linear Prediction (ACELP) [5]. The ACELP algorithm is based on the CELP speech coding algorithm, but having a specific algebraically structured codebook, which is used to generate the excitation codebook vectors. The ACELP algorithm is widely employed in current speech coding standards, one of which is the Adaptive MultiRate (AMR) speech codec [6, 7]. The narrowband AMR speech codec [8] is the most widely used speech codec in the Third Generation (3G) systems [9].

Its wideband relative, the Adaptive MultiRate WideBand (AMR-WB) speech codec [6] is one of the wideband speech codecs developed from the efforts devoted to wideband speech coding, due to an increasing focus on improving the output speech quality. The AMR-WB speech codec provides an excellent speech quality as a benefit of representing the speech bandwidth of 50-7000 Hz, which has been selected as the standard codec by the Third Generation Partnership Project (3GPP) [10, 11]. The AMR-WB codec was also adopted by the International Telecommunications Union (ITU) Telecommunication Standardization Sector (ITU-T) as the G.722.2 standard speech codec [12].

Shannon's classic source- and channel-coding separation principle [13] was derived for the idealized assumptions detailed in Section 14.6.2.1 of [1]. However, in real-time interactive wireless communications, these idealized conditions of Shannon's separation theorem

have limited applicability, which thus motivates the application of Joint Source and Channel Coding (JSCC) for practical speech and audio transmission systems communicating over realistic fading mobile channels.

We may categorize JSCC based on the design procedure used, which may be deemed to fall into two natural categories. The first category of JSCC may be referred to as "source-oriented" design, where the source codes are designed to be robust in the face of channel errors. One of the best examples of this is error-resilient channel-optimized vector quantization [14–16], in which the analog source is quantized by taking into consideration the characteristics of both the source and the channel.

Whilst, the second category of JSCC may be referred to as "channel-oriented" design, which follows the coding approach proposed by Hagenauer, where the inherent residual redundancy of the source code is exploited during the channel decoding process for further error protection [17]. More explicitly, the so-called "channel-oriented" JSCC exploits the known characteristics of the source code in designing the channel code.

One of the subclasses of "channel-oriented" designs is the JSCC scheme known as Unequal Error Protection (UEP), as in [18–23]. More explicitly, in UEP schemes the channel encoder exploits the knowledge of the source bits' error sensitivity for providing the required grade of protection.

Another subclasses of "channel-oriented" designs is the JSCC scheme known as Iterative Source Channel Decoding (ISCD), as in [24–28]. In the literature, UEP and ISCD schemes dominate the research activities in the field of JSCC. Next, a brief historical overview of the UEP and the ISCD JSCC families will be presented in Sections 1.1 and 1.2, respectively.

1.1 Unequal Error Protection Schemes

A limited-complexity, limited-delay source encoder typically exhibits a non-uniform error sensitivity, where the perceptual effect of a channel error may significantly vary from one bit to another [1, 6]. Furthermore, when considering realistic fading mobile channels, the speech and audio encoded bits become vulnerable to transmission errors, hence they may lead to perceptually unacceptable distortions. Therefore, it is desirable to provide stronger error protection for those speech or audio bits, which have a substantial effect on the objective or subjective quality of the reconstructed speech or audio signals. Again,

this philosophy is referred to as UEP.

There has been considerable interest in UEP schemes in the literature, following the work of Masnick and Wolf [29]. In the early study of UEP schemes, a linear UEP code based on block coding was designed for providing UEP within a codeword [29, 30]. Alternatively, UEP of speech, audio and video transmissions may be achieved by Low-Density Parity-Check (LDPC) codes [31–34] or Rate-Compatible Punctured Convolutional (RCPC) channel coding [18–23].

In [23], Hagenauer proposed RCPC codes, which is capable of providing different levels of protection with the aid of different convolutional code rates. Efforts have been invested in investigating UEP schemes, which employ RCPC codes having unequal error protection capabilities for improving the quality of speech, images, audio or video transmission schemes. In [18, 22], RCPC codes were employed for providing UEP, which resulted in beneficial coding gains for speech transmission over a mobile radio channel. More explicitly, the proposed UEP scheme consists of RCPC codes used for unequal error protection, where the resultant bitstream is transmitted using differentially encoded coherent four-phase shift keying over a correlated Rayleigh fading channel. In another proposal by Nazer and Alajaji [19], the authors proposed an UEP scheme employing RCPC codes for a CELP speech transceiver, as well as exploiting the residual redundancy inherent in the CELP-encoded bitstream during the decoding process. A beneficial improvement was achieved [19] in terms of the average speech distortion versus channel E_b/N_0 performance compared to the Equal Error Protection (EEP) benchmark scheme, when communcating over both Additive White Gaussian Noise (AWGN) channels as well as over Rayleigh fading channels. Additionally, the achievable gains of exploiting the residual redundancy inherent in the CELP-encoded bitstream provided a further beneficial improvement compared to the benchmark scheme dispensing with the exploitation of residual redundancy [19]. The application of the UEP scheme employing RCPC codes in image-related transmissions has been proposed in [20, 21]. More explicitly, in [20], a UEP-RCPC scheme was proposed for image tranmission over Rayleigh fading channels, whilst an unequal image protection scheme using RCPC codes for transmission over a wireless Internet Protocal (IP) network was considered in [21].

Considerable efforts have been invested in designing LDPC codes capable of providing UEP [31–33, 35]. In [31, 32], an optimized UEP LDPC codes based on irregular random graphs were proposed for the Binary Erasure Channel (BEC). A twin-class UEP LDPC scheme was proposed in [35]. In [33] a class of LDPC codes exhibiting the UEP property

was proposed, where the source signal was encoded by a variable length code and then unequally protected using irregular LDPC codes.

Alternatively, there has been considerable interest in UEP coded modulation designs. More explicitly, the UEP capability of the Imai-Hirakawa multilevel block coded 8-level Phase Shift Keying (8-PSK) modulation [36] has been studied in [37, 38] for transmission over Rayleigh fading channels. In [39], Seshadri and Sundberg studied the UEP capability of multilevel trellis coded modulation, which may be viewed as an extension of Imai-Hirakawa multilevel coding [36]. However, a key problem in multilevel coded systems is the interleaving delay and thus they are not suitable for real-time speech communications, which requires minimum transmission delay. An example of UEP multilevel coded modulation designed for image transmission over Rayleigh fading channels, can be found in [40]. In another development, recently Aydinlik and Salehi designed UEP TCM that is capable of providing different protection levels with the aid of different coding rates, as well as different locations in the mapping scheme [41]. On the other hand, in [42] the authors investigated the UEP capabilities of multilevel coded modulation designed for an OFDM system, which benefitted from the employment of the adaptive modulation technique.

In some transmission systems, where either only weak channel coding or no channel coding is used such as for example Bluetooth and Digital Enhanced Cordless Telecommunications (DECT) communications, the concept of UEP may be achieved by allocating different transmission power to the individual source bits according to their bit error sensitivities. In [43], Brüggen and Vary proposed a novel UEP concept in order to achieve UEP by periodically allocating unequal transmission power to individual bits, which is known as the so-called modulation with unequal power allocation. A similar UEP scheme was proposed in [44], where the transmission power was adapted according to the current unquantized source sample, while in [43] the correlation between the unquantized and quantized source parameters was exploited in combination with Gray coded bit-to-symbol mapping. Another interesting proposal was reported in [45], where the unequal power allocation applied to the subcarriers of Orthogonal Frequency Division Multiplexing (OFDM) was exploited for providing UEP for different-sensitivity source bits.

Below we have summarised the related contributions on the above-mentioned UEP design approach in Tables 1.1 and 1.2 .

Year	Author(s)	Contribution
UEP-Linear		
1967	Masnick and Wolf [29]	introduced linear UEP codes based on algebraic block coding.
1981	Boyarinov and Katsman [30]	proposed optimal binary UEP codes by combining the parity check matrices of two binary (BCH) codes.
UEP-RCPC		
1990	Hagenauer *et al.* [18]	introduced the UEP concept by employing RCPC codes.
1991	Cox *et al.* [22]	proposed UEP schemes for a subband coded speech transceiver using RCPC codes.
1999	Yap and Ngan [20]	designed UEP scheme employing RCPC codes for image transmission over Rayleigh fading channel.
2002	Nazer and Alajaji [19]	contrived UEP RCPC schemes for a CELP speech codec, while also exploiting the residual redundancy inherent in the CELP-encoded bitstream.
2006	Lamy-Bergot *et al.* [21]	proposed an UEP RCPC scheme for image transmission over wireless IP networks.
UEP-LDPC		
2005	Rahnavard and Fekri [31]	proposed an ad-hoc graph-based approach to provide twin-class UEP.
2007	Rahnavard and Fekri [32]	proposed a scheme for constructing partially regular LDPC codes with UEP property, which was an extension of [31].
2007	Poulliat *et al.* [34]	proposed a hierarchical optimization approach in order to enhance the UEP property of the Irregular LDPC codes.

Table 1.1: Contributions on UEP JSCC schemes (Part 1).

Year	Author(s)	Contribution
UEP-Coded Modulation		
1993	Seshadri and Sundberg [39]	studied the UEP capabilities of multilevel trellis coded modulation, as an extension of the Imai-Hirakawa multilevel coding [36].
1995	Morelos-Zaragoza *et al.* [37]	investigated the UEP capabilities of the Imai-Hirakawa multilevel construction using block coded 8-PSK for AWGN channels.
1997	Morelos-Zaragoza *et al.* [38]	investigated the UEP capabilities of the Imai-Hirakawa multilevel construction using block coded 8-PSK for Rayleigh fading channels.
2001	Yuan *et al.* [40]	proposed an UEP scheme using multilevel coded modulation for image transmission over Rayleigh fading channels.
2008	Aydinlik and Salehi [41]	investigated UEP TCM schemes providing different protection levels with the aid of different coding rates.
2008	Zhou *et al.* [42]	studied the UEP capabilities of multilevel coded modulation assisted adaptive OFDM.
UEP-Power		
1998	Ho [45]	unequal power allocation to OFDM subcarriers exploited for providing UEP to each subchannel.
2003	Goertz and Bresch [44]	UEP is achieved by allocating unequal transmission power based on the current unquantized source sample.
2005	Brüggen and Vary [43]	UEP is achieved by allocating unequal transmission power based on the current unquantized source sample, as well as on exploiting the correlation between the current unquantized and quantized source parameters.

Table 1.2: Contributions on different UEP schemes (Part 2).

1.2 Iterative Source and Channel Decoding

The turbo decoding principle of exchanging *extrinsic* information between the constituent channel decoders has been proposed by Berrou *et al.* [46, 47], which is a particularly powerful and efficient decoding technique designed for parallel concatenated channel codes. It has subsequently been widely employed for other soft-in soft-out (SISO) blocks [6], including ISCD [48].

A realistic, finite-delay lossy source codec leaves some residual redundancy in the encoded parameters, which is not the case for Shannon's ideal entropy codec. Fortunately, this residual redundancy may be beneficially exploited for error protection. Some examples of this approach can be found in [49] and [50], where the residual redundancy of both a CELP [4, 51] and of a Mixed-Excitation Linear Prediction (MELP) [52] speech codec was utilized during the channel decoding stage for enhancing the error resilience of the speech codecs. Additionally, many authors have considered the exploitation of this residual redundancy during the decoding process by the so-called ISCD approach, as in [24–28].

More explicitly, in [24], a bit-level Variable-Length Code (VLC) trellis representation was proposed, where the BCJR [53] algorithm was employed for computing the *A Posteriori* Probabilities (APP). On the other hand, in [25] an iterative source/channel decoding scheme was proposed for exploiting the redundancy imposed by a channel code and the residual redundancy in a symbol-based VLC source, where a significant coding gain was observed compared to the benchmark system having an equivalent transmission rate. Other examples of trellis-based VLC decoding techniques were proposed in [26] and [27], where the APPs were computed based on a Two-Dimensional (2D) and Three-Dimensional (3D) symbol-based VLC trellis representations, respectively. Recently, in [28] a parallel concatenated ISCD scheme employing symbol-based VLC was proposed. The proposed system's performance was studied against a benchmark scheme employing bit-based VLC.

Various schemes employing Variable-Length Error-Correcting (VLEC) [54] codes were studied in [55] and [56], while in [57], Reversible Variable Length Codes (RVLC) [58] were invoked and the associated bit-based trellis structure [24] was used for decoding. In [56] VLEC and RVLC schemes were implemented for transmission over dispersive AWGN channels. More explicitly, a joint iterative turbo equalization and SISO source decoding scheme was proposed for the sake of combating the effects of Inter-Symbol Interference (ISI). It was demonstrated that even without the protection of channel codes, the redundancy in the source code can be exploited for effectively eliminating the ISI. Furthermore, by addi-

itonally employing channel coding, soft source decoding has the potential of significantly improving the attainable system performance, provided that the channel equalization, channel decoding and source decoding are carried out jointly and iteratively.

Recently, in [59, 60] the residual redundancy inherent in the source encoded bitstream was exploited using the innovative concept of soft speech bits, which was further developed to accept the *a priori* information passed to the speech decoder from the channel decoder as *extrinsic* information [61]. Hence, *extrinsic* information is iteratively exchanged between the source and channel decoders for the sake of improving the overall system performance.

As a further development, in [62–64] the ISCD scheme of [48] was improved by exploiting both the intentionally imposed and inherent unintentional residual redundancy found in the source encoded bitstream for mitigating the effects of transmission errors. The performance of the ISCD scheme was characterised using EXtrinsic Information Transfer (EXIT) charts [62], which were proposed by ten Brink [65] for describing the flow of *extrinsic* information between the Soft-Input-Soft-Output (SISO) constituent decoders. Recently, the concept of EXIT chart analysis has been extended to three-stage concatenated schemes [66, 67], which requires the employment of two 3D EXIT charts.

Another interesting development, namely a novel Irregular Variable-Length Code (IrVLC) was proposed by Maunder *et al.* in [68], where the EXIT curve shape of the IrVLC was designed to closely match the EXIT curve of the channel codec with the aid of EXIT charts [65], resulting in near-capacity high-quality source reconstruction.

We have summarised the related contributions on the above-mentioned ISCD design approach in Tables 1.3 and 1.4.

1.3 MIMO Transceiver

There has been considerable interest in the employment of Space-Time Coding (STC) [70] in an effort to exploit the benefits of transmit diversity in Multiple-Input Multiple-Output (MIMO) systems for wireless communications. The concept of STC was introduced by Tarokh *et al.* [70], in order to introduce correlation between the signals transmitted from various antennas in various time slots. More explicitly, in [70] Space-Time Trellis Codes (STTC) employing trellis codes were proposed, which are capable of providing both coding and diversity gains. As a further development, Alamouti [71] proposed a simple STC scheme for transmission using two transmit antennas and an arbitrary number of receiver

Year	Author(s)	Contribution
ISCD-VLC		
1997	Balakirsky [24]	proposed a bit-level VLC trellis representation.
2001	Bauer and Hagenauer [25]	proposed an ISCD scheme employing symbol-based VLC codes.
2002	Kliewer and Thobaben [26]	derived the APPs with the aid of the 2D symbol-based VLC trellis representation,which exploits the residual source correlation of variable-length encoded Markov sources.
2003	Kliewer and Thobaben [27]	derived the APPs with the aid of the 3D symbol-based VLC trellis representation, which exploits the residual source correlation of variable-length encoded Markov sources.
2008	Liu *et al.* [28]	proposed an ISCD scheme of parallel concatenated codes employing VLCs.
ISCD-VLEC		
2003	Hedayat and Nosratinia [55]	proposed ISCD schemes employing VLEC codes.
2005	Wang *et al.* [56]	designed joint iterative turbo equalization and SISO source decoding invoking VLEC and RVLC for combating the ISI.

Table 1.3: Contributions on different ISCD schemes (Part 1).

antennas. Following Alamouti's work, Tarokh *et al.* generalised the transmit diversity scheme to arbitrary number of transmit antennas, leading to the concept of Space-Time Block Codes (STBC) [72]. Unlike the STTCs, STBCs are only capable of providing diversity gain, but no multiplexing gain, although they have a reduced decoding complexity. A similar technique of achieving transmit diversity in Code Division Multiple Access (CDMA) systems proposed by Hochwald *et al.* [73] with the extra benefit of providing a multi-user support capability, which is referred to as Space-Time Spreading (STS).

Since then, there has been considerable efforts in improving space-time modulation design. In [74] Su *et al.* proposed the concept of combining orthogonal transmit diversity designs with the principle of Sphere-Packing (SP), where it was demonstrated that the SP aided STBC scheme was capable of outperforming the conventional orthogonal design based STBC schemes of [71, 72].

Year	Author(s)	Contribution
ISCD-RVLC		
2003	Ng *et al.* [57]	proposed a JSCC scheme invoking RVLC.
ISCD-SBSD		
2001	Adrat *et al.* [48]	proposed an ISCD scheme exploiting the residual redundancy inherent in the source encoded bitstream.
2001	Perkert *et al.* [69]	proposed an ISCD scheme exploiting the residual redundancy of the GSM full-rate speech codec for transmission over burst Rayleigh fading channel.
2003	Adrat *et al.* [61]	used EXIT charts to study the convergence behaviour of the ISCD scheme of [48].
2005	Adrat and Vary [62]	the improved ISCD scheme of [48] by exploiting both the intentionally imposed and inherent unintentional residual redundancy found in the source encoded bitstream.
2006	Clevorn *et al.* [64]	a unity rate block code was employed and the intentionally imposed redundancy of ISCD scheme of [62] was interpreted as a short block.
ISCD-Irregular VLC		
2007	Maunder *et al.* [68]	proposed IrVLC for near-capacity JSCC.

Table 1.4: Contributions on different ISCD schemes (Part 2).

The efficiency of the above-mentioned schemes relies on the availability of channel knowledge, at the receiver. In practice, the Channel Impulse Response (CIR) between each transmit and each receive antenna has to be estimated at the receiver, which is a challenging and high-complexity task. An alternative scheme that eliminates the requirement of CIR knowledge has inspired Tarokh and Jafarkhani [75] to propose a differential encoding and decoding technique for Alamouti's scheme [71]. Similarly, in [76], Hochward and Sweldens proposed a differerential modulation scheme for transmit diversity based on unitary space-time codes [77], which can be employed in conjunction with an arbitrary number of transmit antennas. The differential detection scheme eliminates the requirement of MIMO channel estimation at the cost of the usual 3 dB performance penalty.

The fundamental motivation and rationale of the thesis was that of contriving powerful JSCC schemes designed for MIMO-aided speech and audio transceivers. UEP as well as ISCD schemes were investigated. The ISCD schemes were optimized with the aid of EXIT charts for speech transmission over MIMO-aided wireless channels.

1.4 Outline of the Thesis and Novel Contributions

1.4.1 Outline of the Thesis

The outline of the thesis is as follows:

- **Chapter 2: Unequal Protection Aided Joint Source and Channel Decoding**
  In Chapter 2, two different joint source and channel coding schemes invoking the UEP concept for wireless audio and speech transmissions are presented. The first one is a turbo-detected UEP arrangement using two different-rate convolutional codes, while the second employs Irregular Convolutional Codes (IRCC). We first review the three audio and speech codecs used throughout the thesis, namely, the MPEG-4 Advanced Audio Coding (AAC), the MPEG-4 Transform-Domain Weighted Interleaved Vector Quantization (TwinVQ) and the Adaptive Multi-Rate Wideband (AMR-WB) codecs in Sections 2.3 and 2.4. The error sensitivity of the various bits of the MPEG-4 AAC, the MPEG-4 TwinVQ and the AMR-WB audio/speech codecs is quantified in Sections 2.3.3, 2.3.4 and 2.4.2, respectively.

  In Section 2.5 we proposed a turbo-detected unequal error protection (UEP) arrangement using twin-class convolutional outer coding, as well as joint coding and modulation as the inner constituent components combined with a STTC-based spatial diversity scheme. Specifically, the encoded bitstreams of these codecs are partitioned into two bit-sensitivity classes, based on their bit sensitivity.

  As a design alternative, in Section 2.6 we proposed an UEP turbo transceiver using IRCCs and designed the IRCC to match the bit error sensitivities of the AMR-WB speech codec as well as to accommodate the characteristics of the channel code employed with the aid of EXtrinsic Information Transfer (EXIT) charts.

- **Chapter 3: Soft-Bit Assisted Iterative AMR-WB Source-Decoding**
  In Chapter 3, the AMR-WB speech codec was further developed in order to accept

the *a priori* information provided by the channel decoder. Specifically, a jointly optimized iterative source- and channel-decoding aided Differential Space-Time Spreading (DSTS) assisted Sphere Packing (SP) modulation scheme invoking the AMR-WB speech codec was proposed in Section 3.3. The attainable system performance improvements were demonstrated, when the soft-output AMR-WB speech codec was employed. First of all, we quantified the residual redundancy inherent in the AMR-WB encoded parameters, which was exploited as *a priori* information for computing the *extrinsic* Logarithmic-Likelihood Ratios (LLR) and for estimating the speech parameters. We also investigated the convergence behaviour of the advocated scheme with the aid of EXIT charts. Additionally, the EXIT characteristic of the soft-bit assisted AMR-WB decoder is also analysed, demonstrating that the EXIT curve of the AMR-WB decoder is unable to reach the point of perfect convergence at I_A, I_E=(1,1).

- **Chapter 4: EXIT Chart Optimized Soft-Bit Assisted AMR-WB Transceiver**
  In Chapter 4, we proposed two alternative ways of circumventing the deficient EXIT curve shape of the soft-bit assisted AMR-WB decoder, namely by using the novel technique of over-complete source-mapping [78] as well as by employing a recursive Unity Rate Code (URC). More explicitly, in Section 4.2 we proposed and investigated a jointly optimized Iterative Source and Channel Decoding (ISCD) scheme invoking the AMR-WB speech codec that exploits the intentionally increased residual redundancy of the AMR-WB encoded bitstream by using the novel Over-Complete source-Mapping (OCM) of [78]. Then, in Section 4.3 a turbo transceiver invoking the precoder-aided AMR-WB speech codec was proposed. We investigated the convergence behaviour of the advocated scheme with the aid of EXIT charts.

- **Chapter 5: Conclusions and Future Work**
  Chapter 5 summarizes the main findings of the research and offers suggestions for future research.

1.4.2 Novel Contributions

The novel contributions of the thesis are as follows:

- A sophisticated turbo transceiver capable of providing unequal speech/audio protection using two different-rate Non-Systematic Convolutional (NSC) codes was

proposed. More explicitly, the achievable performance of the MPEG-4 AAC, the MPEG-4 TwinVQ and the AMR-WB audio/speech codecs were studied in conjunction with the systematically designed UEP turbo transceiver using two different-rate NSCs for providing unequal error protection, Trellis Coded Modulation (TCM) for bandwidth-efficient joint coding and modulation and Space-Time Trellis Coding (STTC) for attaining spatial diversity [79, 80].

- A serially concatenated turbo transceiver capable of providing unequal speech protection using an IRCC was proposed. The proposed scheme was optimized using EXIT charts in order to provide unequal AMR-WB speech protection, while maximizing the iteration gain attained [81].

- A jointly optimized ISCD using a DSTS assisted SP modulation scheme [82] invoking the AMR-WB speech codec, was proposed. The Soft-Input-Soft-Output (SISO) AMR-WB decoder was contrived for the sake of accepting the *extrinsic* information passed to it from the channel decoder. The convergence behaviour of the proposed scheme was analysed using both 2D and 3D EXIT charts [83, 84].

- An iteratively detected SP aided DSTS [82] scheme using a Recursive Systematic Convolutional (RSC) code, for protecting the soft-bit assisted AMR-WB decoder's bitstream was proposed, which is also protected by a novel OCM [78,85] scheme [86]. The convergence behaviour of the proposed scheme was analysed and optimized with the aid of 3D EXIT charts and their 2D projections.

- A two-transmit-antenna aided DSTS-SP system associated with a single receive antenna using three-stage iterative detection and invoking the precoder-aided AMR-WB speech codec was proposed and its performance was optimized using EXIT charts [87].

Chapter 2

Unequal Protection Aided Joint Source and Channel Coding

2.1 Introduction

Shannon's classic source- and channel-coding separation principle [13] states that source coding and channel coding can be treated as separate entities under a number of idealized assumptions as detailed in Section 14.6.2.1 of [1]. However, in real-time interactive wireless communications, these idealized conditions of Shannon's separation theorem have limited applicability.

Thus, the limited applicability of Shannon's source- and channel-coding separation theorem [13] motivates the design of joint source and channel decoding for practical speech and audio transmission systems communicating over realistic fading mobile channels. More explicitly, a practical source encoder has both computational complexity and delay limitations. As a consequence, the source encoded parameters will inherently contain residual redundancy and the different encoded bits will exhibit different error sensitivity.

A high compression ratio can be achieved by Code Excited Linear Preditction (CELP)-based speech codecs, such as the Adaptive MultiRate WideBand (AMR-WB) speech codec [6] which is capable of achieving compression ratios ranging from 7 to 24, while operating at 0.5 to 1.5 bits per input speech sample. The AMR-WB speech codec has been selected as the standard codec by the Third Generation Partnership Project (3GPP) for Third Generation (3G) mobile communication systems [88], due to its capability to

provide a superior speech quality in comparison to the conventional telephone-bandwidth voice codecs [10].

A high compression ratio and a high audio quality can be achieved by perceptionally optimized codecs, such as the MPEG-4 Advanced Audio Coding (AAC) and the MPEG-4 Transform-Domain Weighted Interleaved Vector Quantization (TwinVQ) schemes [3], which have found a wide range of communication-based and storage-based applications. As an example, the MPEG-4 AAC has been recommended by 3GPP as the audio codec for Packet-Switched Streaming (PSS) and Multimedia Messaging Service (MMS) applications [89, 90]. However, the speech and audio encoded bits become vulnerable to transmission errors and the different bits exhibit a different error sensitivity.

Thus, the different error sensitivity of the practical source encoded bits motivates the employment of Unequal Error Protection (UEP) schemes, which constitute a specific form of joint source and channel coding. The application of the UEP concept allows us to match the error protection capabilities of channel coding schemes to the different bit-error sensitivities of the speech or audio bits [22].

The UEP concept has been proposed for improving the quality of audio, video or speech transmission schemes. In [91] the concept of UEP schemes using a family of Rate-Compatible Punctured Convolutional (RCPC) codes [23] having as many as 20 different rates ranging from 1/4 to 7/8 was proposed for the Eureka-147 Digital Audio Broadcast (DAB) system [92]. Similarly, Irregular Convolutional Codes (IRCC) [93, 94] which constitute a family of different-rate subcodes are capable of providing an UEP capability. Additionally, designing these IRCCs using the EXtrinsic Information Transfer (EXIT) charts [65] allows us to improve the convergence behaviour of iteratively decoded serially concantenated codes.

Against this background, the concept of UEP turbo transceivers incorporating the audio and speech codecs mentioned above is investigated. In this chapter, two different UEP aided joint source and channel coding philosophies are designed for wireless audio and speech transmissions, namely a turbo-detected UEP scheme using two different-rate convolutional codes and that using more sophisticated IRCCs. In Section 2.3 and 2.4 we provide an overview and the error sensitivity study of the source codecs used. Two different-rate convolutional codes are invoked in Section 2.5, whilst Section 2.6 provides an overview of the UEP transceivers using IRCCs. Finally, the chapter is concluded in Section 2.7.

2.2 Speech and Audio Quality Measure

In this section, the objective speech and audio quality measures used during our investigations of the various speech and audio transceivers are defined. Classic methods of measuring the quality of reconstructed speech have concentrated on the sample to sample difference between the original and reconstructed signals. The most common is the Signal-to-Noise Ratio (SNR) which is also known as an objective measure of reconstructed speech is given by:

$$SNR = \frac{\sum_n s_{in}(n)^2}{\sum_n \{s_{out}(n) - s_{in}(n)\}^2} \tag{2.1}$$

where $s_{in}(n)$ are the original speech samples, $s_{out}(n)$ are the reconstructed samples, and the summations are carried out over the entire waveforms. This measure implies that a listener stores an entire speech utterance before making a quality judgement based on the whole speech. This is clearly an unreasonable proposition. It has also a major drawback, where its value is dominated by the higher-energy speech segments, thus, it does not give fair weighting between the high- and low- energy speech segments.

Better and fairer measurement can be gained by taking the SNR values over short frames or segments of the speech waveform and then sum all the values for all the segments, which is known as Segmental SNR (SegSNR) as given by:

$$SEGSNR = \frac{1}{M_{SegSNR}} \sum_{m=1}^{M_{SegSNR}} 10log_{10} \frac{\sum_{n=1}^{N_{SegSNR}} s_{in}^2(n)}{\sum_{n=1}^{N_{SegSNR}} \{s_{out}(n) - s_{in}(n)\}^2} \tag{2.2}$$

where N_{SegSNR} is the number of speech samples within a segment of typically 15-25 ms, that is, 120-200 samples at a sampling rate of 8 kHz, while M_{SegSNR} is the number of 15-25 ms segments, over which $SEGSNR$ is evaluated.

It is found [95] that the segmental SNR performs much better than the conventional SNR, where it correlates better with subjective speech quality measure, Mean Opinion Score (MOS). MOS is the most commonly used subjective measure [96], where the speech quality of a coder is judged by some listening tests. This is done by asking several listeners to grade the reconstructed speech quality as excellent, good, fair, poor or bad with a mark of 5 given to excellent and down to 1 for bad. Subjective measures such as MOS are very time consuming and expensive to carry out. Hence, the objective measure mentioned above, provides a convenient way to study the performance of different codecs.

2.3 The MPEG-4 Audio Codec

2.3.1 MPEG-4 Audio Standard Overview

The Moving Pictures Experts Group (MPEG) has developed the MPEG-4 standard that aims to define an audiovisual coding solution to support the needs of the emerging interactive and broadcasting services as well as the needs of the mixed service models resulting from their convergence. The audio-related section of the MPEG-4 recommendation referred as the MPEG-4 audio standard specifies a complete 'toolbox' of compression methods that incorporates a wide range of operations from low bitrate speech coding to high-quality audio coding and music synthesis. The audio tools in MPEG-4 consist of audio coding and audio scene description tools, as shown in Figure 2.1. The audio coding tools are further divided into natural and synthetic audio coding [97], which refers to their capability of representing natural sounds such as speech and music as well as to synthesize sounds based on their parametric descriptions.

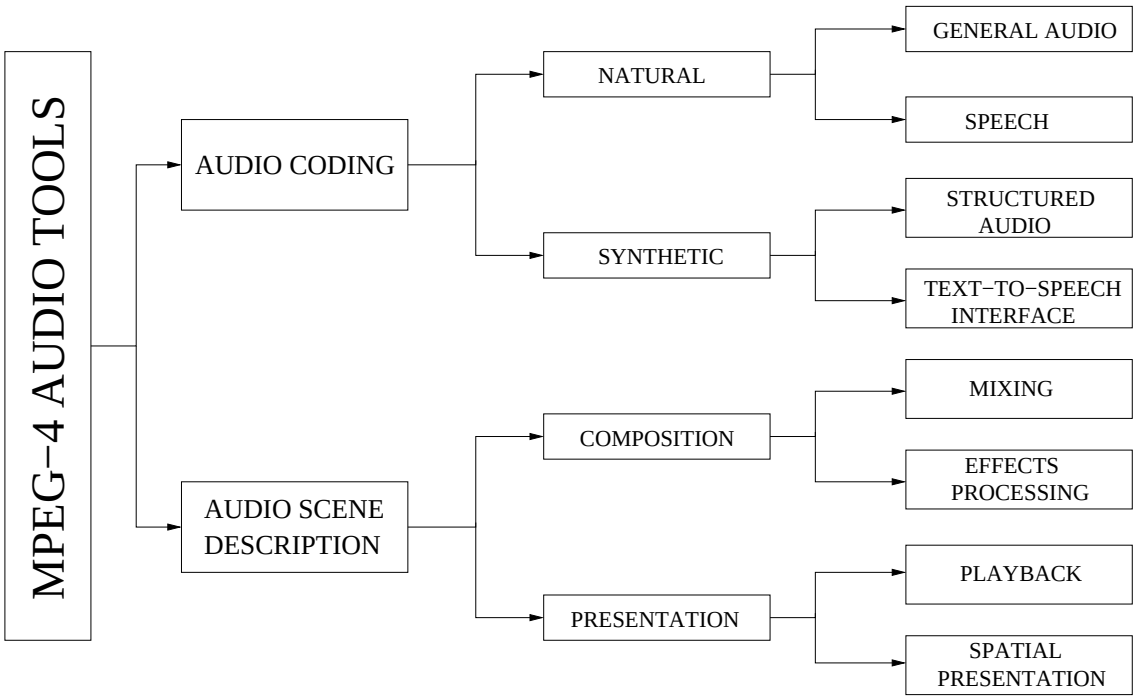

Figure 2.1: The MPEG-4 Audio Tools [97].

However, there is no single source coding algorithm that performs well across the entire range of coding applications, spanning from very low bitrate speech coding at 2 kbps, up to

high quality multi-channel audio coding at 64 kbps per channel. Therefore, the MPEG-4 natural audio coding toolbox consists of a set of different source coding algorithms that guarantees a high coding efficiency for the wide range of anticipated applications, which are classified as general audio coding and speech coding.

The MPEG-4 General Audio (GA) Codec was designed to appropriately process any type of audio material, as it aims at the faithful reproduction of all types of audio input signals. Thus its coding algorithm encompass both Time/Frequency (T/F) processing and parameteric audio coding, using the so-called Harmonic and Individual Lines plus Noise (HILN) technique [97]. It can be observed from Figure 2.2 that for bit rates ranging from 16 to 64 kbps, the MPEG-4 T/F codec based on the MPEG-2 AAC techniques are used. The HILN codec operates at very low bit rates, such as 4 kbps [98] in order to encode an audio signal using a parametric representation of the audio signals based on the decomposition of the input signal into audio objects, such as sinusoids and noise.

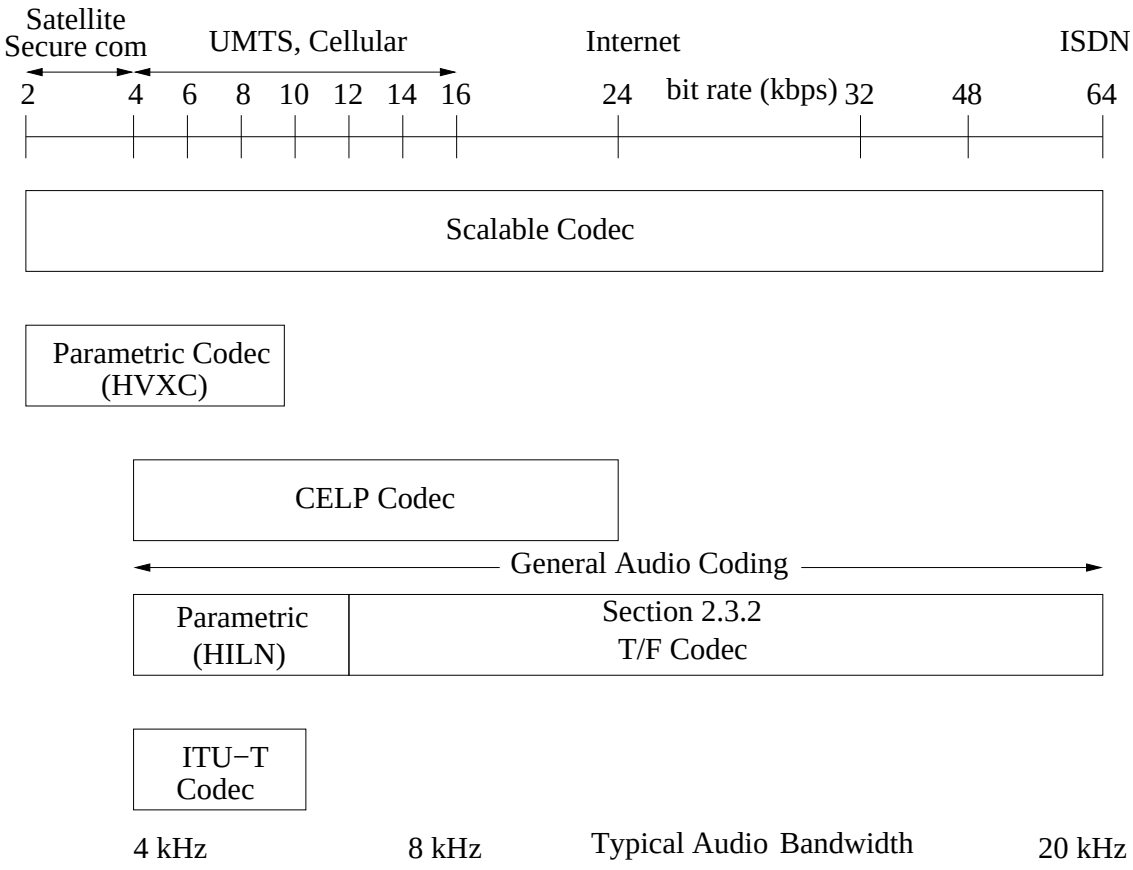

Figure 2.2: The MPEG-4 Audio framework [98, 99].

In contrast to the MPEG-4 GA codec modes designed for natural audio signals, speech codecs are designed to handle human voice as their input signal. In the MPEG-4 Audio scheme, both Code Excited Linear Prediction (CELP)-based speech coding and a parametric coding scheme, namely the so-called Harmonic Vector eXcitation Coding (HVXC) arrangement are used. More explicitly, for encoding speech signals at bit rates between 2 and 4 kbps, a HVXC coding scheme based on parametric coding is used, whilst, CELP-based speech coding is used for encoding narrowband and wideband speech signals ranging from 4 to 24 kbps [97, 99].

Observe in Figure 2.1 that unlike the MPEG-4 natural audio coding scheme which represents the compressed version of an audio signal, the synthetic audio coding mode represents the actual audio signal based on either Structured Audio (SA) or using its Text-To-Speech Interface (TTSI) [99]. More explicitly, an audio signal is generated when the MPEG-4 SA decoder interprets the transmitted descriptions of the sound algorithms and events, whereas, the TTSI generates synthetic speech from textual data.

In the following sections, we will present the T/F-based MPEG-4 GA codec with an emphasis on the MPEG-4 AAC and TwinVQ coding modes. These audio codecs will be used in our investigations in the UEP turbo audio transceiver in Section 2.5. Readers who are specifically interested in other coding tools within the MPEG-4 Audio are referred to the MPEG-4 Standard [99] and to the contributions by Pereira *et al.* [97] for further information.

2.3.2 The MPEG-4 General Audio T/F-based Codec

The MPEG-4 GA T/F-based codec was developed from the MPEG-2 AAC techniques including additional codec functionalities namely, Long Term Prediction (LTP) and Perceptual Noise Substitution (PNS), in order to enhance the achievable coding performance for noise-like and for tonal-like signals, respectively. A further coding and quantization mode, referred to as the TwinVQ was also added to the MPEG-4 GA T/F-based codec, which was designed to provide good coding performance at bitrates below 40 kbps [99]. The MPEG-4 GA T/F-based codec employs a perception based coding algorithm in order to exploit the phsycoacoustic model of human auditory perception, which is also known as the MPEG-4 T/F audio codec. The most important feature of the algorithm employed in the MPEG-4 T/F audio codec is the exploitation of perceptual masking effects within the human auditory system in order to reduce the audibility of quantization noise,

which resulted in transparent and perceptually lossless compression performance. Auditory masking [2] renders a fainter audio signal imperceptible to the human ear due to the presence of a louder masking audio signal within the simultaneously spectral vicinity of the marked signal [100]. The human auditory system behaves like a filter that perceives only frequencies in the range of 20 Hz and 20,000 Hz [101]. The range of audible frequencies can therefore be partitioned into a number of so-called critical bands that reflect similar sound pressure levels in human hearing. In other words, the human auditory system behaves like a bandpass filterbank which consists of overlapping bandpass filters having bandwidths on the order of 100 Hz for signal frequencies below 500 Hz, or having bandwidths as high as 500 Hz for high frequency signals. There are up to twenty five such critical bands in the frequency range up to 20 kHz [102]. In perception based audio compression, these human hearing properties are capatilized by transforming the audio signal to the frequency domain and then partitioning the spectrum into subbands, before quantizing each subband by ensuring that the quantization noise within that subband becomes imperceptible [103].

Figure 2.3 shows the block diagram of the MPEG-4 T/F audio encoder [99]. The MPEG-4 T/F audio encoder tranforms the input audio signal from the time domain to the frequency domain using an analysis filterbank. The analysis filterbank employed in the MPEG-4 T/F codec uses the Modified Discrete Cosine Transform (MDCT) which is preceeded by a Polyphase Quadrature Filter (PQF). In a simplicitic, but plausible approach, the MDCT may be interpreted as a real-valued Fourier-related transform with the additional property of having an overlapping structure, which provides better frequency selectivity owing to avoiding the typical rectangular windowing-induced frequency-domain Gibbs oscillation. Thus, it is widely used in state-of-the-art audio coding schemes. The MDCT supports block lengths of 2048 and 256 samples, which can be switched dynamically, depending on the stationary of the input signal's characteristics. The long block transform based processing using 2048 samples offers an improved coding efficiency for stationary signals. On the other hand, the short block transform based processing relying on 256 samples provides improved coding capabilities for transient signals. Additionally, the MPEG-4 T/F encoder provides the flexibility of different window shapes that can be switched dynamically for the sake of attaining an improved frequency selectivity and again for mitigating the Gibb-oscillation, specifically, the Kaiser-Bessel Derived (KBD) window and the sine window can be used [97, 99, 104]. The input audio signal then passes through the psychoacoustic model block of Figure 2.3 in order to determine the ratio of the signal energy to the *masking threshold*. The *masking threshold* represents the minimum sound

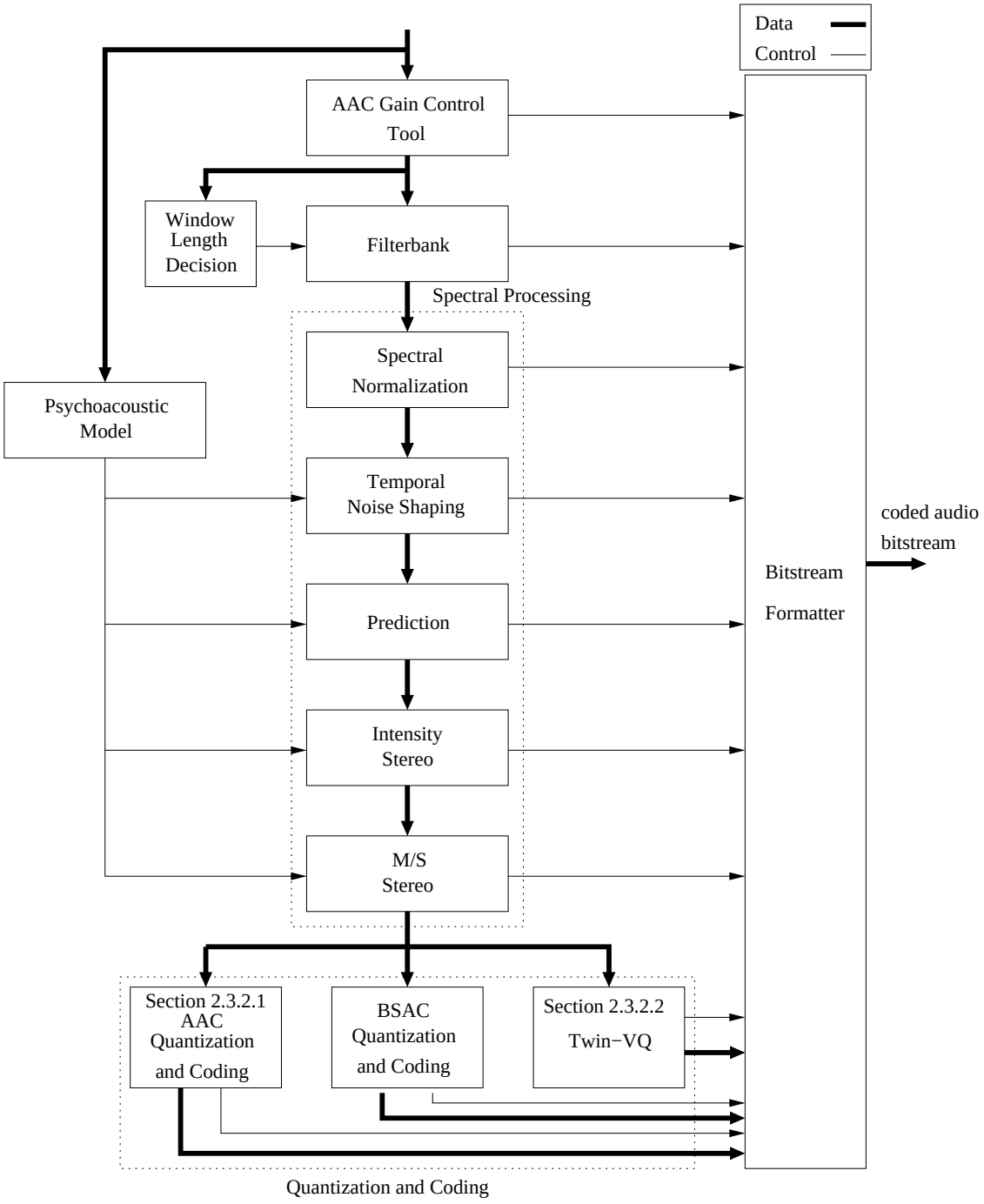

Figure 2.3: Block diagram of the MPEG-4 T/F-based encoder [99].

pressure level required for a signal in order to become audible in the presence of a masking signal. Figure 2.4 depicts an example of the masking threshold of a narrowband masker, having three maskees below the masking threshold. Also shown in the Figure 2.4 is the *threshold in quiet*, whereby signals below this threshold and in the absence of a masking signal, will be inaudible. However, in real life, a source signal may consists of simultaneous maskers, each having its own masking threshold, hence, a *global masking threshold* has to be computed. The psychoacoustic model exploits the masking property, in order to calculate the signal-to-mask ratio. The signal-to-mask ratio is used to decide on the most appropriate bit allocation, in an effort to minimize the effects of the quantization noise.

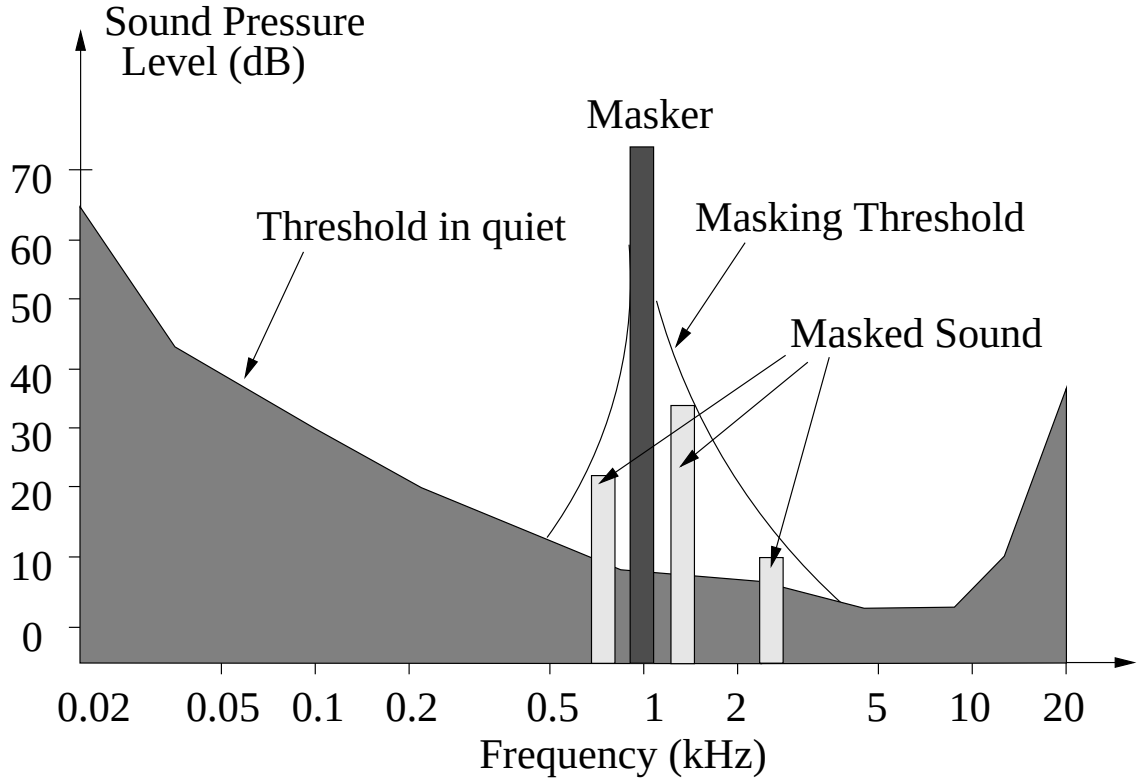

Figure 2.4: Threshold in quiet and masking threshold [105].

After the MDCT is carried out in the filterbank block of Figure 2.3, the spectral coefficients will be processed by the Temporal Noise Shaping (TNS) 'toolbox', except if the TwinVQ mode is used, where the spectral coefficients will be passed to the Spectral Normalization 'toolbox'.

Additionally, the MPEG-4 T/F encoder provides an important tool to further re-

duce the associated signal redundancy, namely LTP. Furthermore, an elaborate set of joint stereo coding techniques designed for multichannel input signals are available in the MPEG-4 T/F encoder, including 'Intensity Stereo' coding and Mid/Side (M/S) stereo coding. The 'Intensity Stereo' coding is a method of replacing the left and right stereo signals by a single channel having embedded directional information, whilst in the M/S stereo coding mode the sum and difference signals are transmitted, instead of transmitting the left and right signals.

The quantization and coding of the spectral values within the MPEG-4 T/F encoder can be performed regardless whether using the AAC, Bit Sliced Arithmetic Coding [99] (BSAC) or the TwinVQ techniques seen in Figure 2.3. The spectral values are quantized and coded with the aim of suppressing the quantization noise below the masking threshold. Finally, the quantized and coded spectral coefficients and all additional coding control parameters are packed into a bitstream and transmitted to the decoder.

The decoder carries out the inverse of the encoder's processing steps, where the coded spectral parameters and the coding control parameters are decoded and then are used to reconstruct the quantized spectral coefficients, before finally transforming the spectral values to their time domain representation using the synthesis filterbank [99].

In the following sections, the MPEG-4 AAC and the TwinVQ quantization and coding process will be reviewed briefly. Readers who are specifically interested in intricate details of the components seen in Figure 2.3 are referred to the MPEG-4 Standard and to the contributions by Pereira *et al.* [97] for further information.

2.3.2.1 AAC Quantization and Coding

The AAC quantization procedure involves the non-uniform quantization of the spectral domain transform coefficients. As known from coding theory [106], transform-domain quantizers have the inherent advantage of facilitating spectral-domain noise shaping in comparison to time-domain quantizers. The quantization noise shaping process is further improved by using scaling-factors, for the sake of improving the achievable subjective audio quality as described below.

More explicitly, the spectral-domain transform coefficients are grouped into scalefactor bands, where each band has its own individual scalefactor. This scalefactor is used for scaling the amplitude of all spectral-domain transform coefficients in the band, in an effort to shape the spectrum of the quantization noise according to the masking threshold

computed from the psychoacoustic model. The scalefactor is adjusted by keeping the quantization noise of each band below the masking threshold. Furthermore, the adjustment of the scalefactor relies on the total number of bits available for audio coding [99].

All the scalefactors are differentially encoded in order to exploit that they do not change dramatically from one scalefactor band to another. The redundancy inherent in the quantized frequency-domain transform coefficients is further reduced by compressing them using their noiseless Huffman coding [97, 99].

2.3.2.2 Transform-Domain Weighted Interleaved Vector Quantization (TwinVQ)

The Transform-Domain Weighted Interleaved Vector Quantization [107] scheme is a newly added quantization and coding tool of the MPEG-4 T/F codec. It is employed for compressing the spectral components, which gives superior performance in comparison to AAC coding at bit rates below 32 kbps per channel [107–109]. The basic idea introduced by TwinVQ coding is to replace the conventional encoding of the scalefactors and other spectral domain data used in MPEG-4 AAC by an interleaved vector quantization scheme applied to the normalized spectrum. More specifically, the quantization of the spectral coefficients involves two main steps, consisting of spectral normalization, followed by their interleaved weighted vector quantization.

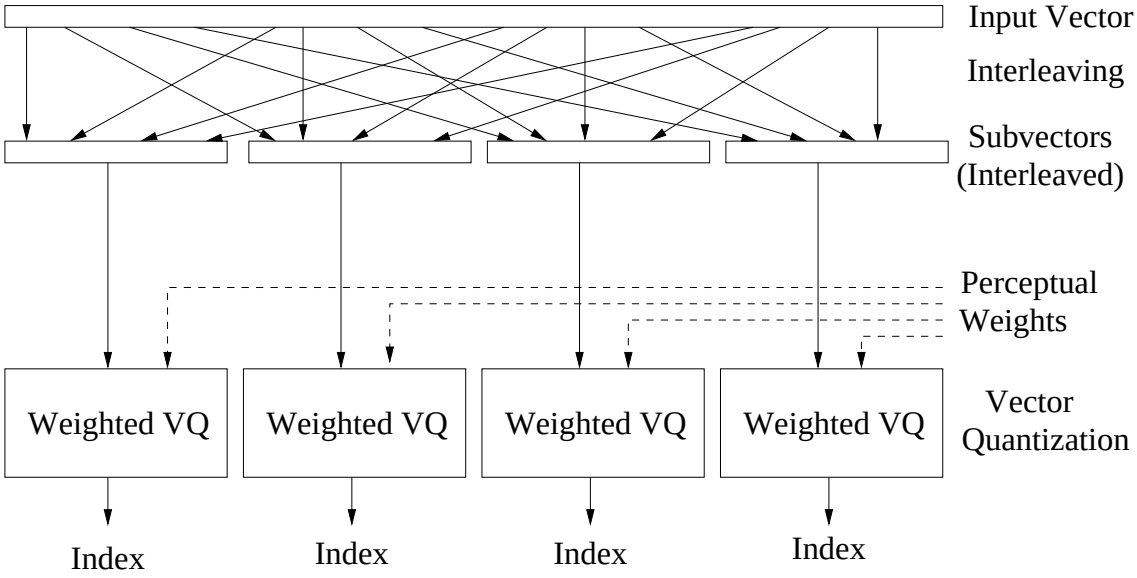

Figure 2.5: TwinVQ interleaved weighted vector quantization process [97].

To elaborate a little further, the spectral normalization process aims at rescaling the spectral coefficients' amplitude to a desired range. This is accomplished in two stages. In the first stage of the spectral normalization process, the audio input signal is subjected to Linear Predictive Coding (LPC) analysis in order to calculate the LPC coefficients. Then, the LPC based spectral envelope is estimated. At this stage, the MDCT coefficients are smoothened using this LPC spectral envelope. The MDCT coefficients are derived as the result of transforming the audio input signal to the frequency domain using the MDCT. The LPC envelope parameters are then encoded by means of their Line Spectral Pair (LSP) representation, which are sent to the receiver for decoding. In the second stage of the spectral normalization process, the smoothened MDCT coefficients are then used for estimating the Bark-scale envelope. This envelope estimation is performed based on each frequency subband, where the bandwidth of each frequency subband is proportional to a Bark-scale. The Bark-scale envelope for each Bark-scale subband is estimated by normalizing the average MDCT coefficient magnitudes by the overall average MDCT value. The Bark-scale envelope is then vector quantized and conveyed to the receiver for decoding.

In the weighted vector quantization process, the smoothened MDCT coefficients are first normalized by a global frequency-domain gain value [3]. The resultant MDCT coefficients are then interleaved and divided into subvectors [3]. In general, much finer quantization and hence more bits are required for the sake of quantizing the lower frequency subband than the higher frequency band, since the quantization errors in the lower frequency region are more objectionable. Thus, the role of the interleaving is to ensure that each subvector has a near-constant power so that an identical bit allocation can be used within each subvector [107]. Similarly, the LPC coefficients are also divided into subvectors interleaved in the same way.

After interleaving, the MDCT coefficients of a subvector are vector quantized using a weighted distortion measure, where the LPC spectral envelope of the corresponding subvector is used as a weighting factor. Thus, the effects of the spectral-domain quantization errors in the perceptually most vulnerable frequency regions are reduced. The employment of the interleaved vector quantization allows for fixed-length codewords to be used, resulting in a robust compression technique [107, 109], which cannot lose synchronization, while classic variable length coding can.

2.3.3 The MPEG-4 AAC Codec's Error Sensitivity

In the advocated UEP wireless system the MPEG-4 AAC of Section 2.3.2 is used for encoding the stereo audio file at a bit rate of 48 kbps. The audio input signal was sampled at 44.1 kHz and decomposed into 1024-sample frames, which hence resulted in an audio frame-length of 1024/(44.1 kHz) = 23.22 ms. The compressed audio information was then mapped to a packetized bitstream, which conveyed a 23.22 ms-duration audio frame. In our system, the number of bits per transmission frame varied from frame-to-frame, but had an average of approximately 48 kbps×23.22 ms=1116 bits per frame. The audio SegSNR of this configuration was found to be 16.28 dB, which guaranteed a transparent audio quality.

As stated in Section 2.1, it is well recognised that in highly compressed audio bitstreams a high Bit Error Ratio (BER) may lead to perceptually unacceptable distortion. In order to prevent the complete loss of transmitted audio frames owing to catastrophic error propagation, the most sensitive bits have to be well protected from channel errors. Hence, in the advocated system UEP is employed, where the compressed audio bitstream was partitioned into two sensitivity classes. More explicitly, an audio bit, which resulted in a SegSNR degradation above 16 dB upon its corruption was classified into protection class-1. The bit error sensitivity of the MPEG-4 AAC codec is investigated by systematically corrupting all of its bits in the encoded audio frame and then evaluating the average SegSNR degradation inflicted. A range of different audio files were used in our work and the results provided are related to a 60 seconds long excerpt of Mozart's "Clarinet Concerto (2nd movement - Adagio)". From the bit sensitivity studies using this audio file as the source, we found that approximately 50% of the total number of MPEG-4 encoded bits falls into class-1.

2.3.4 The MPEG-4 TwinVQ Codec's Error Sensitivity

In this section, the bit error sensitivity of the MPEG-4 TwinVQ codec's source bit sensitivity evaluation is presented. In the UEP turbo transceiver advocated, the MPEG-4 TwinVQ audio codec is used for encoding the stereo audio file at a bit rate of 32 kbps, which was sampled at 44.1 kHz. The analysis frame length is 23.22 ms, which corresponds to 1024 audio input samples, producing exactly 32 kbps×23.22 ms=743 encoded bits in each frame. The resultant MPEG-4 TwinVQ encoded parameters include the window mode, the MDCT coefficients, the vector quantized Bark-envelope, the prediction switch,

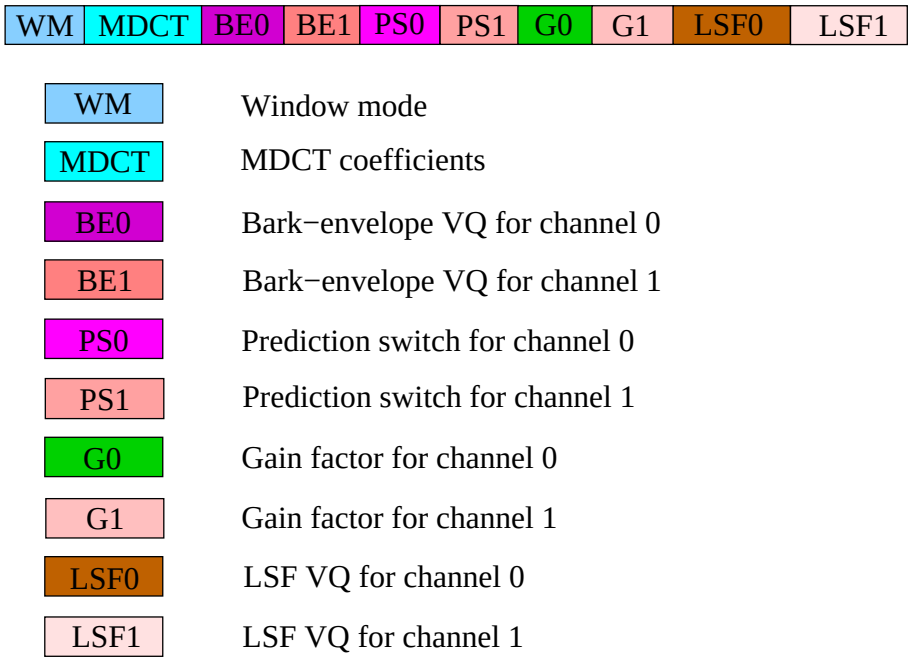

Figure 2.6: The structure of the MPEG-4 TwinVQ encoded stereo audio frame at 32 kbps

the gain factor and the vector quantized LSFs. Figure 2.6 shows the frame structure of the MPEG-4 TwinVQ encoder of Section 2.3.2.2 at 32 kbps.

As stated in Section 2.3.2.2, the MPEG-4 TwinVQ scheme is a transform coder that uses the MDCT [99] for transforming the time-domain input signal samples to the frequency-domain transform coefficients. The input signal is classified into one of three modes, each associated with a different transform window size, namely a long, medium or short window, catering for different input signal characteristics. The MDCT coefficients are normalized by the LPC and Bark-scale spectra. Then the normalized coefficients are interleaved and divided into sub-vectors by using the so-called interleave and division technique of [99], and all sub-vectors are encoded separately by the VQ modules.

The bit error sensitivity of a given bit is quantified by consistently inverting this bit in every speech frame and then evaluating the associated SegSNR degradation [6]. The error sensitivity of the various bits of the MPEG-4 TwinVQ codec averaged over a high number of frames is shown in Figure 2.7. The results provided are based on a 60 seconds long excerpt of Mozart's "Clarinet Concerto (2nd movement - Adagio)".

This figure shows that the bits representing the gain factors, the Line Spectral Frequency (LSF) parameters, and the Bark-envelope are more sensitive to channel errors, compared to the bits representing the MDCT coefficients. The bits signalling the win-

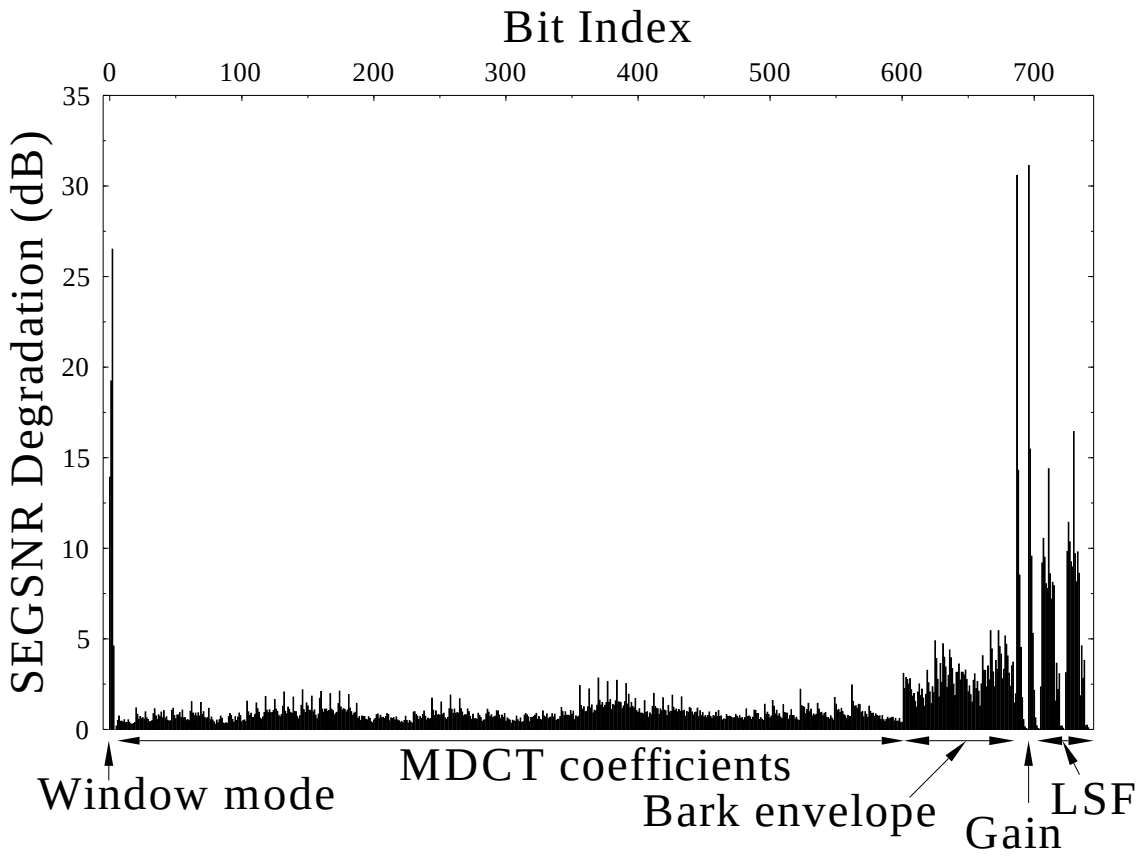

Figure 2.7: SegSNR degradations due to inflicting a 100% BER in the 743-bit, 23.22 ms MPEG-4 TwinVQ frame

dow mode used are also very sensitive to transmission errors and hence have to be well protected. The proportion of highly sensitive bits having SegSNR degradations in excess of 5 dB was only about 10%. This robustness is deemed to be a benefit of the weighted vector-quantization procedure, which uses a fixed-length coding structure as opposed to using an error-sensitive variable-length structure, where transmission errors would result in a loss of synchronisation.

2.4 The Adaptive Multi-Rate Wideband (AMR-WB) Codec

2.4.1 The AMR-WB Codec

The AMR-WB speech codec is capable of supporting nine different speech codec modes having bitrates of 23.85, 23.05, 19.85, 18.25, 15.85, 14.25, 12.65, 8.85 and 6.6 kbps [110].

It has become a 3GPP [111] and the G.722.2 ITU-T standard [12], implying that the same codec has been adopted for both wireless as well as for wireline services. This will eliminate the need for transcoding at the wireline/wireless interface, thus supporting wideband voice applications and services across a wide range of communication systems and platforms. Furthermore, it provides a superior speech quality in comparison to the conventional telephone-bandwith voice codecs [110]. The speech quality improvement is due to the low-frequency speech enhancement representing the low-frequency range spanning from 50 to 200 Hz, which contributes to improved naturalness, presence and comfort. The high-frequency wideband extension spanning from 3400 to 7000 Hz provides better fricative differentiation and therefore higher intelligibility [110]. The AMR-WB codec is based on the Algebraic Code Excited Linear Prediction (ACELP) principle [6, 112]. The principle of the ACELP algorithm [6, 113, 114] is shown in Figure 2.8.

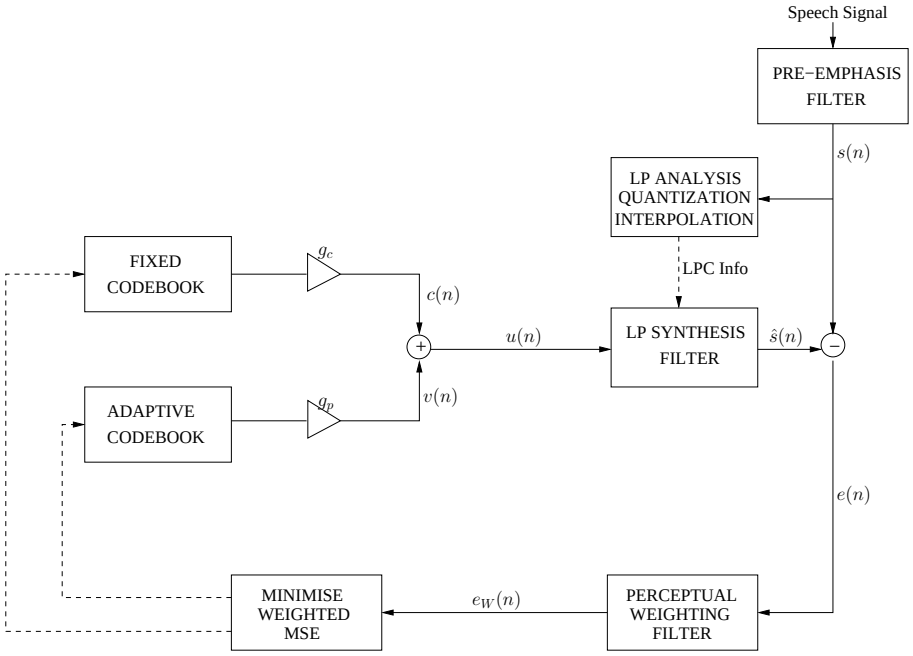

Figure 2.8: The Principle of ACELP Algorithm

The AMR-WB speech codec treats the frequency bands of 50-6400 Hz and 6400-7000 Hz separately in order to decrease the codec's complexity and to tailor the bit allocation by taking into account the subjectively most important frequency ranges. Thus, the input signal is first down-sampled from 16 kHz to 12.8 kHz, which makes the implementation of the interpolation filter simpler, where we have a down-sampling ratio of 12800/16000=4/5.

The human ear has a psycho-acoustic masking property, which renders low-volume quantization noise inaudible in the vicinity of high-energy Frequency-Domain (FD) components. Like most modern audio codecs, the AMR-WB codec exploits this masking property of the human ear by shaping the quantization error so that it has more energy in the speech formant regions, where it will be masked by the high signal energy present in these regions. This perceptual weighting is performed with the aid of the FD phsycho-acoustic error weighting filter. However, in the AMR-WB speech codec's overall spectral envelope, an excessive spectral decay is observed towards higher frequencies [6, 112] when invoking the appropriately up-sampled conventional perceptual weighting filter originally designed for narrowband signals [6, 112]. Therefore, a pre-emphasis filter of the form $P(z) = 1 - \mu z^{-1}$ associated with $\mu = 0.68$ and a modified perceptual weighting filter is used, in order to overcome the above-mentioned spectral tilt observed for wideband signals, when using the conventional perceptual weighting filter transfer function $W(z)$ of Equation (2) in [110].

Like most ACELP-based algorithms, the AMR-WB codec performs Linear Predictive (LP) analysis once per 20 ms segment of speech, producing a set of LP coefficients, which are used for constructing an LP synthesis filter $1/(1 - A(z))$. The LP synthesis filter aims to model the human vocal tract, which describes the spectral envelope of the speech segment [115]. Consequently, the excitation of the LP synthesis filter is determined using the classic analysis-by-synthesis process [6,112]. The excitation signal is found by filtering all candidate excitations through the synthesis filter, and opting for the one, which minimizes the weighted Minimum Mean Square Error (MMSE) between the reconstructed and original speech, as detailed in [6, 112].

In the following sections, we will present a functional description of the AMR-WB codec's operation in the 15.85 kbps and 23.05 kbps modes. These two bit rates will be used in our investigations of an UEP aided wireless turbo speech transceiver designed in this chapter as well as in the turbo speech transceiver using a soft AMR-WB decoder in Chapter 4, respectively.

2.4.1.1 Linear Prediction Analysis

Short-Term Prediction (STP) analysis is performed once for each 20 ms speech frame using the autocorrelation approach within a 30 ms asymmetric window, where a look-ahead of 5 ms is used during the speech signal's autocorrelation computation. The frame structure is depicted in Figure 2.9. The autocorrelation coefficients of the windowed speech

are computed and converted to LP coefficients using the Levinson-Durbin algorithm [6, 112]. A 16th order LP analysis filter is employed for modelling the short term correlation of the speech signal $s(n)$, using the schematic of Figure 2.8. The LP coefficients are then transformed to the Immittance Spectral Frequencies (ISF) domain for quantization and interpolation. The employment of the ISF representation of the LP coefficients is motivated by their advantageous statistical properties, which are similar to those of the Line Spectral Frequency (LSF) representation [116, 117]. Specifically, within each speech frame, there is a strong intra-frame ISF constraint, which manifests itself in terms of the ordering property of the neighbouring ISF values [6, 112]. This essentially motivates the employment of vector quantization. The interpolated quantized and the inverse quantized ISF parameters are converted back to the LP filter coefficients and they are then used for constructing both the sysnthesis and error weighting filters for each 5 ms speech subframe, respectively. More explicitly, the speech frame is divided into four subframes of 5 ms each. The synthesis filter shown in Figure 2.8 is used in the decoder for producing the speech signal reconstructed from the received excitation signal $u(n)$.

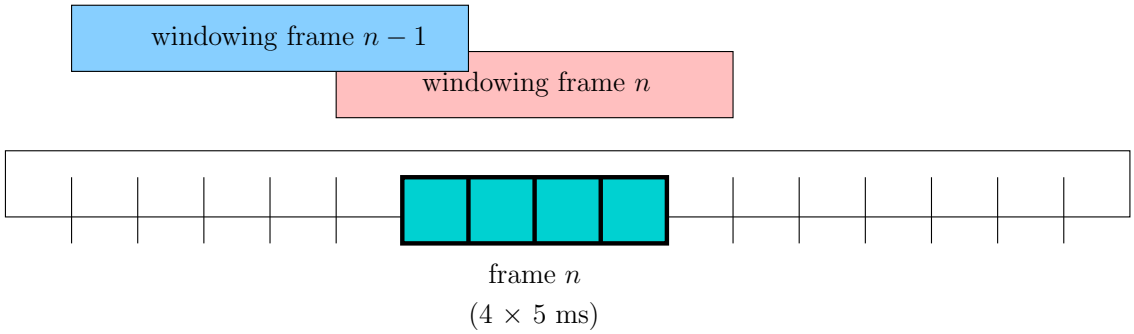

Figure 2.9: The AMR-WB Frame Structure and Windowing [11]

2.4.1.2 ISF Quantization

In the AMR-WB codec, the ISFs are quantized using interframe first order Moving Average (MA) prediction [11, 118] and Split-Multistage Vector Quantization (S-MSVQ) [11, 119, 120]. The resultant 16-element ISF residual vector produced by the first order MA prediction is quantized using S-MSVQ. For the sake of reducing the pattern-matching complexity, the vector is split into two subvectors of 9 and 7 elements, respectively. The two subvectors are quantized in two stages. In the first stage each subvector is quantized using a 256-entry code-book having an index-length of 8 bits. In the next stage, the resul-

tant first-stage quantization error vectors are then split into 3 and 2 subvectors. The five second-stage subvectors generated for the 15.85 kbps speech coding modes have 8, 8, 8, 8 and 16 entries or 3, 3, 3, 3 and 4 codebook index bits. By contrast, in the 23.05 kbps mode they have 64, 128, 128, 32 and 32 entries, represented by 6, 7, 7, 5 and 5 bits, respectively.

2.4.1.3 Pitch Analysis

Pitch analysis is performed on a subframe basis, leading to the adaptive codebook parameters, which are the delay and the gain of the pitch filter. The adaptive codebook contains excitation vectors that model the long-term periodicity corresponding to the pitch of the speech signal. Thus, pitch analysis is also often referred to as LTP analysis, which finds the delay of the pitch filter also referred to as LTP lag, as well as the gain of the pitch filter. As mentioned earlier, the excitation signal, $u(n)$ at the input of the LP synthesis filter seen in the schematic of Figure 2.8, is given by the sum of a gain-scaled and delayed version of the excitation signal and another gain-scaled signal output from a large fixed codebook, which can be written as $u(n) = Gu(n-a) + Gc(n)$ where G_a and a represent the adaptive codebook gain and delay, respectively. The optimum excitation signal $u(n)$ is chosen on the basis of minimising the weighted mean squared error, E_w between the original and synthesized speech over the subframe duration of 5 ms.

The AMR-WB codec employs both so-called open-loop and closed-loop estimation of the adaptive codebook delay parameters G_a and a. The open-loop pitch estimate of the pitch period is employed in order to restrict the search range of the possible adaptive codebook delay values. The open-loop pitch estimate is performed twice per 20 ms speech frame for both the 15.85 kbps and the 23.05 kbps modes. Then the full closed-loop analysis-by-synthesis procedure is used for finding a high-resolution delay at a range of ± 7 samples around the estimated open-loop position. The adaptive codebook gain is then computed directly for the optimal excitation. The pitch delay is encoded in both the 15.85 kbps and the 23.05 kbps modes, with the aid of 9 bits in the first and third subframes, while the relative delay of the other subframes is encoded using 5 bits.

The output of the LP synthesis filter using the optimally delayed and scaled excitation is then subtracted from the desired speech signal and the difference, which is often referred to as the target signal of the optimization process, is then used to find the second component of the excitation. The fixed codebook target is constructed using the unquantised adaptive codebook gain. The second component of the excitation is referred to as the fixed codebook contribution.

2.4.1.4 Fixed Codebook Structure

The algebraic codebook is searched through, in order to minimize the mean-squared error between the weighted input speech and the weighted synthesized speech by taking into account the now known adaptive codebook vector.

The fixed algebraic codebook structure is based on the so-called Interleaved Single-Pulse Permutation (ISPP) code design [11, 110]. In the algebraic codebook of the AMR-WB speech codec, the 64 legitimate positions of each excitation sample corresponding to a 5 ms excitation subframe sampled at a rate of 12.8 kHz is divided into four tracks. Therefore the 64 legitimate positions in a 5 ms subframe are split into 16 positions in each four-fold subsampled track. Note that the input speech is down-sampled to 12.8 kHz during the pre-processing stage, hence, the encoding process is performed at this sampling rate. The fixed, or algebraic codebook structure is specified in Table 2.1 and Table 2.2 for the 15.85 kbps and 23.05 kbps codec modes, respectively.

The algebraic structure of the excitation has non-zero excitation pulse amplitudes of either +1 or -1 in each of the four tracks, and their positions are restricted to the limited number of legitimate excitation pulse positions, as shown in Table 2.1 and Table 2.2 for the AMR-WB speech coding modes of 15.85 kbps and 23.05 kbps, respectively. The exact number of non-zero excitation pulses in each track depends on the corresponding AMR-WB encoding bitrate. The algebraic codebook index or the corresponding excitation codeword represents the non-zero pulse position indices and their polarity in each track. More explicitly, the non-zero pulse position index is given by the excitation pulse position in the subframe divided by the pulse spacing. The resultant integer division based remainder gives the so-called excitation track index. As an example, a non-zero pulse at position 31 in a subframe of the 15.85 kbps codec mode has a position index of $\lfloor 31/4 \rfloor$=7 and it belongs to the track having an index of 3, i.e. to the fourth track, as shown in Table 2.1.

Track	Pulse	Positions
1	i_0, i_4, i_8	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	i_1, i_5, i_9	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	i_2, i_6, i_{10}	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	i_3, i_7, i_{11}	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

Table 2.1: Potential positions of individual pulses in the algebraic codebook of the 15.85 kbps codec mode [11].

Track	Pulse	Positions
1	$i_0, i_4, i_8, i_{12}, i_{16}, i_{20}$	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	$i_1, i_5, i_9, i_{13}, i_{17}, i_{21}$	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	$i_2, i_6, i_{10}, i_{14}, i_{18}, i_{22}$	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	$i_3, i_7, i_{11}, i_{15}, i_{19}, i_{23}$	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

Table 2.2: Potential positions of individual pulses in the algebraic codebook of the 23.05 kbps codec mode [11].

More explicitly, in the 15.85 kbps codec mode, the excitation codebook contains twelve non-zero pulses of amplitudes either +1 or -1, which is split into three non-zero pulses in each track. In the 15.85 kbps codec mode, the three non-zero pulses in the first track are denoted by i_0, i_4 and i_8, as shown in Table 2.1. The algebraic excitation encoding procedure invoked for each track in the 15.85 kbps codec mode contains several steps.

First, each track is divided into two sections, each of which contains eight legitimate positions, which are denoted as Section A for the lower-half section and Section B for the upper-half section. Each section may contain 0, 1, 2 or 3 pulses [11]. Table 2.3 shows the four cases representing the possible number of non-zero pulses in each sections. We can classify the encoding procedure into two cases, namely those having (0,3) and (1,2) excitation pulses.

Case	Pulses in Section A	Pulses in Section B
0	0	3
1	1	2
2	2	1
3	3	0

Table 2.3: Possible number of non-zero pulses in each sections of each track in the algebraic codebook structure of the 15.85 kbps codec mode.

For the cases of having 0 and 3 pulses, the section which contains three non-zero pulses is encoded using a single bit. More explicitly, the lower-half section is represented by bit '0', while the upper-half section is represented by bit '1'. The section which contains three non-zero pulses is now explicitly identified. The position index of two non-zero pulses is encoded using $(2 \cdot 3) = 6$ bits, while their corresponding signs can be encoded with a single

bit, because the sign of one of the non-zero pulses is explicitly signaled to the decoder and the sign of the other non-zero pulse is deduced from the pulse ordering [11]. The remaining non-zero pulse's position index is encoded with the aid of 4 bits, using an additional bit for its corresponding sign. The sign is represented by bit '0' for a positive sign and by bit '1' for a negative sign. Therefore a total of $4 \cdot ((2 \cdot 3) + 1 + 1 + 4 + 1) = 52$ bits are required to represent the excitation pulse positions and signs for the algebraic excitation encoding for each 5 ms subframe in the 15.85 kbps codec mode, as potrayed in Table 2.4. However, for the above-mentioned cases of having 1 and 2 excitation pulses the specific section, which contains two non-zero pulses is identified and encoded using 1 bit, while the remaining procedure is carried out similar to that of the cases having 0 and 3 excitation pulses.

Following the fixed codebook search, a fourth-order MA predictor having fixed coefficients is used, in order to exploit the correlation between the fixed codebook gains of the adjacent frames [110]. The fixed codebook gain is expressed as a weighted version of the predicted gain based on the previous fixed codebook energies using a correlation-related correction factor. The adaptive codebook gain and the correction factor are jointly vector quantised using a 7-bit codebook in each subframe for both the 15.85 kbps and the 23.05 kbps codec modes.

2.4.1.5 Post-Processing

At the decoder, the transmitted parameters are inverse-quantized and the 12.8 kHz bandwidth speech is synthesized by filtering the reconstructed excitation signal through the LP synthesis filter. The synthesised speech is then passed through an adaptive post-processing stage, which consists of two steps. Firstly, the synthesised speech is de-emphasized using the filter $1/P(z) = 1/(1 - 0.68z^{-1})$ and finally, it is upsampled to 16 kHz.

2.4.1.6 The AMR-WB Codec's Bit Allocation

The AMR-WB speech codec's bit allocation is shown in Table 2.4 for the 15.85 kbps and 23.05 kbps speech modes. For both the 15.85 kbps and 23.05 kbps speech modes, 46 bits are used for encoding the ISFs by employing S-MSVQ, as seen in Table 2.4.

The pitch delay is encoded using 9 bits in the first subframe and the relative delays of the other subframes are encoded using 6 bits. The adaptive codebook gain is jointly quantized with the above-mentioned correlation-based correction factor of the fixed code-

Parameter	1st subfr	2nd subfr	3rd subfr	4th subfr	Total per frame
VAD-flag	1 (1)				1
ISPs	8+8+6+7+7+5+5=46 (2-47)				46
LTP Lag	9 (48-56)	6 (116-121)	9 (182-190)	6 (251-256)	30
LTP-filt.-flag	1 (57)	1 (122)	1 (191)	1 (257)	4
Fixed Index	52 (58-109)	52 (123-174)	52 (192-243)	52 (258-310)	208
CB Gains	7 (110-116)	7 (175-181)	7 (244-250)	7 (311-317)	28
Total					317/20 ms =15.85 kbps
VAD-flag	1				1
ISPs	46				46
LTP Lag	9	6	9	6	30
LTP-filt.-flag	1	1	1	1	4
Fixed Index	88	88	88	88	352
CB Gains	7	7	7	7	28
Total					461/20 ms =23.05 kbps

Table 2.4: Bit allocation of the AMR-WB speech codec at 15.85- and 23.05 kbps [11]. The bit positions for the 15.85 kbps mode, which are shown in round bracket assist in identifying the corresponding bits in Figure 2.10.

book gain using 7 bits for every 5 ms. As described in Section 2.4.1.4, 52 bits were used to encode the fixed algebraic codebook indices for every 5 ms subframe, which resulted in a total of 208 bits per 20 ms frame for the 15.85 kbps codec mode. For the 23.05 kbps mode, the algebraic codebook indices are encoded using 88 bits in each subframe, in order to give a total of 352 bits for a 20 ms speech frame.

2.4.2 The AMR-WB Codec's Error Sensitivity

In this section, we present the results of our AMR-WB codec bit-sensitivity study. Most source coded bitstreams contain certain bits that are more sensitive to transmission errors than others. A common approach used for quantifying the sensitivity of a given bit is to consistently invert this bit in every speech frame and evaluate the associated

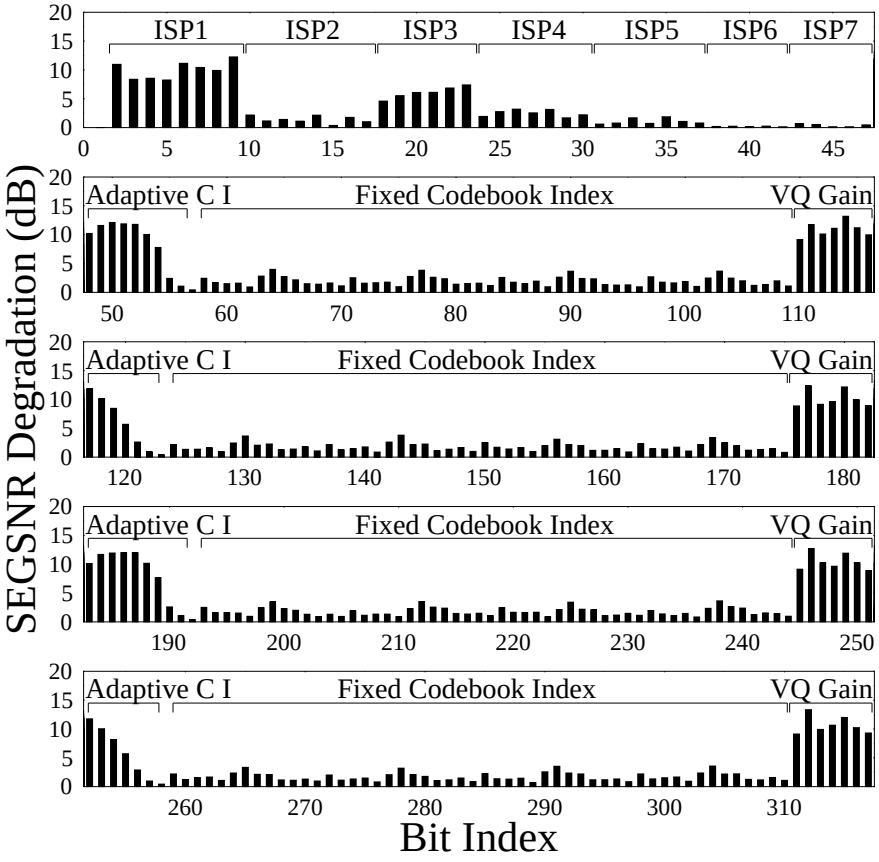

Figure 2.10: SegSNR degradations versus bit index due to inflicting 100% Bit Error Rate (BER) in the 317-bit, 20 ms AMR-WB frame

SegSNR degradation [6]. The SegSNR degradation is computed by subtracting from the SegSNR recorded under error-free conditions the corresponding value when there are channel-induced bit-errors.

The error sensitivity of the various encoded bits in the AMR-WB codec determined in this way is shown in Figure 2.10. The figure shows the bit sensitivities in each speech subframe for the bit rate of 15.85 kbps, with the corresponding bit allocations shown in Table 2.4. The results are based on samples taken from the European Broadcasting Union (EBU) Sound Quality Assessment Material (SQAM) CD [121], sampled at 16 kHz and encoded at 15.85 kbps.

It can be observed that the bits representing the Immittance Spectral Pairs (ISP), the adaptive codebook delay, the algebraic codebook index and the vector quantized gain are fairly error sensitive. The least sensitive bits are related to the fixed codebook's

excitation pulse positions, as shown in Figure 2.10. This is because, when one of the fixed codebook index bits is corrupted, the codebook entry selected at the decoder will differ from that used in the encoder only in the position of one of the non-zero excitation pulses. Therefore the corrupted excitation codebook entry will be similar to the original one. Hence, the algebraic codebook structure used in the AMR-WB codec is quite robust to channel errors. The error sensitivity of various bits in the AMR-WB codec obtained here will be exploited for designing an UEP aided transmission scheme by assigning different-rate channel encoders according to the source bit error sensitivities.

2.5 Unequal Protection Schemes Using Convolutional Codes [1]

2.5.1 System Background

In recent years, Joint Source and Channel Coding (JSCC) has received significant research attention in the context of both delay- and complexity-constrained transmission scenarios [1, 6]. JSCC aims at designing the source codec and channel codec jointly for the sake of achieving the highest possible system performance. As it was argued in Section 1 of [122], this design philosophy does not contradict the classic Shannonian source and channel coding separation theorem. This is because instead of considering perfectly lossless Shannonian entropy coders for source coding and transmitting their bitstreams over Gaussian channels, we consider low-bitrate lossy audio and speech codecs, as well as Rayleigh-fading channels. Since the bitstreams of the speech and audio encoders are subjected to errors during wireless transmission, it is desirable to provide stronger error protection for the audio bits, which have a substantial effect on the objective or subjective quality of the reconstructed speech or audio signals.

[1] This section is based on

1. N. S. Othman, S. X. Ng and L. Hanzo: Turbo-Detected Unequal Protection MPEG-4 Audio Transceiver Using Convolutional Codes, Trellis Coded Modulation and Space-Time Trellis Coding, IEEE 61st Vehicular Technology Conference, Stockholm, Sweden, May 2005, pp. 1600 -1604,

2. N. S. Othman, S. X. Ng and L. Hanzo: Turbo-Detected Unequal Protection Audio and Speech Transceivers Using Serially Concantenated Convolutional Codes, Trellis Coded Modulation and Space-Time Trellis Coding, Proceedings of IEEE 62nd Vehicular Technology Conference, Texas, USA, September 2005, pp. 1044-1048,

and it was based on collaborative research with the co-authors.

Supporting high-quality multimedia services for transmission over wireless communication channels requires the development of techniques for transmitting not only speech, but also video, music and data. Therefore, in the field of audio-coding, high-quality, high-compression and highly error-resilient audio-coding algorithms are required. The MPEG-4 standard [123,124] defines a comprehensive multimedia content representation scheme that is capable of supporting numerous applications - such as streaming multimedia signals over the internet/intranet, content-based storage and retrieval, digital multimedia broadcast or mobile communications. The audio-related section of the MPEG-4 standard [99] defines audio codecs covering a wide variety applications - ranging from narrowband low-rate speech to high quality multichannel audio, and from natural audio signals to synthesized sound effects, as a benefit of its object-based approach used for representing the audio signals.

The MPEG-4 GA encoder is capable of compressing all types of audio signals, at bitrates ranging from 6 kbps to broadcast quality audio at 64 kbps [123].

The MPEG-4 T/F codec is based on the MPEG-2 AAC solutions, with additional functionalities, such as for example the PNS and the LTP techniques invoked for enhancing the achievable compression performance. It also operates a specific coding and quantization mode referred to as TwinVQ for operation at extremely low bit rates, below 40 kbps [99]. In order to render this codec applicable to wireless systems, which typically exhibit a high BER, powerful turbo transceivers are required.

The previously mentioned MPEG-4 AAC has an important operational mode designed for robustness against transmission errors in error-prone propagation channels [125]. The error resilience of the MPEG-4 AAC codec is mainly attributed to the so-called Virtual CodeBook tool [99,126] (VCB11), the Reversible Variable Length Coding (RVLC) tool [58] and the Huffman Codeword Reordering tool [126] (HCR) [97, 125], which facilitate the integration of the MPEG-4 AAC codec into high-BER wireless systems.

Recently, a sophisticated unequal-protection turbo transceiver using twin-class convolutional outer coding, as well as joint coding and modulation as inner coding combined with STTC-based spatial diversity scheme was designed for MPEG-4 video telephony in [122, 127]. Specifically, maximal minimum distance Non-Systematic Convolutional (NSCs) [128] codes having two different code-rates were used as outer encoders for providing unequal MPEG-4 video protection. Good vedio quality was attained at a low SNR and medium complexity by the proposed transceiver. On one hand, Trellis Coded Modulation (TCM) [129–131] constitutes a bandwidth-efficient joint channel cod-

ing and modulation scheme, which was originally designed for transmission over Additive White Gaussian Noise (AWGN) channels. On the other hand, Space-Time Trellis Coding (STTC) [70,130] employing multiple transmit and receive antennas is capable of providing spatial diversity gain. When the spatial diversity order is sufficiently high, the channel's Rayleigh fading envelope is transformed to a Gaussian-like near-constant envelope. Hence, the benefits of a TCM scheme designed for AWGN channels will be efficiently exploited, when TCM is concatenated with STTC.

By contrast, in this section, we study the achievable performance of the AMR-WB speech codec, the MPEG-4 TwinVQ and the MPEG-4 AAC audio codecs in conjunction with the sophisticated unequal-protection turbo transceiver using joint coding and modulation as inner coding, twin-class convolutional outer coding based spatial diversity [122, 127].

The bit error sensitivity studies of the MPEG-4 AAC audio codec, the MPEG-4 TwinVQ audio codec and the AMR-WB speech codec are presented in Section 2.3.3, 2.3.4 and 2.4.2, respectively. In this section, a sophisticated jointly optimised turbo transceiver capable of providing unequal error protection is designed for employment in the MPEG-4 AAC and MPEG-4 TwinVQ audio transceivers, as well as in the AMR-WB speech transceiver. The transceiver advocated consists of STTC employed for mitigating the effects of fading, as well as, TCM and two different-rate NSC codes used for unequal error protection. A benchmarker scheme combining STTC and a single-class protection NSC is used for comparison with the proposed scheme. We will demonstrate that significant iteration gains are attained with the aid of the proposed turbo transceiver. The architecture of the turbo transceiver is described in Section 2.5.2. We elaborate further by characterising the achievable system performance in Section 2.5.3.

2.5.2 System Overview

The block diagram of the serially concatenated STTC-TCM-2NSC turbo scheme using a STTC, a TCM and two different-rate NSCs as its constituent codes is depicted in Figure 2.11. The STTC-TCM-2NSC scheme's turbo decoder structure is illustrated in Figure 2.12, where there are four constituent decoders, each labelled with a round-bracketed index. The Maximum A-Posteriori (MAP) algorithm [130] operating in the logarithmic-domain is employed by the STTC, TCM and the twin-class NSC decoders, respectively. The notations $P(.)$ and $L(.)$ in Figure 2.12 represent the logarithmic-domain symbol prob-

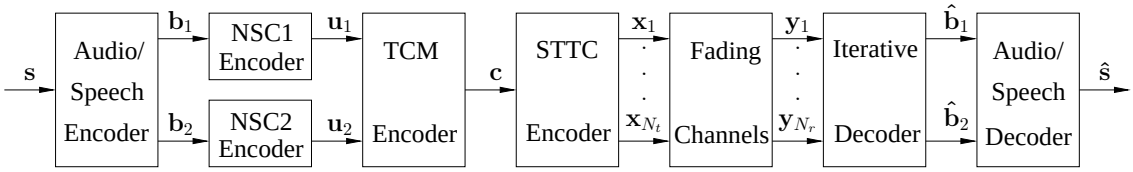

Figure 2.11: Block diagram of the serially concatenated STTC-TCM-2NSC assisted audio/speech scheme. The notations $\mathbf{s}$, $\hat{\mathbf{s}}$, $\mathbf{b}_i$, $\hat{\mathbf{b}}_\mathbf{i}$, $\mathbf{u}_i$, $\mathbf{c}$, $\mathbf{x}_j$ and $\mathbf{y}_k$ denote the vector of the audio/speech source symbol, the estimate of the audio/speech source symbol, the class-i audio/speech bits, the estimates of the class-i audio/speech bits, the encoded bits of class-i NSC encoders, the TCM coded symbols, the STTC coded symbols for transmitter j and the received symbols at receiver k, respectively. Furthermore, N_t and N_r denote the number of transmit and receive antennas, 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 2.12.

abilities and the Logarithmic-Likelihood Ratio (LLR) of the bit probabilities, respectively. The notations c, u and b_i in the round brackets (.) of Figure 2.12 denote TCM coded symbols, TCM information symbols and the class-i audio 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 $3b$.

The iterative turbo-detection scheme shown in Figure 2.12 enables an efficient information exchange between STTC, TCM and NSCs constituent codes for the sake of achieving spatial diversity gain, coding gain, unequal error proctection and a near-channel-capacity performance. The information exchange mechanism between each pair of constituent decoders is detailed in [127], which is restated here for convenience. As we can observe from Figure 2.12, 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+1}-ary TCM input information symbols emerging from the NSC decoders of block 3, namely

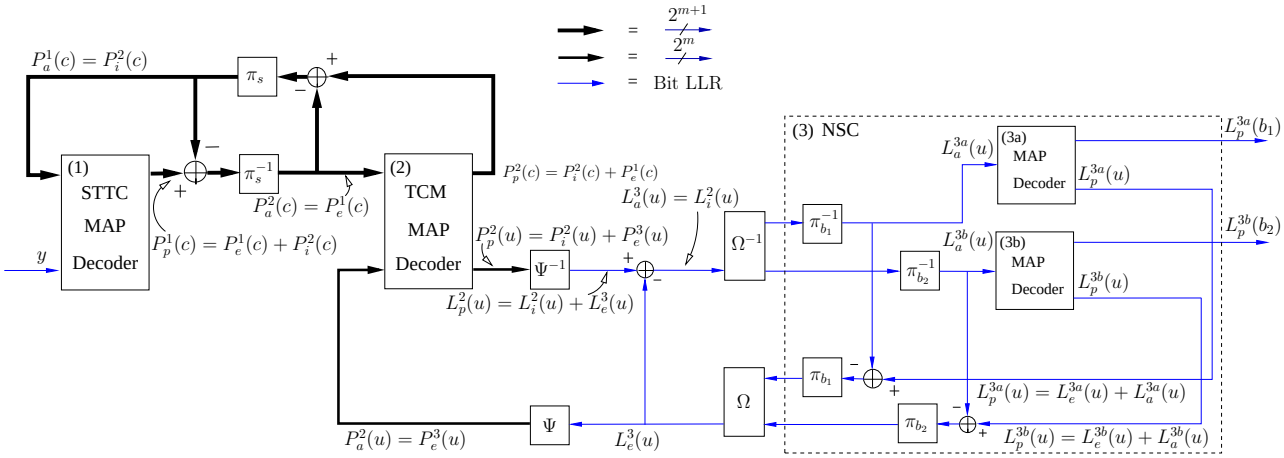

Figure 2.12: Block diagram of the STTC-TCM-2NSC turbo detection scheme seen at the right of Figure 2.11. The notations $\pi_{(s,b_i)}$ and $\pi^{-1}_{(s,b_i)}$ denote the interleaver and deinterleaver, while the subscript s denotes the symbol-based interleaver of TCM and the subscript b_i denotes the 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 [127] ©IEE, 2004, Ng, Chung and Hanzo.

$P_a^2(u) = P_e^3(u)$. 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 $L_a^3(u) = L_a^2(u)$ 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 2.12.

In the following sections, the detailed architecture of the MPEG-4 AAC and MPEG-4 TwinVQ audio transceivers, as well as the AMR-WB speech transceiver will be presented.

2.5.2.1 Turbo-Detected Unequal Protection AAC Audio Transceiver

From the bit sensitivity studies of the MPEG-4 AAC audio coder presented in Section 2.3.3 we found that approximately 50% of the total number of MPEG-4 encoded bits falls into class-1. Since the number of class-1 audio bits is approximately the same as that of the class-2 audio bits and there are an average of approximately 48 kbps×23.22 ms=1116 bits per audio frame, we protect the 558-bit class-1 audio sequence using a rate-R_1 NSC encoder and the 558-bit class-2 sequence using a rate-R_2 NSC encoder. Let us denote the turbo scheme as STTC-TCM-2NSC-1, when the NSC coding rates of $R_1 = k_1/n_1 = 1/2$ and $R_2 = k_2/n_2 = 3/4$ are used. Furthermore, when the NSC coding rates of $R_1 = 2/3$ and $R_2 = 3/4$ are used, we denote the turbo scheme as STTC-TCM-2NSC-2. The code memory of the class-1 and class-2 NSC encoders is $L_1 = 3$ and $L_2 = 3$, respectively. More explicitly, in the STTC-TCM-2NSC-1 scheme the number of coded bits emerging from the $R_1 = 1/2$ NSC encoder is $(558 + k_1L_1)/R_1 = 1122$ bits, while that generated by the $R_2 = 3/4$ NSC encoder is $(558 + k_2L_2)/R_2 = 756$ bits. However, in the STTC-TCM-2NSC-2 scheme the number of coded bits emerging from the $R_1 = 2/3$ NSC encoder is $(558 + k_1L_1)/R_1 = 846$ bits, while that generated by the $R_2 = 3/4$ NSC encoder is $(558 + k_2L_2)/R_2 = 756$ bits.

The class-1 and class-2 NSC coded bit sequences are interleaved by two separate bit interleavers. The two interleaved bit sequences are then concatenated to form a bit sequence of $1122 + 756 = 1878$ bits and $846 + 756 = 1602$ bits for the STTC-TCM-2NSC-1 and STTC-TCM-2NSC-2 schemes, respectively. This bit sequences is then fed to the rate-$R_3 = 3/4$ TCM [129–131] scheme having a code memory of $L_3 = 3$. Code termination was employed for the NSCs, TCM [129–131] and STTC codecs [70, 130]. Hence, at the TCM encoder's output of the STTC-TCM-2NSC-1 scheme, we have $(1878 + k_3L_3)/R_3 = 2516$ bits or $2516/4 = 629$ symbols, while that of the STTC-TCM-2NSC-2 scheme we have $(1602 + k_3L_3)/R_3 = 2148$ bits or $2148/4 = 537$ 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 $M_{TCM} = 16$-level Quadrature Amplitude Modulation (16QAM) [131]. The STTC employing $N_t = 2$ requires one 16QAM-based termination symbol. Therefore, at each transmit antenna of the STTC-TCM-2NSC-1 scheme, we have $629+1 = 630$ 16 QAM symbols or $4(630) = 2520$ bits in a transmission frame. However, at each transmit antenna of the STTC-TCM-2NSC-2 scheme, we have $537+1 = 538$ 16 QAM symbols or $4(538) = 2152$ bits in a transmission frame.

Hence, the overall coding rate is given by $R_{s1} = 1116/2520 \approx 0.4429$ and $R_{s2} = 1116/2152 \approx 0.5186$ for the STTC-TCM-2NSC-1 and STTC-TCM-2NSC-2 schemes, respectively. The effective throughput of the STTC-TCM-2NSC-1 and STTC-TCM-2NSC-2 schemes is $\log_2(M_{TCM})R_{s1} \approx 1.77$ Bits Per Symbol (BPS) and $\log_2(M_{TCM})R_{s2} \approx 2.07$ BPS, respectively.

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 audio bit sequences in both class-1 and class-2, as seen in Figure 2.11. The output of the turbo transceiver is decoded using the MPEG-4 AAC decoder. An undecodable received audio frames are considered as erroneous. During the decoding process, the erroneously received audio frames were dropped and replaced by the previous error-free audio frame for the sake of avoiding an even more dramatic error-infested audio-quality degradation [1, 6].

For the sake of benchmarking the scheme advocated, we created a powerful benchmark scheme by replacing the TCM and NSC encoders of Figure 2.11 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 audio 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 $(1116 + k_0 L_0)/R_0 = 2244$, which corresponds to 561 16QAM symbols. Again, a 16-state STTC scheme having $N_t = 2$ transmit antennas is employed. After code termination, we have $561 + 1 = 562$ 16QAM symbols or $4(562) = 2248$ bits in a transmission frame at each transmit antenna. The overall coding rate is given by $R = 1116/2248 \approx 0.4964$ and the effective throughput is $\log_2(16)R \approx 1.99$ BPS, both of which are very close to the corresponding values of the STTC-TCM-2NSC-2 scheme. A decoding iteration of the STTC-NSC benchmarker scheme is comprised of a STTC decoding and a NSC decoding step.

2.5.2.2 Turbo-Detected Unequal Protection TwinVQ Audio and AMR-WB Speech Transceivers

Let us denote the turbo scheme using the AMR-WB codec as STTC-TCM-2NSC-AMR-WB, whilst STTC-TCM-2NSC-TVQ refers to the turbo scheme using the MPEG-4 TwinVQ as the source codec. For comparison, both schemes protect the most sensitive 25% of the encoded bits (which we refer to as the class-1 bits) using an NSC code rate of

$R_1 = k_1/n_1 = 1/2$. By contrast, the remaining 75% of the bits (class-2) are protected by an NSC scheme having a rate of $R_2 = k_2/n_2 = 3/4$. The code memory of the class-1 and class-2 encoders is $L_1 = 3$ and $L_2 = 3$, respectively. More explicitly, in the STTC-TCM-2NSC-AMR-WB scheme the number of coded bits emerging from the $R_1 = 1/2$ NSC encoder is $(85 + k_1L_1)/R_1 = 176$ bits, while that generated by the $R_2 = 3/4$ NSC encoder is $(255 + k_2L_2)/R_2 = 352$ bits. However, in the STTC-TCM-2NSC-TVQ scheme the number of coded bits emerging from the $R_1 = 1/2$ NSC encoder is $(186 + k_1L_1)/R_1 = 378$ bits, while that generated by the $R_2 = 3/4$ NSC encoder is $(558 + k_2L_2)/R_2 = 756$ bits.

The class-1 and class-2 NSC coded bit sequences are interleaved by two separate bit interleavers. The two interleaved bit sequences are then concatenated to form a bit sequence of $176 + 352 = 528$ bits and $378 + 756 = 1134$ bits for the STTC-TCM-2NSC-AMR-WB and STTC-TCM-2NSC-TVQ schemes, respectively. This bit sequences is then fed to the rate-$R_3 = 3/4$ TCM scheme [129–131] having a code memory of $L_3 = 3$. Code termination was employed for the NSCs, as well as for the TCM [129–131] and STTC codecs [70, 130]. Hence, at the TCM encoder's output of the STTC-TCM-2NSC-AMR-WB scheme, we have $(528 + k_3L_3)/R_3 = 716$ bits or $716/4 = 179$ symbols, while that of the STTC-TCM-2NSC-TVQ scheme we have $(1134 + k_3L_3)/R_3 = 1524$ bits or $1524/4 = 381$ symbols. The TCM symbol sequence is then symbol-interleaved and fed to the STTC encoder as seen in Figure 2.12. We invoke a 16-state STTC scheme having a code memory of $L_4 = 4$ and $N_t = 2$ transmit antennas, employing M_{TCM} =16-level Quadrature Amplitude Modulation (16QAM) [131]. The STTC scheme employing $N_t = 2$ transmit antennas requires a single 16QAM-based termination symbol. Therefore, at each transmit antenna of the STTC-TCM-2NSC-AMR-WB scheme, we have $179 + 1 = 180$ 16 QAM symbols or $4(180) = 720$ bits in a transmission frame. However, at each transmit antenna of the STTC-TCM-2NSC-TVQ scheme, we have $381 + 1 = 382$ 16 QAM symbols or $4(382) = 1528$ bits in a transmission frame. In the STTC-TCM-2NSC-AMR-WB scheme the 25% of the bits that are classified into class-1 includes 23 header bits, which gives a total of $317 + 23 = 340$ AMR-WB-encoded bits. In the International Telecommunications Union (ITU) stream format [111], the header bits of each frame include the frame types and the window-mode used. For comparison, both STTC-TCM-2NSC-AMR-WB and STTC-TCM-2NSC-TVQ schemes employ the same simulation parameters. Therefore, in our simulation, the STTC-TCM-2NSC-TVQ scheme includes one doping bit, which gives a total of $743 + 1 = 744$ bits.

Hence, the overall coding rate of the STTC-TCM-2NSC-AMR-WB scheme becomes

$R_{AMRWB} = 340/720 \approx 0.4722$. By contrast, the overall coding rate of the STTC-TCM-2NSC-TVQ scheme is $R_{TVQ} = 744/1528 \approx 0.4869$. The effective throughput of the STTC-TCM-2NSC-AMR-WB and STTC-TCM-2NSC-TVQ schemes is $\log_2(M_{TCM}) \cdot R_{AMRWB} \approx 1.89$ BPS and $\log_2(M_{TCM}) \cdot R_{TVQ} \approx 1.95$ BPS, respectively.

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 audio bit sequences in both class-1 and class-2, as seen in Figure 2.11.

For the sake of benchmarking both audio schemes advocated, we created a powerful benchmark scheme for each of them by replacing the TCM and NSC encoders of Figure 2.11 by a single-class NSC codec having a coding rate of $R_0 = k_0/n_0 = 1/2$ and a code memory of $L_0 = 6$. Note that if we reduce the code memory of the NSC constituent code of the STTC-NSC benchmarker arrangement from L_0=6 to 3, the achievable performance becomes poorer, as expected. If we increased L_0 from 6 to 7 (or higher), the decoding complexity would double, while the attainable 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. We will refer to this benchmarker scheme as the STTC-NSC-TVQ and the STTC-NSC-AMR-WB arrangement designed for the audio and the speech transceiver, respectively. Again, all audio and speech 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 of the STTC-NSC-TVQ benchmarker scheme is $(744+k_0L_0)/R_0 = 1500$, which corresponds to 375 16QAM symbols. By contrast, in the STTC-NSC-AMR-WB scheme the number of output bits after taking into account the bits required for code termination becomes $(340 + k_0L_0)/R_0 = 692$, which corresponds to 173 16QAM symbols. Again, a 16-state STTC scheme having $N_t = 2$ transmit antennas is employed. After code termination, we have 375+1 = 376 16QAM symbols or 4(376) = 1504 bits in a transmission frame at each transmit antenna for the STTC-NSC-TVQ. The overall coding rate is given by $R_{TVQ-b} = 744/1504 \approx 0.4947$ and the effective throughput is $\log_2(16)R_{TVQ-b} \approx 1.98$ BPS, both of which are very close to the corresponding values of the STTC-TCM-2NSC-TVQ scheme. Similary, for the STTC-NSC-AMR-WB scheme, after code termination, we have $173 + 1 = 174$ 16QAM symbols or $4(174) = 696$ bits in a transmission frame at each transmit antenna. This gives the overall coding rate as $R_{AMRWB-b} = 340/696 \approx 0.4885$ and the effective throughput becomes $\log_2(16)R_{AMRWB-b} \approx 1.95$ BPS. Again, both of the

values are close to the corresponding values of the STTC-TCM-2NSC-AMR-WB scheme. A decoding iteration of the STTC-NSC benchmarker scheme is comprised of a STTC decoding and a NSC decoding step.

2.5.2.3 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 decoding step for the proposed STTC-TCM-2NSC scheme. Similarly, a decoding iteration of each of the STTC-NSC benchmarker schemes 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 for 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 given by

$$S = 2^{L_1} + 2^{L_2} + 2^{L_3} + 2^{L_4} = 40. \tag{2.3}$$

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

$$S = 2^{L_0} + 2^{L_4} = 80. \tag{2.4}$$

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.

2.5.3 Performance of UEP Schemes Using Convolutional Codes

In this section we evaluate the performance of the MPEG-4 AAC audio transceiver seen in Figure 2.11, as well as comparatively study the performance of the MPEG-4 TwinVQ audio and the AMR-WB speech transceivers using the SegSNR audio quality metric. Table 2.5 gives an overview of the STTC-TCM-2NSC-1 and the STTC-TCM-2NSC-2 assisted MPEG-4 AAC schemes' parameters, whilst Table 2.6 summarizes the parameters of the STTC-TCM-2NSC-TVQ and the STTC-TCM-2NSC-AMR-WB schemes.

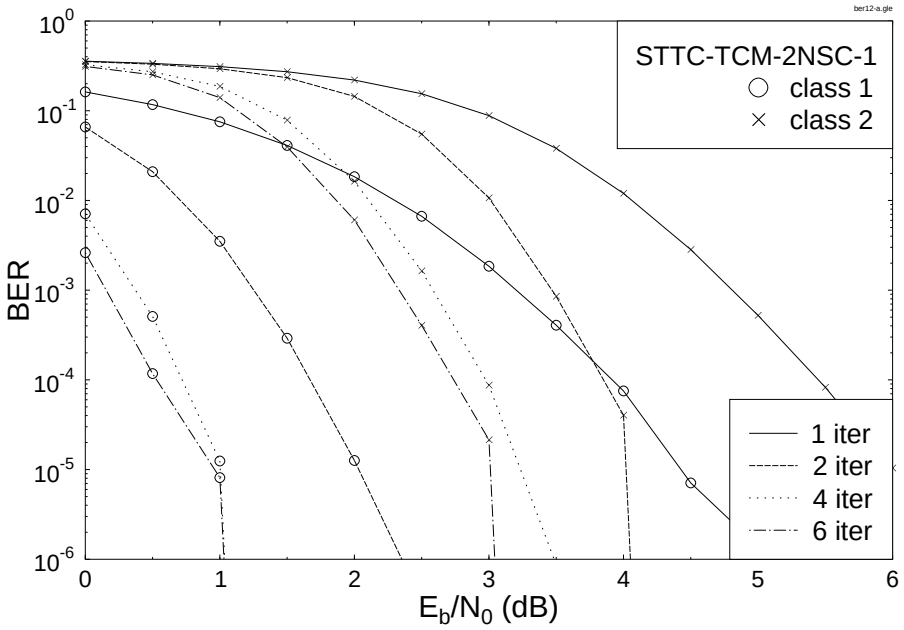

Figure 2.13: BER versus E_b/N_0 performance of the 16QAM-based STTC-TCM-2NSC-1 assisted MPEG-4 audio scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput is **1.77 BPS**.

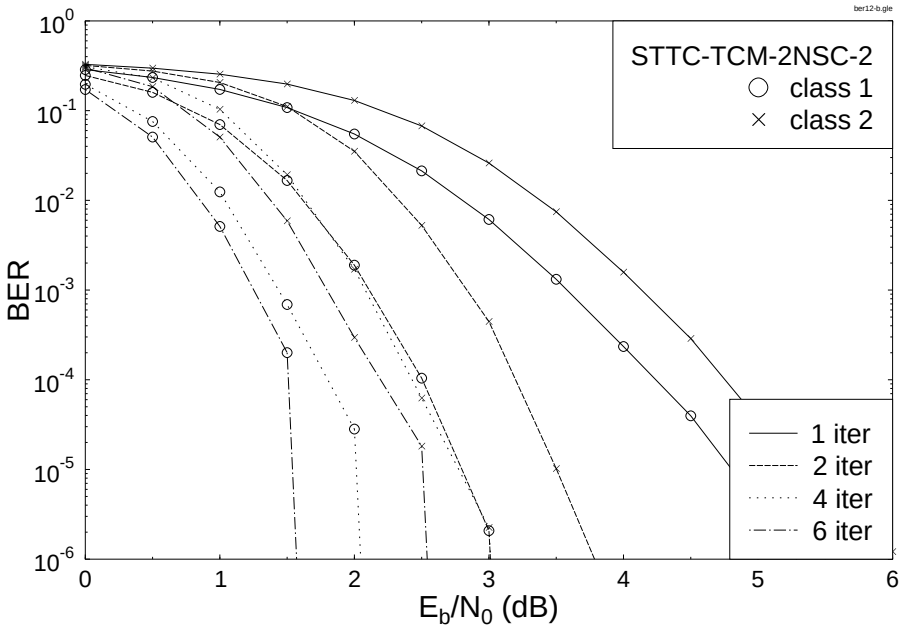

Figure 2.14: BER versus E_b/N_0 performance of the 16QAM-based STTC-TCM-2NSC-2 assisted MPEG-4 audio scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput is **2.07 BPS**.

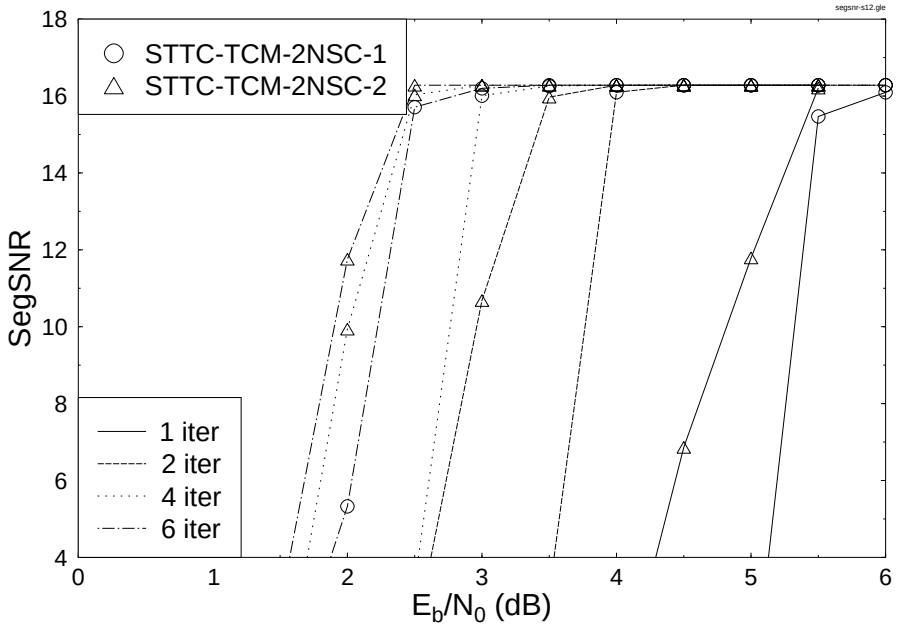

Figure 2.15: Average SegSNR versus E_b/N_0 performance of the 16QAM-based STTC-TCM-2NSC assisted MPEG-4 audio scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput of STTC-TCM-2NSC-1 and STTC-TCM-2NSC-2 is **1.77** and **2.07 BPS**, respectively.

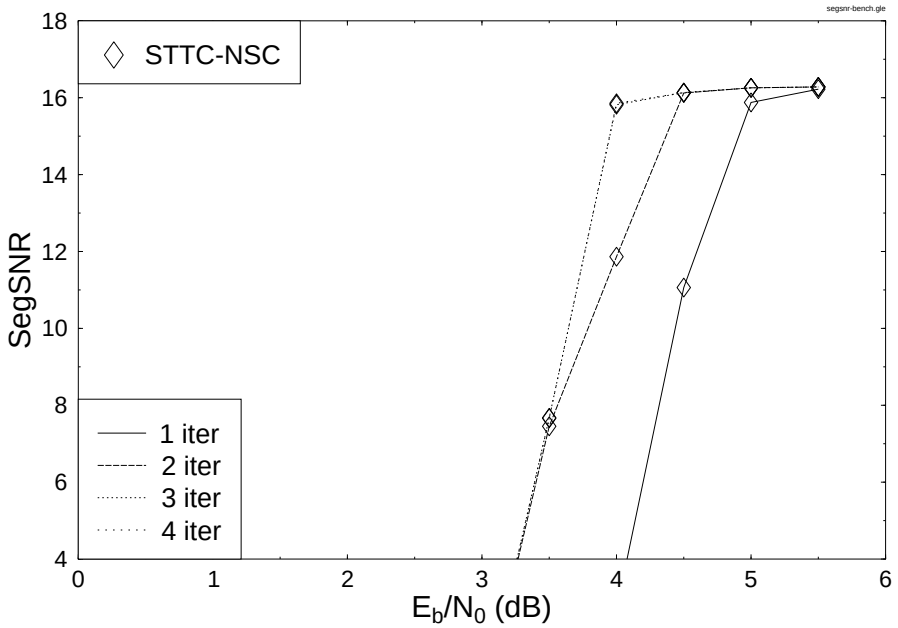

Figure 2.16: Average SegSNR versus E_b/N_0 performance of the 16QAM-based STTC-NSC assisted MPEG-4 audio benchmarker scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput is **1.99 BPS**.

Systems Parameters	**Value**	
Source Coding	MPEG-4 AAC	
Bit Rates (kbps)	48.00	
Audio/Speech Frame Length (ms)	23.22	
Sampling Rate(kHz)	44.1	
UEP Channel Coding	Non-Systematic Convolutional Code	
	STTC-TCM-2NSC-1	STTC-TCM-2NSC-2
The Code Rate and Code Memory of Class-1	R_1=1/2;L_1=3	R_1=2/3;L_1=3
The Code Rate and Code Memory of Class-2	R_2=3/4;L_2=3	R_2=3/4;L_2=3
Modulation Scheme	Trellis Coded Modulation (16QAM)	
The Code Rate, R_3	3/4	
The Code Memory, L_3	3	
MIMO Scheme	STTC	
Number of Transmitters, N_t	2	
Number of Receiver, N_r	2	
The Code Memory, L_4	4	

Table 2.5: System parameters of the STTC-TCM-2NSC-1 and the STTC-TCM-2NSC-2 assisted MPEG-4 AAC schemes.

2.5.3.1 Performance of Turbo-Detected Unequal Protection AAC Audio Transceiver

Figures 2.13 and 2.14 depict the BER versus SNR per bit, namely E_b/N_0, performance of the 16QAM-based STTC-TCM-2NSC-1 and STTC-TCM-2NSC-2 schemes, respectively, when communicating over uncorrelated Rayleigh fading channels. As we can observe from Figures 2.13 and 2.14, the gap between the BER performance of the class-1 and class-2 audio bits is wider for STTC-TCM-2NSC-1 compared to the STTC-TCM-2NSC-2 scheme. More explicitly, the class-1 audio bits of STTC-TCM-2NSC-1 have a higher protection at the cost of a lower throughput compared to the STTC-TCM-2NSC-2 scheme. However, the BER performance of the class-2 audio bits of the STTC-TCM-2NSC-1 arrangement is approximately 0.5 dB poorer than that of STTC-TCM-2NSC-2 at BER=10^{-5}.

Let us now study the audio SegSNR performance of the schemes in Figures 2.15 and 2.16. As we can see from Figure 2.15, the SegSNR performance of STTC-TCM-2NSC-1 is inferior in comparison to that of STTC-TCM-2NSC-2, despite providing a

Systems Parameters	**Value**	
Source Coding	MPEG-4 TwinVQ	AMR-WB
Bit Rates (kbps)	32.00	15.85
Audio/Speech Frame Length (ms)	23.22	20.00
Sampling Rate(kHz)	44.1	16.0
UEP Channel Coding	Non-Systematic Convolutional Code	
The Code Rate and Code Memory of Class-1	R_1=1/2;L_1=3	
The Code Rate and Code Memory of Class-2	R_2=3/4L_2=3	
Modulation Scheme	Trellis Coded Modulation (16QAM)	
The Code Rate, R_3	3/4	
The Code Memory, L_3	3	
MIMO Scheme	STTC	
Number of Transmitters, N_t	2	
Number of Receiver, N_r	2	
The Code Memory, L_4	4	

Table 2.6: System parameters of the STTC-TCM-2NSC-TVQ and the STTC-TCM-2NSC-AMR-WB schemes.

higher protection for the class-1 audio bits. More explicitly, STTC-TCM-2NSC-2 requires $E_b/N_0 = 2.5$ dB, while STTC-TCM-2NSC-1 requires $E_b/N_0 = 3$ dB, when having an audio SegSNR in excess of 16 dB after the fourth turbo iteration. Hence the audio SegSNR performance of STTC-TCM-2NSC-1 is 0.5 dB poorer than that of STTC-TCM-2NSC-2 after the fourth iteration. Note that the BER of the class-1 and class-2 audio bits for the corresponding values of E_b/N_0, SegSNR and iteration index is less than 10^{-7} and 10^{-4}, respectively, for the two different turbo schemes. After the sixth iteration, the SegSNR performance of both turbo schemes becomes quite similar since the corresponding BER is low. These results demonstrate that the MPEG-4 audio decoder requires a very low BER for both class-1 and class-2 audio bits, when aiming for a SegSNR above 16 dB.

Figure 2.16 portrays the SegSNR versus E_b/N_0 performance of the STTC-NSC audio benchmarker scheme, when communicating over uncorrelated Rayleigh fading channels. Note that if we reduce the code memory of the NSC constituent code of the STTC-NSC benchmarker arrangement from L_0=6 to 3, the achievable performance becomes poorer, as expected. If we increased L_0 from 6 to 7 (or higher), the decoding complexity would

increase 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. It is shown in Figures 2.15 and 2.16 that the first iteration based performance of the STTC-NSC benchmarker scheme is better than that of the proposed STTC-TCM-2NSC arrangements. However, at the same decoding complexity of 160 (240) trellis decoding states STTC-TCM-2NSC-2 having 4 (6) iterations performs approximately 2 (1.5) dB better than the STTC-NSC arrangement having 2 (3) iterations. The decoding complexity of the proposed STTC-TCM-2NSC arrangement is quantified in terms of the number of trellis decoding states using Equation (2.3), while that of the STTC-NSC arrangement is quantified using Equation (2.4), as defined in Section 2.5.2.3.

It is worth mentioning that other joint coding and modulation schemes directly designed for fading channels, such as for example Bit Interleaved Coded Modulation (BICM) [130–132], were outperformed by the TCM-based scheme, since the STTC arrangement rendered the error statistics more Gaussian-like [122].

2.5.3.2 Performance of Turbo-Detected Unequal Protection TwinVQ Audio and AMR-WB Speech Transceivers

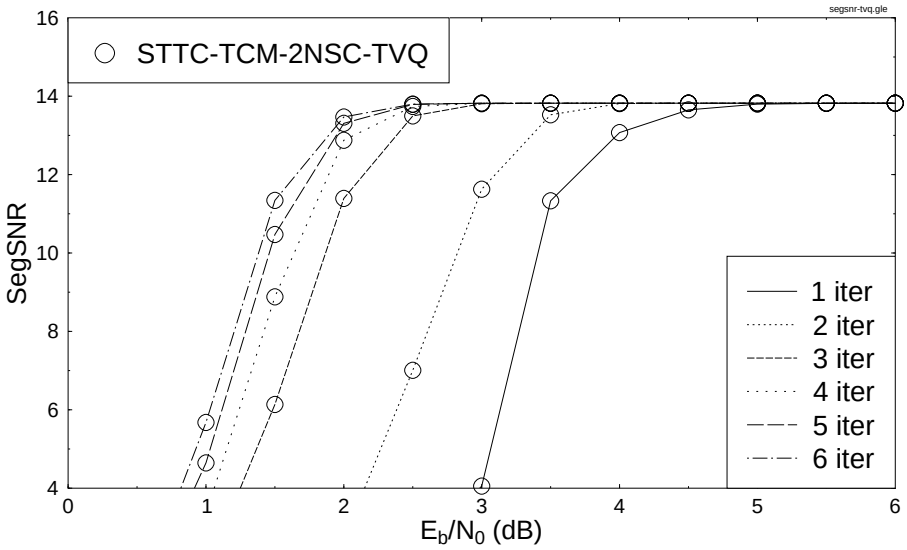

Figure 2.17: Average SegSNR versus E_b/N_0 performance of the 16QAM-based STTC-TCM-2NSC assisted MPEG-4 TwinVQ audio scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was **1.95 BPS**.

Figures 2.17 and 2.18 depict the audio SegSNR performance of the STTC-TCM-2NSC-

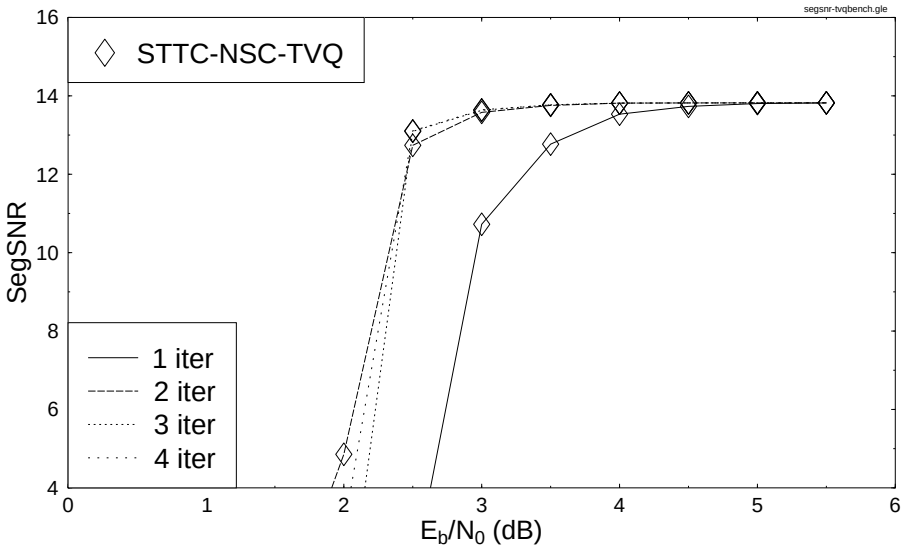

Figure 2.18: Average SegSNR versus E_b/N_0 performance of the 16QAM-based STTC-NSC assisted MPEG-4 TwinVQ audio benchmarker scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was **1.98 BPS**.

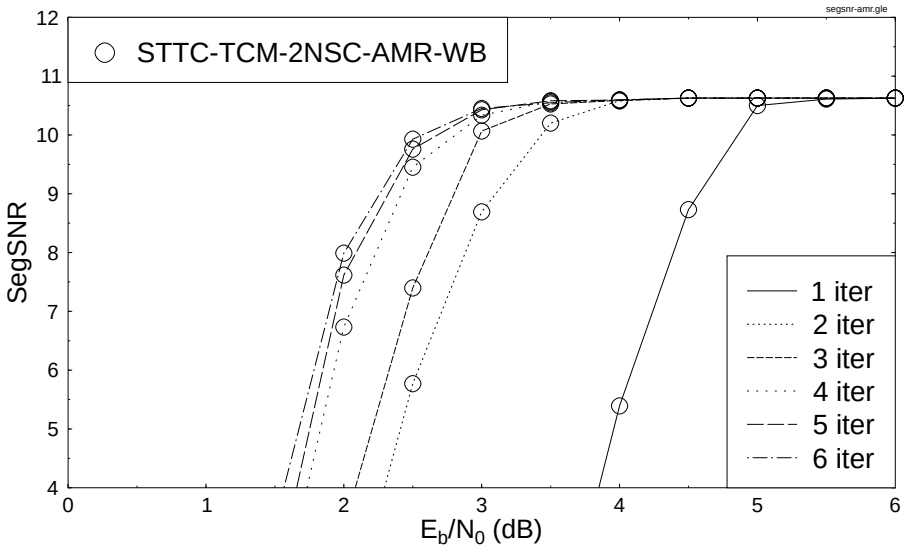

Figure 2.19: Average SegSNR versus E_b/N_0 performance of the 16QAM-based STTC-TCM-2NSC assisted AMR-WB speech scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was **1.89 BPS**.

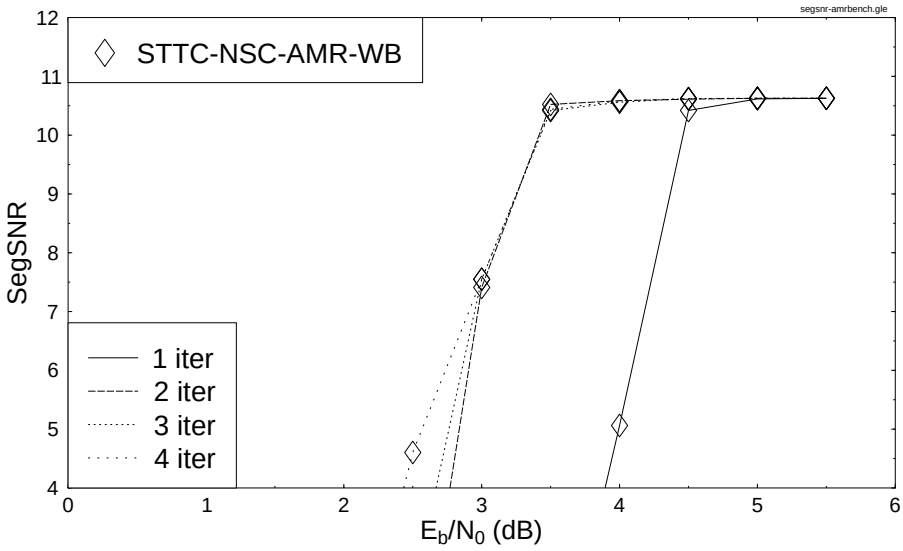

Figure 2.20: Average SegSNR versus E_b/N_0 performance of the 16QAM-based STTC-NSC assisted AMR-WB speech benchmarker scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was **1.95 BPS**.

TVQ and that of its corresponding STTC-NSC-TVQ benchmarker schemes, respectively, when communicating over uncorrelated Rayleigh fading channels. It can be seen from Figures 2.17 and 2.18 that the non-iterative single-detection based performance of the STTC-NSC-TVQ benchmarker scheme is better than that of the STTC-TCM-2NSC assisted MPEG-4 TwinVQ audio scheme. However, at the same decoding complexity quantified in terms of the number of trellis decoding states the STTC-TCM-2NSC-TVQ arrangement performs approximately 0.5 dB better in terms of the required channel E_b/N_0 value than the STTC-NSC-TVQ benchmarker scheme, both exhibiting a SegSNR of 13.8 dB. For example, at the decoding complexity of 160 trellis decoding states, this corresponds to the STTC-TCM-2NSC-TVQ scheme's 4th iteration, whilst in the STTC-NSC-TVQ scheme this corresponds to the 2nd iteration. Therefore, we observe in Figures 2.17 and 2.18 that the STTC-TCM-2NSC-TVQ arrangement performs by 0.5 dB better in terms of the required channel E_b/N_0 value than its corresponding benchmarker scheme.

Similarly, it can be observed from Figures 2.19 and 2.20 that at the decoding complexity of 160 trellis decoding states the STTC-TCM-2NSC-AMR-WB arrangement performs 0.5 dB better in terms of the required channel E_b/N_0 value than the STTC-NSC-AMR-WB scheme, when targetting a SegSNR of 10.6 dB. By comparing Figures 2.17 and 2.19, we observe that the SegSNR performance of the STTC-TCM-2NSC-AMR-WB scheme is inferior in comparison to that of STTC-TCM-2NSC-TVQ.

More explicitly, the STTC-TCM-2NSC-TVQ system requires an E_b/N_0 value of 2.5 dB, while the STTC-TCM-2NSC-AMR-WB arrangement necessitates $E_b/N_0 = 3.0$ dB, when having their respective maximum attainable average SegSNRs. The maximum attainable average SegSNRs for STTC-TCM-2NSC-TVQ and STTC-TCM-2NSC-AMR-WB are 13.8 dB and 10.6 dB, respectively.

This discrepancy is due to the reason that both schemes map the most sensitive 25% of the encoded bits to class-1. By contrast, based on the bit error sensitivity study of the MPEG-4 TwinVQ codec outlined in Section 2.3.4, only 10% of the MPEG-4 TwinVQ encoded bits were found to be gravely error sensitive. Therefore, the 25% class-1 bits of the MPEG-4 TwinVQ also includes some bits, which were found to be only moderately sensitive to channel errors. However, in the case of the AMR-WB codec all the bits of the 25%-partition were found to be quite sensitive to channel errors. Furthermore, the frame length of the STTC-TCM-2NSC-TVQ scheme is longer than that of the STTC-TCM-2NSC-AMR-WB arrangement and hence benefits from a higher coding gain.

2.5.4 Conclusions on UEP Schemes Using Convolutional Codes

In conclusion, a jointly optimised audio source-coding, outer twin-class NSC channel-coding, inner TCM and spatial diversity aided STTC turbo transceiver was proposed for employment in a MPEG-4 wireless audio transceiver. Furthermore, we comparatively studied the performance of the MPEG-4 TwinVQ and AMR-WB audio/speech codecs incorporated in our sophisticated unequal-protection turbo transceivers.

With the aid of two different-rate NSCs the audio and the speech bits were protected differently according to their error 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 joint decoding structure. The MPEG-4 AAC audio decoder was found to require a very low BER for both classes of audio bits in order to attain a perceptually pleasing, artefact-free audio quality. On one hand, the proposed twin-class STTC-TCM-2NSC assisted MPEG-4 AAC audio scheme performs approximately 2 dB better in terms of the required E_b/N_0 than the single-class STTC-NSC assisted MPEG-4 AAC audio benchmarker. On the other hand, both proposed twin-class STTC-TCM-2NSC assisted MPEG-4 TwinVQ audio and AMR-WB speech schemes perform approximately 0.5 dB better in terms of the required E_b/N_0 than the corresponding single-class STTC-

NSC assisted MPEG-4 TwinVQ audio and AMR-WB speech benchmarker schemes. This relatively modest advantage of the twin-class protected transceiver was a consequence of having a rather limited turbo-interleaver length. This can be observed in the longer interleaver length of the MPEG-4 AAC audio transceiver, where an approximately 2 dB E_b/N_0 gain was achieved. For a longer-delay non-realtime audio streaming scheme a similar performance would be achieved to that of the video system proposed in [127].

2.6 Unequal Protection Schemes Using Irregular Convolutional Codes [2]

2.6.1 System Background

Source encoded information sources, such as speech, audio or video, typically exhibit a non-uniform error sensitivity, where the effect of a channel error may significantly vary from one bit to another [1, 6]. Hence UEP is applied to ensure that the perceptually more important bits benefit from more powerful protection. In [22], the speech bits were protected by a family of RCPC codes [23] whose error protection capabilities had been matched to the bit-sensitivity of the speech codec. Different-rate RCPC codes were obtained by puncturing the same mother code, while satisfying the rate-compatibility restriction. However, they were not designed in the context of turbo detection. Other schemes using a serially concatenated system and turbo processing were proposed in [79, 127], where the UEP was provided by two different-rate convolutional codes.

Recently, Tüchler *et al.* [93,94] studied the construction of IRCCs and proposed several design criteria. These IRCCs consisted of a family of convolutional codes having different code rates and were specifically designed with the aid of EXIT charts [65] invoked, for the sake of improving the convergence behaviour of iteratively decoded serially concatenated systems. In general, EXIT chart analysis assumes having a long interleaver block length. However, it was shown in [94] that by using an appropriate optimization criterion, the concatenated system is capable of performing well even for short interleaver block lengths. Since the constituent codes have different coding rates, the resultant IRCC is capable of providing UEP.

[2] This section is based on J. Wang, N. S. Othman, J. Kliewer, L-L. Yang and L. Hanzo: Turbo Detected Unequal Error Protection General Configuration Irregular Convolutional Codes Designed for the Wideband Advanced Multirate Speech Codec, Proceedings of IEEE 62nd Vehicular Technology Conference, Texas, USA, September 2005, pp. 1044-1048, and it was based on collaborative research with the co-authors.

In this section we present a novel jointly optimised turbo transceiver based on EXIT charts. The turbo transceiver advocated consists of a serially concatenated IRCC scheme used as the outer code for the transmission of AMR-WB-encoded speech. Rather than being decoded separately, the constituent codes of the IRCC are decoded jointly and iteratively by exchanging *extrinsic* information with the inner code. The IRCC is optimized to match the characteristics of both the speech source codec and those of the channel, so that UEP is achieved while maximizing the iteration gain attained.

The bit error sensitivity study of the AMR-WB speech codec was provided in Section 2.4.2. The system overview of the UEP turbo transceiver using the AMR-WB codec is provided in Section 2.6.2. Section 2.6.3 describes the design procedure of IRCCs, followed by an application example in Section 2.6.4. The achievable system performance is characterised in Section 2.6.5, while our conclusions are offered in Section 2.6.6.

2.6.2 System Overview

Figure 2.21 shows the system's schematic diagram. At the transmitter, each of the k-bit speech frame is protected by a serially concatenated channel code consisting of an outer code (Encoder I) and an inner code (Encoder II) before transmission over the channel, resulting in an overall coding rate of R. At the receiver, iterative decoding is performed with advent of *extrinsic* information exchange between the inner code (Decoder II) and the outer code (Decoder I). Both decoders employ the *A Posteriori* Probability (APP) decoding algorithm, e.g., the BCJR algorithm [53]. After F number of iterations, the speech decoder is invoked in order to reconstruct the speech frame.

According to the design rules of [133], the inner code of a serially concatenated system should be recursive to enable iterative decoding convergence towards the Maximum Likelihood (ML) decoding performance. Furthermore, it has been shown in [134] that for Binary Erasure Channels (BECs) and block lengths tending to infinity the inner code should have rate-1 to achieve capacity. Experiments have shown that this approximately holds also for AWGN channels [93, 94]. For the sake of simplicity, we opted for employing a memory-1 recursive convolutional code having a generator polynomial of $1/(1 + D)$, which is actually a simple accumulator. Hence the decoding complexity of the inner code is extremely low. In the system advocated, we use an IRCC as the outer code, while in the benchmarker system, we use a regular non-systematic convolutional (NSC) code as the outer code. Binary Phase Shift Keying (BPSK) modulation and encountering an AWGN

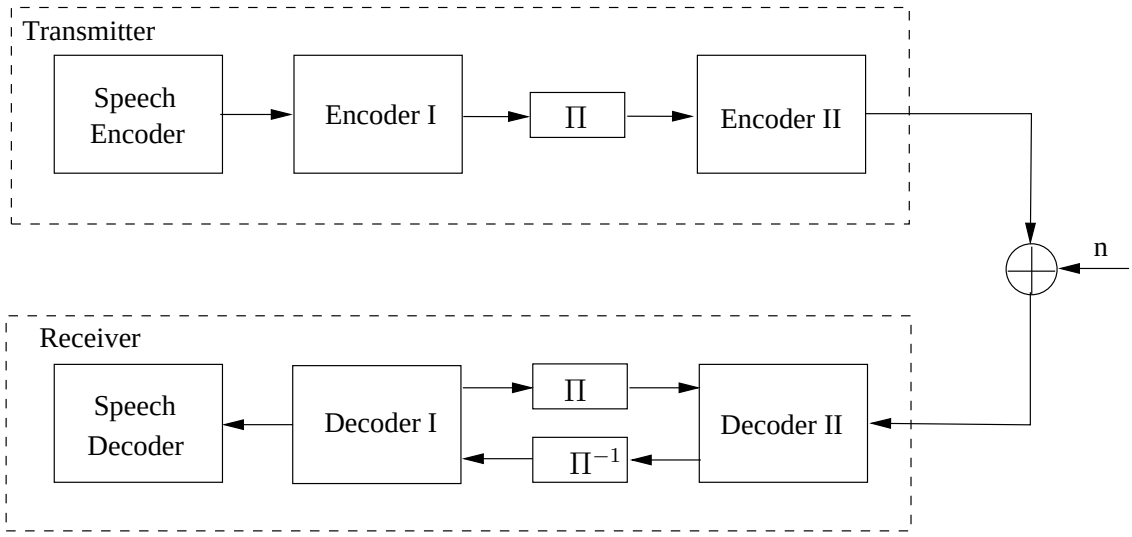

Figure 2.21: System Model

channel are assumed.

2.6.3 Design of Irregular Convolutional Codes

An IRCC is constructed from a family of P subcodes. First, a rate-r convolutional mother code C_1 is selected and the $(P-1)$ other subcodes C_i of rate $r_i > r$ are obtained by puncturing. Let n denote the total number of encoded bits generated from the k input information bits. Each subcode encodes a fraction of $\alpha_i r_i n$ information bits and generates $\alpha_i n$ encoded bits. Given the target code rate of $R \in [0,1]$, the weighting coefficient α_i has to satisfy:

$$1 = \sum_{i=1}^{P} \alpha_i, \; R = \sum_{i=1}^{P} \alpha_i r_i, \text{ and } \alpha_i \in [0,1], \; \forall i. \tag{2.5}$$

For example, in [94] a family of $P = 17$ subcodes were constructed from a systematic, rate-1/2, memory-4 mother code defined by the generator polynomial $(1, g_1/g_0)$, where $g_0 = 1 + D + D^4$ is the feedback polynomial and $g_1 = 1 + D^2 + D^3 + D^4$ is the feedforward one. Higher code rates may be obtained by puncturing, while lower rates are created by adding more generators and by puncturing under the constraint of maximizing the achievable free distance. The two additional generators used are $g_2 = 1 + D + D^2 + D^4$ and $g_3 = 1 + D + D^3 + D^4$. The resultant 17 subcodes have coding rates spanning from $0.1, 0.15, 0.2, \cdots$, to 0.9.

The IRCC constructed has the advantage that the decoding of all subcodes may be performed using the same mother code trellis, except that at the beginning of each block of $\alpha_i r_i n$ trellis sections corresponding to the subcode C_i, the puncturing pattern has to be restarted. Trellis termination is necessary only after all of the k information bits have been encoded.

We now optimize the iterative receiver by means of EXIT charts [65], which is capable of predicting the performance of an iterative receiver by examining the *extrinsic* information transfer function of each of the component devices independently.

For the outer decoder (Decoder I), denote the mutual information between the a *priori* input A and the transmitted code bits C as $I_{A1} = I(C; A)$, while the mutual information between the *extrinsic* output E and the transmitted code bits C is denoted as $I_{E1} = I(C; E)$. Then the transfer function of Decoder I can be defined as:

$$I_{E1} = T_I(I_{A1}), \tag{2.6}$$

which maps the input variable I_{A1} to the output variable I_{E1}. Similarly, for the inner decoder (Decoder II), we denote the mutual information between the a *priori* input A and the transmitted information bits X as $I_{A2} = I(X; A)$. Furthermore, we denote the mutual information between the *extrinsic* output E and the transmitted information bits X as $I_{E2} = I(X; E)$. Note that the *extrinsic* output of the inner code also depends on the channel SNR or E_b/N_0. Hence the transfer function of the inner code is defined as

$$I_{E2} = T_{II}(I_{A2}, E_b/N_0). \tag{2.7}$$

The transfer functions can be obtained by using the histogram-based LLR measurements as proposed in [65] or the simplified method as proposed in [135].

When using IRCCs, the transfer function of an IRCC can be obtained from those of its subcodes. Denote the transfer function of the subcode i as $T_{I,i}(j)$. Assuming that the trellis fractions of the subcodes do not significantly interfere with each other, which might change the associated transfer characteristics, the transfer function $T_I(j)$ of the target IRCC is the weighted superposition of the transfer function $T_{I,i}(j)$ [94], yielding,

$$T_I(j) = \sum_{i=1}^{P} \alpha_i T_{I,i}(j). \tag{2.8}$$

Note that in iterative decoding, the *extrinsic* output $E2$ of Decoder II becomes the a *priori* input $A1$ of Decoder I and vice versa. Given the transfer function, $T_{II}(j, E_b/N_0)$,

of the inner code, and that of the outer code $T_I(j)$, the *extrinsic* information I_{E1} at the output of Decoder I after the jth iteration can be calculated using the recursion of:

$$\mu_j = T_I(T_{II}(\mu_{j-1}, E_b/N_0)), \quad j = 1, 2, \ldots, \tag{2.9}$$

with $\mu_0 = 0$, i.e., assuming the absence of a *priori* input for Decoder II at the commencement of iterations.

Generally, interactive speech communication systems require a low delay, and hence a short interleaver block length. And the number of iterations for the iterative decoder is also limited due to the constraint of complexity. It has been found [94] that EXIT charts may provide a reasonable convergence prediction for the first couple of iterations even in the case of short block lengths. Hence, we fixed the transfer function of the inner code for a given E_b/N_0 value yielding $T_{II}(j) = T_{II}(j, E_b/N_0)$, and optimized the weighting coefficients $\{\alpha_i\}$ of the outer IRCC for the sake of obtaining a transfer function $T_I(j)$ that specifically maximizes the *extrinsic* output after exactly F number of iterations [94], which is formulated as:

$$\text{maximize} \quad \mu_j = T_I(T_{II}(\mu_{j-1})), \quad j = 1, 2, \ldots, F, \tag{2.10}$$

with $\mu_0 = 0$.

Additionally, considering the non-uniform error sensitivity of the speech source bits, we may intentionally enhance the protection of the more sensitive source data bits by using strong subcodes, thus imposing the source constraints of:

$$\sum_{i=i_1}^{i_2} \alpha_i r_i / R \geq x\%, \quad 1 \leq i_1 \leq i_2 \leq P, \quad 0 \leq x \leq 100, \tag{2.11}$$

which implies that the percentage of the speech source bits protected by the subcodes i_1 to i_2 is at least $x\%$.

Finally, our task is to find a weight vector $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_P]^T$, so that Equation (2.10) is maximized, while satisfying the constraints of Equations (2.5) and (2.11). This optimization problem can be solved by slightly modifying the procedure proposed in [94], as it will be illustrated by the following example.

2.6.4 An Example of Irregular Convolutional Codes

We assume the overall system coding rate to be $R = 0.5$. As stated in Section 2.6.2, the inner code has a unitary code rate, hence all the redundancy is assigned to the outer code.

We use a half-rate, memory-4, maximum free distance NSC code having the generator polynomials of $g_0 = 1 + D + D^2 + D^4$, and $g_1 = 1 + D^3 + D^4$. The *extrinsic* information transfer functions of the inner code and the outer NSC code are shown in Figure 2.22. It can be seen that the minimum convergence SNR threshold for the benchmarker system using the NSC outer code is about 1.2 dB, although we note that these curves are based on the assumption of having an infinite interleaver length and a Gaussian LLR distribution. In the case of short block lengths, the actual SNR convergence threshold might be higher.

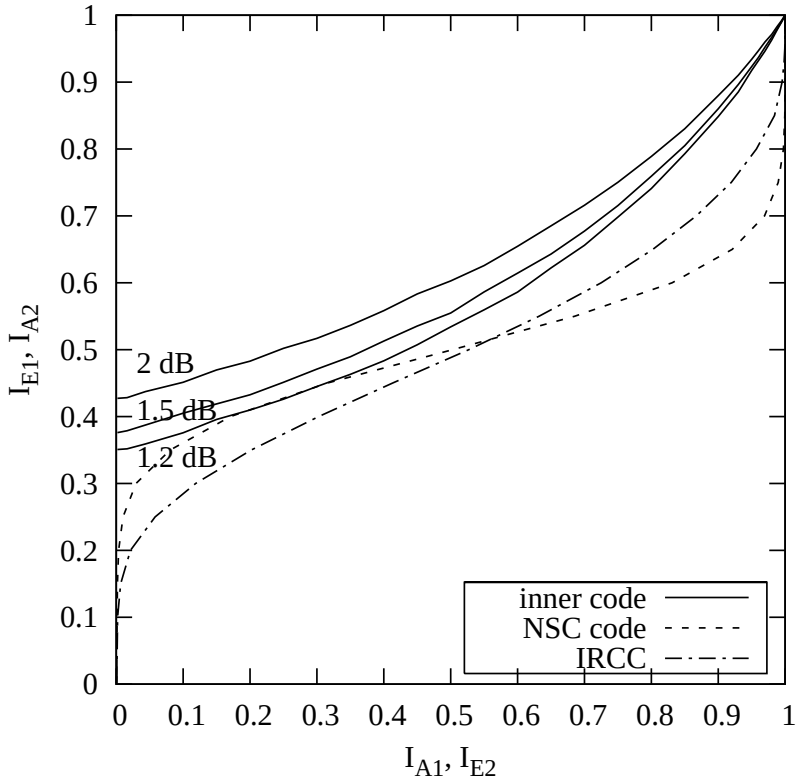

Figure 2.22: *Extrinsic* information transfer functions of the outer NSC code and the designed IRCC, as well as those of the inner code at $E_b/N_0 = 1.2$, 1.5 and 2 dB.

Hence, when constructing the IRCC, we choose the target inner code transfer function $T_{II}(j)$ at $E_b/N_0 = 1.5$ dB, and the number of iterations $F = 6$. For the constituent subcodes, we use those proposed in [94] except that code rates of $r_i > 0.75$ are excluded from our design for the sake of avoiding significant error floors. The resultant code rates of the subcodes span the range of $r_1 = 0.1, r_2 = 0.15, \ \cdots, \ r_{14} = 0.75$.

Initially the source constraint of Equation (2.11) was not imposed. By using the op-

timization procedure of [94], we arrive at the weight vector of $\boldsymbol{\alpha}_0 = [0\ 0\ 0\ 0\ 0.01\ 0.13\ 0.18\ 0.19\ 0.14\ 0.12\ 0.10\ 0.01\ 0.03\ 0.10]^T$, and the percentage of the input speech data bits protected by the different subcodes becomes [0, 0, 0, 0, 0.6%, 9.0%, 14.4%, 16.7%, 14.0%, 13.0%, 11.5%, 1.6%, 4.2%, 15.0%]^T. The *extrinsic* output of Decoder I after 6 iterations becomes $\mu_6 = 0.98$.

Observe in the context of the vector containing the corresponding speech bit fractions that only 0.6% of the source bits are protected by the $r_5 = 0.3$-rate subcode, whereas a total of 23.4% of the speech bits is protected by the $r_6 = 0.35$ and $r_7 = 0.4$-rate subcodes. In order to enhance the protection of the more sensitive speech bits, we impose now the source constraint of Equation (2.11) by requiring all the header information bits in a speech frame to be protected by the relatively strong $r_5 = 0.3$-rate subcode. More explicitly, we impose the constraint of $\alpha_5 r_5/0.5 \geq 7\%$, resulting in a new weight vector of $\boldsymbol{\alpha}_1 = [0\ 0\ 0\ 0\ 0.12\ 0.06\ 0.14\ 0.16\ 0.13\ 0.12\ 0.10\ 0.02\ 0.04\ 0.11]^T$, and the new vector of speech bit fractions becomes [0, 0, 0, 0, 7.1%, 4.0%, 10.9%, 14.8%, 13.5%, 13.3%, 12.2%, 2.7%, 5.5%, 16%]^T. The *extrinsic* output after 6 iterations is now slightly reduced to $\mu_6 = 0.97$, which is close to the maximum value of 0.98. Furthermore, now, 14.9% of the speech bits are protected by the $r_6 = 0.35$ and $r_7 = 0.4$-rate subcodes.

The *extrinsic* information transfer function of this IRCC is also shown in Figure 2.22. As seen from the EXIT chart, the convergence SNR threshold for the system using the IRCC is lower than 1.2 dB and there is a wider EXIT chart tunnel between the inner code's curve and the outer code's curve which is particularly so at the low I_A values routinely encountered during the first couple of iterations. Hence, given a limited number of iterations, we would predict that the system using the IRCC may be expected to perform better than that using the NSC outer code in the range of $E_b/N_0 = 1.5{\sim}2.5$ dB.

2.6.5 Performance of UEP Schemes Using Irregular Convolutional Codes

Finally, the achievable system performance was evaluated for a $k = 340$ speech bit per 20 ms transmission frame, resulting in an interleaver length of $n = 688$ bits, including 8 tail bits. This wideband-AMR speech coded [110] frame was generated at a bit rate of 15.85 kbps in the codec's mode 4. Before channel encoding, each frame of speech bits is rearranged according to the descending order of the error sensitivity of the bits, so that the more important data bits are protected by stronger IRCC subcodes. An S-random interleaver [136] was employed with $S = 15$, where all of the subcodes' bits are interleaved

together, and 10 iterations were performed by the iterative decoder.

The BER performance of the UEP system using IRCCs and that of the Equal Error Protection (EEP) benchmarker system using the NSC code are depicted in Figure 2.23. It can be seen that the UEP system outperforms the EEP system in the range of $E_b/N_0 = 1.5 \sim 2.5$ dB, which matches our performance prediction inferred from the EXIT chart analysis of Section 2.6.3.

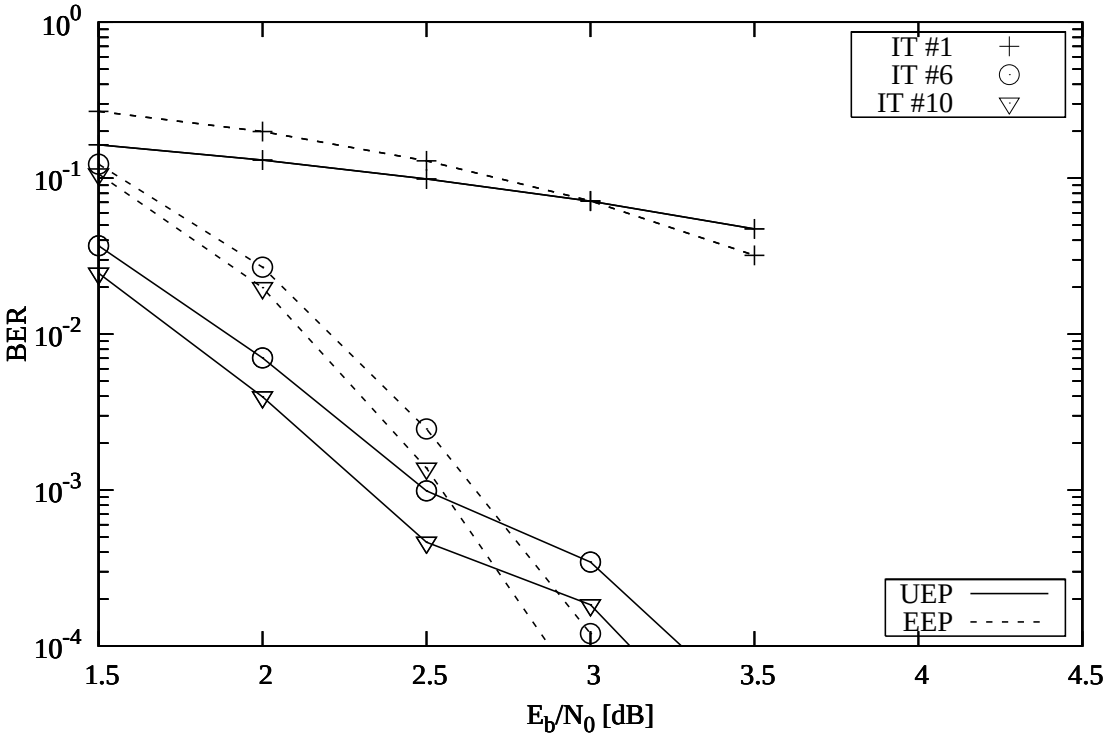

Figure 2.23: BER performance of both the UEP system employing the IRCC and the EEP system using the NSC code

The actual decoding trajectories of both the UEP system and the EEP system recorded at E_b/N_0=1.5 and 2 dB are shown in Figure 2.24 and 2.25, respectively. These are obtained by measuring the evolution of mutual information at the input and output of both the inner decoder and the outer decoder as the iterative decoding algorithm is simulated. Due to the relatively short interleaver block length of 688 bits, the actual decoding trajectories do not closely follow the transfer functions especially when increasing the number of iterations. Nonetheless, the UEP system does benefit from having a wider open tunnel during the first

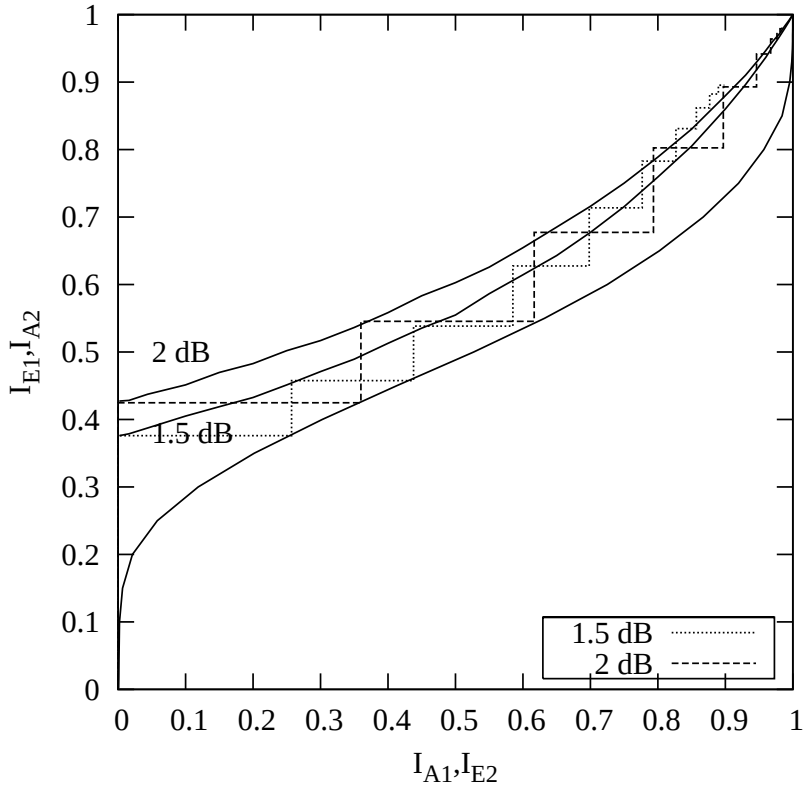

Figure 2.24: The EXIT chart and the simulated decoding trajectories of the UEP system using our IRCC as the outer code and a rate-1 recursive code as the inner code at both E_b/N_0=1.5 and 2 dB

couple of iterations and hence it is capable of reaching a higher *extrinsic* output in the end, resulting in a lower BER. On the other hand, observe in Figure 2.25 that the EXIT curve of the NSC decoder closely approaches the point of $I_{E2} = 1.0$ at $I_{A2} \approx 0.8$, while that of the IRCC decoder requires $I_{A1} = 1.0$ in order to reach the point of $I_{E1} = 1.0$, as shown in Figure 2.24. Therefore, due to the short-interleaver length the BER performance of the EEP system becomes superior in comparison to that of the UEP system for $E_b/N_0 >$ 2.5 dB, where both systems exhibit the benefit of having a wide open EXIT tunnel during the first couple of iterations. However, the audio quality is not severely affected at low BERs, particularly since the UEP scheme's errors are likely to be in the less protected, but less error sensitive class.

The BER profiles of the UEP system at E_b/N_0=1.5, 2 and 2.5 dB are plotted in Figure 2.26. As intended, different fractions of the speech frame benefitted from different degrees of IRCC-aided protection. The first 60 bits represent the header information bits and the most sensitive speech bits, which require the lowest BER. The SegSNR

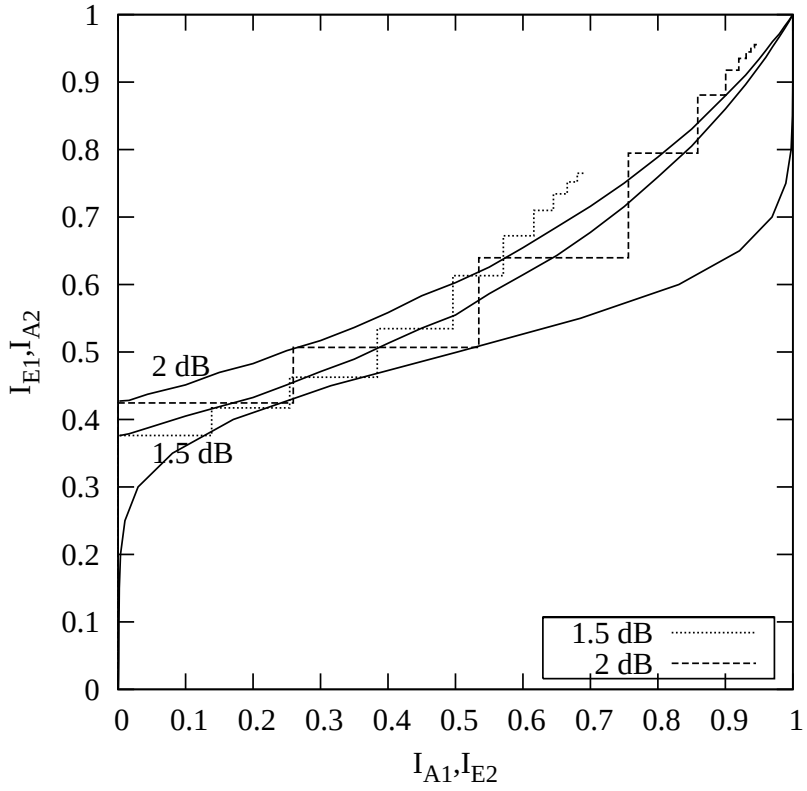

Figure 2.25: The EXIT chart and the simulated decoding trajectories of the EEP system using our NSC code as the outer code and a rate-1 recursive code as the inner code at both E_b/N_0=1.5 and 2 dB

performances of both the UEP and EEP system are depicted in Figure 2.27. The UEP system is seen to outperform the EEP system at $E_b/N_0 \leq 2.5$ dB. Above this E_b/N_0 point, the two systems attained almost the same SegSNRs. To achieve a good speech quality associated with SegSNR>9 dB, the UEP system requires $E_b/N_0 \geq 2$ dB, about 0.3 dB less than the EEP system.

2.6.6 Conclusions on UEP Schemes Using Irregular Convolutional Codes

We investigated the application of IRCCs for the sake of providing UEP for the AMR-WB speech codec. The IRCCs were optimized with the aid of EXIT charts and the design procedure used was illustrated with the aid of an example.

In the design of IRCCs, we aimed for matching the *extrinsic* information transfer function of the outer IRCC to that of the inner code, where that of the latter is largely determined by the channel SNR. At the same time, we imposed certain source constraints

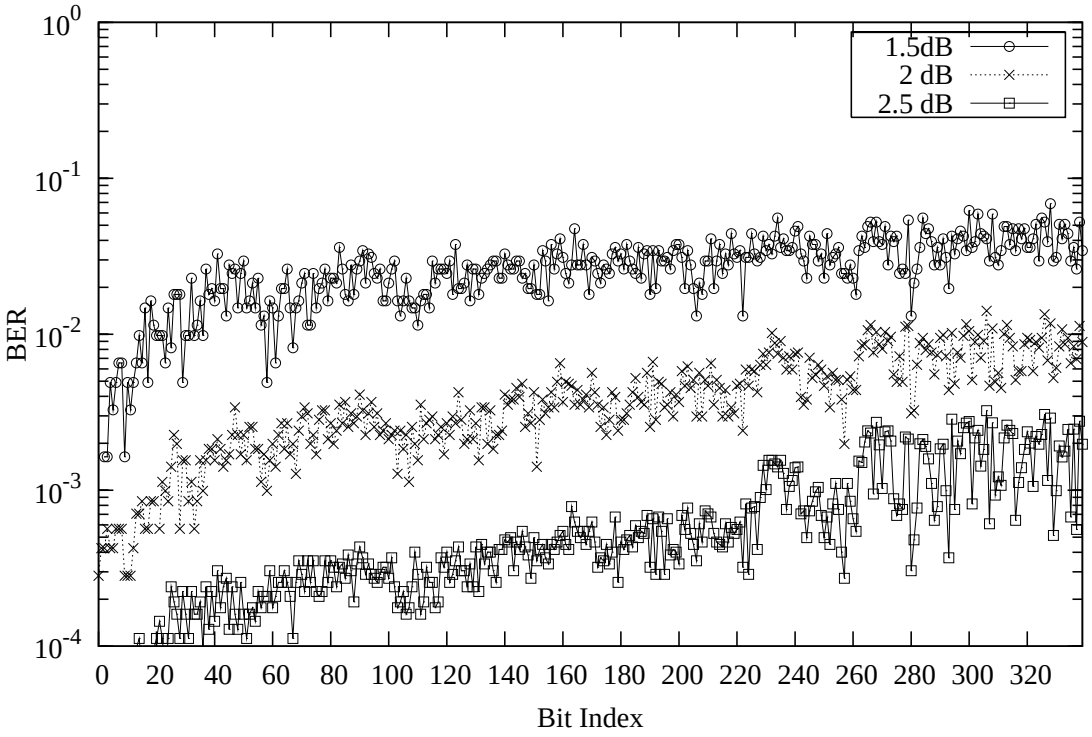

Figure 2.26: Bit error rate of the different speech bits after ten iterations at both E_b/N_0=1.5, 2 and 2.5dB recorded by transmitting 10^5 speech frames.

determined by the error sensitivity of the AMR-WB source bits. Hence the design method proposed here may be viewed as an attractive joint source/channel codec optimization.

The concatenated system using an IRCC benefits from having a low convergence SNR threshold. Owing to its design flexibility, various transfer functions can be obtained for an IRCC. We have shown that our IRCC was capable of achieving better convergence than a regular NSC code having the same constraint length and code rate. Hence the system using IRCCs has the potential of outperforming the corresponding arrangement using regular NSC codes in the low SNR region.

Furthermore, IRCCs are capable of providing UEP, since it is constituted by various subcodes having different code rates and hence different error protection capabilities. Multimedia source information, such as speech, audio and video source can benefit from this property, when carefully designing the IRCC to match the source's bit sensitivity. Our future research aims for exchanging soft speech bits between the speech and channel

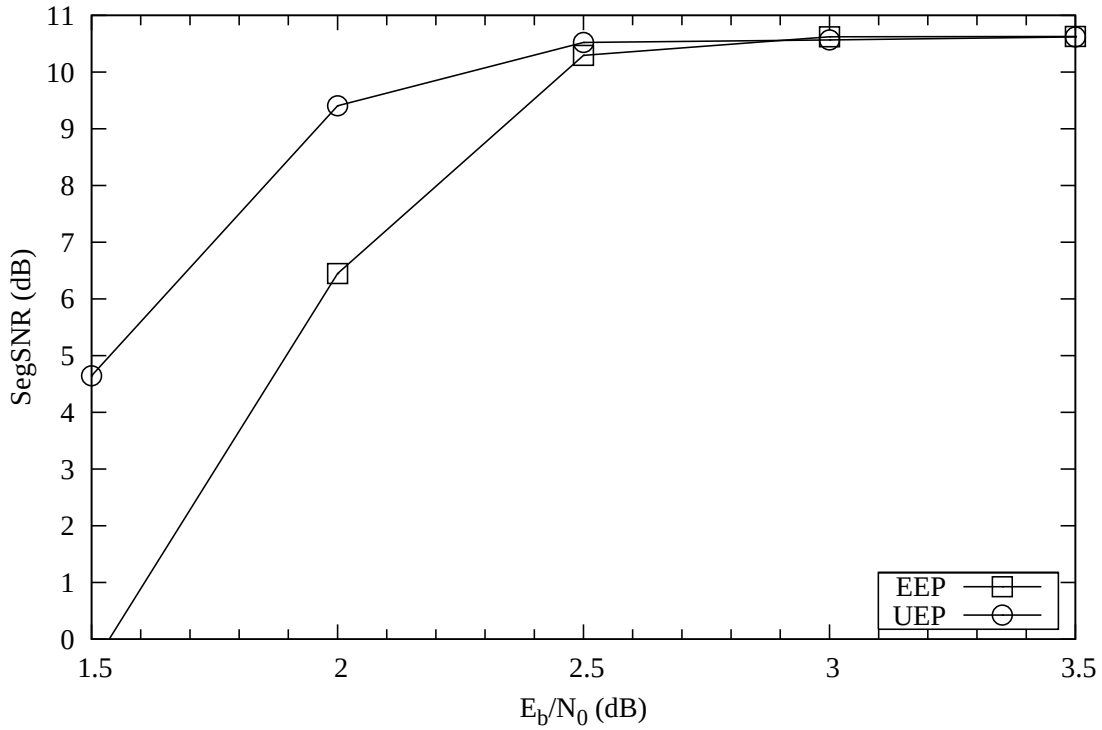

Figure 2.27: Comparison of SegSNRs of the AMR-WB speech codec using both EEP and UEP decoders.

2.7 Chapter Conclusion

In this chapter, we proposed jointly optimised turbo transceivers capable of providing unequal error protection for employment in audio and speech transceivers. The transceiver advocated consists of a serially concatenated STTC, TCM and two different-rate NSCs used for unequal twin-class error protection. In our investigations, the MPEG-4 AAC, the MPEG-4 TwinVQ and the AMR-WB audio/speech codecs were incorporated in the sophisticated UEP turbo transceiver of Figure 2.11, where the bit error sensitivity of each parameter was investigated in Section 2.3.3, as well as in Figures 2.7 and 2.10 of Section 2.3.4 and Section 2.4.2, respectively. Consequently, the audio/speech bits were protected differently according to their error sensitivity with the aid of two different-rate NSCs. A benchmarker scheme combining a STTC scheme and a single-class NSC code

was used for comparison with the proposed scheme. The audio/speech performance of the scheme was evaluated, when communicating over uncorrelated Rayleigh fading channels. Our investigations presented in Figures 2.15, 2.17 and 2.19 demonstrated that the proposed twin-class STTC-TCM-2NSC assisted audio/speech codec schemes outperform the single-class STTC-NSC benchmark schemes, when targetting the maximum attainable average SegSNRs. The maximum attainable average SegSNRs recorded in Figures 2.15, 2.17 and 2.19 for the STTC-TCM-2NSC assisted MPEG-4 AAC, MPEG-4 TwinVQ and AMR-WB audio/speech schemes are 16.28 dB, 13.8 dB and 10.6 dB, respectively. The corresponding system components, as well as the audio/speech SegSNR performance of the twin-class STTC-TCM-2NSC audio/speech schemes expressed in terms of their E_b/N_0 gain, are summarised in Table 2.7 when compared to their corresponding single-class STTC-NSC benchmarkers studied in this chapter. In Section 2.5.2.1 a comparative study of two different STTC-TCM-2NSC assisted MPEG-4 AAC was carried out. The class-1 audio bits of the STTC-TCM-2NSC-1 system are protected using a 1/2-rate NSC code, while those of the STTC-TCM-2NSC-2 arrangement are protected using a 2/3-rate NSC code. In Table 2.7, the STTC-TCM-2NSC-1 and the STTC-TCM-2NSC-2 schemes are referred to as System 1 and System 2 of the MPEG-4 AAC audio transceiver, respectively.

We can observe from Table 2.7 that System 1 and System 2 constructed of the twin-class STTC-TCM-2NSC assisted MPEG-4 AAC outperformed their corresponding single-class STTC-NSC audio/speech benchmarkers by about 1.5 dB and 2.0 dB, respectively, in terms of the required E_b/N_0, when aiming for their respective maximum attainable average SegSNRs. This was achieved at the cost of a similar complexity. The decoding complexity is quantified in terms of the number of trellis decoding states, as defined in Section 2.5.2.3. More explicitly, the class-1 audio bits of System 1 have a higher protection at the cost of a lower throughput compared to the System 2. It is also observed in Figure 2.15 that the audio SegSNR performance of System 1 is approximately 0.5 dB poorer than that of the System 2, as discussed in Section 2.5.3.1. On the other hand, in Section 2.5.2.2 a comparative study of the MPEG-4 TwinVQ and AMR-WB audio/speech schemes was carried out, where both of the proposed twin-class schemes outperformed their corresponding single-class STTC-NSC audio/speech benchmarkers by about 0.5 dB, in terms of the required E_b/N_0, when aiming for their respective maximum attainable average SegSNRs. Again, this was attained at the cost of a similar complexity. The relatively modest advantage of the twin-class protected transceiver using the MPEG-4 TwinVQ and AMR-WB audio/speech codecs was a consequence of having a rather limited turbo-interleaver length. As a benefit of the longer interleaver of the higher-rate audio

Source Codec	**MPEG-4 AAC**		**MPEG-4**	**AMR-WB**
	System 1	System 2	**TwinVQ**	
Bit Rates (kbps)	48.00		32.00	15.85
Interleaver Length (bits)	1116		743	317
Channel Codec	**Non-Sytematic Convolutional Code**			
Class1	$R_1 = 1/2$	$R_1 = 2/3$	$R_1 = 1/2$	$R_1 = 1/2$
	$L_1 = 3$	$L_1 = 3$	$L_1 = 3$	$L_1 = 3$
Class2	$R_2 = 3/4$			
	$L_2 = 3$			
Modulation Scheme	**Trellis Coded Modulation (16QAM)**			
The Code Rate	$R_3 = 3/4$			
The Code Memory	$L_3 = 3$			
MIMO Scheme	**STTC**			
Number of Transmitters	N_t=2			
Number of Receivers	N_r=2			
The Code Memory	$L_4 = 4$			
Effective Throughput (BPS)	1.77	2.07	1.95	1.98
E_b/N_0 **Gain** (dB)	1.5	2.0	0.5	0.5

Table 2.7: General overview of the UEP schemes using convolutional codes in Section 2.5 and their corresponding E_b/N_0 gain, when compared to their corresponding single-class STTC-NSC benchmarker audio/speech schemes. The remaining parameters of these schemes were summarised in Tables 2.5 and 2.6.

transceiver used for the MPEG-4 AAC codec, a significant performance improvement is achieved without increasing the system's audio-delay.

In Section 2.6, we proposed a turbo transceiver employing IRCCs for the sake of providing UEP for the AMR-WB speech codec. The EXIT chart based design procedure of Section 2.6.3 was used to design the IRCCs, with the aim of matching the *extrinsic* information transfer function of the outer IRCC to that of the inner code. More explicitly, an IRCC was constructed with the aid of EXIT charts, to ensure that its EXIT curve matched that of the inner code, which leads to an UEP system capable of converging at lower E_b/N_0 values in comparison to the EEP system, as discussed in Section 2.6.4. At the same time, joint source/channel coding was used for satisfying the source bit protection requirements of the AMR-WB codec. Our investigations demonstrated that the IRCC of

the UEP system was capable of achieving a better BER compared to the EEP system employing regular NSC codes, despite having the same constraint length and code rate. At the same time, the IRCC was capable of providing UEP, since its constituent codes have different coding rates and hence different error protection capabilities. As a benefit, the UEP aided system using IRCCs exhibits an SNR advantage of about 0.4 dB over the equal-protection system employing regular convolutional codes, when communicating over a Gaussian channel, at the point of tolerating a SegSNR degradation of 1 dB. Again, this relatively modest advantage of the IRCC protected transceiver was a consequence of having a rather limited turbo-interleaver length, as discussed in Section 2.6.5.

Chapter 3

Soft-Bit Assisted Iterative AMR-WB Source-Decoding

3.1 Introduction

In Chapter 2, joint source and channel coding schemes capable of providing unequal error protection were proposed for employment in the MPEG-4 Advanced Audio Coding (AAC), the MPEG-4 Transform-Domain Weighted Interleaved Vector Quantization (TwinVQ) and the Adaptive MultiRate WideBand (AMR-WB) audio/speech transceivers. The performance of the UEP turbo transceivers was shown to be better than that of the corresponding equal-protection benchmarker schemes, which was a benefit of protecting the audio/speech bits differently according to their error sensitivity. In this chapter, we present a particular joint source and channel coding philosophy which exploits the residual-redundancy inherent in the source encoded bitstream for enhancing the error resilience of the system by exchanging soft information amongst the various receiver components.

More specifically, a realistic, finite-delay lossy source codec leaves some residual redundancy in the encoded parameters, which is not the case for Shannon's ideal entropy codec. Vary and his team has developed [59,60] the innovative concept of soft speech bits, which exploits this residual redundancy, leading to the formulation of Iterative Souce and Channel Decoding (ISCD) [61]. The ISCD philosophy may be beneficially combined with iterative soft demapping in the context of multilevel modulation and amalgamated with a number of other sophisticated wireless transceiver components. The resultant multi-stage scheme consists of three serially concatenated decoder stages, namely the demodulator,

the channel decoder and the soft-input source decoder, in the spirit of [137].

On the other hand, in [17], Hagenauer proposed the so-called source-controlled channel coding approach, where the inherent residual redundancy of the source code is exploited during the channel decoding process for further error protection. Other examples of this approach can be found in [49] and [50], where the residual redundancy of both a Code Excited Linear Prediction (CELP) [4, 51] and of a Mixed-Excitation Linear Prediction (MELP) [52] speech codec was quantified and utilized during the channel decoding stage for enhancing the error resilience of the speech codecs. In [49], the residual redundancy inherent in the Line Spectral Pair (LSP) parameters of the CELP speech codec was exploited during channel decoding, which resulted in a significant speech quality improvement, both subjectively and objectively, especially during hostile channel conditions.

EXtrinsic Information Transfer (EXIT) charts have been proposed by ten Brink [65] for analysing the convergence behaviour of iterative decoding schemes, based on the exchange of mutual information amongst the constituent receiver components. Recently, the concept of EXIT chart analysis has been extended to three-stage concatenated codes [66,67], which require the employment of Three-Dimensional (3D) EXIT charts.

In Section 2.4, we briefly introduced the AMR-WB speech codec. In this chapter, the AMR-WB speech decoder of [110] is further developed for the sake of accepting the *a priori* information passed to it from the channel decoder as *extrinsic* information, in the spirit of [61]. More specifically, in this chapter we propose and investigate a novel system that invokes jointly optimised iterative source- and channel-decoding for enhancing the error resilience of the AMR-WB speech codec.

This chapter is structured as follows. In Section 3.2, the residual redundancy inherent in the AMR-WB encoded parameters is quantified. The proposed soft-bit assisted iterative AMR-WB source and channel decoding transceiver will be described in Section 3.3, where the motivation and the overview of our system will be provided. Furthermore, the convergence behaviour of the proposed scheme will be evaluated with the aid of both 3D and Two-Dimensional (2D) EXIT charts. Finally, the chapter is concluded in Section 3.4.

3.2 Residual Redundancy in the AMR-WB Speech Codec

The AMR-WB speech codec is capable of supporting nine different bit rates [11], each of which may be activated in conjunction with different-rate channel codecs and different-

Parameter	$\mathcal{R}_D = \mathcal{B} - H(\mathcal{U})$	Residual Redundancy Percentage
$ISP_1(t)$	0.57	7.1
$ISP_2(t)$	0.45	5.6
$ISP_3(t)$	0.10	1.7
$ISP_4(t)$	0.06	0.8
$ISP_5(t)$	0.14	2.0
$ISP_6(t)$	0.03	0.6
$ISP_7(t)$	0.07	1.4
$LTPLag_{sub_1}(t)$	0.23	2.6
$LTPLag_{sub_2}(t)$	0.40	6.6
$LTPLag_{sub_3}(t)$	0.20	2.2
$LTPLag_{sub_4}(t)$	0.39	6.5
$FixedInd._{sub_1}(t)$	0.33	3.0
$FixedInd._{sub_2}(t)$	0.31	2.8
$FixedInd._{sub_3}(t)$	0.31	2.8
$FixedInd._{sub_4}(t)$	0.31	2.8
$CBGain_{sub_1}(t)$	0.12	1.7
$CBGain_{sub_2}(t)$	0.12	1.7
$CBGain_{sub_3}(t)$	0.13	1.9
$CBGain_{sub_4}(t)$	0.14	2.0

Table 3.1: Unequal Redundancy in the AMR-WB codec's parameters.

throughput adaptive modem modes [138]. Similar near-instantaneously adaptive speech and video systems were designed in [1, 6]. In our prototype system investigated here the AMR-WB codec of Section 2.4 operated at 23.05 kbps, generates a set of speech parameters encoded by a total of 461 bits per 20 ms frame for representing the 8 kHz bandwidth speech signal sampled at 16 kHz. Similar to any other CELP based codecs [6], it performs Short-Term Prediction (STP), Long-Term Prediction (LTP) and generates the excitation CodeBook (CB) parameters [110]. The resultant bit-allocation scheme is summarized in Table 2.4.

The STP coefficients produced for the 20 ms speech frame are converted to the so-called Immittance Spectrum Pair (ISP) representation [112] and they are vector-quantised

using Split-MultiStage Vector Quantisation (S-MSVQ) producing 7 ISP parameters using 46 bits.

The LTP analysis [6] is performed for each new 5 ms subframe, which produces the LTP lag and gain parameters. The LTP lag is encoded using 9 bits in the first and third subframes. By contrast, in the second and fourth subframes it is differentially encoded with respect to the first and third subframes using 6 bits, respectively. The LTP and fixed CB gains are jointly vector-quantized using 7 bits per subframe, whilst the fixed excitation CB parameters of each 5 ms subframe are encoded using a total of 88 bits [11], as summarized in Table 2.4 of Section 2.4.

The AMR-WB codec includes a Voice Activity Detector (VAD), which indicates whether a 20 ms frame contains an active speech spurt that should be encoded and transmitted. The binary VAD-flag indicates the presence of an active speech spurt, if it is set to "1". In the Global System of Mobile Communications (GSM), the AMR-WB codec includes the so-called Discontinuous Transmission (DTX) functionality, which disables transmissions with the aid of the VAD in order to increase the achievable network capacity and to reduce the power consumption of the mobile terminal.

The ideal Shannonian entropy-coding based source encoder would produce a stream of independent and identically distributed (i.i.d) equiprobable bits. However, since the AMR-WB encoder is not an ideal high-delay lossless entropy encoder, but a realistic, finite-delay lossy CELP codec [11], it leaves some residual redundancy in the encoded parameters. The first manifestation of the residual redundancy may be referred to as unequal-probability-related redundancy, while the second manifestation of the residual redundancy may be referred to as either inter-frame first-order correlation or intra-frame first-order correlation. These manifestations of the residual redundancy inherent in the encoded parameters will be presented in the following sections.

3.2.1 Unequal-probability-related Redundancy

The so-called unequal-probability-related redundancy manifests itself in terms of the unequal probability of occurence of the different values of a specific parameter in each 20 ms AMR-WB-encoded frame.

For the sake of quantifying the residual redundancy inherent in the bitstream, a large training sequence of 2,133,035 20 ms speech frames was applied to the AMR-WB encoder, which produces 52 different encoded parameters for each 20 ms frame. The probability

of occurrence was recorded for each of the AMR-WB encoded parameters. The resultant residual redundancy referred to as residual unequal-probability-related redundancy of each AMR-WB codec parameter is summarized in Table 3.1 in terms of $\mathcal{R}_D = \mathcal{B} - H(\mathcal{U})$, where $H(\mathcal{U})$ is the entropy of the quantized parameter $\mathcal{U}$ and $\mathcal{B}$ is the number of bits actually used for quantizing the parameter $\mathcal{U}$.

Example 3.2.1 *Let us provide an example for quantifying the residual unequal-probability-related redundancy of the ISP parameters. Assuming that the ISP parameters can be approximated by a stationary stochastic process, let the random process, $U_{i,\tau} : i = 1, 2, ..., 7$, represent the i^{th} vector-quantized ISP of each AMR-WB frame, which are encoded using 8, 8, 6, 7, 7, 5 and 5 bits, respectively, whilst τ denotes the time index referring to the current encoded frame index.*

Below we will provide an example of quantifying the residual unequal-probability-related redundancy of the first ISP denoted as ISP_1. Therefore the entropy $H(U_{1,\tau})$ quantifies the amount of information conveyed by the parameter ISP_1 in terms of the number of bits per parameter and hence directly represents the minimum number of bits required to represent the information content of $U_{1,\tau}$. As expected, for all of the 2,133,035 speech frames we found that $\mathcal{B} > H(U_{1,\tau})$, which indicates that there is unequal-probability-related redundancy in the ISP_1 parameter, which is quantified as:

$$\begin{aligned} \mathcal{R}_D &= \mathcal{B} - H(U_{1,\tau}) \\ &= 8 - H(U_{1,\tau}), \end{aligned} \tag{3.1}$$

where again $H(U_{1,\tau})$ is the entropy of the parameter $U_{1,\tau}$, given by:

$$H(U_{1,\tau}) = -\sum_{u_{1,\tau}} P(U_{1,\tau} = u_{1,\tau}).\log_2[P(U_{1,\tau} = u_{1,\tau})] \tag{3.2}$$

Again, the unequal-probability-related redundancy of the AMR-WB codec parameters is quantified in Table 3.1.

3.2.2 Inter-frame Redundancy

Let us define the inter-frame first-order correlation as that constituted by the correlation of the corresponding parameters in the current and the immediately preceeding 20 ms AMR-WB-encoded frames as illustrated in Figure 3.1. More explicitly, the parameters extracted from speech, such as the Linear Predictive Coding (LPC) coefficients for example, show

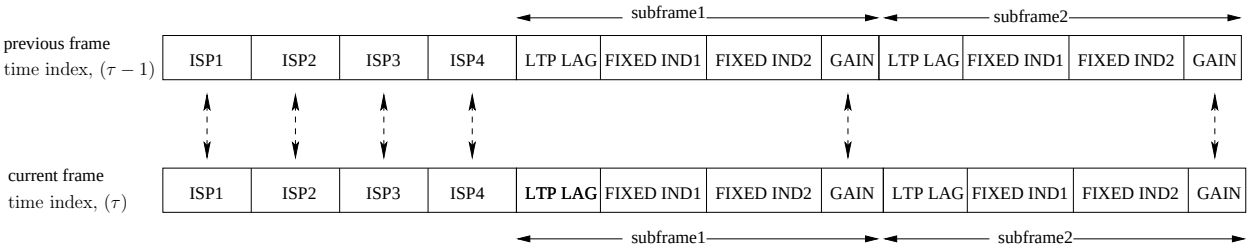

Figure 3.1: Inter-frame First-Order Correlation

a significant correlation between successive frames, especially for voiced segments. The two popular LPC parameter representations used in a wideband speech codecs are the Line Spectral Frequency (LSF)s [139] and ISPs [112]. It is shown in [6, 140] that the LSF traces versus time indicate the presence of high correlation between successive frames. As mentioned in Section 2.4, the AMR-WB codec uses the so-called immittance spectral pair to represent the LPC parameters. Thus, it is expected that the correlation of the ISPs between successive 20 ms speech frames is also high. Recall from Section 2.4 that the remaining AMR-WB encoded parameters, namely the LTP lags, the CB gains and the fixed CB indices are encoded on a per subframe or 5 ms basis. These parameters of the subsequent 5 ms subframes within a 20 ms frame also exhibit correlations. However, the odd and even index subframe LTP lags have been encoded differently. By contrast, the fixed CB indices representing the random excitation vectors are expected to exhibit no inter-frame first-order correlation. Hence, it is justifiable that we only quantify the inter-frame first-order correlation of the ISP and the CB gains parameters.

Similarly to the unequal-probability-related redundancy, for the sake of quantifying the inter-frame first-order correlation inherent in the bitstream, a large training sequence of 2,133,035 20 ms frames was applied to the AMR-WB encoder, which produces 52 different encoded parameters for each 20 ms frame, as summarized in Table 2.4. The relative frequency of each individual legitimate AMR-WB encoded parameter transition was computed for the sake of estimating the transition probabilities of those parameters, which did exhibit non-negligible inter-frame first-order correlation in two consecutive 20 ms frames. The resultant residual redundancy recorded for the inter-frame correlated parameters, namely the ISP and the CB gain parameters are summarized in Table 3.2 in terms of the mutual information $\mathcal{R}_M$ between the corresponding parameters of two consecutive 20 ms-spaced speech frames.

Example 3.2.2 *Let us now provide an example for quantifying the inter-frame first-order correlation inherent in the consecutive 20 ms frames of the ISP parameters. Assuming that the ISP parameters can be approximated by a stationary stochastic process, let the random process, $U_{i,\tau} : i = 1, 2, ..., 7$, represent the i^{th} vector-quantized ISP of each AMR-WB frame, which are encoded using 8, 8, 6, 7, 7, 5 and 5 bits, respectively, whilst τ denotes the time index referring to the current encoded frame index.*

Below we will give an example of quantifying the inter-frame first-order correlation of the first ISP denoted as ISP_1 in the current and the immediately preceeding 20 ms AMR-WB-encoded frames. Therefore the mutual information $\mathcal{I}(U_{1,\tau}; U_{1,\tau-1})$ quantifies the information $\mathcal{R}_M$ shared by ISP_1 of two consecutive 20 ms-spaced speech frames as follows:

$$\begin{aligned} \mathcal{R}_M &= \mathcal{I}(U_{1,\tau}; U_{1,\tau-1}) \\ &= H(U_{1,\tau}) - H(U_{1,\tau}|U_{1,\tau-1}), \end{aligned} \tag{3.3}$$

where $H(U_{1,\tau}|U_{1,\tau-1})$ is given by:

$$\begin{aligned} & H(U_{1,\tau}|U_{1,\tau-1}) \\ &= -\sum_{u_{1,\tau}} \sum_{u_{1,\tau-1}} P(U_{1,\tau} = u_{1,\tau}, U_{1,\tau-1} = u_{1,\tau-1}) \times \log_2[P(U_{1,\tau} = u_{1,\tau}|U_{1,\tau-1} = u_{1,\tau-1})] \\ &= -\mathcal{E}_{p(u_{1,\tau}, u_{1,\tau-1})} \times \log_2[P(U_{1,\tau} = u_{1,\tau}|U_{1,\tau-1} = u_{1,\tau-1})] \end{aligned} \tag{3.4}$$

and $H(U_{1,\tau})$ is the entropy of the parameter $U_{1,\tau}$, given by:

$$H(U_{1,\tau}) = -\sum_{u_{1,\tau}} P(U_{1,\tau} = u_{1,\tau}). \log_2[P(U_{1,\tau} = u_{1,\tau})]. \tag{3.5}$$

3.2.3 Intra-frame Redundancy

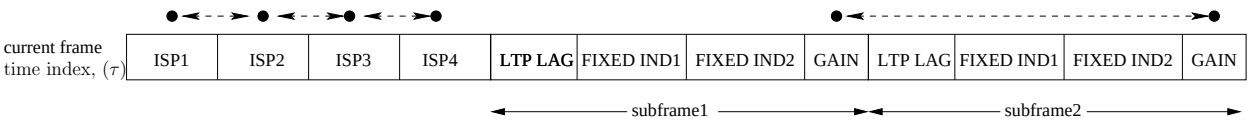

Figure 3.2: Intra-frame First Order Correlation

In contrast to the above-mentioned inter-frame first-order correlation, let us now define the *intra-frame* first-order correlation as that constituted by the correlation of a parameter

Parameter	$\mathcal{R}_M = \mathcal{I}(x;y)$	Residual Redundancy Percentage
$ISP_1(t); ISP_1(t-1)$	1.67	20.9
$ISP_2(t); ISP_2(t-1)$	1.26	15.8
$ISP_3(t); ISP_3(t-1)$	0.29	4.8
$ISP_4(t); ISP_4(t-1)$	0.12	1.7
$ISP_5(t); ISP_5(t-1)$	0.12	1.7
$ISP_6(t); ISP_6(t-1)$	0.05	1.0
$ISP_7(t); ISP_7(t-1)$	0.06	1.0
$CBGain_{sub_1}(t); CBGain_{sub_1}(t-1)$	0.97	13.9
$CBGain_{sub_2}(t); CBGain_{sub_2}(t-1)$	1.00	14.3
$CBGain_{sub_3}(t); CBGain_{sub_3}(t-1)$	1.05	13.1
$CBGain_{sub_4}(t); CBGain_{sub_4}(t-1)$	1.04	14.9

Table 3.2: Inter-frame First-Order Correlation in the AMR-WB codec's parameters

to its immediately preceeding parameter within each 20 ms AMR-WB-encoded frame, as illustrated in Figure 3.2. More explicitly, it has been shown in [6, 112] that the seven ISP parameters within a given speech frame obey the so-called ordering property, which implies that $ISP_1 < ISP_2 < \ldots < ISP_7$. The remaining AMR-WB encoded parameters, namely the LTP lags, the CB gains and the fixed CB indices are encoded on a per subframe or 5 ms basis. Additionally, these parameters of the subsequent 5 ms subframes within a 20 ms frame also exhibit correlations. Recall however from Section 2.4 that the odd and even index subframe LTP lags have been encoded differently. Recall also that the fixed CB indices do not exhibit correlation, because they represent random excitation vectors. Hence again, justifiably only the intra-frame first-order correlation of the ISP and the CB gains parameters are quantified.

Similarly to the inter-frame first-order correlation, the intra-frame first-order correlation of each legitimate encoded parameter pair is quantified in the same way, as described in the previous section. More explicitly, for the sake of quantifying the residual redundancy inherent in the bitstream, a large training sequence of 2,133,035 20 ms frames was applied to the AMR-WB encoder, which produces 52 different encoded parameters for each frame. The relative frequency of each legitimate encoded parameter transition was computed for the sake of estimating the transition probabilities of the intra-frame first-

order correlation of those parameters. The resultant residual redundancy recorded for the intra-frame correlated parameters, namely the ISP and CB gain parameters are summarized in Table 3.3 in terms of the mutual information $\mathcal{R}_M$ between the pairs of intra-frame correlated parameters.

Parameter	$\mathcal{R}_M = \mathcal{I}(x;y)$	Residual Redundancy Percentage
$ISP_1(t); ISP_2(t)$	0.96	12.00
$ISP_2(t); ISP_3(t)$	0.10	1.25
$ISP_3(t); ISP_4(t)$	0.06	1.00
$ISP_4(t); ISP_5(t)$	0.05	0.71
$ISP_5(t); ISP_6(t)$	0.03	0.43
$ISP_6(t); ISP_7(t)$	0.04	0.80
$CBGain_{sub_1}; CBGain_{sub_2}$	1.10	15.71
$CBGain_{sub_2}; CBGain_{sub_3}$	1.12	16.00
$CBGain_{sub_3}; CBGain_{sub_4}$	1.12	16.00

Table 3.3: Intra-frame First-Order Correlation in the AMR-WB codec's parameters

Example 3.2.3 *Let us now continue by providing an example for quantifying the intra-frame first-order correlation inherent in a pair of ISP parameters within each 20 ms AMR-WB-encoded frame. Assuming that the ISP parameters can be approximated by a stationary stochastic process, let the random process, $U_{i,\tau} : i = 1, 2, ..., 7$, represent the i^{th} vector quantized ISP of each AMR-WB frame, which are encoded using 8, 8, 6, 7, 7, 5 and 5 bits, respectively, whilst τ denotes the time index referring to the current encoded frame index.*

In what follows, we will provide an example of quantifying the intra-frame first-order correlation between the first two of the eight ISP parameters seen in Figure 3.2 denoted as ISP_1 and ISP_2, within each 20 ms AMR-WB-encoded frames. Therefore the mutual information $\mathcal{I}(U_{2,\tau}; U_{1,\tau})$ quantifies the information $\mathcal{R}_M$ shared by the pair ISP_1 and ISP_2 within each 20 ms-spaced speech frame,which is formulated as:

$$\begin{aligned} \mathcal{R}_M &= \mathcal{I}(U_{2,\tau}; U_{1,\tau}) \\ &= H(U_{2,\tau}) - H(U_2|U_{1,\tau}), \end{aligned} \tag{3.6}$$

where $H(U_{2,\tau}|U_{1,\tau})$ *is given by:*

$$\begin{aligned} & H(U_{2,\tau}|U_{1,tau}) \\ = & -\sum_{u_{2,\tau}}\sum_{u_{1,\tau}} P(U_{2,\tau}=u_{2,\tau}, U_{1,\tau}=u_{1,\tau}) \times \log_2[P(U_{2,\tau}=u_{2,\tau}|U_{1,\tau}=u_{1,\tau})] \\ = & -\mathcal{E}_{p(u_{2,\tau},u_{1,\tau})} \times \log_2[P(U_{2,\tau}=u_{2,\tau}|U_{1,\tau}=u_{1,\tau})] \end{aligned} \tag{3.7}$$

and $H(U_{2,\tau})$ *is the entropy of parameter* $U_{2,\tau}$*, given by:*

$$H(U_{2,\tau}) = -\sum_{u_{2,\tau}} P(U_{2,\tau}=u_{2,\tau}).\log_2[P(U_{2,\tau}=u_{2,\tau})]. \tag{3.8}$$

3.2.4 Residual Redundancy in the AMR-WB Codec: Discussions

The resultant residual redundancy quantified in Sections 3.2.1, 3.2.2 and 3.2.3 may be beneficially exploited as a source of valuable *a priori* information, when computing the soft information and estimating the speech parameters.

More explicitly, this residual redundancy may be exploited as a source of valuable *a priori* information concerning certain AMR-WB codec parameters. As an example, we can observe from Table 3.2 that the speech-energy-related CB gain parameters have a high inter-frame first-order mutual information, where the CB gain parameters of the first subframe contain 0.97 bits/parameter information, i.e. 13.8% of residual redundancy. By contrast, the efficient employment of the S-MSQV in the encoding process of the ISP parameters removes most of the redundancy and hence only the first two ISP parameters have a relatively high first-order inter-frame mutual information, where they contain 20.9% and 15.8% redundancy, respectively. As a counter-example, the fixed CB index representing random excitation vectors is expected to exhibit no inter-frame first-order correlation and hence provides no substantial inter-frame first-order *a priori* information.

These results suggest that the high-correlation CB gain parameters and the first two ISP parameters would benefit from exploiting the non-negligible inter-frame first-order correlation based *extrinsic* information. By contrast, the rest of the parameters would benefit from exploiting the unequal-probability-based *extrinsic* information during the decoding process, at the cost of increasing the complexity imposed only marginally.

On the other hand, when considering the third column of Table 3.3, it becomes clear that the AMR-WB codec is highly efficient in terms of removing the redundancy from each 20 ms speech frame. Hence, we may conclude from Table 3.3 that the intra-frame first-order correlation of all parameters within a given frame is insignificant. Therefore its

exploitation is unlikely to be justifiable, when considering the significant complexity imposed in terms of the required trellis decoding, because the attainable system performance would only be marginally increased.

More explicitly, the resultant redundancy quantified in Sections 3.2.1 and 3.2.2 will be exploited by our three-stage iterative detector as *a priori* information concerning the ISP and the LTP-lag of the AMR-WB codec parameters and hence assists the remaining two constituent decoders in their decisions, as detailed in Section 3.3. It is important to note that this is achieved without any increase of the system's delay, since no *extrinsic* information is gleaned from the 20 ms AMR-WB speech frames yet to arrive in the future. By contrast, the previously decoded frames are explicitly exploited, but this is achieved without any extra delay.

Specifically, in order to realize a transmission scheme that does not introduce any extra inter-frame-coding induced delay, we only exploit the *A Posteriori* Probability (APP) gleaned from the previously received speech frames. The algorithm used for computing the APPs quantified in terms of their Logarithmic-Likelihood Ratios (LLR) and their Maximum *A Posteriori* (MAP) decoding will be briefly reviewed in Section 3.3.2. It will be shown in Section 3.3.4 that the exploitation of this residual redundancy at the decoder has the potential of providing useful performance gains, when compared to the less sophisticated receiver dispensing with this *a priori* knowledge.

3.3 Soft-Bit Assisted Iterative AMR-WB Source-Decoding and Turbo-Detection Scheme [1]

3.3.1 System Background

The classic Shannonian source and channel coding separation theorem [13] has limited applicability in the context of finite-complexity, finite-delay lossy speech [6]. These arguments are particularly valid, when the limited-complexity, limited-delay source encoders fail to remove all the redundancy from the correlated speech source signal. Fortunately, this residual redundancy may be beneficially exploited for error protection by intelligently ex-

[1] This section is based on N. S. Othman, M. El-Hajjar, O. Alamri and L. Hanzo: "Iterative AMR-WB Source and Channel-Decoding of Differential Space-Time Spreading Assisted Sphere Packing Modulation", to appear in IEEE Transactions on Vehicular Technology, and it was based on collaborative research with the co-authors.

changing soft information amongst the various receiver components. More explicitly, it was demonstrated in [122] that the video performance of a twin-class protected MPEG4 video transceiver substantially benefitted from a multi-stage turbo detection process, which exchanged soft-information across three Soft-In-Soft-Out (SISO) blocks.

These powerful turbo principles may be further enhanced by exploiting the innovative concept of soft speech bits, which was developed by Vary and his team [59, 60], culminating in the formulation of ISCD [48]. More explicitly, in ISCD the source- and channel-decoder iteratively exchange *extrinsic* information for the sake of improving the overall system performance. As a further development, in [141] the turbo principle [130] was employed for iterative soft demapping in multilevel modulation [131] schemes combined with channel coding, which resulted in an enhanced Bit Error Rate (BER) performance. Thus, ISCD may be beneficially combined with iterative soft demapping in the context of multilevel modulation and amalgamated with a number of other sophisticated wireless transceiver components. In the resultant multi-stage scheme *extrinsic* information is exchanged amongst three receiver components, namely the demodulator, the channel decoder and the soft-input source decoder in the spirit of [69, 137].

Explicitly, we propose and investigate the jointly optimised ISCD scheme of Figure 3.3 invoking the AMR-WB speech codec [110], which is protected by a Recursive Systematic Convolutional (RSC) code. The resultant bitstream is transmitted using Differential Space-Time Spreading (DSTS) amalgamated with Sphere Packing (SP) modulation [82] over a narrowband temporally correlated Rayleigh fading channel. An efficient iterative turbo-detection scheme is utilised for exchanging *extrinsic* information between the constituent codes. In an effort to mitigate the effects of the hostile Rayleigh fading channel, DSTS [82] employing two transmit and one receive antennas was invoked for the sake of providing a spatial diversity gain. This powerful wireless transceiver is advocated here in conjunction with SP modulation, since it was demonstrated in [74] that the employment of SP modulation combined with the orthogonal transmit diversity designs outperformed its conventional counterpart of [71, 72]. We will refer to this three-stage iteratively detected scheme as the DSTS-SP-RSC-AMRWB scheme.

The convergence behaviour of this iterative process is studied using EXIT charts [65] by visualizing the input/output mutual information exchange of the individual constituents of the Soft-Input-Soft-Output (SISO) decoder. Recently, the 2D EXIT charts of two-stage concatenated systems have been extended to three-stage turbo receivers [66, 67], where soft-information is exchanged across three SISO blocks.

The novelty and rationale of the proposed system can be summarised as follows:

1. A Soft-Input Soft-Output AMR-WB Decoder is contrived, which is capable of accepting the *extrinsic* information passed to it from the channel decoder, and subsequently exchanges its *extrinsic* information with the channel decoder. More explicitly, the residual redundancy inherent in the AMR-WB speech codec parameters is quantified and this redundancy is exploited as *a priori* knowledge for achieving further performance gains, when compared to the less sophisticated receiver dispensing with this *a priori* knowledge.

2. EXIT chart analysis has been used for designing the optimum combination of receiver components. More explicitly, the EXIT curve of the AMR-WB source decoder never reaches the (1,1) point of perfect convergence and hence the achievable BER remains high. However, if the intermediate RSC decoder and the AMR-WB decoder are viewed as a combined outer SISO module, then the joint EXIT function of this module becomes capable of reaching the point of convergence at (1,1).

3. Conventional coherent Space-Time Spreading (STS) requires the estimation of the Channel Impulse Responses (CIR) of all the multiple antenna links. For the sake of eliminating the potentially high complexity of Multi-Input Multi-Output (MIMO) channel estimation in the proposed scheme, the employment of non-coherently detected DSTS using two transmit antennas and a single receive antenna is advocated for achieving a transmit diversity gain.

4. The employment of SP modulation facilitates the joint design of several DSTS timeslots' signal, which allows the direct minimisation of the DSTS symbol error probability.

The residual redundancy inherent in the AMR-WB encoded parameters was quantified in Section 3.2 and will be exploited by the three-stage turbo detection aided DSTS-SP-RSC-AMRWB scheme of Figure 3.3. In Section 3.3.2, the overall system model is described, while the system's convergence behaviour is analyzed in Section 3.3.3 with the aid of 3D EXIT charts and their 2D projections. Section 3.3.4 quantifies the overall performance of our proposed three-stage scheme, while our conclusions are offered in Section 3.3.5.

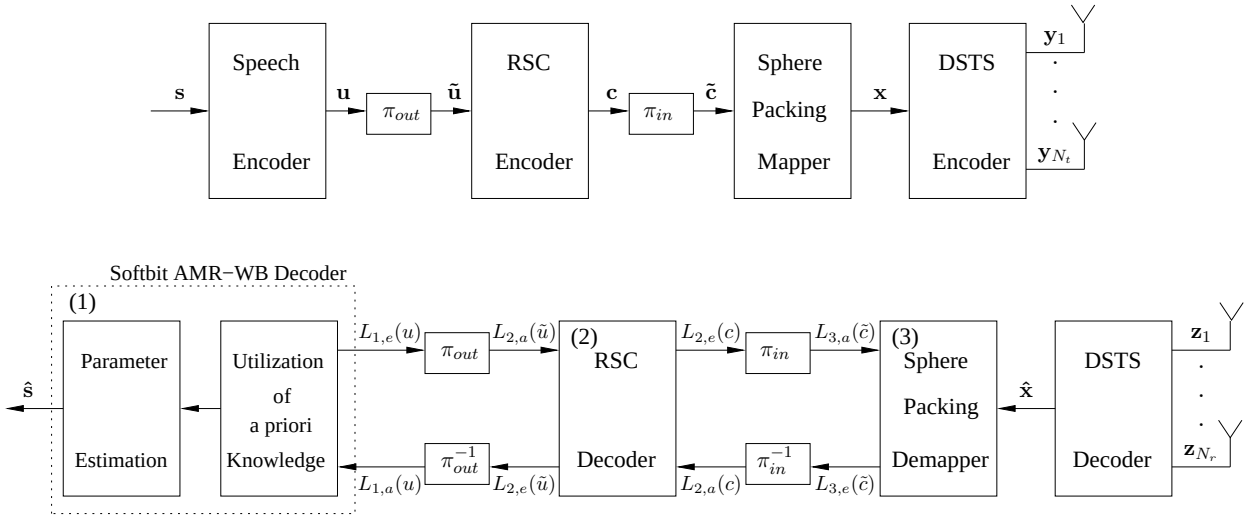

Figure 3.3: Block diagram of the DSTS-SP-RSC-AMRWB scheme.

3.3.2 System Overview

Figure 3.3 shows the iterative decoder structure of the DSTS-SP-RSC-AMRWB scheme, where the *extrinsic* information gleaned is exchanged amongst all three constituent decoders, namely the AMR-WB decoder, RSC decoder and the SP demapper. The *inner* iterative loop corresponds to the iterative soft SP demapping and RSC channel decoding [82], while the *outer* iterative loop represents the *extrinsic* information exchange between the AMR-WB speech decoder and the RSC decoder.

3.3.2.1 Transmitter

The AMR-WB speech encoder produces a frame of speech codec parameters, namely $\{\mathbf{v}_{1,\tau},\mathbf{v}_{2,\tau},\ldots,\mathbf{v}_{\kappa,\tau},\ldots,\mathbf{v}_{52,\tau}\}$, where $\mathbf{v}_{\kappa,\tau}$ denotes an encoded parameter, with $\kappa = 1, ..., K$ denoting the index of each parameter in the encoded speech frame and $K = 52$, whilst τ denotes the time index referring to the current encoded frame index. Then, $\mathbf{v}_{\kappa,\tau}$ is quantised and mapped to the bit sequence $\mathbf{u}_{\kappa,\tau} = [u(1)_{\kappa,\tau}\ u(2)_{\kappa,\tau}\ \ldots\ u(M)_{\kappa,\tau}]$, where M is the total number of bits assigned to the κth parameter. Then, the outer interleaver π_{out} permutes the bits of the sequence $\mathbf{u}$, yielding $\tilde{\mathbf{u}}$ of Figure 3.3.

The bit sequence c of Figure 3.3 is generated by a $\frac{1}{2}$-rate RSC code having a code memory of 3 and octally represented generator polynomials of $(G_1, G_2) = (13, 6)$. The

DSTS-SP modulator of Figure 3.3 first maps B number of channel-coded bits $\tilde{\mathbf{c}}$=[$\tilde{c}_0$ $\tilde{c}_1$ ... $\tilde{c}_{B-1}$] $\in$ 0,1 to a sphere packing symbol $x \in X$, using the mapping function $x = map_{sp}(\tilde{\mathbf{c}})$. Furthermore, we have $B = \log_2(L_{SP}) = \log_2(16) = 4$, where L_{SP} represents the set of legitimate SP constellation points, as outlined in [74]. This set of SP symbols is transmitted using DSTS in conjunction with two transmit antennas, as detailed in [74]. STS was first proposed by Hochwald *et. al* [73] for the sake of providing space-time-coding-style spatial transmit diversity gains for Code Division Multiple Access (CDMA) systems [142]. However, the CIR of all transmit-receive antenna links has to be estimated in coherently detected STS systems, which is a challenging and high-complexity task. Hence, in [82] the philosophy of differentially detected DSTS was developed, which dispenses with the need of estimating the CIR and thus eliminates the complexity of MIMO channel estimation at the cost of the usual 3 dB performance penalty. In this study, we consider transmissions over a narrowband temporally correlated Rayleigh fading channel, associated with a normalised Doppler frequency of $f_D = 0.01$, while the spatial channel coefficients are independent.

The notations $L(.)$ in Figure 3.3 denote the LLRs of the bit probabilities. The notations $\tilde{c}$, c, $\tilde{u}$ and u in the round brackets (.) in Figure 3.3 denote the SP bits, RSC coded bits, RSC data bits and the AMR-WB encoded bits, respectively. The specific nature of the LLRs is represented by the subscripts of $L_{.,a}$, $L_{.,p}$ and $L_{.,e}$, which denote in Figure 3.3 *a priori*, *a posteriori* and *extrinsic* information, respectively. The LLRs associated with one of the three constituent decoders having a label of {1,2,3} are differentiated by the corresponding subscripts (.) of {1,2,3}. Note that the subscript 2 is used for representing the RSC decoder of Figure 3.3.

3.3.2.2 Receiver

Inner Iterations: The complex-valued received symbols, $\mathbf{z}$ are demapped to their LLR [130] representation for each of the B number of RSC-encoded bits per DSTS-SP symbol. The *a priori* LLR values $L_{3,a}(\tilde{c})$ provided by the RSC decoder are subtracted from the *a posteriori* LLR values $L_{3,p}(\tilde{c})$ at the output of the SP demapper for the sake of generating the *extrinsic* LLR values $L_{3,e}(\tilde{c})$. Then the LLRs $L_{3,e}(\tilde{c})$ are deinterleaved by a soft-bit deinterleaver. Next, the deinterleaved soft-bits $L_{2,a}(c)$ of Figure 3.3 are passed to the RSC decoder in order to compute the *a posteriori* LLR values $L_{2,p}(c)$ provided by the MAP algorithm [143] for all the RSC-encoded bits. The *extrinsic* information $L_{2,e}(c)$ seen in Figure 3.3 is generated by subtracting the *a priori* information $L_{2,a}(c)$ from the *a posteriori* information $L_{2,p}(c)$, which is then fed back to the SP demapper as the *a priori*

information $L_{3,a}(\tilde{c})$ after appropriately reordering them using the inner soft-value interleaver. The SP demapper of Figure 3.3 exploits the *a priori* information $L_{3,a}(\tilde{c})$ for the sake of providing improved *a posteriori* LLR values $L_{3,p}(\tilde{c})$ which are then passed to the RSC decoder and in turn, back to the SP demapper for further iterations.

Outer Iterations: The *extrinsic* LLR values $L_{2,e}(\tilde{u})$ seen in Figure 3.3 are generated by subtracting the *a priori* LLR values $L_{2,a}(\tilde{u})$ of the RSC decoder from the LLR values $L_{2,p}(\tilde{u})$. Then, the LLRs $L_{2,e}(\tilde{u})$ are deinterleaved by the outer soft-bit deinterleaver. The resultant soft-bits $L_{1,a}(u)$ are passed to the AMR-WB decoder that was further developed for handling soft input bits in order to compute the *extrinsic* LLR values $L_{1,e}(u)$ with the aid of Soft-Bit Source Decoding (SBSD), as proposed in [48] and detailed during our further discourse. These *extrinsic* LLR values are then fed back to the RSC decoder after appropriately reordering them in the specific order required by the RSC decoder for the sake of completing an outer iteration.

We define one inner iteration followed by two outer iterations as having one "system iteration" denoted as $I_{system} = 1$. The residual redundancy quantified in Section 3.2 is exploited as *a priori* information for computing the *extrinsic* LLR values and for estimating the speech parameters. Specifically, we exploited the inter-frame first order correlation of both the CB gain and of the first two ISP parameters, as well as the unequal-probability-related redundancy of the remaining AMR-WB encoded parameters. The probability of occurence and the transition probabilities of each individual legitimate AMR-WB encoded parameters are derived from the same large 2,133,035 20 ms speech frame based training sequence as the one used in Section 3.2. The required probabilities were estimated off-line in advance and were made available at the receiver as additional *a priori* input to the soft-bit assisted AMR-WB decoder. More explicitly, the details of the algorithm used for computing the *extrinsic* LLR values $L_{1,e}(u)$ of the speech parameters can be found in [48, 61], which are briefly reviewed below. Firstly, the channel decoder's output information related to each speech parameter is given by the product of each of the constituent bits as follows:

$$p(\hat{\mathbf{u}}_{\kappa,\tau}|\mathbf{u}_{\kappa,\tau}) = \prod_{i=1}^{M} p[\hat{u}_{\kappa,\tau}(i)|u_{\kappa,\tau}(i)], \tag{3.9}$$

where $\hat{\mathbf{u}}_{\kappa,\tau} = [\hat{u}(1)_{\kappa,\tau}\ \hat{u}(2)_{\kappa,\tau}\ \ldots\ \hat{u}(M)_{\kappa,\tau}]$ represents the received bit sequence of the κth parameter, while $\mathbf{u}_{\kappa,\tau}$ is the corresponding transmitted bit sequence, provided that all these bits are independent of each other, although in reality the M bits of the κ-th parameter are not entirely independent of each other at the output of a practical source codec. The effects of this approximation are eliminated by the iterative detector during

its consecutive iterations.

***Extrinsic* LLR of soft speech bit generation for exploiting the parameters' unequal probability:** As usual, we exclude the bit under consideration from the present bit sequence within each of the κth parameter where $\kappa = 1, ..., K$ and $K = 52$, namely from $\mathbf{u}_{\kappa,\tau}$=$[u_{\kappa,\tau}(\lambda)\ \mathbf{u}_{\kappa,\tau}^{[ext]}]$. The *extrinsic* channel decoder output information $u_{\kappa,\tau}(\lambda)$ of each desired bit, is expressed as:

$$p(\hat{\mathbf{u}}_{\kappa,\tau}^{[ext]}|\mathbf{u}_{\kappa,\tau}^{[ext]}) = \prod_{i\neq\lambda, i=1}^{M} p[\hat{u}_{\kappa,\tau}(i)|u_{\kappa,\tau}(i)], \tag{3.10}$$

where the term $\mathbf{u}_{\kappa,\tau}^{[ext]}$ denotes all elements of the bit pattern $\mathbf{u}_{\kappa,\tau}$, but excludes the desired bit $u_{\kappa,\tau}(\lambda)$ itself. Finally, the *extrinsic* LLR value $L_{1,e}(u)$ generated for each bit can be obtained by combining the corresponding channel decoder output information and the *a priori* knowledge $p(\mathbf{u}_{\kappa,\tau})$ concerning the κth parameter, which is given by [48, 144]:

$$L_{1,e}(u_{\kappa,\tau}(\lambda)) = \log \frac{\sum_{\mathbf{u}_{\kappa,\tau}^{[ext]}} p(\mathbf{u}_{\kappa,\tau}^{[ext]}|u_{\kappa,\tau}(\lambda)=0).\exp A}{\sum_{\mathbf{u}_{\kappa,\tau}^{[ext]}} p(\mathbf{u}_{\kappa,\tau}^{[ext]}|u_{\kappa,\tau}(\lambda)=1).\exp A}, \tag{3.11}$$

where

$$A = \sum_{u_{\kappa,\tau}(l) \text{ of } \mathbf{u}_{\kappa,\tau}^{[ext]}} \frac{u_{\kappa,\tau}(l)}{2}(L_{1,a}[u_{\kappa,\tau}(l)])$$

and $L_{1,a}$ represents the *a priori* LLR values of the AMR-WB decoder, which is the deinterleaved counterpart of $L_{2,e}$ generated by the RSC decoder.

***Extrinsic* LLR of soft speech bit generation for exploiting the inter-frame first-order correlation:** As seen in Figure 3.3, the *extrinsic* LLRs $L_{1,e}(u)$ can be generated by subtracting the *a priori* information, $L_{1,a}(u)$ from the *a posteriori* information, $L_{1,p}(u)$. Again, in order to realize a transmission scheme imposing no extra latency we generate and exploit only the forward APPs by exploiting the *a priori* knowledge expressed in terms of $p(\mathbf{u}_{\kappa,\tau}|\mathbf{u}_{\kappa,\tau-1})$, yielding:

$$\alpha_{\tau-1}(\mathbf{u}_{\kappa,\tau-1}) = \mathcal{C}p(\hat{\mathbf{u}}_{\kappa,\tau-1}|\mathbf{u}_{\kappa,\tau-1}) \sum_{\mathbf{u}_{\kappa,\tau-2}} p(\mathbf{u}_{\kappa,\tau-1}|\mathbf{u}_{\kappa,\tau-2}).\alpha_{\tau-2}(\mathbf{u}_{\kappa,\tau-2}), \tag{3.12}$$

where $\alpha_{\tau-1}(\mathbf{u}_{\kappa,\tau-1})$ represents a forward recursive values and $\mathcal{C}$ represents a normalization constant. Finally, the *a posteriori* LLR values, $L_{1,p}(u)$ generated by each bit are given by

$$L_{1,p}(u_{\kappa,\tau}(\lambda)) = \log \frac{\sum_{u_{\kappa,\tau}(\lambda)=0} p(\hat{\mathbf{u}}_{\kappa,\tau}|\mathbf{u}_{\kappa,\tau}) \sum_{\mathbf{u}_{\kappa,\tau-1}} p(\mathbf{u}_{\kappa,\tau}|\mathbf{u}_{\kappa,\tau-1})\alpha_{\tau-1}(\mathbf{u}_{\kappa,\tau-1})}{\sum_{u_{\kappa,\tau}(\lambda)=1} p(\hat{\mathbf{u}}_{\kappa,\tau}|\mathbf{u}_{\kappa,\tau}) \sum_{\mathbf{u}_{\kappa,\tau-1}} p(\mathbf{u}_{\kappa,\tau}|\mathbf{u}_{\kappa,\tau-1})\alpha_{\tau-1}(\mathbf{u}_{\kappa,\tau-1})}. \tag{3.13}$$

The proposed scheme of Figure 3.3 is compared to the benchmark scheme carrying out joint channel decoding and DSTS aided SP demodulation in conjunction with separate AMR-WB decoding.

3.3.3 EXIT Chart Analysis

EXIT charts have been widely used in the design of iterative schemes, which facilitate the prediction of the associated convergence behaviour, based on the exchange of *extrinsic* information amongst the constituent receiver components.

The application of EXIT charts is based on two assumptions. Firstly, we assume that the *a priori* LLR values are fairly uncorrelated, when employing a sufficiently high interleaver length. Secondly, we assume that the *a priori* LLR values exhibit a Gaussian Probability Density Function (PDF).

More explicitly, let $I_{.,A}(x)$ denote the Mutual Information (MI) [13] between the *a priori* value $A(x)$ and the bit x, whilst $I_{.,E}(x)$ denote the MI between the *extrinsic* value $E(x)$ and the bit x. The MI associated with one of the three constituent decoders having a label of {1,2,3} is differentiated by the corresponding subscripts (.) of {1,2,3}.

3.3.3.1 EXIT Characteristics of the Soft-bit Assisted AMR-WB Decoder

The *extrinsic* information transfer characteristic of the source decoder describes the relationship between the channel coded input $L_{1,A}(u)$ and the source decoded *extrinsic* output $L_{1,E}(u)$. The source decoder has a single input, namely the *a priori* input $L_{1,A}(u)$ represented in form of the LLRs provided by the channel decoder. Therefore, the *extrinsic* information transfer characteristic of the source decoder is independent of the channel's E_b/N_0-value and hence may be written as:

$$I_{1,E}(u) = T_u[I_{1,A}(u)], \tag{3.14}$$

where $I_{1,A}(u) = I[u; L_{1,A}(u)], 0 \leq I_{1,A}(u) \leq 1$, is the mutual information between the source coded bits u and the channel decoder's output LLR values $L_{1,A}(u)$ and similarly $I_{1,E}(u) = I[u; L_{1,E}(u)], 0 \leq I_{1,E}(u) \leq 1$, is the mutual information between the source coded bits u and the *extrinsic* LLR values $L_{1,E}(u)$.

However, as mentioned earlier in Section 3.2.4, the residual redundancy quantified in Section 3.2 may be beneficially exploited for error enhanced protection by intelligently exchanging soft information amongst the various receiver components. In Section 3.3.2 the

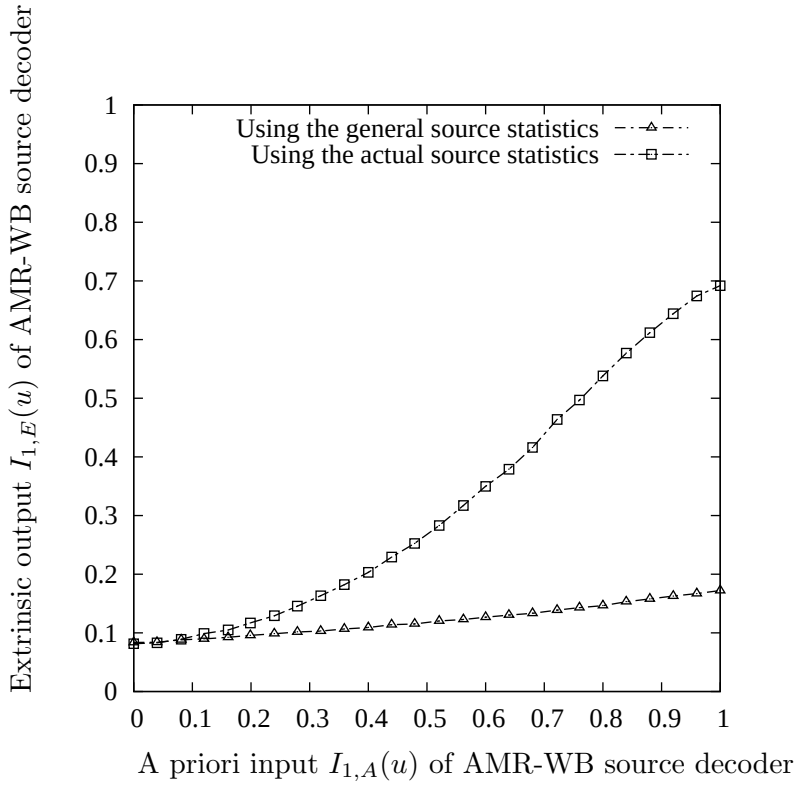

Figure 3.4: *Extrinsic* information transfer characteristics of the soft-bit assisted AMR-WB decoder when the residual unequal-probability-related redundancy of each AMR-WB encoded parameters is exploited.

procedure used for computing the APPs quantified in terms of the LLRs, which exploits both the residual unequal-probability-related redundancy and the inter-frame first-order correlation was presented. More explicitly, the APP computation includes the probability of occurence evaluation and the transition probability calculation for each individual legitimate AMR-WB encoded parameter. As detailed in Section 3.3.2.2, these source statistics can be derived from the large general training sequence as described in Section 3.2, which may be referred to as the general source statistics. These general source statistics can be evaluated in an off-line fashion in advance and then made available for the receiver as additional *a priori* for the source decoder, resulting in an enhanced BER performance, as we will demonstrate in Section 3.3.4.

On the other hand, these source statistics can also be derived in an on-line fashion from the actual transmitted source signal, which may be referred to as the actual source statistics, because it is evaluated for each frame and then sent to the receiver as side information. Thus, it provides more accurate *a priori* information. However, this requires an increased bitrate for the encoding of the source statistics.

Figure 3.4 illustrates the *extrinsic* information transfer characteristics of the AMR-WB decoder when only the residual unequal-probability-related redundancy of each AMR-WB encoded parameter is exploited. As we can observe from Figure 3.4, the EXIT curve of the AMR-WB decoder utilising the general source statistics reaches the point $I_{1,E} \approx 0.17$ at $I_{1,A} = 1.0$. By contrast, the EXIT curve of the AMR-WB decoder reaches the point $I_{1,E} \approx 0.69$ at $I_{1,A} = 1.0$, when utilising the actual source statistics. This behaviour is due to the fact that the actual statistics provide more accurate information about the received bits. However, the value of $I_{1,E}$ still fails to reach $I_{1,E} = 1$ because it is shown in Section 3.2.1 that the residual unequal-probability-related redundancy of the AMR-WB encoded parameters is insufficiently high.

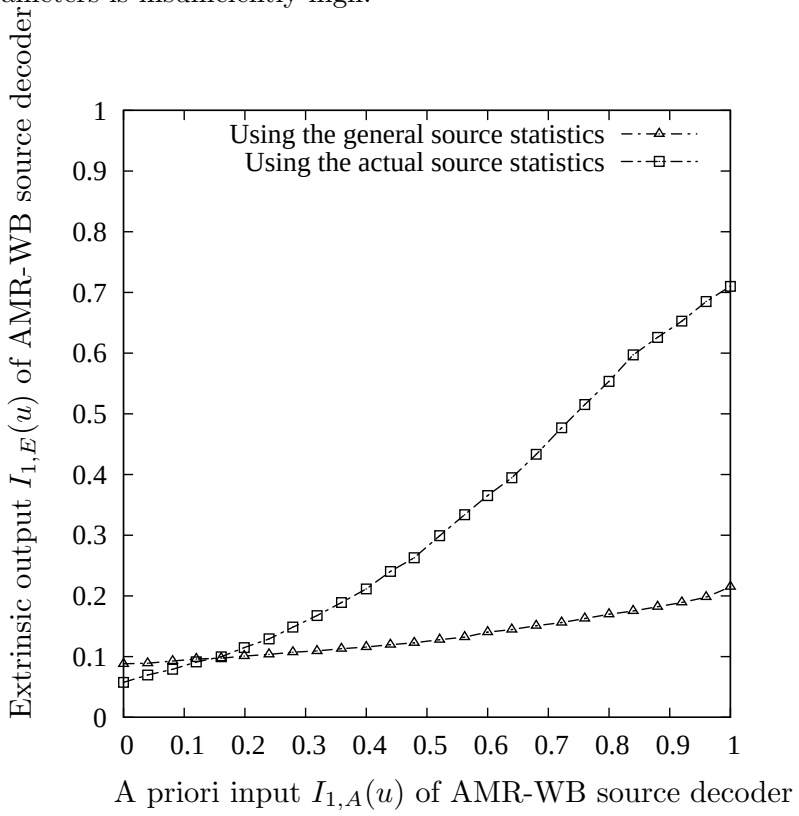

Figure 3.5: *Extrinsic* information transfer characteristics of the soft-bit assisted AMR-WB decoder when both the residual unequal-probability-related redundancy as well as the inter-frame first-order correlation of the CB gain parameters and of the first two ISP parameters are exploited.

On the other hand, it can be seen from Figure 3.5 that the exploitation of the residual unequal-probability-related redundancy as well as that of the inter-frame first-order correlation of the CB gain parameters and the first two ISP parameters facilitates reaching the

point $I_{1,E} \approx 0.71$ at $I_{1,A} = 1.0$, when utilising the actual source statistics. By contrast, the EXIT curve of the soft-bit assisted AMR-WB decoder reaches the point $I_{1,E} \approx 0.22$ at $I_{1,A} = 1.0$, when utilising the general source statistics. Thus, we conclude that the EXIT curve of the soft-bit assisted AMR-WB decoder is capable of reaching a higher $I_{1,E}$ value when the inter-frame first-order correlation of the CB gain parameters and that of the first two ISP parameters is exploited. This behaviour is due to the fact that the inter-frame first-order correlation of the CB gain parameters and of the first two ISP parameters indicates the presence of residual redundancy, as quantified in Section 3.2.2. Although the resultant EXIT curve fails to reach the point of perfect iterative decoding convergence at $(I_A, I_E) = (1, 1)$, where typically an infinitesimally low BER is attained, nonetheless, an improved BER may be expected. Therefore, Figures 3.4 and 3.5 demonstrate that the *extrinsic* information transfer characteristics of the AMR-WB decoder depend on the amount of residual redundancy inherent in the bitstream.

3.3.3.2 3D EXIT Chart

Let us now study the iterative detection convergence of the three-stage concatenated scheme of Figure 3.3 employing the parameters described in Section 3.3.2 with the aid of 3D EXIT charts [67], when communicating over a narrowband temporally correlated Rayleigh fading channel.

As seen from Figure 3.3, the RSC decoder receives inputs from and provides outputs for both the SP demapper and the AMR-WB decoder. Thus, the input of the RSC decoder is constituted by the *a priori* input, $I_{2,A}(c)$ corresponding to the coded bits c originating from the *extrinsic* output of the SP demapper as well as the *a priori* input, $I_{2,A}(\tilde{u})$, available for the data bits $\tilde{u}$, which was generated from the *extrinsic* output of the AMR-WB decoder. Note that the subscript 2 is used for representing the RSC decoder of Figure 3.3.

Correspondingly, the RSC decoder generates both the *extrinsic* output, $I_{2,E}(\tilde{u})$, which represents the data bits $\tilde{u}$ as well as the *extrinsic* output, $I_{2,E}(c)$, representing the coded bits c. Therefore, the EXIT characteristic of the RSC decoder can be described by the following two EXIT functions [66]:

$$I_{2,E}(\tilde{u}) = T_{\tilde{u}}[I_{2,A}(\tilde{u}), I_{2,A}(c)], \tag{3.15}$$

$$I_{2,E}(c) = T_c[I_{2,A}(\tilde{u}), I_{2,A}(c)], \tag{3.16}$$

which are illustrated by the 3D surfaces seen in Figures 3.6 and 3.7, respectively.

By contrast, the AMR-WB decoder as well as the SP demapper only receive input from and provide output for the RSC decoder. Thus, the corresponding EXIT functions are:

$$I_{1,E}(u) = T_u[I_{1,A}(u)], \tag{3.17}$$

for the AMR-WB decoder and

$$I_{3,E}(\tilde{c}) = T_{\tilde{c}}[I_{3,A}(\tilde{c}), E_b/N_0], \tag{3.18}$$

for the SP demapper. Equations (3.17) and (3.18) are illustrated in Figures 3.6 and 3.7, respectively.

The EXIT chart [65] analysis of the iterative decoding scheme's convergence behaviour indicates that an infinitesimally low bit-error rate (BER) may only be achieved by an iterative receiver, if an open tunnel exists between the EXIT curves of the two SISO components.

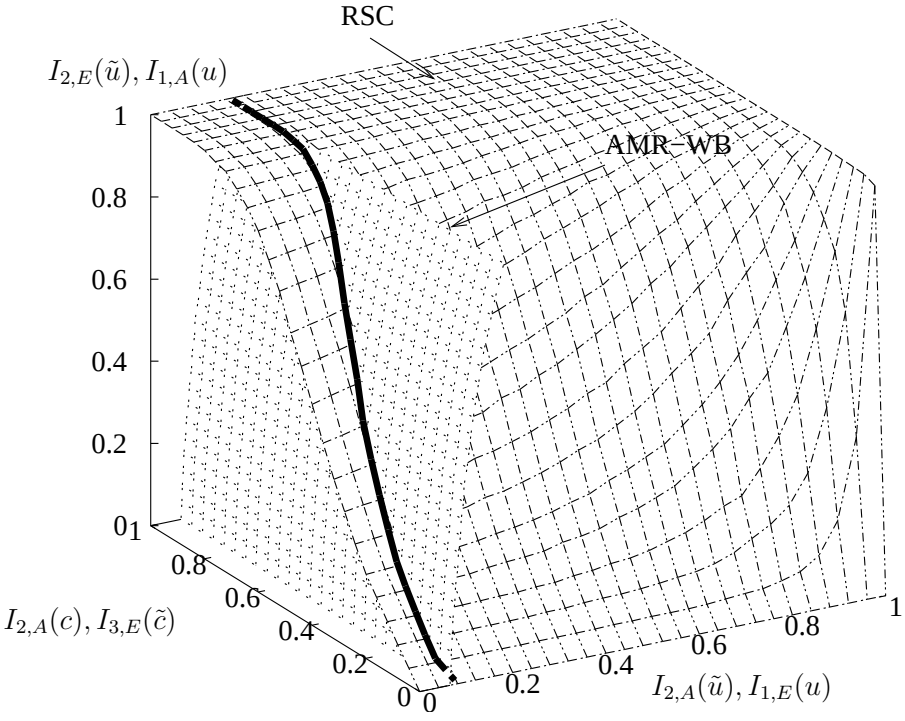

Figure 3.6: 3D EXIT chart of the RSC decoder and of the AMR-WB decoder.

More explicitly, the intersection of the surfaces seen in Figure 3.6 characterizes the best possible attainable performance, when exchanging information between the RSC decoder and the AMR-WB decoder of Figure 3.3 for different fixed values of $I_{2,A}(c)$, which is shown

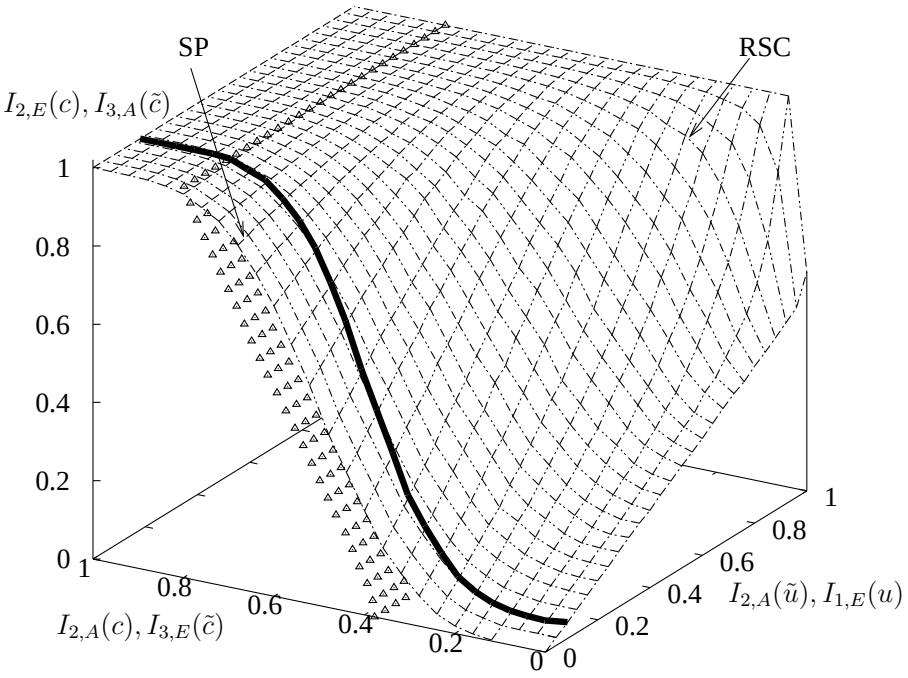

Figure 3.7: 3D EXIT chart of the RSC decoder and of the SP demapper at E_b/N_0=5.0 dB with projection from Figure 3.6.

as a thick solid line. For each point $[I_{2,A}(\tilde{u}), I_{2,A}(c), I_{2,E}(\tilde{u})]$ of this line on the 3D space of Figure 3.7, there is a specific value of $I_{2,E}(c)$ determined by $I_{2,A}(\tilde{u})$ and $I_{2,A}(c)$ according to the EXIT function of Equation (3.16). Therefore the solid line on the surface of the EXIT function of the RSC decoder seen in Figure 3.6 is mapped to the solid line shown in Figure 3.7.

3.3.3.3 2D EXIT Chart

In order to avoid the somewhat cumbersome interpretation of 3D EXIT charts, we project the bold EXIT curve of Figure 3.7 onto the 2D plane at $I_{2,A}(\tilde{u}) = 0$, yielding the bold dashed-dotted line in Figure 3.8. As seen from Figure 3.8, the EXIT curve of the AMR-WB decoder denoted by the line marked with triangles, fails to reach the point of convergence at (1,1) and intersects with the EXIT curve of the inner SP demapper, which implies that residual errors persists. On the other hand, if the intermediate RSC decoder and the AMR-WB decoder are viewed as a combined outer SISO module, then the joint EXIT function of this module becomes capable of reaching the point of convergence at (1,1), which can be described by $T_c^p[I_{3,E}(\tilde{c}), I_{2,A}(c)]$. The joint EXIT function is denoted by the bold dotted line in Figure 3.8, which corresponds to the 2D EXIT curve projected

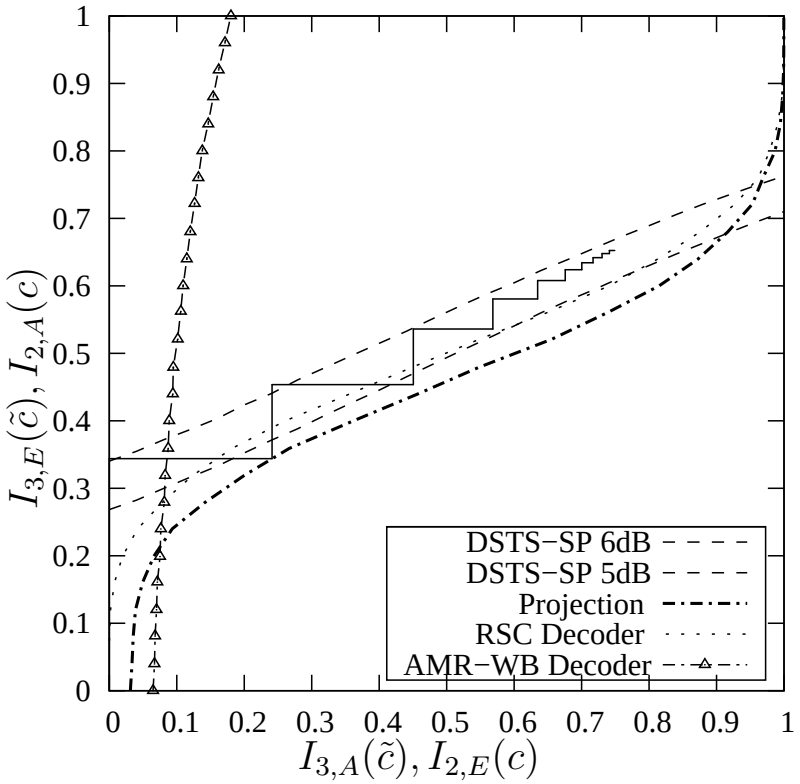

Figure 3.8: 2D projection of the EXIT chart of the three-stage DSTS-SP-RSC-AMRWB scheme and the 2D EXIT chart of the two-stage benchmarker scheme.

from Figure 3.7. Also shown are the EXIT curves of the SP demapper for various E_b/N_0 values and the EXIT curve of the RSC decoder used in the DSTS-SP-RSC benchmarker scheme. This 2D projection-based EXIT chart can therefore be used to determine the convergence threshold in terms of the minimum E_b/N_0 value required. It can be seen in Figure 3.8 that there is an open tunnel between the projected EXIT curve and that of the SP demapper at E_b/N_0=5.0 dB. By contrast, the EXIT curve of the SP demapper and that of the RSC decoder of the benchmarker scheme employing no outer iterations exhibit an open tunnel at E_b/N_0=6.0 dB. Thus according to the EXIT chart predictions the 3-stage system outperforms its benchmark scheme.

3.3.4 System Performance

In this section we characterize the attainable performance of the proposed scheme using both the BER and the Segmental Signal to Noise Ratio (SegSNR) [6] evaluated at the speech decoder's output as a function of the channel SNR. We consider a two-transmit-

antenna aided DSTS-SP system associated with $L_{SP} = 16$ and a single receiver antenna, while the remaining simulation parameters described in Section 3.3.2 are listed in Table 3.3.4. In our simulations each inner iteration was followed by two outer iterations, which together formed a three-stage "system iteration" and the LLR values were unquantized. Having two outer iterations implied that the *extrinsic* information was exchanged twice between the intermediate RSC decoder and the outer AMR-WB decoder. As characterized by the intersection of the surfaces seen in Figure 3.6 of Section 3.3.3, this resulted in the best attainable *extrinsic* LLR values $L_{2,e}(c)$ before being interleaved and fed back to the SP demapper as the *a priori* information $L_{3,a}(\tilde{c})$. Any further increase of the number of outer iterations resulted in an increased complexity, but only marginally increased the attainable *extrinsic* LLR values.

System Parameters	**DSTS-SP-RSC-AMR-WB**
Source Coding	AMR-WB
Bit Rates (kbit/s)	23.05
Speech Frame Length (ms)	20
Sampling Rate (kHz)	16
Channel Coding	Recursive Sys. Conv. Code
Code Rate	R=1/2
Code Memory	L=3
Code Generator (G_1, G_2)	$(13, 6)_8$
Modulation Scheme	Sphere Packing (L_{SP}=16)
MIMO Scheme	DSTS
Number of Transmitters, N_t	2
Number of Receivers, N_r	1
Spreading Code	Walsh Code
Spreading Factor	8
Number of users	4
Channel	Correlated Rayleigh Fading
Normalised Doppler frequency	0.01
System Throughput (bit/symbol)	1

Table 3.4: System Parameters

3.3.4.1 Effect of Interleaver Depth

Figure 3.8 illustrates the actual decoding trajectory of the DSTS-SP-RSC-AMRWB scheme of Figure 3.3 at E_b/N_0=6 dB, when using an interleaver length of 461 bits, namely the number of bits in each 20 ms AMR-WB-encoded frame. A system-iteration I_{system} was defined as having a single inner iteration between the SP demapper and the RSC decoder, followed by two outer-iterations between the RSC decoder and the AMR-WB decoder. The zigzag-path in Figure 3.8 represents the actual *extrinsic* information transfer between the SP demapper and the projected EXIT curve of Figure 3.8. The Monte-Carlo simulation-based iterative decoding trajectories do not closely follow the EXIT characteristics due to the short interleaver length employed. This behaviour is owing to the fact that the assumptions outlined at the beginning of Section 3.3.3 are not fully satisfied and hence the EXIT chart based performance becomes inaccurate because it is assumed during the EXIT chart evaluation that the *extrinsic* information is uncorrelated, although in reality this assumption has a limited applicability. However, the decoding trajectories shown in Figures 3.9 and 3.10 become clearly more accurate, when an interleaver length of four (1844 bits) or 10 AMR-WB-encoded frames (4610 bits), are employed respectively. More explicitly, it can be observed that the decoding trajectories follow the EXIT characteristics more closely, as the interleaver length increases and reach a higher I_E value in both Figures 3.9 and 3.10 after $I_{system} = 10$ iterations.

3.3.4.2 BER Performance

Figure 3.11 depicts the BER versus Signal to Noise Ratio (SNR) per bit, namely versus E_b/N_0 performance of the DSTS-SP-RSC-AMRWB scheme and that of its corresponding DSTS-SP-RSC benchmarker, when communicating over narrowband temporally correlated Rayleigh fading channels. It can be seen from Figure 3.11 that the DSTS-SP-RSC-AMRWB scheme outperforms the DSTS-SP-RSC benchmarker scheme by about 1 dB at BER=4.0×10^{-5} after $I_{system} = 2$ iterations, where again we define a system iteration I_{system} as having one inner iterations followed by two outer-iteration, as mentioned in Section 3.3.2. The AMR-WB decoded scheme has a lower BER at its speech-decoded output than its benchmarker dispensing with iterative speech decoding, because the *extrinsic* information exchange between the AMR-WB decoder and the RSC decoder has the potential of improving the attainable BER.

From Figure 3.8, it is expected that the DSTS-SP-RSC-AMRWB scheme outperforms

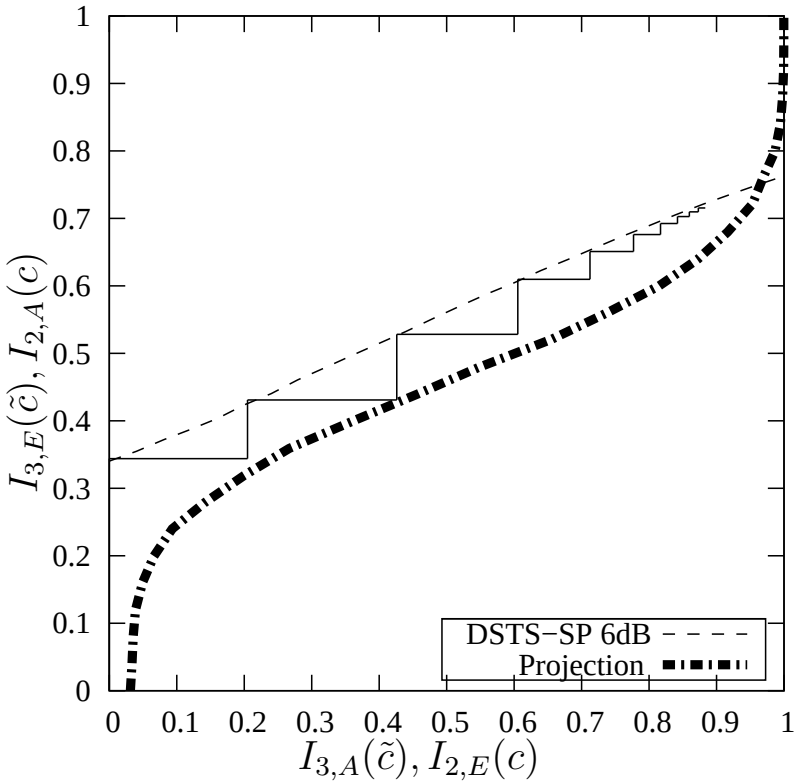

Figure 3.9: Decoding trajectory of the three-stage DSTS-SP-RSC-AMRWB scheme with an interleaver length of 1844 bits at E_b/N_0=6 dB.

the DSTS-SP-RSC benchmarker scheme at E_b/N_0=5.0 dB. This is indeed expected, since there is an open EXIT tunnel for the DSTS-SP-RSC-AMRWB scheme at E_b/N_0=5.0 dB, which is expected to lead to a low BER. However, due to the short interleaver length of 461 bits, the actual iterative decoding trajectories do not closely follow the EXIT characteristics, especially when increasing the number of iterations [94], as seen in Figure 3.8. More explicitly, the actual decoding trajectory of Figure 3.8 recorded for $I_{system} = 10$ iterations at E_b/N_0=5.0 dB was unable to reach $I_{2,E}(c)$=1.0, and hence the combined system's actual BER failed to reach an infinitesimally low value.

3.3.4.3 SegSNR Performance

Let us now study the speech SegSNR performance of the proposed scheme in Figure 3.12. It can be seen from Figure 3.12 that the exploitation of the residual source redundancy during the parameter estimation in soft-bit speech decoding [60] provides valuable *a priori* information. More explicitly, at the point of tolerating a SegSNR degradation of 1 dB, the

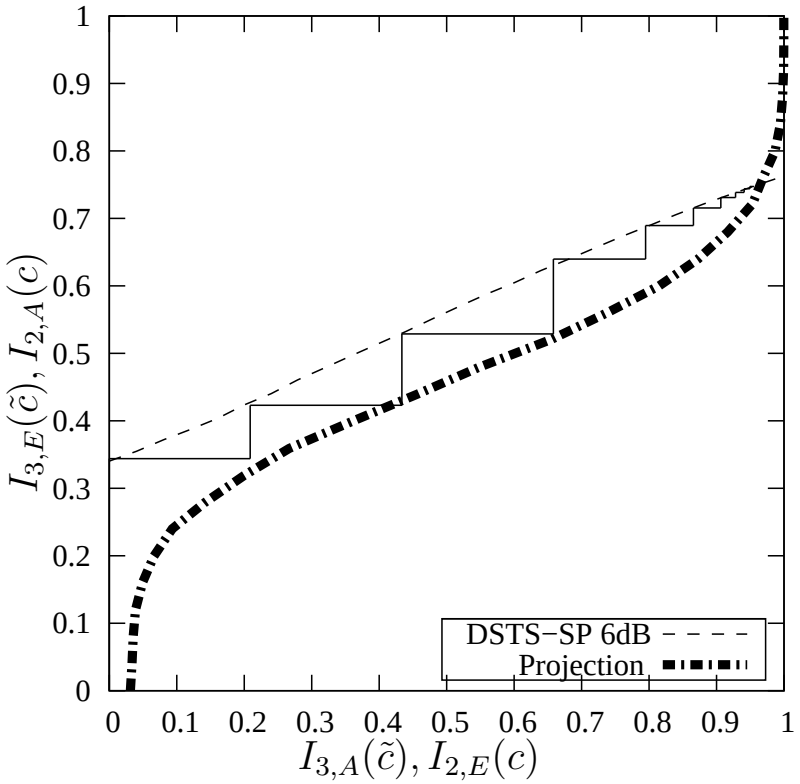

Figure 3.10: Decoding trajectory of the three-stage DSTS-SP-RSC-AMRWB scheme with an interleaver length of 4610 bits at E_b/N_0=6 dB.

employment of soft-bit assisted AMR-WB decoder performs approximately 0.5 dB better in terms of the required channel E_b/N_0 value, than its corresponding hard speech decoding based counterpart. Additionally, iteratively exchanging the soft-information amongst the three receiver components of the amalgamated DSTS-SP-RSC-AMRWB scheme has resulted in a further E_b/N_0 gain of about 3.0 dB after $I_{system} = 2$ iterations, again, at the point of tolerating a SegSNR degradation of 1 dB.

3.3.5 System Conclusions

In this contribution the three-stage turbo detection aided DSTS-SP-RSC-AMRWB scheme of Figure 3.3 and Table 3.3.4 was proposed for transmission over a narrowband temporally correlated Rayleigh fading channel. The employment of the soft-output AMR-WB speech codec, which exploits the residual redundancy inherent in the encoded bitstream demonstrates a significant improvement in terms of the average SegSNR versus channel E_b/N_0 performance compared to its corresponding hard decoding based benchmarker. The per-

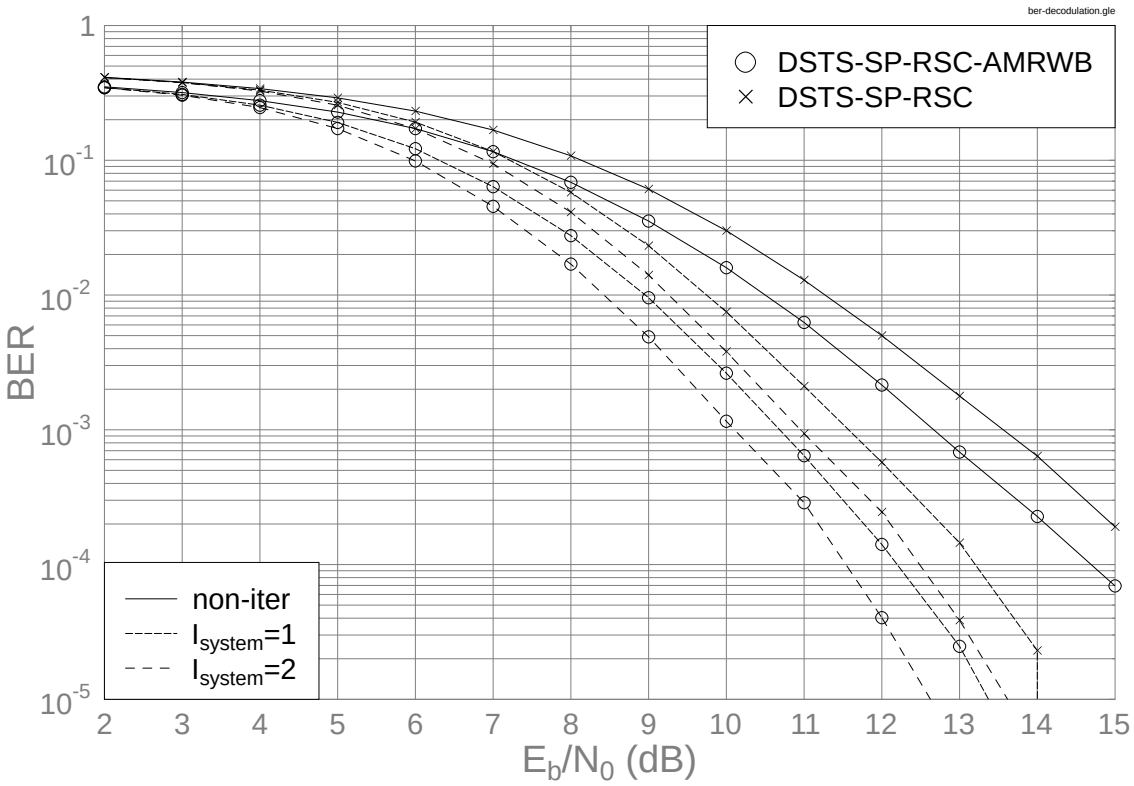

Figure 3.11: BER versus E_b/N_0 performance of the jointly optimised DSTS-SP-RSC-AMRWB scheme of Figure 3.3 and Table 3.3.4, when communicating over correlated non-dispersive Rayleigh fading channels.

formance of the three-component turbo receiver is about 1 dB better in terms of the E_b/N_0 required in comparison to the benchmarker scheme also employing joint iterative channel decoding and DSTS aided SP demodulation, but using separate non-iterative AMR-WB decoding.

3.4 Chapter Conclusion

In this chapter, we proposed a novel system that invokes jointly optimised iterative source and channel decoding for enhancing the error resilience of the AMR-WB speech codec.

In Section 3.2, we characterized the residual redundancy inherent in the AMR-WB encoded parameters. It was observed that the CB gain and the first two ISP parameters exhibit significant inter-frame first-order correlation. Thus, the inter-frame first-order

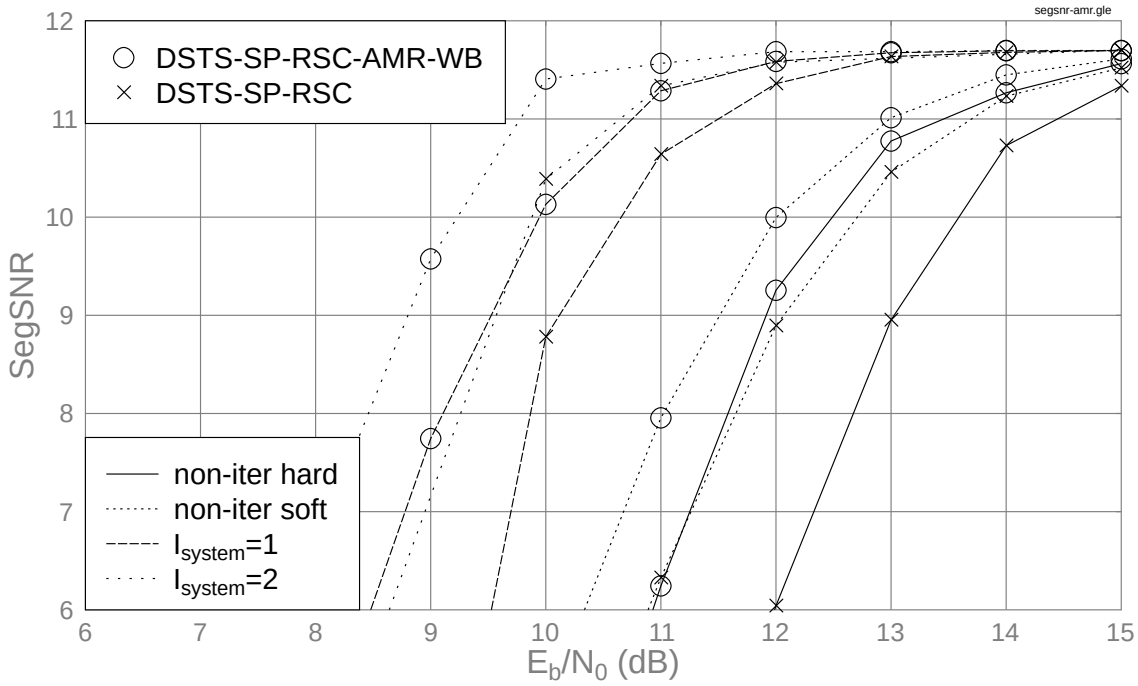

Figure 3.12: Average SegSNR versus E_b/N_0 performance of the jointly optimised DSTS-SP-RSC-AMRWB scheme of Figure 3.3 and Table 3.3.4 in comparison to the DSTS-SP-RSC benchmarker scheme, when communicating over correlated non-dispersive Rayleigh fading channels.

correlation of the CB gain and of the first two ISP parameters were exploited during the decoding process in our proposed scheme of Figure 3.3. Furthermore, we also exploited the unequal-probability-related redundancy of the remaining AMR-WB encoded parameters with the aid of the forward BCJR algorithm, as outlined in Section 3.3.2 because this was achieved at a marginally increased complexity. In Section 3.3.2, we showed how the AMR-WB decoder was modified for exploiting the *a priori* knowledge provided by the channel decoder, which is essential for the employment of iterative source and channel decoding, in the spirit of [48, 61].

Subsequently, EXIT chart analysis was invoked in Section 3.3.3 in order to study the convergence behaviour of the scheme outlined in Figure 3.3. Furthermore, the EXIT transfer characteristics of the source decoder were investigated, when exploiting both the so-called general source statistics, as well as the so-called actual source statistics. It was observed that the EXIT curve of the AMR-WB decoder can reach a higher $I_{1,E}$ value at $I_{1,A} = 1$, when exploiting the actual source statistics.

The BER and SegSNR performance of the soft-bit assited AMR-WB source- and channel-decoder was presented in Section 3.3.4.2 and Section 3.3.4.3, respectively, in conjunction with a two-transmit-antenna aided DSTS-SP system communicating over narrowband temporally correlated Rayleigh fading channels, which does not require sophisticated channel estimation techniques. Hence, it reduces the associated implementational complexity at the cost of about 3 dB E_b/N_0 loss, when compared to the equivalent coherently detected scheme using perfect channel knowledge. Based on our EXIT chart analysis outlined in Section 3.3.3 with the aid of the actual decoding trajectories shown in Figure 3.8 of Section 3.3.2, it is predicted that the proposed scheme will outperform the benchmark scheme carrying out joint channel decoding and DSTS aided SP demodulation in conjunction with separate AMR-WB decoding. The effect of interleaver depth was also addressed in Section 3.3.4.1, where - as expected - it was found that an improved performance is achieved, when employing a larger interleaver size. This is due to the fact that matching the predictions of the EXIT chart analysis is only guaranteed, when employing large interleaver depths. The BER performance of the proposed DSTS-SP-RSC-AMRWB scheme of Figure 3.3 was compared to that of the DSTS-SP-RSC benchmark scheme, where the advocated scheme exhibits an approximately 1 dB signal-to-noise ratio gain in comparison to the benchmark scheme, when using I_{system}=2 system iterations and when communicating over narrowband temporally correlated Rayleigh fading channels.

In the next chapter, we consider the optimization of the soft-bit assisted AMR-WB source decoder with the aid of the EXIT charts.

Chapter 4

EXIT Chart Optimized Soft-Bit Assisted AMR-WB Transceivers

4.1 Introduction

In Chapter 3, we proposed a jointly optimised iterative source- and channel-decoding scheme for enhancing the error resilience of the Adaptive MultiRate WideBand (AMR-WB) speech codec. The employment of the soft-output AMR-WB speech codec, which exploits the residual redundancy inherent in the encoded bitstream demonstrates a beneficial performance improvement in comparison to the benchmark scheme carrying out AMR-WB decoding separately. Subsequently, EXtrinsic Information Transfer (EXIT) charts [65] were used to analyse the convergence behaviour of the proposed system. Ideally, in order for the *extrinsic* information exchange between the inner code and the AMR-WB decoder to converge at a specific E_b/N_0 value, the EXIT curve of the inner code at the E_b/N_0 value of interest and the EXIT curve of the AMR-WB decoder should only intersect at the (1,1) point, which allows the system to exhibit an open *convergence tunnel* [65]. However, it was observed in Figure 3.4 and 3.5 of Section 3.3.3 that the EXIT curve of the AMR-WB decoder is incapable of reaching the (I_A, I_E)=(1,1) point, thus preventing the creation of an open convergence tunnel.

In [133], a recursive inner code was employed in the iterative decoding scheme advocated for the sake of maximising the interleaver gain and avoiding the formation of a BER floor. This principle may be adopted to circumvent the defficiency of the EXIT curve of the soft-bit assisted AMR-WB decoder by serially concatenating the AMR-WB decoder

with a recursive precoder, which indeed resulted in reaching the point of convergence at (1,1). The employment of the precoder is attractive because it has a unity-rate and hence it does not reduce the overall coding rate. To be specific, the inner decoder, the intermediate precoder and the outer source decoder constitute a three-stage serially concatenated system, which performs *extrinsic* information exchange aided iterative decoding across the three receiver components.

It was demonstrated in [62,63] that the innovative concept of soft speech bits employed in the Iterative Source and Channel Decoding (ISCD) scheme [48] can be further improved. Both the intentionally imposed and the inherent unintentional residual redundancy found in source encoded bitstream were exploited for mitigating the effects of transmission errors and the performance of the ISCD scheme was demonstrated also characterised using ten Brink's EXIT charts [65]. Therefore, this principle may be adopted as an alternative to circumvent the defficiencies of the EXIT curve of the soft-bit assisted AMR-WB decoder for the sake of reaching the point of convergence at (I_A, I_E)=(1,1). As a further development, in [78] a novel so-called Over-Complete source-Mapping (OCM) algorithm was developed for intentionally imposing redundancy, on the source-coded bitstream with the intention of enhancing the EXIT curve shape of the source decoder and hence to reach the point of convergence at (1,1).

Therefore, our aim in this chapter is to enhance the soft-bit assisted AMR-WB decoder's EXIT characteristic. More specifically, in this chapter we propose and investigate two turbo transceivers, demonstrating that the performance of the advocated schemes can be further improved by employing an enhanced over-complete mapping aided soft-bit assisted AMR-WB decoder.

More explicitly, in Section 4.2 we propose and investigate a jointly optimized ISCD scheme invoking the AMR-WB speech codec that exploits the intentionally increased residual redundancy of the AMR-WB encoded bitstream by using the novel OCM of [78]. The convergence behaviour of the advocated scheme will be analysed using both Three-Dimensional (3D) as well as Two-Dimensional (2D) EXIT charts [65–67] and the achievable overall system performance will be characterised. In Section 4.3 a turbo transceiver invoking the precoder-aided AMR-WB speech codec is proposed. Finally, this chapter is concluded in Section 4.4.

4.2 Over-Complete Source-Mapping Aided AMR-WB MIMO Transceiver Using Three-Stage Iterative Detection [1]

4.2.1 System Background

The employment of joint source and channel coding techniques has been motivated by the fact that the classic Shannonian source and channel coding separation theorem [13] has limited applicability in practical speech systems [6]. This is due to the delay- and complexity constraints of practical speech transmission systems. For example, iterative turbo decoding can be used to exploit the residual redundancy found in the encoded bitstream of finite-delay lossy speech codecs. This residual redundancy is inherently present owing to the limited-complexity, limited-delay source encoders' failure to remove all the redundancy from the correlated speech source signal.

Vary and his team [59,60] developed the concept of soft speech bits exploiting the residual redundancy. Their work culminated in the formulation of ISCD [48]. More explicitly, in order to improve the overall system performance, *extrinsic* information is exchanged between the constituent decoders including the source decoder. As a further development, in [62] and [63] the inherent residual redundancy of the encoded bitstream was deliberately increased using redundant index assignments and multi-dimensional mapping schemes, which resulted in an enhanced soft-bit source decoder performance.

In [62], the iterative decoding behaviour of an ISCD scheme was studied using EXIT charts [65], for characterizing the achievable performance of the ISCD scheme exploiting the residual redundancy inherent in the source encoded bitstream. In this contribution we propose and investigate the jointly optimised ISCD scheme of Figure 4.1 invoking the AMR-WB speech codec [110] exploiting the intentionally increased residual redundancy of the AMR-WB encoded bitstream by using the novel OCM of [78], which is protected by a Recursive Systematic Convolutional (RSC) code. The resultant bitstream is transmitted using Differential Space-Time Spreading (DSTS) combined with Sphere Packing (SP) modulation [82] over a narrowband temporally correlated Rayleigh fading channel. More explicitly, in the resultant multi-stage scheme *extrinsic* information is exchanged amongst the three constituent decoders, namely the source decoder, the channel decoder

[1] This section is based on N. S. Othman, M. El-Hajjar, A. Q. Pham, O. Alamri, S. X. Ng and L. Hanzo: "Over-Complete Source-Mapping Aided AMR-WB MIMO Transceiver Using Three-Stage Iterative Detection", IEEE International Conference on Communications, Beijing, China, 19-23 May 2008, pp. 751-755 and it was based on collaborative research with the co-authors.

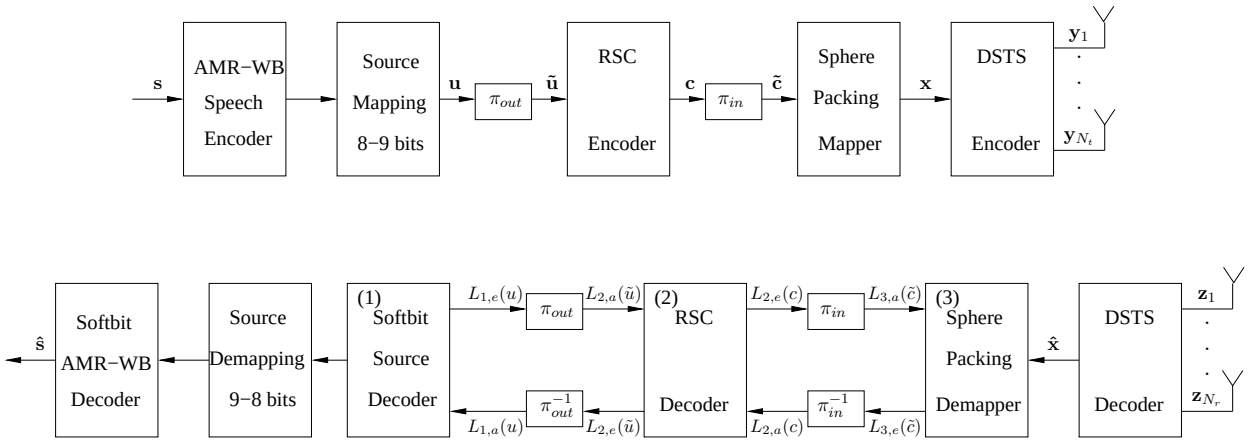

Figure 4.1: Block diagram of the DSTS-SP-RSC-AMRWB-OCM scheme. The notations $\mathbf{s}$, $\hat{\mathbf{s}}$, $\mathbf{u}$, $\mathbf{c}$, $\mathbf{x}$, $\mathbf{y}_j$, $\mathbf{z}_k$, π_{out} and π_{in} represent the frame of the speech source samples, the estimate of the speech source samples, the over-complete source-mapped bits of the encoded speech parameters, the encoded bits of the RSC encoder, the SP-mapped symbols, the DSTS coded symbols of transmitter j, the received symbols at receiver k, the outer bit interleaver, and the inner bit interleaver, respectively. Furthermore, N_t and N_r denote the number of transmit and receive antennas, respectively.

and the DSTS-SP demapper. On the other hand, DSTS employing two transmit and a single receive antenna was invoked for the sake of providing spatial diversity gain with the aid of non-coherent detection, without the potentially high complexity of channel estimation. Moreover, this powerful wireless transceiver benefits from the employment of SP modulation introduced for the sake of increasing the coding gain of the DSTS scheme.

In Section 3.3 the soft-bit assisted AMR-WB codec exploiting the concept of soft speech bits was employed in a multi-stage turbo detection process, which resulted in an enhanced Bit Error Ratio (BER) performance [83]. By contrast, in this treatise we study the achievable performance of the AMR-WB speech codec exploiting the intentionally increased residual redundancy of the AMR-WB encoded bitstream using over-complete source-mapping [78], while employing a 3D EXIT chart based procedure and its 2D EXIT chart projection technique [66, 67] for designing the optimum combination of receiver components. We will refer to this three-stage system as the DSTS-SP-RSC-AMRWB-OCM scheme.

4.2.2 System Overview

4.2.2.1 Transmitter

The DSTS-SP-RSC-AMRWB-OCM system model is depicted in Figure 4.1. As shown in Figure 4.1, *extrinsic* information is exchanged amongst all three constituent decoders, namely the source decoder, the RSC decoder and the SP demapper. The AMR-WB speech codec is capable of supporting nine different bit rates [11], each of which may be activated in conjunction with different-rate channel codecs and different-throughput adaptive modem modes [138]. Similar near-instantaneously adaptive speech and video systems were designed in [1, 6]. In our prototype system investigated here the AMR-WB codec operates at 23.05 kbps, generating a set of 52 speech parameters encoded by a total of 461 bits per 20 ms frame for representing the 8 kHz bandwidth speech signal sampled at 16 kHz.

Each AMR-WB-encoded frame consists of a set of 52 parameters denoted by $\{\mathbf{v}_{1,\tau},\mathbf{v}_{2,\tau},\ldots,\mathbf{v}_{\kappa,\tau},\ldots,\mathbf{v}_{52,\tau}\}$, where $\mathbf{v}_{\kappa,\tau}$ represents an encoded parameter, $\kappa = 1, ..., K$ denotes the index of each parameter in the encoded speech frame, $K = 52$ and τ denotes the time index referring to the current encoded frame index.

However, in the advocated system we employ the over-complete source-mapping philosophy of [78] using a rate of $R_{mapping}$=8/9, which is detailed in Appendix A. Thus, the AMR-WB encoded bitstream is divided into 8-bit source symbols $\tilde{\mathbf{v}}_{\kappa,\tau} = [\tilde{v}(1)_{\kappa,\tau}\ \tilde{v}(2)_{\kappa,\tau}\ \ldots\ \tilde{v}(M)_{\kappa,\tau}]$, where $M = 8$ is the total number of bits assigned to the κth parameter. Then, $\tilde{\mathbf{v}}_{\kappa,\tau}$ is mapped to the bit sequence, $\mathbf{u}_{\kappa,\tau} = [u(1)_{\kappa,\tau}\ u(2)_{\kappa,\tau}\ \ldots\ u(N)_{\kappa,\tau}]$ using over-complete source-mapping, where $N = 9$. Then, the outer interleaver, π_{out} permutes the bits of the sequence $\mathbf{u}$, yielding $\tilde{\mathbf{u}}$ of Figure 4.1.

The bit sequence c of Figure 4.1 is the output of the RSC encoder, where a $\frac{3}{4}$-rate RSC code having a code memory of 3 and octally represented generator polynomials of $(G_1, G_2, G_3, G_4) = (11, 2, 4, 10)_8$ is employed. The RSC encoded bits are interleaved by interleaver π_{in} of Figure 4.1, which are then transmitted by using DSTS-SP. The SP demapper maps B number of channel-coded bits $\tilde{\mathbf{c}}$=$[\tilde{c}_0\ \tilde{c}_1\ \ldots\ \tilde{c}_{B-1}] \in \{0,1\}$ to a SP symbol $x \in X$ as detailed in [82]. Furthermore, we have $B = \log_2(L_{SP}) = \log_2(16) = 4$, where L_{SP} represents the set of legitimate SP constellation points. Subsequently, we have a set of SP symbols that can be transmitted using DSTS and two transmit antennas, where one SP symbol is transmitted in two time slots, hence we have $R_{DSTS-SP}$=1/2, as detailed in [82]. In this study, we consider transmissions over a narrowband temporally correlated

Rayleigh fading channel, associated with a normalised Doppler frequency of $f_D = 0.01$.

Hence, the overall coding rate of the DSTS-SP-RSC-AMRWB-OCM scheme becomes R_{system} =464/708 ≈0.66. The effective spectral efficiency of the DSTS-SP-RSC-AMRWB-OCM scheme is $\log_2(L_{SP}) \cdot R_{system} \cdot R_{DSTS-SP}$ ≈1.31 bits per channel use.

4.2.2.2 Receiver

The notation $L(.)$ in Figure 4.1 denotes the LLRs of the bit probabilities. The notations $\tilde{c}$, c, $\tilde{u}$ and u in the round brackets (.) of Figure 4.1 denote the SP bits, RSC coded bits, RSC data bits and the over-complete source-mapping aided AMR-WB encoded bits, respectively. The specific nature of the LLRs is represented by the subscripts of $L_{.,a}$, $L_{.,p}$ and $L_{.,e}$, denoting the *a priori*, *a posteriori* and *extrinsic* information, respectively, as shown in Figure 4.1. The LLRs associated with one of the three constituent decoders having a label of {1,2,3} are differentiated by the corresponding subscripts (.) of {1,2,3}. Note that the subscript 2 is used for representing the RSC decoder of Figure 4.1.

Inner Iterations: The complex-valued received symbols **z** are demapped to their LLR [130] representation for each of the B number of RSC-encoded bits per DSTS-SP symbol. As seen in Figure 4.1, the *a priori* LLR values $L_{3,a}(\tilde{c})$ provided by the RSC decoder are subtracted from the *a posteriori* LLR values $L_{3,p}(\tilde{c})$ at the output of the SP demapper for the sake of generating the *extrinsic* LLR values $L_{3,e}(\tilde{c})$. Then the LLRs $L_{3,e}(\tilde{c})$ are deinterleaved by a soft-bit deinterleaver. Next, the deinterleaved soft-bits $L_{2,a}(c)$ of Figure 4.1 are passed to the RSC decoder in order to compute the *a posteriori* LLR values $L_{2,p}(c)$ provided by the Maximum *A Posteriori* (MAP) algorithm [143] for all the RSC-encoded bits. The *extrinsic* information $L_{2,e}(c)$ seen in Figure 4.1 is generated by subtracting the *a priori* information $L_{2,a}(c)$ from the *a posteriori* information $L_{2,p}(c)$ according to $L_{2,e}(c) = L_{2,p}(c) - L_{2,a}(c)$, which is then fed back to the SP demapper as the *a priori* information $L_{3,a}(\tilde{c})$ after appropriately reordering them using the inner soft-value interleaver. The SP demapper of Figure 4.1 exploits the *a priori* information $L_{3,a}(\tilde{c})$ for the sake of providing improved *a posteriori* LLR values $L_{3,p}(\tilde{c})$ which are then passed to the RSC decoder and in turn, back to the SP demapper for further iterations.

Outer Iterations: As seen in Figure 4.1, the *extrinsic* LLR values $L_{2,e}(\tilde{u})$ of the original uncoded systematic information bits are generated by subtracting the *a priori* LLR values $L_{2,a}(\tilde{u})$ of the RSC decoder from the LLR values $L_{2,p}(\tilde{u})$ of the original uncoded non-systematic information bits. Then, the LLRs $L_{2,e}(\tilde{u})$ are deinterleaved by the outer

soft-bit deinterleaver. The resultant soft-bits $L_{1,a}(u)$ are passed to the Soft-Bit Source Decoding (SBSD) [48] that computes the *extrinsic* LLR values $L_{1,e}(u)$, as detailed during our further discourse. These *extrinsic* LLR values are then fed back to the RSC decoder after appropriately reordering them in the specific order required by the RSC decoder for the sake of completing an outer iteration.

We define two inner iterations followed by one outer iteration as having one "system iteration" denoted as $I_{system} = 1$. The residual redundancy, which manifests itself in terms of the unequal probability of occurence of the M-ary source symbols is exploited as *a priori* information for computing the *extrinsic* LLR values.

The details of the algorithm used for computing the *extrinsic* LLR values $L_{1,e}(u)$ of the speech parameters can be found in [48], which are briefly reviewed below. Firstly, the channel decoder's output information related to each speech parameter is given by the product of each of the constituent bits as follows:

$$p(\hat{\mathbf{u}}_{\kappa,\tau}|\mathbf{u}_{\kappa,\tau}) = \prod_{i=1}^{N} p(\hat{u}_{\kappa,\tau}(i)|u_{\kappa,\tau}(i)), \tag{4.1}$$

where $\hat{\mathbf{u}}_{\kappa,\tau} = [\hat{u}(1)_{\kappa,\tau}\ \hat{u}(2)_{\kappa,\tau}\ \ldots\ \hat{u}(N)_{\kappa,\tau}]$ is the received bit sequence of the κth parameter, while $\mathbf{u}_{\kappa,\tau}$ is the corresponding transmitted bit sequence provided that all these bits are independent of each other. Hence, by excluding the bit under consideration from the present bit sequence within each of the κth parameter where $\kappa = 1, ..., K$, namely from $\mathbf{u}_{\kappa,\tau} = [u_{\kappa,\tau}(\lambda)\ \mathbf{u}_{\kappa,\tau}^{[ext]}]$, we obtain the *extrinsic* channel output information for each desired bit, $u_{\kappa,\tau}(\lambda)$:

$$p(\hat{\mathbf{u}}_{\kappa,\tau}^{[ext]}|\mathbf{u}_{\kappa,\tau}^{[ext]}) = \prod_{i\neq\lambda, i=1}^{N} p(\hat{u}_{\kappa,\tau}(i)|u_{\kappa,\tau}(i)), \tag{4.2}$$

where the term $\mathbf{u}_{\kappa,\tau}^{[ext]}$ denotes all elements of the bit pattern $\mathbf{u}_{\kappa,\tau}$, but excludes the desired bit $u_{\kappa,\tau}(\lambda)$ itself. Finally, the *extrinsic* LLR value $L_{1,e}(u)$ generated for each bit can be obtained by combining its channel decoder output information and the *a priori knowledge* concerning the κth parameter, $p(\mathbf{u}_{\kappa,\tau})$, which is given by [48, 144]:

$$L_{1,e}(u_{\kappa,\tau}(\lambda)) = \log \frac{\sum_{\mathbf{u}_{\kappa,\tau}^{[ext]}} p(\mathbf{u}_{\kappa,\tau}^{[ext]}|u_{\kappa,\tau}(\lambda) = 0).\exp A}{\sum_{\mathbf{u}_{\kappa,\tau}^{[ext]}} p(\mathbf{u}_{\kappa,\tau}^{[ext]}|u_{\kappa,\tau}(\lambda) = 1).\exp A}, \tag{4.3}$$

where

$$A = \sum_{u_{\kappa,\tau}(l)\ \text{of}\ \mathbf{u}_{\kappa,\tau}^{[ext]}} \frac{u_{\kappa,\tau}(l)}{2}(L_{1,a}[u_{\kappa,\tau}(l)])$$

and $L_{1,a}$ represents the *a priori* LLR values of the SBSD decoder, which is the deinterleaved counterpart of $L_{2,e}$ generated by the RSC decoder.

The proposed scheme's performance was studied against its benchmark scheme, which does not employ the over-complete source-mapping. We will refer to the benchmarker as the DSTS-SP-RSC-AMRWB scheme. The AMR-WB-encoded bitstream is protected by a $\frac{2}{3}$-rate RSC code having a code memory of 4 and octally represented generator polynomials of $(G_1, G_2, G_3) = (23, 2, 10)_8$. Thus, the overall coding rate of the DSTS-SP-RSC-AMRWB scheme dispensing with over-complete source-mapping becomes $R_{benchmark} = 464/708 \approx 0.66$. The effective throughput of the DSTS-SP-RSC-AMRWB scheme dispensing over-complete source-mapping is $\log_2(L_{SP}) \cdot R_{benchmark} \cdot R_{DSTS-SP} \approx 1.31$ bit per channel use. In the benchmark scheme advocated, the soft-bit assisted AMR-WB speech decoder exploiting the natural residual redundancy, which manifests itself in terms of the unequal probability of occurence of the different values of a specific parameter in each 20 ms AMR-WB-encoded frame was invoked, as detailed in [83]. Thus, both the proposed DSTS-SP-RSC-AMRWB-OCM and the DSTS-SP-RSC-AMRWB benchmark schemes have the same overall coding rate and hence the same spectral efficiency.

4.2.3 EXIT-Chart Analysis

EXIT charts have been widely used in the design of iterative schemes, since they facilitate the prediction of the associated convergence behaviour, based on the exchange of mutual information amongst the constituent receiver components.

As seen from Figure 4.1, the RSC decoder receives inputs from and provides outputs for both the SP demapper and the SBSD. More explicitly, let $I_{.,A}(x)$ denote the mutual information (MI) [13] between the *a priori* value $A(x)$ and the symbol x, whilst $I_{.,E}(x)$ denotes the MI between the *extrinsic* value $E(x)$ and the symbol x. The MI associated with one of the three constituent decoders having a label of {1,2,3} is differentiated by the corresponding subscripts (.) of {1,2,3}. Thus, the input of the RSC decoder is constituted by the *a priori* input, $I_{2,A}(c)$ corresponding to the coded bits c originating from the *extrinsic* output of the SP demapper as well as the *a priori* input, $I_{2,A}(\tilde{u})$, available for the data bits $\tilde{u}$, which was generated from the *extrinsic* output of the SBSD. Note that the subscript 2 is used for representing the RSC decoder of Figure 4.1.

Correspondingly, the RSC decoder generates both the *extrinsic* output, $I_{2,E}(c)$, representing the coded bits c as well as the *extrinsic* output, $I_{2,E}(\tilde{u})$ representing the data

bits $\tilde{u}$. Therefore, the EXIT characteristic of the RSC decoder can be described by the following two EXIT functions [66]:

$$I_{2,E}(c) = T_c[I_{2,A}(\tilde{u}), I_{2,A}(c)], \tag{4.4}$$

$$I_{2,E}(\tilde{u}) = T_{\tilde{u}}[I_{2,A}(\tilde{u}), I_{2,A}(c)], \tag{4.5}$$

which are illustrated by the 3D surfaces seen in Figures 4.2 and 4.3, respectively.

By contrast, the SP demapper as well as the soft-bit source decoder only receive input from and provide output for the RSC decoder. Thus, the corresponding EXIT functions are:

$$I_{3,E}(\tilde{c}) = T_{\tilde{c}}[I_{3,A}(\tilde{c}), E_b/N_0], \tag{4.6}$$

for the SP demapper and

$$I_{1,E}(u) = T_u[I_{1,A}(u)], \tag{4.7}$$

for the SBSD. Equations (4.6) and (4.7) are illustrated in Figures 4.2 and 4.3, respectively.

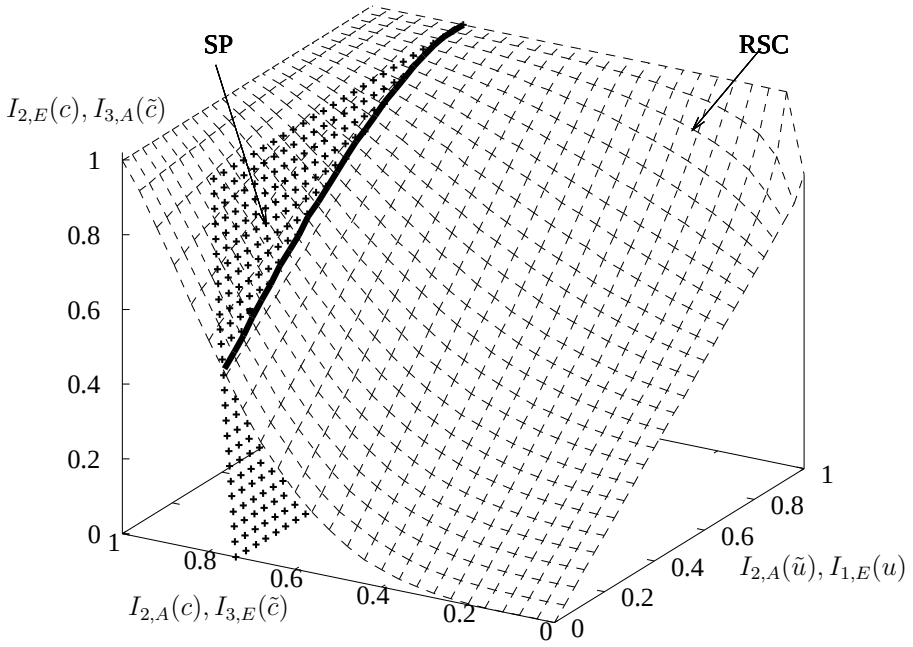

Figure 4.2: 3D EXIT chart of the RSC decoder and the SP demapper at E_b/N_0=8.0 dB.

The EXIT chart analysis [65] of the iterative decoding scheme's convergence behaviour indicates that an infinitesimally low BER may only be achieved by an iterative receiver,

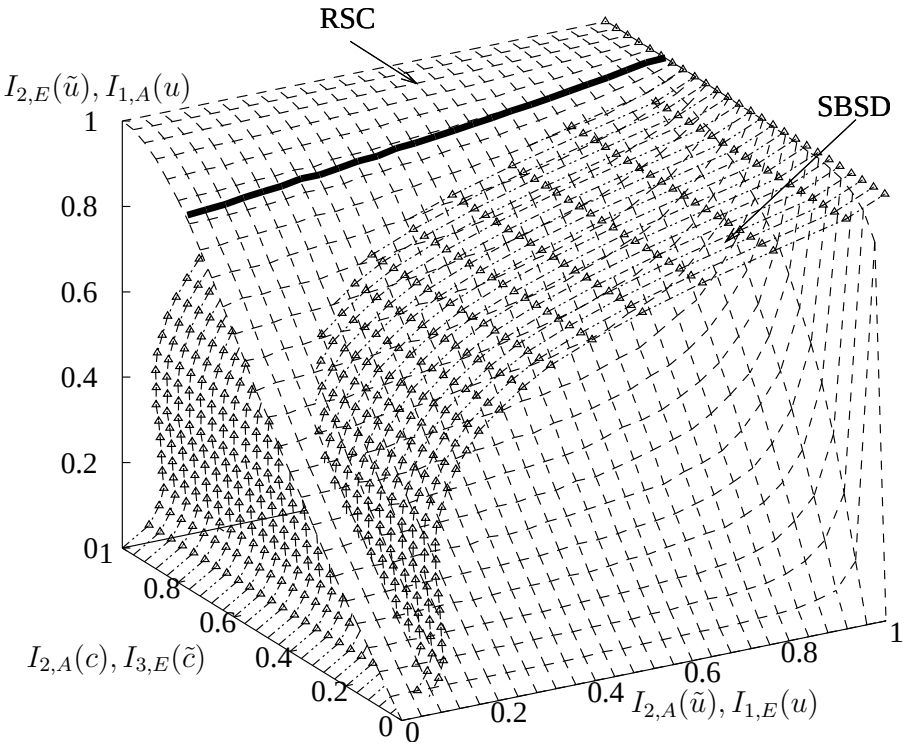

Figure 4.3: 3D EXIT chart of the RSC decoder and the soft-bit source decoder with projection from Figure 4.2.

if an open tunnel exists between the EXIT curves of the two Soft-In-Soft-Out (SISO) components.

More explicitly, the intersection of the surfaces seen in Figure 4.2 characterizes the best possible attainable performance, when exchanging information between the RSC decoder and the SP demapper of Figure 4.1 for different fixed values of $I_{2,A}(\tilde{u})$, which is shown as a thick solid line. For each point $[I_{2,A}(\tilde{u}), I_{2,A}(c), I_{2,E}(c)]$ of this line on the 3D space of Figure 4.2, there is a specific value of $I_{2,E}(\tilde{u})$ determined by $I_{2,A}(\tilde{u})$ and $I_{2,A}(c)$ according to the EXIT function of Equation (4.5). Therefore the solid line on the surface of the EXIT function of the RSC decoder seen in Figure 4.2 is mapped to the solid line shown in Figure 4.3.

In order to avoid the somewhat cumbersome 3D representation, we project the bold EXIT curve of Figure 4.3 onto the 2D plane at $I_{2,A}(c) = 0$, yielding the line indicated by the squares in Figure 4.4. Also shown is the EXIT curve of the AMR-WB decoder employing over-complete source-mapping used in the advocated DSTS-SP-RSC-AMRWB-OCM scheme, which is denoted by the line marked with triangles.

We also carried out the EXIT chart analysis of the DSTS-SP-RSC-AMRWB benchmark scheme. More explicitly, the intersection of the RSC decoder and the SP demapper's 3D surfaces results in a line, which characterizes the best possible attainable performance, when exchanging information between them. This line is then projected onto the 2D plane at an abscissa value of $I_{2,A}(c) = 0$ yielding the dotted line denoted with squares in Figure 4.4. However, in the DSTS-SP-RSC-AMRWB benchmark scheme a 2/3-rate RSC was invoked, as opposed to a 3/4-rate RSC employed in the DSTS-SP-RSC-AMRWB-OCM scheme. The soft-bit assisted AMR-WB decoder dispensing with the over-complete source-mapping is denoted by the dotted line marked with triangles in Figure 4.4.

As seen in Figure 4.4 the EXIT curve of the soft-bit assisted AMR-WB decoder cannot reach the point of convergence at (1,1) and intersects with the EXIT curve of the projected curve, which implies that residual errors persist, regardless of both the number of iteration used and the size of the interleaver. On the other hand, by exploiting the intentionally imposed redundancy of the AMR-WB encoded bitstream using over-complete source-mapping resulted in reaching the point of convergence at (1,1). Thus, there is an open tunnel between the projected EXIT curve and that of the over-complete source-mapping assisted AMR-WB decoder at E_b/N_0=8.0 dB, as seen in Figure 4.4. Thus according to the EXIT chart predictions, the proposed system outperforms its benchmark scheme.

4.2.4 System Performance

In this section, the attainable performance of the proposed scheme is characterised in terms of BER and Segmental Signal to Noise Ratio (SegSNR) [6] evaluated at the speech decoder's output as a function of the channel Signal to Noise Ratio (SNR) per bit.

We consider a two-transmit-antenna aided DSTS-SP system associated with $L_{SP} = 16$ and a single receive antenna. The remaining simulation parameters were described in Section 4.2.2. In our simulations, a single three-stage "system iteration" is constituted by two inner iterations followed by an outer iteration.

Figure 4.5 depicts the BER versus SNR per bit, namely versus E_b/N_0 performance of the DSTS-SP-RSC-AMRWB-OCM scheme and that of its corresponding DSTS-SP-RSC-AMRWB benchmark scheme. It can be seen from Figure 4.5 that the DSTS-SP-RSC-AMRWB-OCM scheme outperforms the DSTS-SP-RSC-AMRWB benchmark scheme by about 3 dB at BER=1×10^{-4} after $I_{system} = 4$ iterations, where again we define a "sys-

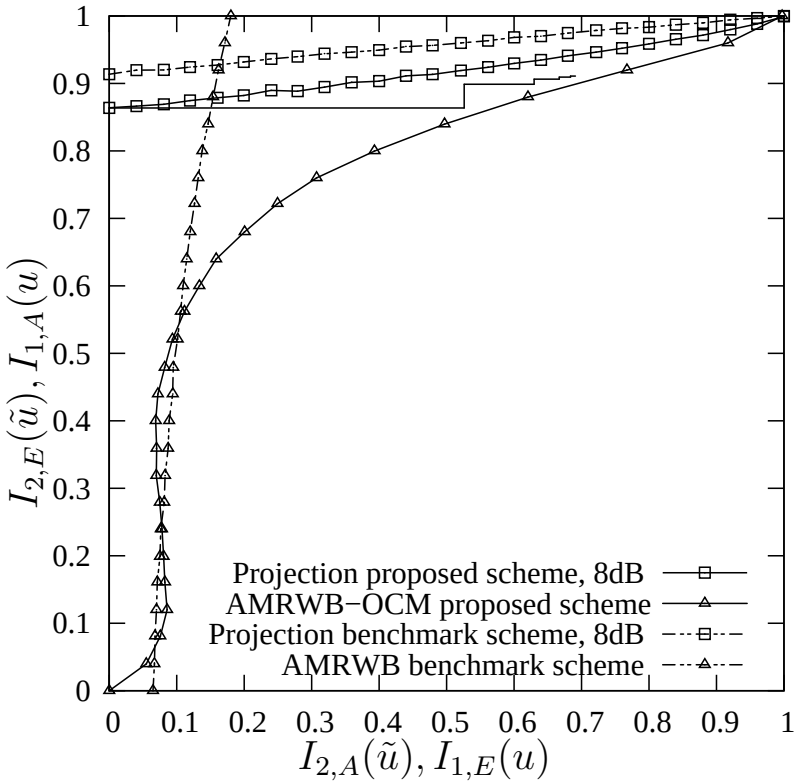

Figure 4.4: 2D projection of the EXIT chart of the proposed DSTS-SP-RSC-AMRWB-OCM scheme at E_b/N_0=8.0 dB.

tem iteration" I_{system} as having two inner iterations followed by a single outer-iteration, as mentioned in Section 4.2.2. The AMR-WB-decoded scheme employing over-complete source-mapping has a lower BER at its speech-decoded output than its benchmarker dispensing with over-complete source-mapping, because the intentionally added residual redundancy of the AMR-WB-encoded bitstream has imposed the EXIT characteristics of the soft-bit source decoder, which resulted in an enhanced attainable BER.

In Figure 4.6 we plot the speech SegSNR performance of the proposed scheme and the benchmark scheme versus E_b/N_0. It can be seen from Figure 4.6 that the exploitation of the deliberately increased residual redundancy in the AMR-WB encoded bitstream has resulted in a E_b/N_0 gain of about 2 dB after $I_{system} = 4$ iterations, when tolerating a SegSNR degradation of 1 dB.

More explicitly, the DSTS-SP-RSC-AMRWB scheme has no OCM scheme, only a rate $R_2 = \frac{2}{3}$ channel encoder, while the DSTS-SP-RSC-AMRWB-OCM scheme employs a rate $R_1 = \frac{8}{9}$ OCM scheme combined with a rate $R_2 = \frac{3}{4}$ channel encoder. Although

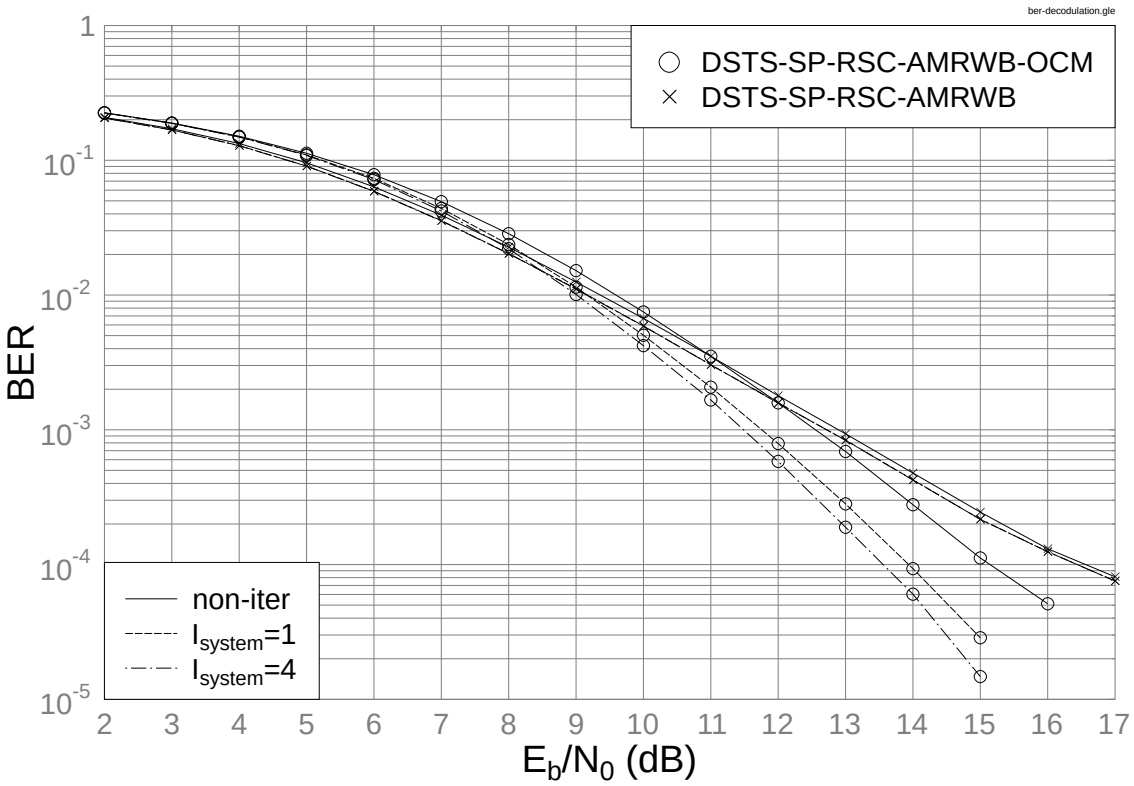

Figure 4.5: BER versus E_b/N_0 performance of the jointly optimised DSTS-SP-RSC-AMRWB-OCM scheme of Figure 4.1, when communicating over narrowband temporally correlated Rayleigh fading channels.

both schemes have the same overall coding rate of $R_{system} = R_{benchmark} = 2/3$, the latter assigns part of its channel encoder's redundancy to the OCM scheme and this is in addition to the source residual redundancy inherited in the source-encoded bitstream. It was shown in Section 4.2.3 that the assignment of channel encoder's redundancy to the OCM created an open EXIT chart tunnel right through to the point of convergence at (1,1) even at a low SNR. The DSTS-SP-RSC-AMRWB-OCM scheme, which benefits from "an early" convergence outperforms the benchmark scheme having the same effective spectral efficiency.

In Section 4.2, we showed that the joint design of source and channel coding was beneficial where the redundancy allocation was appropriately apportioned for the OCM and channel encoders. The achievable performance was contrasted to that of the benchmark scheme where the redundancy was assigned entirely to the channel encoder. Our results also demonstrated that both iterative detection and the appropriate redundancy alloca-

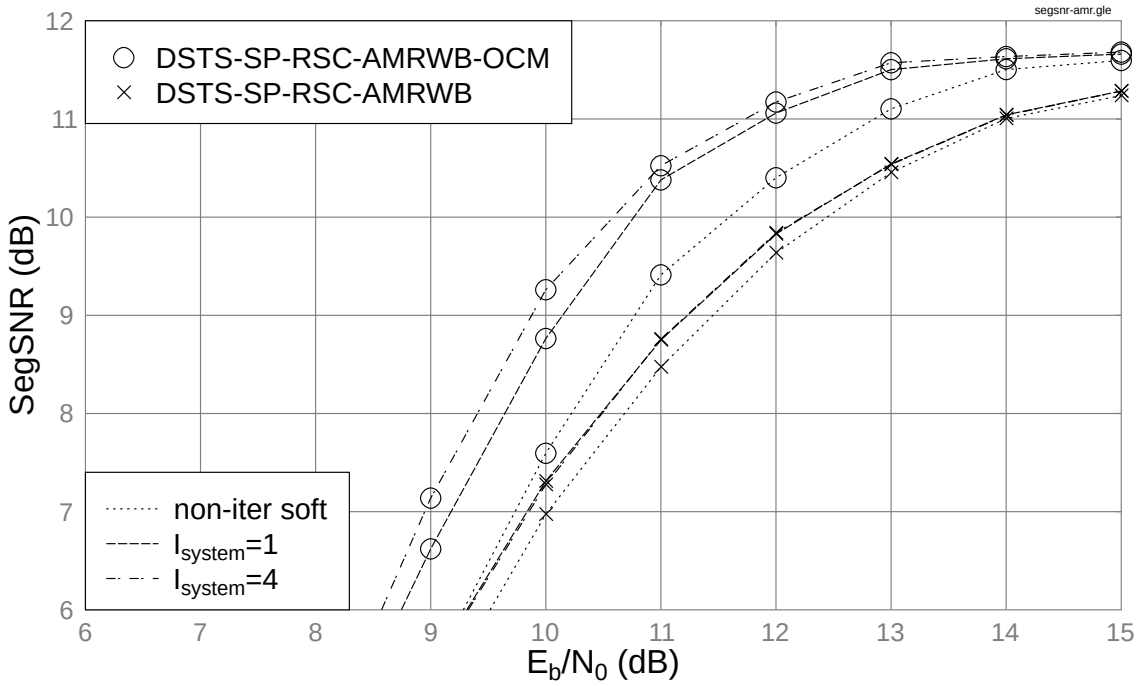

Figure 4.6: Average SegSNR versus E_b/N_0 performance of the jointly optimised DSTS-SP-RSC-AMRWB-OCM scheme of Figure 4.1 in comparison to the DSTS-SP-RSC-AMRWB benchmark scheme, when communicating over narrowband temporally correlated Rayleigh fading channels.

tion between the OCM and channel codecs is crucial in the design of powerful joint source and channel coding schemes.

4.2.5 Effect of Different OCM Rates

In this section, we investigate the effect of employing different combinations of RSC and OCM rates on the DSTS-SP-RSC-AMRWB-OCM scheme of Figure 4.1. More explicitly, we investigate the effect of apportioning the redundancy among the channel encoder and OCM scheme.

Table 4.1 summarises the DSTS-SP-AMRWB-OCM schemes employing different combinations of RSC and OCM rates, while fixing the overall code rate R_{system}. More explicitly, we investigate the DSTS-SP-AMRWB-OCM schemes having two different overall coding rates, namely R_{system}=1/2 and R_{system}=2/3, but employing different combinations of RSC and OCM rates. The RSC-Scheme 2 of Table 4.1 is outlined in Section 4.2.2.

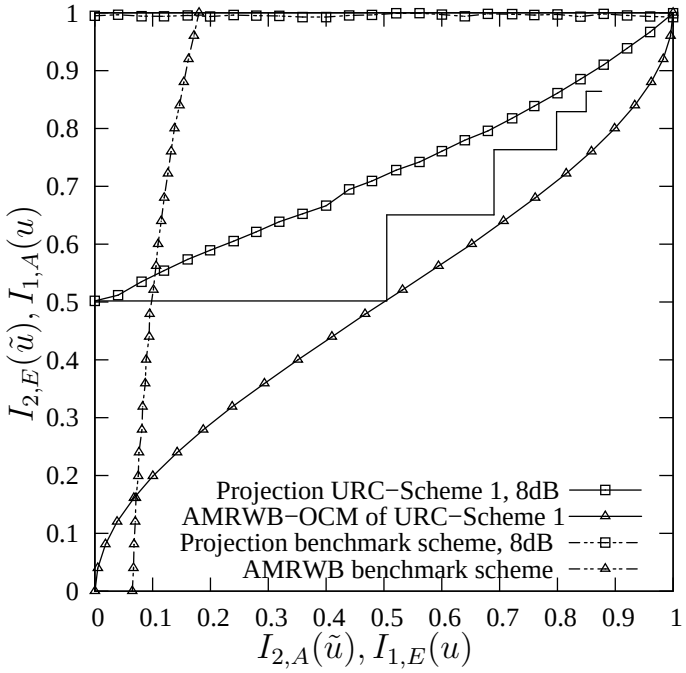

Figure 4.7: 2D EXIT chart of the URC-Scheme 1 of Table 4.1 at E_b/N_0=8.0 dB

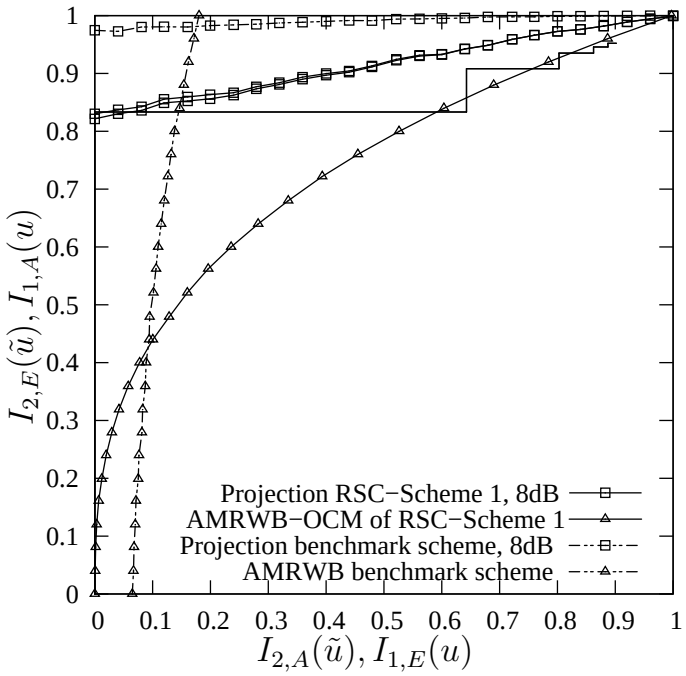

Figure 4.8: 2D EXIT chart of the RSC-Scheme 2 of Table 4.1 at E_b/N_0=8.0 dB.

	System with R_{system}=1/2		**System with R_{system}=2/3**	
	URC-Scheme 1	RSC-Scheme 1	URC-Scheme 2	RSC-Scheme 2
RSC Rate	1	2/3	1	3/4
OCM Rate	1/2	3/4	2/3	8/9
E_b/N_0 **Gain** (dB)	0.3	1.0	1.0	2.0

Table 4.1: The DSTS-SP-RSC-AMRWB-OCM scheme of Figure 4.1 having different combinations of RSC and OCM rates, when R_{system}=1/2 and R_{system}=2/3.

The EXIT charts of the advocated systems of Table 4.1, namely the URC-Scheme 1, the RSC-Scheme 1 and the URC-Scheme 2, are shown in Figures 4.7, 4.8 and 4.9, respectively, while that of the RSC-Scheme 2 is shown in Figure 4.4 of Section 4.2.3. As discussed in Section 4.2.3, the projected EXIT function of both the inner and of the intermediate SISO modules of the DSTS-SP-RSC-AMRWB-OCM schemes are denoted by the line marked with squares, while the lines indicated by the triangles represent the EXIT curve of the AMR-WB decoder assisted OCM schemes, referred as AMRWB-OCM. We will refer to the projected EXIT function of the inner and the intermediate SISO modules of the DSTS-SP-RSC-AMRWB-OCM schemes as the EXIT curve of the DSTS-SP-RSC. Also shown in Figures 4.7, 4.8, 4.9 and 4.4 are the EXIT charts of the corresponding DSTS-SP-RSC-AMRWB benchmark schemes indicated by the dotted line.

As discussed in Section 4.2.3, the EXIT curve of the AMR-WB decoder becomes capable of reaching the point of convergence at (1,1) with the aid of the OCM scheme, which resulted in having an open EXIT tunnel between the DSTS-SP-RSC's projected 2D EXIT curve of Figure 4.3 and that of the AMRWB-OCM arrangement. However, in the DSTS-SP-RSC-AMRWB benchmark scheme, the projected EXIT curve of Figure 4.3 intersects with the EXIT curve of the AMR-WB decoder, which resulted in persistent residual errors. Therefore, it is predicted that the proposed scheme outperforms its benchmark arrangement.

The actual decoding trajectories of the DSTS-SP-RSC-AMRWB-OCM schemes of Table 4.1 recorded at E_b/N_0=8 dB and $I_{system} = 4$ iterations are also illustrated in the corresponding EXIT charts. As discussed in Section 3.3.4.1, the Monte-Carlo simulation-based iterative decoding trajectories do not closely follow the EXIT characteristics due to the short interleaver length employed.

We can observe in Figures 4.7 and 4.8 for the system having the overall code-rate

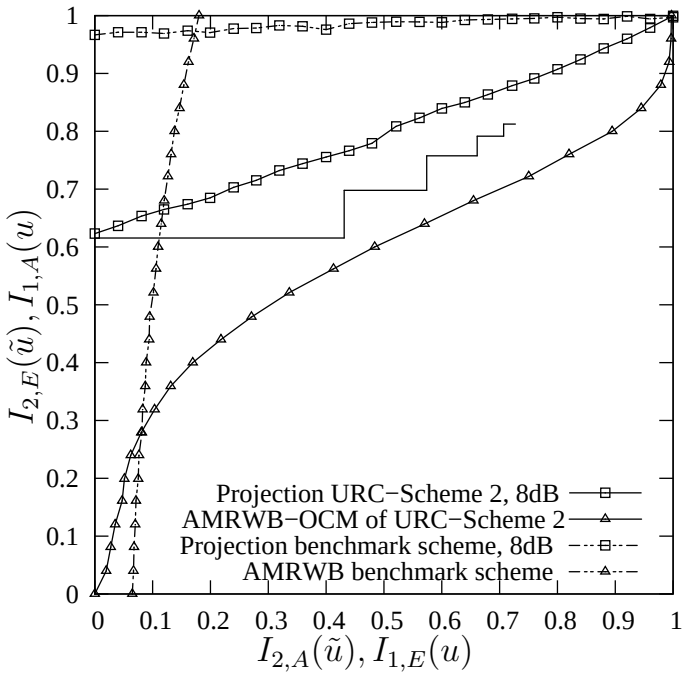

Figure 4.9: 2D EXIT chart of the URC-Scheme 2 of Table 4.1 at E_b/N_0=8.0 dB.

of R_{system}=1/2, that both the URC-Scheme 1 and the RSC-Scheme 1 exhibit an open convergence tunnel at E_b/N_0=8 dB. However, as seen in Figure 4.7, the actual decoding trajectory of the URC-Scheme 1 at E_b/N_0=8 dB and $I_{system} = 4$ iterations reaches the point $(I_{1,A}, I_{1,E}) = (0.86, 0.88)$, while that of the RSC-Scheme 1 is capable of reaching a point closer to the $(I_{1_A}, I_{1,E}) = (1, 1)$, namely $(I_{1,A}, I_{1,E}) = (0.95, 0.90)$. Thus, according to the EXIT chart prediction as well as to the actual decoding trajectory, the RSC-Scheme 1 outperforms the URC-Scheme 1 after $I_{system} = 4$ iterations. Similar observations may be made for the system having the overall code-rate of R_{system}=2/3. Therefore, it was found to be beneficial to apportion the redundancy among the channel encoder and the OCM scheme, rather than assigning all the redundancy to the OCM scheme, despite the fact that they both exhibit an open convergence tunnel.

Figures 4.10, 4.11 and 4.12 depict the BER versus the E_b/N_0 performance of the URC-Scheme 1, of the RSC-Scheme 1 and of the URC-Scheme 2 of Table 4.1, as well as that of their corresponding DSTS-SP-RSC-AMRWB benchmark schemes. It can be seen from Figure 4.10, 4.11 and 4.12 that the URC-Scheme 1, the RSC-Scheme 1 and the URC-Scheme 2 outperform their corresponding benchmark schemes in terms of E_b/N_0 by about 0.3 dB, 1.0 dB and 2.0 dB at BER=1×10^{-4}, after I_{system}=4 iterations. Similarly, the RSC-Scheme 2 outperforms the corresponding benchmark scheme in terms of the required

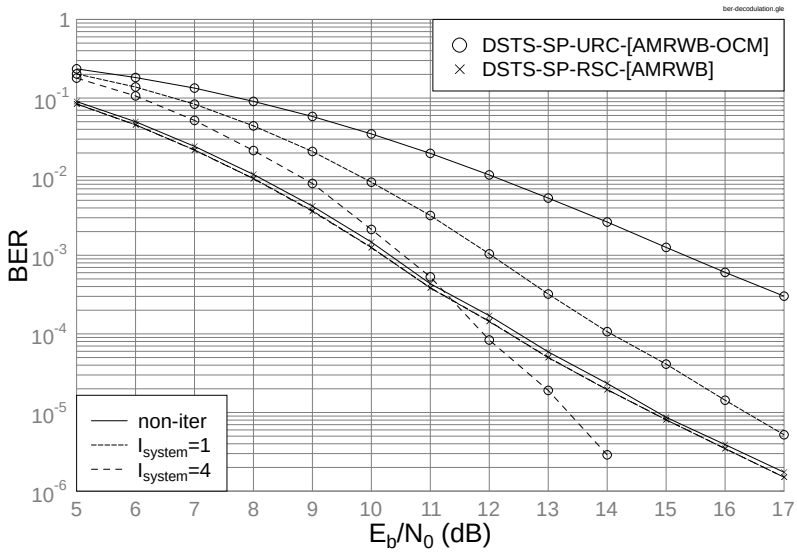

Figure 4.10: BER versus E_b/N_0 performance of the URC-Scheme 1 of Table 4.1, when communicating over narrowband temporally correlated Rayleigh fading channels, having overall coding rate of 1/2.

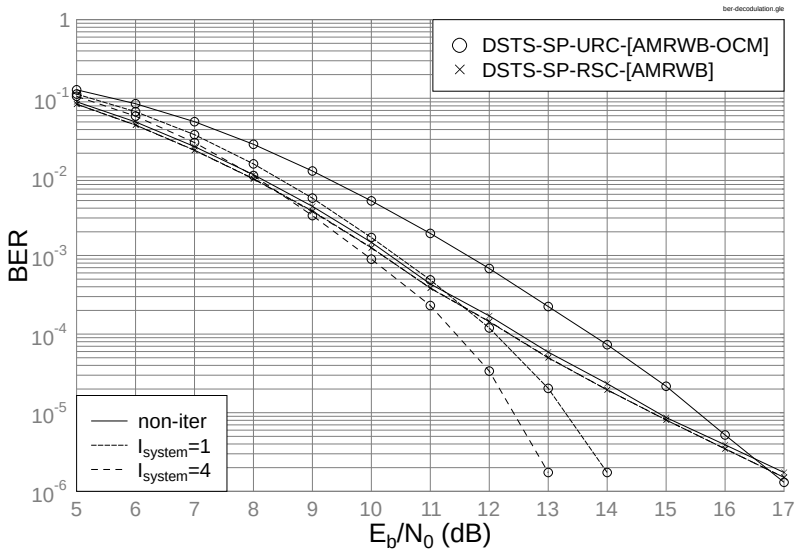

Figure 4.11: BER versus E_b/N_0 performance of the RSC-Scheme 1 of Table 4.1, when communicating over narrowband temporally correlated Rayleigh fading channels, having overall code rate of 1/2.

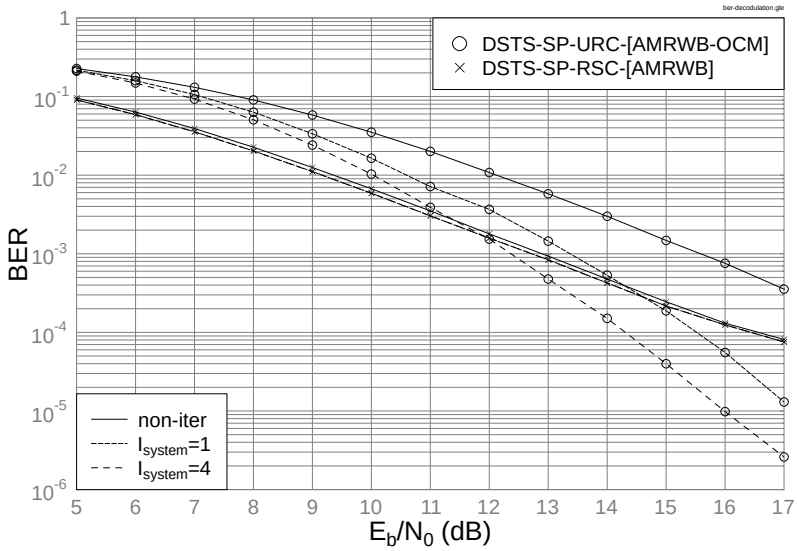

Figure 4.12: BER versus E_b/N_0 performance of the URC-Scheme 2 of Table 4.1, when communicating over narrowband temporally correlated Rayleigh fading channels, having overall coding rate of 2/3.

E_b/N_0 by about 3.0 dB, as shown in Figure 4.5 of Section 4.2.4.

It can be seen from Figures 4.7 and 4.9 that a higher-rate OCM requires a higher channel SNR for an open tunnel to exist between the projected EXIT curve and that of the AMR-WB decoder assisted OCM, which is a prerequisite for the sake of avoiding persistent residual errors. This is reflected in the BER curve shown in Figures 4.10 and 4.12, respectively, where the URC-Scheme 1 and the URC-Scheme 2 required E_b/N_0 values of about 12 dB and 14.2 dB, respectively, for achieving a BER of 1×10^{-4}.

The corresponding SegSNR performances are shown in Figures 4.13, 4.14, 4.15 and 4.6, where the URC-Scheme 1, the RSC-Scheme 1, the URC-Scheme 2 and the RSC-Scheme 2, outperformed their corresponding DSTS-SP-RSC-AMRWB benchmark schemes by approximately 0.3 dB, 1.0 dB, 1.0 dB and 2.0 dB, respectively, when tolerating a SegSNR degradation of 1 dB, after I_{system}=4 iterations.

In the case when the advocated schemes had a fix overall code rate of R_{system}=1/2, it can be observed that the specific scheme which apportions the redundancy among the channel encoder and OCM scheme resulted in a beneficial improvement. More explicitly, the URC-Scheme 1 assigned all the redundancy to the OCM scheme, while, the RSC-Scheme 1 employed a 1/2-rate RSC code combined with a 3/4-rate OCM scheme. Similar observations may be made for the scheme having a fixed overall code rate of R_{system}=2/3.

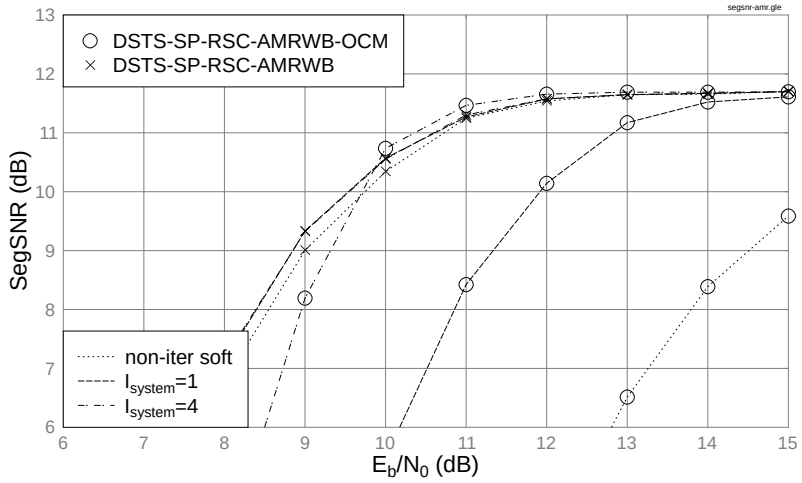

Figure 4.13: SegSNR versus E_b/N_0 performance of the URC-Scheme 1 of Table 4.1, when communicating over narrowband temporally correlated Rayleigh fading channels, having overall coding rate of 1/2

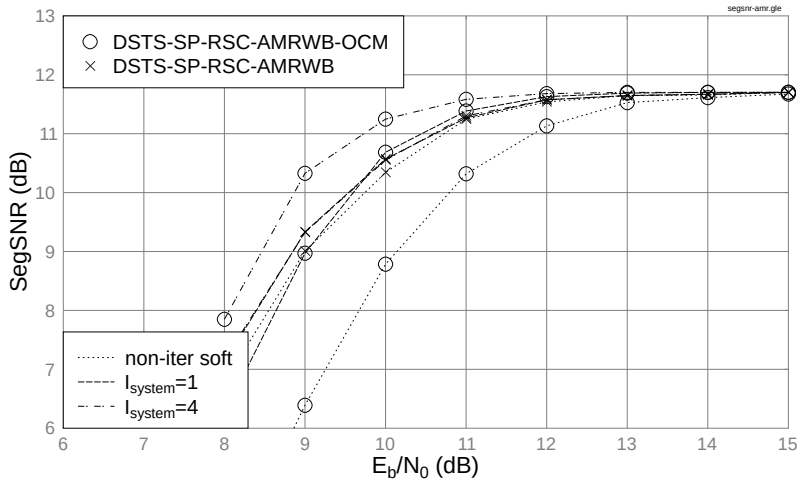

Figure 4.14: SegSNR versus E_b/N_0 performance of the RSC-Scheme 1 of Table 4.1, when communicating over narrowband temporally correlated Rayleigh fading channels, having overall code rate of 1/2.

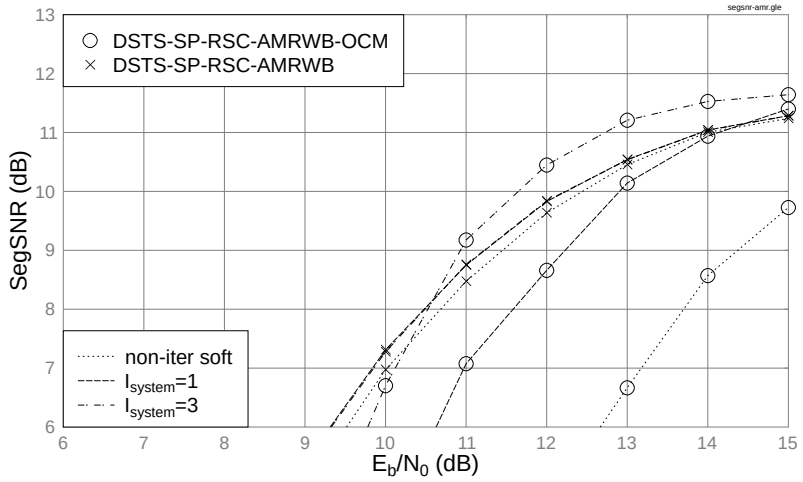

Figure 4.15: SegSNR versus E_b/N_0 performance of the URC-Scheme 2 of Table 4.1, when communicating over narrowband temporally correlated Rayleigh fading channels, having overall code rate of 2/3

CSNR (dB)	**SegSNR Degradation of DSTS-SP-RSC-AMRWB-OCM**			
	System with R_{system}=1/2		**System with R_{system}=2/3**	
	URC-Scheme 1	RSC-Scheme 1	URC-Scheme 2	RSC-Scheme 2
5.0	11.68	11.68	11.68	11.68
6.0	11.68	10.27	11.68	11.68
7.0	11.68	7.28	11.68	9.67
8.0	7.88	3.85	11.68	7.21
9.0	3.50	1.37	8.12	4.54
10.0	0.96	0.45	4.94	2.42
11.0	0.23	0.12	2.47	1.16
12.0	0.04	0	1.19	0.51
13.0	0	0	0.43	0.11
14.0	0	0	0.11	0
15.0	0	0	0	0

Table 4.2: SegSNR degradation of schemes of Table 4.1 having different combinations of RSC and OCM rates, when R_{system}=1/2 and R_{system}=2/3. The above results were extracted from Figures 4.13, 4.14, 4.15 and 4.6.

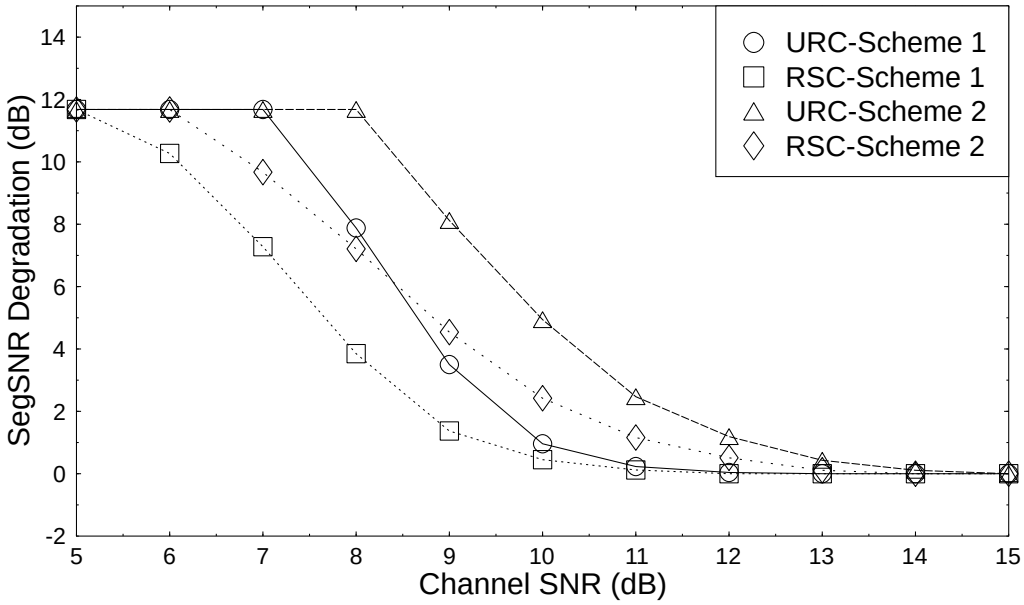

Figure 4.16: SegSNR degradation of schemes of Table 4.1 having different combinations of RSC and OCM rates, when R_{system}=1/2 and R_{system}=2/3. The above results were extracted from Figures 4.13, 4.14, 4.15 and 4.6.

Based on Figures 4.13, 4.14, 4.15 and 4.6, the SegSNR degradation versus Channel SNR (CSNR) performance of the systems of Table 4.1 are summarised in Table 4.2 and plotted in Figure 4.16. More explicitly, the results recorded in Table 4.2 and plotted in Figure 4.16 are based on the performance of the systems of Table 4.1 recorded for the DSTS-SP-RSC-AMRWB-OCM scheme's 4th iteration. The RSC-Scheme 1 and RSC-Scheme 2 require CSNRs of 12.0 dB and 14.0 dB, respectively, when aiming for an error-free audio performance associated with a 0 dB SegSNR degradation, while the URC-Scheme 1 and the URC-Scheme 2 require CSNRs of 13.0 dB and 15.0 dB, respectively. The maximum attainable average SegSNR recorded in Figures 4.13, 4.14, 4.15 and 4.6 for the DSTS-SP-RSC-AMRWB-OCM schemes of Table 4.1 is 11.68 dB. Therefore, both the RSC-Scheme 1 and the RSC-Scheme 2 benefitted from the carefully split redundancy apportioning amongst the channel encoder and OCM scheme, rather than assigning all the redundancy to the OCM schemes.

4.2.6 Conclusions on OCM Aided AMR-WB MIMO Transceiver

In this contribution the three-stage DSTS-SP-RSC-AMRWB-OCM scheme of Figure 4.1 was proposed for transmission over a narrowband temporally correlated Rayleigh fading

channel. The employment of the over-complete source-mapping scheme, which delibrately imposed redundancy on the AMR-WB-encoded bitstream provided a significant improvement in terms of the BER versus channel E_b/N_0 performance compared to its corresponding benchmark scheme dispensing with over-complete source-mapping. The performance of the proposed transceiver is about 3.0 dB better in terms of the E_b/N_0 in comparison to the three-stage benchmark scheme, but dispensing with over-complete source-mapping.

4.3 Three-Stage Iterative Detection of Precoded Soft-Bit AMR-WB for Speech MIMO Transceiver [2]

4.3.1 System Background

Recently, considerable interest has been devoted to the employment of joint source and channel coding in both delay- and complexity-constrained speech transmission systems [6]. This is justified by the limited applicability of Shannon's classic source and channel coding separation theorem [13] in practical speech systems. A beneficial example of this technique exploits the residual redundancy found in the encoded bitstream of finite-delay lossy speech codecs as *a priori* information in an interactive turbo decoding process, as detailed in Chapter 3. More explicitly, the limited-complexity, limited-delay source encoders fail to remove all the redundancy from the correlated speech source signal, thus leaving some residual redundancy in the encoded parameters.

The innovative concept of soft speech bits exploiting the residual redundancy was developed by Vary and his team [59,60], which culminated in the formulation of ISCD. More explicitly, in ISCD *extrinsic* information is iteratively exchanged between the source- and channel- decoder, which results in overall system performance improvements. The convergence behaviour of this iterative decoding scheme can be studied using EXIT charts [65], which characterizes the achievable performance of the ISCD scheme exploiting the residual redundancy inherent in the source encoder bitstream.

In this section we propose the jointly optimised ISCD scheme of Figure 4.17, invoking the AMR-WB-speech codec [110], which employs a Unity Rate Code (URC) [145] referred

[2]This section is based on N. S. Othman, M. El-Hajjar, O. Alamri, S. X. Ng and L. Hanzo: "Three-Stage Iterative Detection of Precoded Soft-Bit AMR-WB for Speech MIMO Transceiver", to be submitted to IEEE 69th Vehicular Technology Conference 2009, and it was based on collaborative research with the co-authors.

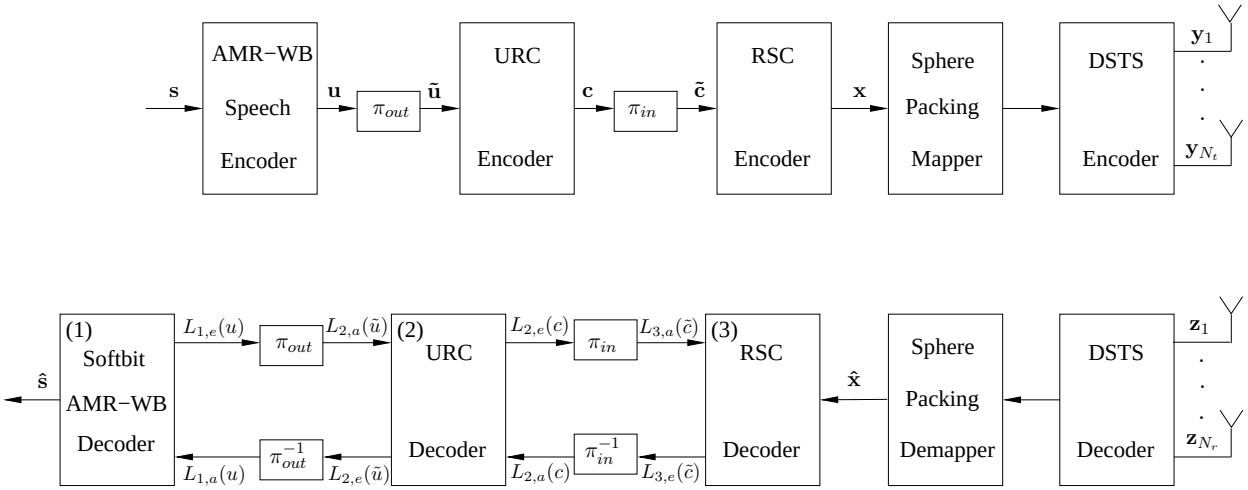

Figure 4.17: Block diagram of the DSTS-SP-RSC-URC-AMRWB scheme. The notations $\mathbf{s}$, $\hat{\mathbf{s}}$, $\mathbf{u}$, $\mathbf{c}$, $\mathbf{x}$, $\hat{\mathbf{x}}$, π_{out} and π_{in} represent the frame of the speech source samples, the estimate of the speech source samples, the bit-representation of the encoded speech parameters, the encoded bits of the URC encoder, the RSC-coded symbols, the received symbols at the receiver, the outer bit interleaver, and the inner bit interleaver, respectively.

to as a precoder. The precoder has a recursive structure, corresponding to an infinite impulse response for the sake of achieving a precoding-aided convergence characteristics enhancement. The resultant bitstream is protected by a RSC code and transmitted using DSTS combined with SP modulation. Similar to the systems of Figures 3.3 and 4.1 in Sections 3.3 and 4.2, the advocated scheme invokes DSTS employing two transmit and a single receive antenna, which provides spatial diversity gain without the potentially high complexity of channel estimation required by coherently detected MIMOs, which is an explicit benefit of using non-coherent detection, while tolerating a typical 3 dB performance loss. We will refer to this three-stage system as the DSTS-SP-RSC-URC-AMRWB arrangement.

This section is structured as follows. Section 4.3.2 provides an overview of the system considered, while in Section 4.3.3 we invoked 3D EXIT charts and their 2D projections for characterizing the iterative detection aided convergence behaviour of the advocated scheme. Section 4.3.4 quantifies the performance of our proposed three-stage scheme, while our conclusions are offered in Section 4.3.5.

4.3.2 System Overview

4.3.2.1 Transmitter

Figure 4.17 shows the schematic of the DSTS-SP-RSC-URC-AMRWB scheme. As shown in Figure 4.17, *extrinsic* information is exchanged amongst the AMR-WB decoder, the URC decoder and the RSC decoder. In the advocated scheme, the soft-bit assisted AMR-WB of Chapter 3 is employed. Recall however from Section 3.3.2 that the AMR-WB speech encoder produces a frame of speech codec parameters, namely $\{\mathbf{v}_{1,\tau},\mathbf{v}_{2,\tau},\ldots,\mathbf{v}_{\kappa,\tau},\ldots,\mathbf{v}_{52,\tau}\}$, where $\mathbf{v}_{\kappa,\tau}$ denotes an encoded parameter, with $\kappa = 1,...,K$ denoting the index of each parameter in the encoded speech frame and $K = 52$, whilst τ denotes the time index referring to the current encoded frame index. Then, $\mathbf{v}_{\kappa,\tau}$ is quantised and mapped to the bit sequence $\mathbf{u}_{\kappa,\tau} = [u(1)_{\kappa,\tau}\ u(2)_{\kappa,\tau}\ \ldots\ u(M)_{\kappa,\tau}]$, where M is the total number of bits assigned to the κth parameter. The outer interleaver π_{out} permutes the bits of the sequence $\mathbf{u}$, yielding $\tilde{\mathbf{u}}$ of Figure 4.17.

The interleaved AMR-WB-encoded bit sequence $\tilde{\mathbf{u}}$ of Figure 4.17 is then encoded by a rate-1 recursive precoder having a code memory of 5. The URC coded bit sequence $\mathbf{c}$ is then permuted by the inner interleaver π_{in}, before it is fed to the $\frac{1}{2}$-rate RSC encoder having a code memory of 3. The RSC encoded bits are transmitted by using DSTS-SP. Recall from Section 4.2.2 that the SP mapper maps B number of channel-coded bits $\tilde{\mathbf{x}}$=$[\tilde{x}_0\ \tilde{x}_1\ \ldots\ \tilde{x}_{B-1}] \in \{0,1\}$ to a SP symbol, as detailed in [82]. Therefore, we have $B = \log_2(L_{SP}) = \log_2(16) = 4$, where L_{SP} represents the set of legitimate SP constellation points. Similar to the system of Figure 4.1 outlined in Section 4.2.2, in this investigation we consider transmissions over a narrowband temporally correlated Rayleigh fading channel, associated with a normalised Doppler frequency of $f_D = 0.01$.

4.3.2.2 Receiver

At the receiver, the *extrinsic* information gleaned is exchanged amongst all three constituent decoders of Figure 4.17, namely the AMR-WB decoder, URC decoder and the RSC decoder in a number of consecutive iterations. The *inner* iterative loop corresponds to the iterative RSC and URC decoders, while the *outer* iterative loop represents the *extrinsic* information exchange between the AMR-WB speech decoder and the URC decoder.

We also use the same notations as in Section 4.2. More explicitly, the notations $L(.)$

in Figure 4.17 represent the LLRs of the bit probabilities, while $\tilde{c}$, c, $\tilde{u}$ and u in the round brackets (.) of Figure 4.17 denote the RSC data bits, URC coded bits, URC data bits and the AMR-WB encoded bits, respectively. The specific nature of the LLRs is represented by the subscripts of $L_{.,a}$, $L_{.,p}$ and $L_{.,e}$, which denote in Figure 4.17 *a priori*, *a posteriori* and *extrinsic* information, respectively. The LLRs associated with one of the three constituent decoders having a label of {1,2,3} are differentiated by the corresponding subscripts (.) of {1,2,3}. Note that the subscript 2 is used for representing the URC decoder of Figure 4.17.

We define an inner iteration between RSC and URC decoders followed by two outer iterations between URC and AMR-WB decoders as having one "system iteration" denoted as $I_{system} = 1$. The advocated scheme invokes the soft-bit assisted AMR-WB of Chapter 3, where the residual redundancy quantified in Section 3.2 is exploited as *a priori* information for computing the *extrinsic* LLR values and for estimating the speech parameters. The details of the algorithm used for computing the *extrinsic* LLR values $L_{1,e}(u)$ of the speech parameters can be found in [48, 61], which was briefly outlined in Section 3.3.2.

The proposed scheme's performance was studied against its benchmark scheme, which does not employ the URC. We will refer to the benchmark scheme as the DSTS-SP-RSC-AMRWB scheme.

4.3.3 EXIT Chart Analysis

Recall from Section 4.2.3 that the employment of EXIT charts [65] in the design of the proposed scheme facilitates the prediction of its convergence behaviour, based on the mutual information exchange amongst the constituent receiver components.

For the readers' convenience, let us recall from Section 4.2.3 the notations used for describing the EXIT charts. Let $I_{.,A}(x)$ denote the mutual information (MI) [13] between the *a priori* value $A(x)$ and the symbol x, whilst $I_{.,E}(x)$ denote the MI between the *extrinsic* value $E(x)$ and the symbol x. The MI associated with one of the three constituent decoders having a label of {1,2,3} is differentiated by the corresponding subscripts (.) of {1,2,3}.

As seen from Figure 4.17, the URC decoder receives inputs from and provides outputs for both the RSC and the AMR-WB decoders. More explicitly, the URC decoder generates both the *extrinsic* output, $I_{2,E}(\tilde{u})$ representing the data bits $\tilde{u}$ as well as the *extrinsic* output, $I_{2,E}(c)$, representing the coded bits c. Both of the *extrinsic* MI outputs are functions of the *a priori* MI inputs of the URC decoder, namely $I_{2,A}(\tilde{u})$ and $I_{2,A}(c)$ which

corresponds to the data bits $\tilde{u}$ originating from the *extrinsic* output of the RSC decoder and the coded bits c generated from *extrinsic* output of the AMR-WB decoder, respectively. Therefore, the EXIT characteristic of the URC decoder can be described by the following two EXIT functions [66]:

$$I_{2,E}(\tilde{u}) = T_{\tilde{u}}[I_{2,A}(\tilde{u}), I_{2,A}(c)], \tag{4.8}$$

$$I_{2,E}(c) = T_{c}[I_{2,A}(\tilde{u}), I_{2,A}(c)], \tag{4.9}$$

which are illustrated by the 3D surfaces seen in Figures 4.18 and 4.19, respectively.

By contrast, the AMR-WB decoder as well as the RSC decoder only receive input from and provide output for the URC decoder. Thus, the corresponding EXIT functions are:

$$I_{1,E}(u) = T_{u}[I_{1,A}(u)], \tag{4.10}$$

for the AMR-WB decoder and

$$I_{3,E}(\tilde{c}) = T_{\tilde{c}}[I_{3,A}(\tilde{c}), E_b/N_0], \tag{4.11}$$

for the RSC decoder for the two-transmit-antenna aided DSTS-SP system associated with a single receive antenna. Equations (4.10) and (4.11) are illustrated in Figures 4.18 and 4.19, respectively.

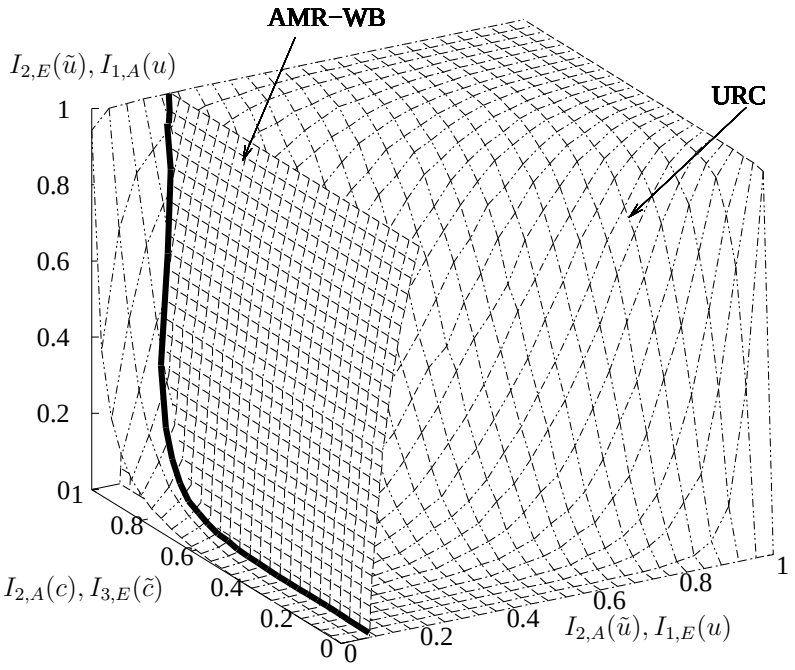

Figure 4.18: 3D EXIT chart of the URC decoder and of the AMR-WB decoder.

The EXIT chart analysis [65] of the iterative decoding scheme's convergence behaviour indicates that an infinitesimally low BER may only be achieved by an iterative receiver, if an open tunnel exists between the EXIT curves of the two Soft-In-Soft-Out (SISO) components.

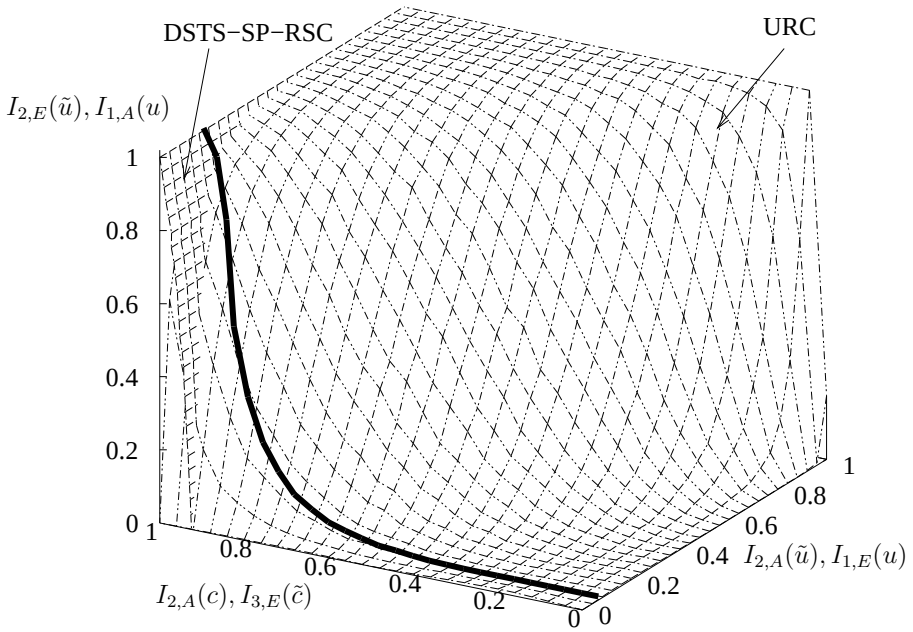

Figure 4.19: 3D EXIT chart of the URC and the RSC decoders for the two-transmit-antenna aided DSTS-SP system at E_b/N_0=10.0 dB with projection from Figure 4.18.

More explicitly, the intersection of the surfaces seen in Figure 4.18 characterizes the best possible attainable performance, when exchanging information between the URC decoder and the AMR-WB decoder of Figure 4.17 for different fixed values of $I_{2,A}(c)$, which is shown as a thick solid line. For each point $[I_{2,A}(\tilde{u}), I_{2,A}(c), I_{2,E}(\tilde{u})]$ of this line on the 3D space of Figure 4.19, there is a specific value of $I_{2,E}(c)$ determined by $I_{2,A}(\tilde{u})$ and $I_{2,A}(c)$ according to the EXIT function of Equation (4.9). Therefore the solid line on the surface of the EXIT function of the URC decoder seen in Figure 4.18 is mapped to the solid line shown in Figure 4.19.

In order to avoid the somewhat cumbersome 3D representation, we project the bold EXIT curve of Figure 4.18 onto the 2D plane at $I_{2,A}(\tilde{u}) = 0$, yielding the line indicated by the squares in Figure 4.20. Also shown is the EXIT curve of the RSC decoder in the two-transmit antenna aided DSTS-SP system associated with a single receive antenna for various E_b/N_0 values. Whilst, in the DSTS-SP-RSC-AMRWB benchmark scheme, the soft-bit assisted AMR-WB decoder dispensing with the URC is denoted by the dotted line

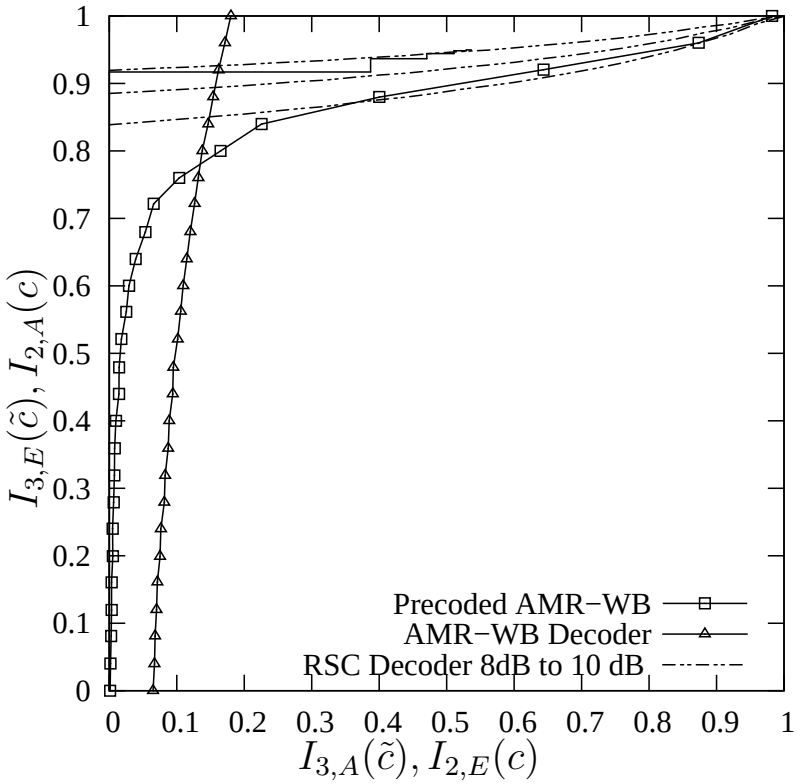

Figure 4.20: 2D projection of the EXIT chart of the three-stage DSTS-SP-RSC-URC-AMRWB scheme and the 2D EXIT chart of the two-stage DSTS-SP-RSC-AMRWB benchmarker scheme.

marked with triangles in Figure 4.20.

As seen in Figure 4.20 the EXIT curve of the soft-bit assisted AMR-WB decoder cannot reach the point of convergence at (1,1) and intersects with the EXIT curve of the RSC decoder, which implies that residual errors persist, regardless of both the number of iteration used and the size of the interleaver. On the other hand, the precoder-aided AMR-WB decoder is capable of reaching close to the point of convergence at (1,1). Thus, there is an open tunnel between the EXIT curve of the RSC decoder and that of the precoder-aided AMR-WB decoder at E_b/N_0=9.0 dB, as seen in Figure 4.20. Thus according to the EXIT chart predictions, the proposed system outperforms its benchmark scheme.

4.3.4 System Performance

In this section we evaluate the attainable overall system performance of the proposed scheme using both the BER and the SegSNR [6]. The simulation parameters were de-

scribed in Section 4.3.2. In our simulations each inner iteration between RSC and URC decoders was followed by two outer iterations between URC and AMR-WB decoders of Figure 4.17, which together formed a three-stage "system iteration".

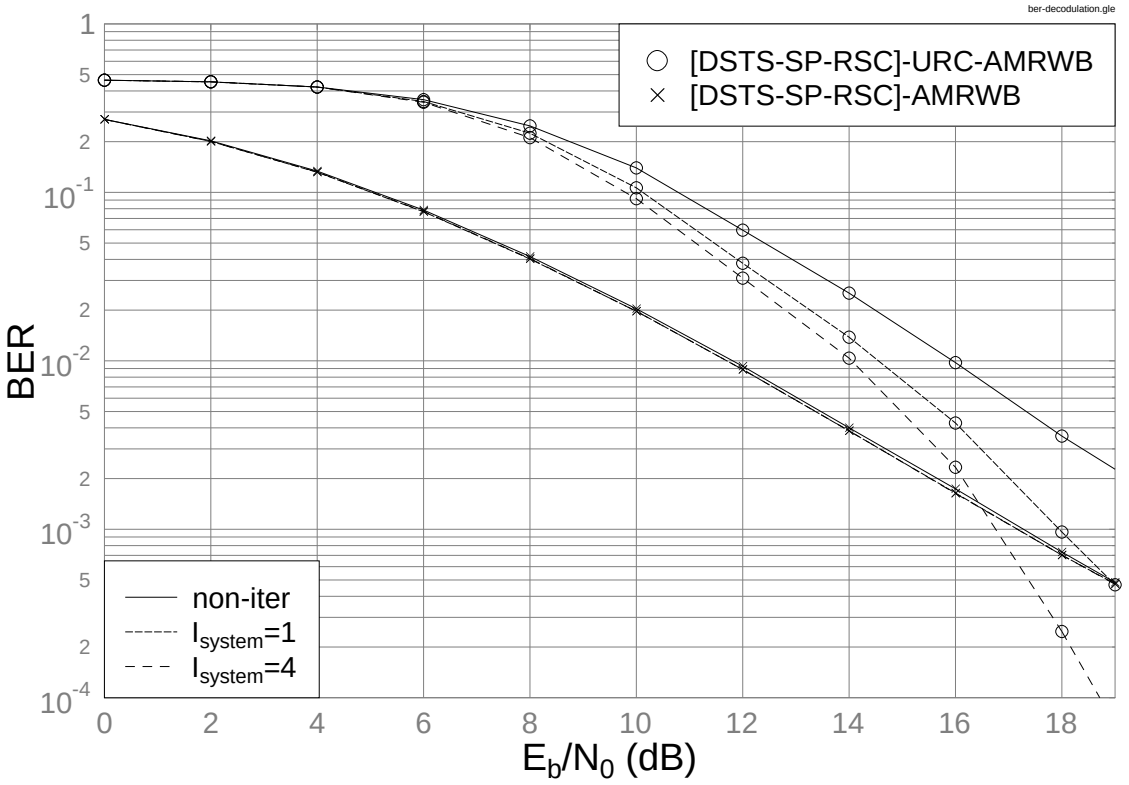

Figure 4.21: BER versus E_b/N_0 performance of the jointly optimised DSTS-SP-RSC-URC-AMRWB scheme of Figure 4.17, when communicating over narrowband temporally correlated Rayleigh fading channels.

Figure 4.21 compares the BER performance of the DSTS-SP-RSC-URC-AMRWB scheme and that of its corresponding DSTS-SP-RSC-AMRWB benchmark scheme, when communicating over narrowband temporally correlated Rayleigh fading channels. It can be seen from Figure 4.21 that the DSTS-SP-RSC-URC-AMRWB scheme outperforms the DSTS-SP-RSC-AMRWB benchmarker by about 1.5 dB at BER=4×10^{-4} after $I_{system} = 4$ iterations, where again, we define a "system iteration" I_{system} as having an inner iteration followed by two outer-iterations, as mentioned in Section 4.3.2. Observe in Figure 4.20 that with the advent of introducing a recursive precoder, the EXIT characteristic of the AMR-WB decoder is enhanced, which resulted in an open EXIT tunnel for the DSTS-SP-RSC-URC-AMRWB scheme at E_b/N_0=9.0 dB. Thus, the system is expected to achieve a

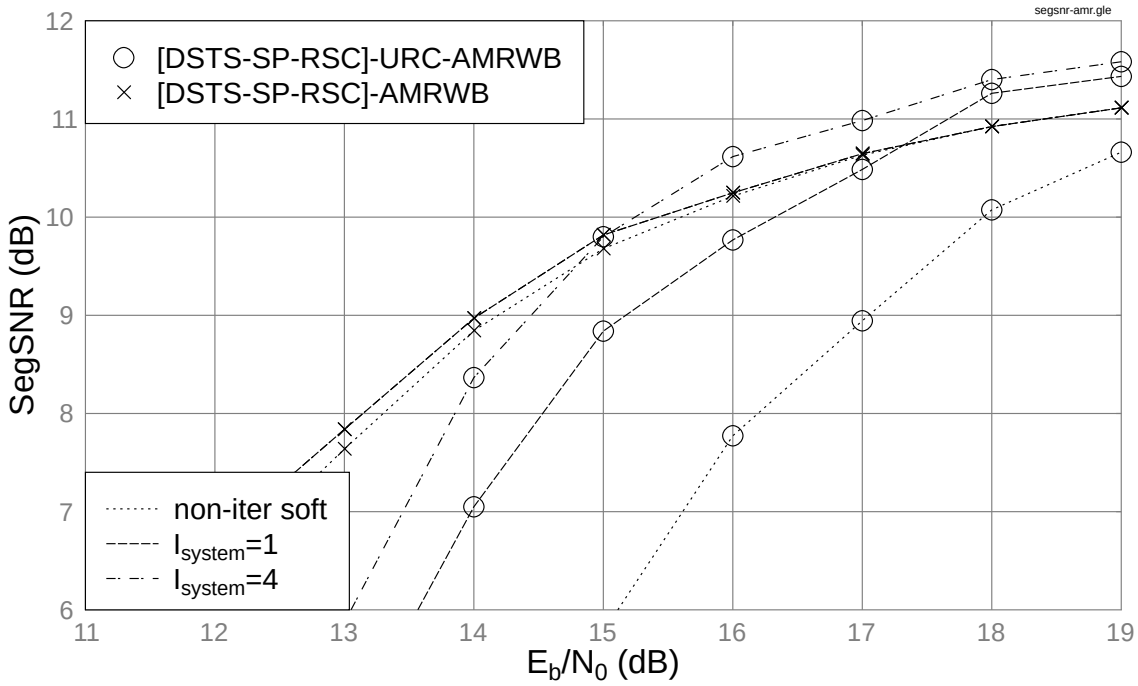

Figure 4.22: Average SegSNR versus E_b/N_0 performance of the jointly optimised DSTS-SP-RSC-URC-AMRWB scheme of Figure 4.17 in comparison to the DSTS-SP-RSC-AMRWB benchmark scheme, when communicating over narrowband temporally correlated Rayleigh fading channels.

low BER. However, the actual decoding trajectory of Figure 4.20 recorded for $I_{system} = 4$ iterations at E_b/N_0=9.0 dB was unable to reach $I_{2,E}(c)$=1.0, and hence the precoded system's actual BER failed to reach an infinitesimally low value.

Figure 4.22 illustrates the speech SegSNR performance of both the proposed scheme and of the benchmark scheme versus SNR per bit, namely E_b/N_0. It is shown in Figure 4.22 that the precoded scheme performs by 1 dB better in terms of the required channel E_b/N_0 value than its corresponding benchmark scheme, when tolerating a SegSNR degradation of 1 dB.

4.3.5 Conclusions on Precoder Aided AMR-WB MIMO Transceiver

In this section, we proposed the three-stage serial concantenated DSTS-SP-RSC-URC-AMRWB scheme of Figure 4.17, where the iterative receiver was constituted by three SISO modules, namely the RSC decoder, the URC decoder and the AMR-WB decoder.

Furthermore, the convergence behaviour of our advocated scheme was analyzed using 3D EXIT charts and their 2D projections. It has been demonstrated that the DSTS-SP-RSC-URC-AMRWB scheme achieves a significant BER performance improvement compared to that of the conventional two-stage DSTS-SP-RSC-AMRWB scheme, when communicating over correlated Rayleigh fading channels. This is achieved by the employment of the recursive URC, which enhanced the EXIT characteristics of the AMR-WB decoder, resulting in an improved BER performance. Explicitly, at BER=4×10^{-4} the precoder was capable of enhancing the achievable E_b/N_0 performance by 1.5 dB, when communicating over correlated Rayleigh fading channels. Our future research aims for combining the benefits of a recursive URC and over-complete-mapping [78] scheme, both of which are capable of improving the EXIT characteristics of the AMR-WB decoder in order to approach the point of convergence at (1,1).

4.4 Chapter Conclusion

In this chapter we presented two jointly optimized turbo transceivers invoking the soft-output AMR-WB speech codec. EXIT charts were used to analyse the convergence behaviour of the proposed turbo transceivers using the soft-bit assisted AMR-WB decoder.

Our novel system benefited both from the residual redundancy inherent in the AMR-WB encoded bitstream as well as from the intentional redundancy imposed by the OCM scheme of [78]. The redundancy inherent in the AMR-WB encoded bitstream was intentionally increased using the OCM scheme in order to shape the EXIT characteristics of the AMR-WB source decoder. Our investigations demonstrated that significant performance improvements may be achieved, when using an appropriate over-complete source-mapper. To be specific, at the same overall code-rate, the proposed system exhibits an approximately 3.0 dB signal-to-noise ratio gain in comparison to the corresponding benchmark scheme dispensing with OCM.

Our investigations studying the effects of having different combinations RSC and OCM rates on the DSTS-SP-RSC-AMRWB-OCM scheme, while fixing the overall code-rate demonstrated that an improved system performance was achieved, when the total affordable redundancy was beneficially apportioned between the channel encoder and the OCM scheme, rather than assigning all the redundancy to the OCM scheme. To be specific, at the same overall code-rate of R_{system}=1/2, the RSC-Scheme 1 of Table 4.1 exhibited an approximately 1.0 dB SNR gain in comparison to the corresponding benchmarker dispensing

with the OCM scheme. On the other hand, the URC-Scheme 1 of Table 4.1 exhibited an approximately 0.3 dB SNR gain, when compared to its corresponding benchmark scheme, despite exhibitting an open convergence tunnel. This investigation was repeated for the DSTS-SP-RSC-AMRWB-OCM scheme having an overall code-rate of R_{system}=2/3. It was observed that the RSC-Scheme 2 exhibited an approximately 2.0 dB SNR gain, while the URC-Scheme 2 had a more modest 1.0 dB E_b/N_0 gain in comparison to the benchmark scheme dispensing with the OCM scheme, when tolerating a SegSNR degradation of 1 dB after I_{system}=4 iterations.

Finally, we proposed a further novel system that benefits from precoding-aided convergence characteristic enhancements. The convergence behaviour of the advocated system was analysed with the aid of 3D EXIT charts and their 2D projections. Our investigations demonstrated that useful performance improvements may be achieved, when performing iterative decoding exchanging *extrinsic* information between the URC and the soft-bit assisted AMR-WB decoders, over the DSTS-SP-RSC-AMRWB arrangement using no URCs. More explicitly, the proposed DSTS-SP-RSC-URC-AMRWB scheme exhibits an approximately 1.5 dB signal-to-noise ratio gain in comparison to the benchmark scheme dispensing with the precoder.

Chapter 5

Conclusions and Future Work

The summary of our investigations in this thesis is presented in Section 5.1, while, possible extensions of our work are proposed in Section 5.2.

5.1 Summary and Conclusions

In this thesis, the Joint Source and Channel Coding (JSCC) schemes designed for speech and audio Multi-Input Multi-Output (MIMO) transceivers have been investigated. Specifically, two different types of JSCC schemes, namely those employing the Unequal Error Protection (UEP) concept, as well as the Iterative Source and Channel Decoding (ISCD) concept, were evaluated in the context of speech and audio MIMO transceivers.

In Chapter 2, we conducted research into jointly optimised turbo transceivers capable of providing unequal error protection for employment in audio and speech transceivers. Specifically, speech and audio codecs, namely the MPEG-4 Advanced Audio Coding (AAC), the MPEG-4 Transform-Domain Weighted Interleaved Vector Quantization (TwinVQ) and the Adaptive MultiRate WideBand (AMR-WB) audio/speech codecs were evaluated in the context of UEP aided turbo transceivers.

The bitstream sensitivity of the MPEG-4 AAC audio codec was investigated in Section 2.3.3, while the MPEG-4 TwinVQ and the AMR-WB audio/speech codecs were characterised in Figures 2.7 and 2.10 of Section 2.3.4 and Section 2.4.2, respectively. Based on these bit sensitivity studies, we observed that each of the source encoded bits has a different impact on the attainable audio/speech quality, when it is corrupted. For example, we found from Figures 2.10 of Section 2.4.2, that the AMR-WB encoded bits represent-

ing the Immittance Spectrum Pairs (ISPs), the adaptive codebook delay, the algebraic codebook index and the vector quantized gain are fairly error sensitive, while the least sensitive bits are related to the fixed codebook's excitation pulse positions. As a remedy to this problem, two different UEP aided JSCC philosophies were proposed, namely a turbo-detected UEP scheme using two different-rate convolutional codes and that using more sophisticated Irregular Convolutional Codes (IRCCs).

In Section 2.5 the audio/speech bits were partitioned into two bit-sensitivity classes and protected accordingly with two different-rate Non-Systematic Convolutional (NSC) codes. The study of an UEP scheme in conjunction with a turbo-detected Trellis Coded Modulation (TCM) and Space-Time Trellis Coding (STTC) scheme was carried out in Section 2.5. More specifically, the MPEG-4 AAC, the MPEG-4 TwinVQ and the AMR-WB audio/speech codecs were incorporated in the sophisticated UEP turbo transceiver of Figure 2.11. The proposed scheme's performance was studied against its benchmark scheme, which consists of STTC and a single-class NSC code. In our simulations we observed that the proposed STTC-TCM-2NSC assisted audio/speech codec schemes outperform their single-class STTC-NSC benchmarker, when communicating over uncorrelated Rayleigh fading channels. More specifically, the audio/speech Segmental SNR (SegSNR) performance of the proposed twin-class schemes was presented in Figures 2.15, 2.17 and 2.19. The maximum attainable average SegSNRs recorded in Figures 2.15, 2.17 and 2.19 for the STTC-TCM-2NSC assisted MPEG-4 AAC, MPEG-4 TwinVQ and AMR-WB audio/speech schemes are 16.28 dB, 13.8 dB and 10.6 dB, respectively. In Section 2.5.2.1, a comparative study of the STTC-TCM-2NSC assisted MPEG-4 AAC audio codec was provided. The class-1 audio bits in the STTC-TCM-2NSC-1 scheme were protected using a 1/2-rate NSC code, while the bits of the STTC-TCM-2NSC-2 scheme were protected using a 2/3-rate NSC. The corresponding system components of the UEP aided audio and speech transceivers are summarised in Table 5.1. The audio/speech SegSNR performance of the twin-class STTC-TCM-2NSC audio/speech schemes expressed in terms of their E_b/N_0 gain, when compared to their corresponding single-class STTC-NSC benchmarker audio/speech were studied in Section 2.5.

We can observe from Table 5.1 that the STTC-TCM-2NSC-1 and STTC-TCM-2NSC-2 schemes of the MPEG-4 AAC audio transceivers, referred as System 1 and System 2, outperformed their corresponding single-class STTC-NSC audio benchmarkers by approximately 1.5 dB and 2.0 dB, respectively, when aiming for their respective maximum attainable average SegSNRs, while maintaining a similar complexity. The decoding com-

Source Codec	**MPEG-4 AAC**		**MPEG-4**	**AMR-WB**
	System1	System2	**TwinVQ**	
Bit Rates (kbps)	48.00		32.00	15.85
Interleaver Length (bits)	1116		743	317
Channel Codec	**Non-Sytematic Convolutional Code**			
	the most sensitive 50%		the most sensitive 25%	
Class1	$R_1 = 1/2$	$R_1 = 2/3$	$R_1 = 1/2$	$R_1 = 1/2$
	$L_1 = 3$	$L_1 = 3$	$L_1 = 3$	$L_1 = 3$
Class2	$R_2 = 3/4$			
	$L_2 = 3$			
Modulation Scheme	**Trellis Coded Modulation (16QAM)**			
The Code Rate	$R_3 = 3/4$			
The Code Memory	$L_3 = 3$			
MIMO Scheme	**STTC**			
Number of Transmitters	N_t=2			
Number of Receivers	N_r=2			
The Code Memory	$L_4 = 4$			
Effective Throughput (BPS)	1.77	2.07	1.95	1.98
E_b/N_0 **Gain** (dB)	1.5	2.0	0.5	0.5

Table 5.1: General overview of the UEP schemes using convolutional codes in Section 2.5 and their corresponding E_b/N_0 gain, when compared to their corresponding single-class STTC-NSC benchmarker audio/speech schemes. The remaining parameters of these schemes were summarised in Tables 2.5 and 2.6.

plexity was quantified in terms of the number of trellis decoding states, as defined in Section 2.5.2.3. As discussed in Section 2.5.3.1, the class-1 audio bits of System 1 have a higher protection at the cost of a lower throughput compared to System 2. Furthermore, the SNR performance of the class-2 audio bits of System 1 is approximately 0.5 dB poorer than that of System 2 at BER=10^{-5}. Thus, this is reflected in the SegSNR performance of both System 1 and System 2 as shown in Figure 2.15, where the average SegSNR performance of System 1 is seen to be inferior in comparison to that of the System 2, despite providing a higher protection for the class-1 audio bits.

In Section 2.5.2.2, a comparative study of the STTC-TCM-2NSC assisted MPEG-4 TwinVQ and AMR-WB audio/speech codecs was provided, where both schemes protect

the most sensitive 25% of the bits in class-1 using a 1/2-rate NSC code. By contrast, the remaining 75% of the bits in class-2 are protected by a 3/4-rate NSC code. We can observe from Table 5.1, that the twin-class STTC-TCM-2NSC assisted MPEG-4 TwinVQ and AMR-WB audio/speech schemes outperformed their corresponding single-class STTC-NSC audio/speech benchmarkers by approximately 0.5 dB, in terms of the required E_b/N_0, when aiming for their respective maximum attainable average SegSNRs. These findings were recorded at a similar complexity. Again, the relatively modest advantage of the twin-class protected transceiver using the MPEG-4 TwinVQ and AMR-WB audio/speech codecs was a consequence of having a rather limited turbo-interleaver length. As a benefit of the longer interleaver of the audio transceiver used for the MPEG-4 AAC codec, a significant performance improvement is achieved.

Based on Figures 2.15, 2.17 and 2.19, the SegSNR degradation versus Channel SNR (CSNR) performance of the systems studied in Section 2.5 are summarised in Table 5.2 and plotted in Figure 5.1. More explicitly, the results recorded in Table 5.2 and plotted in Figure 5.1 are based on the STTC-TCM-2NSC scheme's 4th iteration. The System 2 configuration of both the MPEG-4 AAC audio transceiver, as well as of the MPEG-4 TwinVQ audio transceiver require a CSNR of 3.0 dB, when aiming for an error-free audio performance associated with a 0 dB SegSNR degradation, despite having the benefit of a longer interleaver in the System 2 configuration of the MPEG-4 AAC audio transceiver. As discussed in Section 2.3.2.2, the employment of the interleaved vector quantization allows for fixed-length codewords to be used, resulting in a robust, error-resilient compression technique [107,109], which does not lose synchronization, as variable length coding would. Furthermore, based on the error sensitivity of the various bits of the MPEG-4 TwinVQ codec in each frame shown in Figure 2.7, only 10% of the MPEG-4 TwinVQ encoded bits were found to be gravely error sensitive. However, the MPEG-4 TwinVQ audio transceiver maps the most sensitive 25% of the encoded bits to class-1, therefore the 25% class-1 bits of the MPEG-4 TwinVQ protected by a 1/2-rate NSC code also includes some bits, which were found to be only moderately sensitive to channel errors. However, in the case of the MPEG-4 AAC audio transceiver, the 50% class-1 bits of the MPEG-4 AAC protected by a 2/3-rate NSC code was found to have SegSNR degradations in excess of 16 dB.

On the other hand, the AMR-WB speech transceiver requires a CSNR of 4.5 dB, when aiming for an error-free speech performance associated with a 0 dB SegSNR degradation. This is reflected by the fact that all the bits of the class-1 25%-partition were found to be quite sensitive to channel errors. Additionally, the inferiority of the AMR-WB speech

transceiver was a consequence of having the shortest turbo-interleaver length.

CSNR (dB)	**SegSNR Degradation of STTC-TCM-2NSC**			
	MPEG-4 AAC		**MPEG-4 TwinVQ**	**AMR-WB**
	System1	System2		
0.0	16.28	16.28	13.82	10.63
0.5	16.28	16.28	13.59	10.63
1.0	16.28	16.28	10.49	10.63
1.5	16.28	16.28	4.94	9.00
2.0	16.28	6.34	0.95	3.89
2.5	12.99	0.25	0.08	1.18
3.0	0.27	0	0	0.29
3.5	0.02	0	0	0.06
4.0	0	0	0	0.05
4.5	0	0	0	0

Table 5.2: SegSNR degradation of the UEP schemes using convolutional codes in Section 2.5 assisted MPEG-4 AAC, MPEG-4 TwinVQ and AMR-WB audio/speech codecs. The above results were extracted from Figures 2.15, 2.17 and 2.19.

In Section 2.6, a turbo transceiver employing Irregular Convolutional Codes (IRCCs) was proposed for the sake of providing UEP for the AMR-WB speech codec. In Section 2.6.4, the IRCC employed in the UEP system was constructed with the aid of EXIT charts, based on the design procedure of Section 2.6.3. As discussed in Section 2.6.4, the employment of the IRCCs in the UEP system leads to a system which capable of achieving a convergence SNR threshold of about 1.2 dB. The IRCC was designed to ensure that its EXIT curve matched the EXIT curve of the inner decoder. At the same time, the concept of JSCC was incorporated in an effort to meet the source sensitivity constraints determined by the error sensitivity of the AMR-WB source bits.

Our investigations demonstrated that the proposed UEP system employing the IRCC was capable of achieving a better BER than the corresponding Equal Error Protection (EEP) based benchmark system, which employs an identical-complexity regular NSC code. This trend was expected, based on the EXIT chart analysis of Section 2.6.4. At the same time, the IRCC arrangement was capable of providing UEP, since its constituent codes have different coding rates and hence different error protection capabilities. As a benefit,

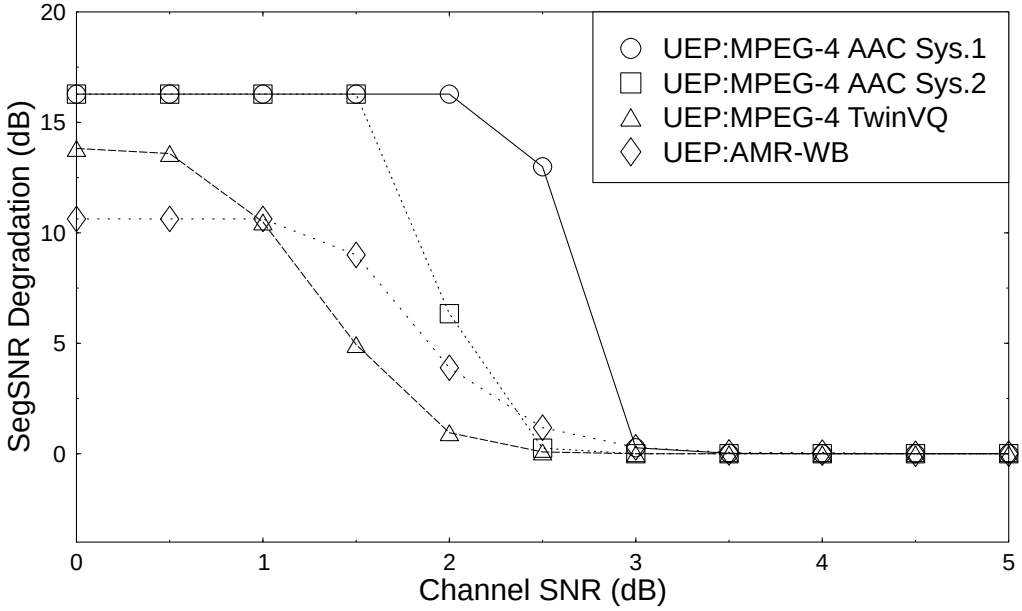

Figure 5.1: SegSNR degradation of the UEP schemes using convolutional codes in Section 2.5 assisted MPEG-4 AAC, MPEG-4 TwinVQ and AMR-WB audio/speech codecs. The above results were extracted from Figures 2.15, 2.17 and 2.19.

the UEP aided system using IRCCs exhibits an Signal to Noise Ratio (SNR) advantage of about 0.4 dB over the equal-protection system employing regular convolutional codes, when communicating over a Gaussian channel, at the point of tolerating a SegSNR degradation of 1 dB. Again, this relatively modest advantage of the IRCC protected transceiver was a consequence of having a rather limited turbo-interleaver length, as discussed in Section 2.6.5. In conclusion, the employment of the IRCC in the proposed UEP aided speech transceiver in the context of a real-time application resulted in a superior SegSNR performance, when compared to the EEP benchmark scheme. The proposed UEP system was designed to achieve low convergence SNR with the aid of EXIT charts analyis, as well as to meet the source sensitivity constraints.

As discussed in Chapter 1, the limited validity of Shannon's source and channel coding separation theorem for practical wireless speech communication systems motivates the application of JSCC. An attractive technique is offered by the concept of ISCD, where the source decoder is further developed for the sake of accepting the *a priori* information passed to it from the channel decoder as *extrinsic* information, in the spirit of [59–61]. More explicitly, a realistic, finite-delay lossy source codec leaves some residual redundancy in the encoded parameters and this redundancy may be beneficially exploited during the

decoding process of an ISCD scheme [61].

Chapter 3 commenced with the characterization of the residual redundancy inherent in the AMR-WB encoded parameters. More explicitly, in Section 3.2, we commenced our discussions by outlining the manifestation of this residual redundancy, namely the so-called unequal-probability-related redundancy, as well as the inter-frame or intra-frame first-order correlation. As discussed in Section 3.2, it was observed that the CB gain and the first two ISP parameters exhibit significant inter-frame first-order correlation. However, our investigations demonstrated that the intra-frame first-order correlation of all parameters within a given frame is insignificant, as shown in Table 3.3. To elaborate a little further, their trellis decoding would be required for exploiting the related intra-frame first-order correlation, yet the attainable performance would only be marginally increased. On the other hand, the complexity imposed when exploiting the unequal-probability-based *extrinsic* information would only be increased marginally. Therefore, we exploited the inter-frame first-order correlation of the CodeBook (CB) gain and of the first two Immittance Spectrum Pair (ISP) parameters, as well as the unequal-probability-related redundancy of the remaining AMR-WB encoded parameters during the decoding process of our proposed system of Figure 3.3 in Section 3.3.

In Section 3.3.2, we showed how the AMR-WB decoder was modified for exploiting the *a priori* knowledge provided by the channel decoder with the aid of the forward BCJR algorithm. Subsequently, EXtrinsic Information Transfer (EXIT) chart analysis was invoked in Section 3.3.3 in order to study the convergence behaviour of the scheme outlined in Figure 3.3 of Section 3.3. Furthermore, the EXIT transfer characteristics of the source decoder were investigated, when exploiting both the general source statistics, as well as the so-called actual source statistics. It was observed that the EXIT curve of the AMR-WB decoder can reach a higher $I_{1,E}$ value at $I_{1,A} = 1$, when exploiting the actual source statistics. As discussed in Section 3.3.3.1, the actual source statistics gives more accurate information since it was derived in an on-line fashion from the actual transmitted source signal. However, this requires an increased bitrate for the encoding of the source statistics, since it is evaluated for each frame and sent to the receiver as side information. Therefore, our research in Section 3.3 was conducted by exploiting the so-called general source statistics.

In Section 3.3, we proposed a novel system that invokes jointly optimised iterative source and channel decoding for enhancing the error resilience of the AMR-WB speech codec, denoted as the DSTS-SP-RSC-AMRWB scheme. The performance of the soft-bit

assited iterative AMR-WB source-decoding and turbo-detection scheme was presented in Section 3.3.4 in conjunction with a two-transmit-antenna aided DSTS-SP system communicating over narrowband temporally correlated Rayleigh fading channels, which does not require sophisticated channel estimation techniques. Hence, it reduces the associated implementational complexity at the cost of about 3 dB E_b/N_0 loss. The proposed scheme's performance was evaluated against the benchmark scheme, which carries out joint channel decoding and Differential Space-Time Spreading (DSTS) aided Sphere-Packing (SP) demodulation in conjunction with separate AMR-WB decoding.

The BER and SegSNR performance of the turbo scheme outlined in Section 3.3.2 was discussed in Section 3.3.4.2 and Section 3.3.4.3, respectively. The predictions of our EXIT chart analysis outlined in Section 3.3.3 were verified by generating the actual decoding trajectories seen in Figure 3.8 of Section 3.3.2. The effect of interleaver depth was also addressed in Section 3.3.4.1, where - as expected - it was found that an improved performance is achieved, when employing a larger interleaver size. This is due to the fact that matching the predictions of the EXIT chart analysis is only guaranteed, when employing large interleaver depths. The BER performance of the proposed DSTS-SP-RSC-AMRWB scheme of Figure 3.3 was compared to that of the DSTS-SP-RSC benchmark scheme, where the advocated scheme exhibits an approximately 1 dB SNR gain in comparison to the benchmark scheme, when using I_{system}=2 system iterations and when communicating over narrowband temporally correlated Rayleigh fading channels.

The proposed scheme also benefitted from the exploitation of the residual redundancy inherent in the AMR-WB encoded parameters, where an approximately 0.5 dB E_b/N_0 gain was achieved in comparison to its corresponding hard speech decoding based counterpart, as shown in Figure 3.12. As discussed in Section 3.3.4.3, at the point of tolerating a SegSNR degradation of 1 dB, the amalgamated DSTS-SP-RSC-AMRWB scheme achieved a further E_b/N_0 gain of about 3.0 dB after $I_{system} = 2$ iterations by iteratively exchanging the soft-information amongst its three receiver components.

In Chapter 4, we evaluated the performance of the transceivers when employing the EXIT-chart optimized soft-bit assisted AMR-WB source decoder. More explicitly, it was observed in Figures 3.4 and 3.5 of Section 3.3.3 in Chapter 3 that the soft-bit assisted AMR-WB decoder is incapable of reaching the (I_A, I_E)=(1,1) point, regardless of whether the unequal-probability-related redundancy or the inter-frame first-order correlation are exploited. This condition resulted in an intersection between the EXIT curve of the soft-bits assisted AMR-WB decoder and that of the inner code before reaching the point of

convergence at (I_A, I_E)=(1,1), thus preventing the creation of an open convergence tunnel. Therefore, in Chapter 4 the EXIT curve of the soft-bit assisted AMR-WB decoder was further shaped using the proposed Over-Complete source Mapping (OCM) scheme, as well as, a Unity Rate Code (URC) referred to as a precoder, in an effort for the soft-bit assisted AMR-WB decoder EXIT curve to reach the point of convergence at (1,1), which resulted in an improved system performance.

In Section 4.2, we proposed a novel system that benefitted both from the residual redundancy inherent in the AMR-WB encoded bitstream as well as from the intentional redundancy imposed by the OCM scheme of [78]. Specifically, we artificially introduced redundancy in the AMR-WB encoded bitstream using the proposed OCM scheme, in order to shape the EXIT characteristics of the AMR-WB source decoder. The performance of the OCM aided AMR-WB Multiple-Input-Multiple-Output (MIMO) transeiver using three-stage iterative detection was presented in Section 4.2.4 in conjunction with a two-transmit-antenna aided DSTS-SP system communicating over narrowband temporally correlated Rayleigh fading channels. The proposed scheme benefitted from the employment of the differential detection scheme, which eliminates the associated implementational complexity of channel estimation, but at the cost of about 3 dB E_b/N_0 loss, when compared to the equivalent coherently detected scheme using perfect channel knowledge. The proposed scheme's performance was evaluated against the benchmark scheme, which does not employ the OCM scheme.

The BER and SegSNR performance of the turbo detected transceiver employing the OCM scheme outlined in Section 4.2.2 was discussed in Section 4.2.4. Our investigations demonstrated that significant performance improvements may be achieved, when using an appropriate OCM scheme. To be specific, at the same overall code-rate and at BER=1×10^{-4} after $I_{system} = 4$ iterations, the proposed system exhibited an approximately 3.0 dB SNR gain in comparison to the benchmark scheme dispensing with over-complete source mapping.

The effect of employing different combinations of Recursive Systematic Convolutional (RSC) and OCM rates on the DSTS-SP-RSC-AMRWB-OCM scheme was addressed in Section 4.2.5, while fixing the overall code-rate. More specifically, two systems having two different overall code rates, namely R_{system}=1/2 and R_{system}=2/3, but having different combinations of RSC and OCM rates were investigated, as summarised in Table 4.1. As discussed in Section 4.2.5, it was found that the system that carefully apportioned the redundancy among the channel encoder and the OCM scheme provided a high system

performance, when compared to the system that assigned all the redundancy to the OCM scheme, despite the fact that they both exhibited an open convergence tunnel.

In Section 4.3, a system that benefits from precoding-aided convergence characteristic enhancements was proposed. The convergence behaviour of the advocated system was analysed with the aid of Three-Dimensional (3D) EXIT charts and their Two-Dimensional (2D) projections. Our investigations demonstrated that useful performance improvements may be achieved over the uncoded DSTS-SP-RSC-AMRWB arrangement, when performing iterative decoding exchanging *extrinsic* information between the URC and the soft-bit assisted AMR-WB decoders. More explicitly, the proposed DSTS-SP-RSC-URC-AMRWB scheme exhibits an approximately 1.5 dB SNR gain in comparison to the benchmark scheme dispensing with the precoder at BER=4×10^{-4} after $I_{system} = 4$ iterations.

5.2 Future Work

Our suggestions for future research are outlined as follows:

- Comparative Study of Precoding-aided and Overcomplete Source Mapping Aided AMR-WB Transceivers

  It was shown in Chapter 4 that the EXIT transfer characteristic of the soft-bit assisted AMR-WB was further enhanced by employing a recursive precoder, as well as by intentionally imposing redundancy using overcomplete source mapping. It would be interesting to comparatively study the performance of both techniques in conjunction with a turbo transceiver, allowing us to draw qualitative conclusions.

- Iterative Precoding-aided Source-Decoding and Turbo-Detection Scheme for AMR-WB+

  In Chapter 4, we proposed a precoding-aided AMR-WB turbo-detection scheme, which was designed for communicating over narrowband temporally correlated Rayleigh fading channels. It would be interesting to consider a two-transmit-antenna aided DSTS-SP system communicating over correlated narrowband temporally Rayleigh fading channels, which does not require sophisticated channel estimation techniques. Furthermore, the employment of the Adaptive MultiRate WideBand Plus (AMR-WB+) audio codec is expected to benefit from longer interleaver lengths, when considering either high-quality, high-rate or non-realtime applications. More explicitly, the AMR-WB+ audio codec has been selected by the Third Generation Partnership

Project (3GPP) as a new audio coding standard for Third Generation (3G) mobile audio services [146, 147].

- Joint Source and Channel Decoding that Exploits the Residual Source Redundancy in the Analogue Domain
  It was shown in Chapter 3 that the EXIT transfer characteristics of the soft-bit assisted AMR-WB decoder is affected by the accuracy of the source statistics used as *a priori* knowledge for computing the *extrinsic* information. We found that the EXIT curve shape will be enhanced if the actual statistics are exploited for computing the *extrinsic* information. However, this can only be realised by signalling these statistics as side information to the receiver, which requires additional bandwidth. Recently, the residual source redundancy was also estimated at the receiver, as presented in [148].

  In the work of Kliewer *et. al* [149–151], an explicit symbol-based redundancy was inserted into the source signal for error protection by applying structured overcomplete signal expansions prior to quantization. This approach is in contrast to the conventional one, which employs an error protection scheme after the quantization scheme. The resultant frame expansions can be interpreted as real-valued block codes.

  The explicit redundancy introduced in the continuous-amplitude domain is exploited by the decoder in two stages. First, the symbol transition probabilities that are present due to the overcomplete signal representation are exploited for computing the *A Posteriori* Probabilities (APPs) of the different symbols. In a second stage, the structure of the redundancy introduced is exploited by using the reliability information computed in the previous decoding stage. In the final stage of the decoding process, the waveform is reconstructed with the aid of the pseudo-inverse of the overcomplete block transform.

  However, in our proposal, we do not insert redundancy, instead the existing redundancy observed in terms of source correlation may be exploited as error protection in the spirit of [149–151]. Additionally, the residual source redundancy can be estimated at the receiver in the spirit of [148].

Appendix A

Over-Complete Source Mapping of Various Rates

In this appendix, the over-complete source mapping of rate 1/2, 2/3, 3/4 and 8/9 used in Chapter 4 are presented.

A.1 Over-Complete Source Mapping of Rate 1/2

Table A.1 lists the over-complete source symbol mapping scheme of rate 1/2 using a 3/6 mapping scheme, where the index i in the first column of Table A.1 represents the quantization level of each 3-bit AMR-WB encoded source symbol. More specifically, each 3-bit AMR-WB encoded source symbol $\tilde{\mathbf{v}}_{\kappa,\tau}(i)$ is mapped to the bit sequence, $\mathbf{u}_{\kappa,\tau}$, where $\tilde{\mathbf{v}}_{\kappa,\tau}(i) = [\tilde{v}(1)_{\kappa,\tau}\ \tilde{v}(2)_{\kappa,\tau}\ \ldots\ \tilde{v}(M)_{\kappa,\tau}]$, $M = 3$ is the total number of bits assigned to the κth parameter and the index i is the quantization level of each 3-bit AMR-WB encoded source symbol, whilst, $\mathbf{u}_{\kappa,\tau} = [u(1)_{\kappa,\tau}\ u(2)_{\kappa,\tau}\ \ldots\ u(N)_{\kappa,\tau}]$ and $N = 6$ is the corresponding over-complete source-mapped symbol.

Example A.1 :
For $\tilde{\mathbf{v}}_{\kappa,\tau}(7)$ the corresponding over-complete source-mapped symbol is $\mathbf{u}_{\kappa,\tau}$=52. More explicitly, i=7 $\rightarrow$ $\mathbf{u}_{\kappa,\tau}$=52.

Index i	Over-Complete Source Mapping Symbol $\mathbf{u}_{\kappa,\tau}$							
0-7	0	7	25	30	42	45	51	52

Table A.1: Over-Complete Source Mapping of Rate 1/2. The 6-bit over-complete source-mapped symbol is represented decimally.

A.2 Over-Complete Source Mapping of Rate 2/3

Table A.2 lists the over-complete source symbol mapping scheme of rate 2/3 using a 8/12 mapping scheme, where the index i in the first column of Table A.2 represents the quantization level of each 8-bit AMR-WB encoded source symbol. More specifically, each 8-bit AMR-WB encoded source symbol $\tilde{\mathbf{v}}_{\kappa,\tau}(i)$ is mapped to the bit sequence, $\mathbf{u}_{\kappa,\tau}$, where $\tilde{\mathbf{v}}_{\kappa,\tau}(i) = [\tilde{v}(1)_{\kappa,\tau}\ \tilde{v}(2)_{\kappa,\tau}\ \ldots\ \tilde{v}(M)_{\kappa,\tau}]$, $M = 8$ is the total number of bits assigned to the κth parameter and the index i is the quantization level of each 8-bit AMR-WB encoded source symbol, whilst, $\mathbf{u}_{\kappa,\tau} = [u(1)_{\kappa,\tau}\ u(2)_{\kappa,\tau}\ \ldots\ u(N)_{\kappa,\tau}]$ and $N = 12$ is the corresponding over-complete source-mapped symbol.

Example A.2 :

For $\tilde{\mathbf{v}}_{\kappa,\tau}(15)$ *the corresponding over-complete source-mapped symbol is* $\mathbf{u}_{\kappa,\tau}$*=127. More explicitly,* i*=15* $\rightarrow$ $\mathbf{u}_{\kappa,\tau}$*=127.*

Index i	Over-Complete Source Mapping Symbol $\mathbf{u}_{\kappa,\tau}$							
0-7	0	7	25	30	42	45	51	52
8-15	75	76	82	85	97	102	120	127
16-23	385	390	408	415	427	428	434	437
24-31	458	461	467	468	480	487	505	510
32-39	642	645	667	668	680	687	689	694
40-47	713	718	720	727	739	740	762	765
48-55	771	772	794	797	809	814	816	823
56-63	840	847	849	854	866	869	891	892
64-71	1155	1156	1178	1181	1193	1198	1200	1207
72-79	1224	1231	1233	1238	1250	1253	1275	1276
80-87	1282	1285	1307	1308	1320	1327	1329	1334
88-95	1353	1358	1360	1367	1379	1380	1402	1405
96-103	1537	1542	1560	1567	1579	1580	1586	1589
104-111	1610	1613	1619	1620	1632	1639	1657	1662
112-119	1920	1927	1945	1950	1962	1965	1971	1972
120-127	1995	1996	2002	2005	2017	2022	2040	2047
128-135	2184	2191	2193	2198	2210	2213	2235	2236
136-143	2243	2244	2266	2269	2281	2286	2288	2295
144-151	2313	2318	2320	2327	2339	2340	2362	2365
152-159	2370	2373	2395	2396	2408	2415	2417	2422
160-167	2570	2573	2579	2580	2592	2599	2617	2622
168-175	2625	2630	2648	2655	2667	2668	2674	2677
176-183	2955	2956	2962	2965	2977	2982	3000	3007
184-191	3008	3015	3033	3038	3050	3053	3059	3060
192-199	3083	3084	3090	3093	3105	3110	3128	3135
200-207	3136	3143	3161	3166	3178	3181	3187	3188
208-215	3466	3469	3475	3476	3488	3495	3513	3518
216-223	3521	3526	3544	3551	3563	3564	3570	3573
224-231	3721	3726	3728	3735	3747	3748	3770	3773
232-239	3778	3781	3803	3804	3816	3823	3825	3830
240-247	3848	3855	3857	3862	3874	3877	3899	3900
248-255	3907	3908	3930	3933	3945	3950	3952	3959

Table A.2: Over-Complete Source Mapping of rate 2/3. The 12-bit over-complete source-mapped symbol is represented decimally.

A.3 Over-Complete Source Mapping of Rate 3/4

Table A.3 lists the over-complete source symbol mapping scheme of rate 3/4 using a 3/4 mapping scheme, where the index i in the first column of Table A.3 represents the quantization level of each 3-bit AMR-WB encoded source symbol. More specifically, each 3-bit AMR-WB encoded source symbol $\tilde{\mathbf{v}}_{\kappa,\tau}(i)$ is mapped to the bit sequence, $\mathbf{u}_{\kappa,\tau}$, where $\tilde{\mathbf{v}}_{\kappa,\tau}(i) = [\tilde{v}(1)_{\kappa,\tau}\ \tilde{v}(2)_{\kappa,\tau}\ \ldots\ \tilde{v}(M)_{\kappa,\tau}]$, $M = 3$ is the total number of bits assigned to the κth parameter and the index i is the quantization level of each 3-bit AMR-WB encoded source symbol, whilst, $\mathbf{u}_{\kappa,\tau} = [u(1)_{\kappa,\tau}\ u(2)_{\kappa,\tau}\ \ldots\ u(N)_{\kappa,\tau}]$ and $N = 4$ is the corresponding over-complete source-mapped symbol.

Example A.3 :

For $\tilde{\mathbf{v}}_{\kappa,\tau}(7)$ the corresponding over-complete source-mapped symbol is $\mathbf{u}_{\kappa,\tau}$=12. More explicitly, i=4 $\rightarrow$ $\mathbf{u}_{\kappa,\tau}$=12.

Index i	Over-Complete Source Mapping Symbol $\mathbf{u}_{\kappa,\tau}$							
0-7	0	9	10	3	12	5	6	12

Table A.3: Over-Complete Source Mapping of Rate 3/4. The 4-bit over-complete source-mapped symbol is represented decimally.

A.4 Over-Complete Source Mapping of Rate 8/9

Table A.4 lists the over-complete source symbol mapping scheme of rate 8/9 using a 8/9 mapping scheme, where the index i in the first column of Table A.4 represents the quantization level of each 8-bit AMR-WB encoded source symbol. More specifically, each 8-bit AMR-WB encoded source symbol $\tilde{\mathbf{v}}_{\kappa,\tau}(i)$ is mapped to the bit sequence, $\mathbf{u}_{\kappa,\tau}$, where $\tilde{\mathbf{v}}_{\kappa,\tau}(i) = [\tilde{v}(1)_{\kappa,\tau}\ \tilde{v}(2)_{\kappa,\tau}\ \ldots\ \tilde{v}(M)_{\kappa,\tau}]$, $M = 8$ is the total number of bits assigned to the κth parameter and the index i is the quantization level of each 8-bit AMR-WB encoded source symbol, whilst, $\mathbf{u}_{\kappa,\tau} = [u(1)_{\kappa,\tau}\ u(2)_{\kappa,\tau}\ \ldots\ u(N)_{\kappa,\tau}]$ and $N = 9$ is the corresponding over-complete source-mapped symbol.

Example A.4 :

For $\tilde{\mathbf{v}}_{\kappa,\tau}(7)$ the corresponding over-complete source-mapped symbol is $\mathbf{u}_{\kappa,\tau}$=15. More explicitly, i=7 $\rightarrow$ $\mathbf{u}_{\kappa,\tau}$=15.

Index i	Over-Complete Source Mapping Symbol $\mathbf{u}_{\kappa,\tau}$							
0-7	0	3	5	6	9	10	12	15
8-15	17	18	20	23	24	27	29	30
16-23	33	34	36	39	40	43	45	46
24-31	48	51	53	54	57	58	60	63
32-39	65	66	68	71	72	75	77	78
40-47	80	83	85	86	89	90	92	95
48-55	96	99	101	102	105	106	108	111
56-63	113	114	116	119	120	123	125	126
64-71	129	130	132	135	136	139	141	142
72-79	144	147	149	150	153	154	156	159
80-87	160	163	165	166	169	170	172	175
88-95	177	178	180	183	184	187	189	190
96-103	192	195	197	198	201	202	204	207
104-111	209	210	212	215	216	219	221	222
112-119	225	226	228	231	232	235	237	238
120-127	240	243	245	246	249	250	252	255
128-135	257	258	260	263	264	267	269	270
136-143	272	275	277	278	281	282	284	287
144-151	288	291	293	294	297	298	300	303
152-159	305	306	308	311	312	315	317	318
160-167	320	323	325	326	329	330	332	335
168-175	337	338	340	343	344	347	349	350
176-183	353	354	356	359	360	363	365	366
184-191	368	371	373	374	377	378	380	383
192-199	384	387	389	390	393	394	396	399
200-207	401	402	404	407	408	411	413	414
208-215	417	418	420	423	424	427	429	430
216-223	432	435	437	438	441	442	444	447
224-231	449	450	452	455	456	459	461	462
232-239	464	467	469	470	473	474	476	479
240-247	480	483	485	486	489	490	492	495
248-255	497	498	500	503	504	507	509	510

Table A.4: Over-Complete Source Mapping of rate 8/9. The 9-bit over-complete source-mapped symbol is represented decimally.

List of Symbols

General notation

- The superscript ^{T} is used to indicate matrix transpose operation.

Special symbols

B: The number of binary bits corresponding to Sphere Packing constellation symbol.

C_i: The i-th subcode of a family of subcodes of IRCCs.

E_b: Bit energy.

f_D: The normalised Doppler frequency.

G: The feedforward generator polynomial of recursive systematic convolutional (RSC) codes.

G_r: The feedback generator polynomial of recursive systematic convolutional (RSC) codes.

I_A: The mutual information associated with the *a priori* information.

I_E: The mutual information associated with the *extrinsic* information.

I_{system}: The number of system iterations.

L: The code memory.

k: The number of input bits of convolutional codes.

K: The total number of source encoded parameters of 20 ms speech frame.

K_{ldpc}: LDPC output block length.

L_{SP}: The size of the legitimate Sphere Packing modulation constellation.

L_a: The *a priori* LLR values.

L_p: The *a posteriori* LLR values.

L_e: The *extrinsic* LLR values.

L_i: The *intrinsic* LLR values.

m: The number of information bits in a TCM coded symbol.

M: The total number of bits assigned to the κth source encoded parameters.

M_{TCM}: The number of levels of a QAM modulation scheme.

M_{SegSNR}: The number of 15-25 ms segments

n: The number of output bits of convolutional codes.

N: The number of binary bits corresponding to Over-Complete source-Mapping symbol.

N_t: Number of transmit antennas.

N_r: Number of receive antennas.

N_0: Single-sided power spectral density of white noise.

N_{SegSNR}: The number of speech samples within a segment of typically 15-25 ms.

P: Number of subcodes in a family of subcodes (e.g. IRCCs).

P_a: The *a priori* logarithmic-domain symbol probabilities.

P_p: The *a posteriori* logarithmic-domain symbol probabilities.

P_e: The *extrinsic* logarithmic-domain symbol probabilities.

P_i: The *intrinsic* logarithmic-domain symbol probabilities.

r_i: Coding rate of the i-th subcode.

R: Coding rate.

$\mathcal{R}_D$: Residual unequal-probability-related redundancy of each source encoded parameter.

$\mathcal{R}_M$: The mutual information $\mathcal{R}_M$ between the pairs of correlated source encoded parameters.

S: The total number of decoding trellis states per iteration.

α_i: Weight coefficient of the i-th subcode.

α_τ: Forward recursive value of the current encoded frame index of the soft-bit assisted AMR-WB decoder.

π: Interleaver.

π^{-1}: Deinterleaver.

Ψ: LLR-to-symbol probability conversion.

Ψ^{-1}: Symbol probability-to-LLR conversion.

Ω: Parallel-to-serial converter.

Ω^{-1}: Serial-to-parallel converter.

Glossary

16QAM	16-level Quadrature Amplitude Modulation
2D	Two-Dimensional
3D	Three-Dimensional
3G	Third Generation
3GPP	Third Generation Partnership Project
8-PSK	8-level Phase Shift Keying
AAC	Advanced Audio Coding
ACELP	Algebraic Code Excited Linear Prediction
AMR	Adaptive MultiRate
AMR-WB	Adaptive MultiRate WideBand
AMR-WB+	Adaptive MultiRate WideBand Plus
APP	A Posteriori Probability
AWGN	Additive White Gaussian Noise
BCH	Bose-Chaudhuri-Hocquenghem. A class of forward error correcting codes (FEC)
BEC	Binary Erasure Channel
BER	Bit Error Ratio, the number of the bits received incorrectly

BICM	Bit Interleaved Coded Modulation
BPS	Bits Per Symbol
BPSK	Binary Phase Shift Keying
BSAC	Bit Sliced Arithmetic Coding
CDMA	Code Division Multiple Access
CELP	Code Excited Linear Prediction
CIR	Channel Impulse Response
DAB	Digital Audio Broadcasting
DECT	Digital Enhanced Cordless Telecommunications
DSTS	Differential Space-Time Spreading
DTX	Discontinuous Transmission
EBU	European Broadcasting Union
EEP	Equal Error Protection
EXIT	EXtrinsic Information Transfer
FD	Frequency-Domain
FEC	Forward Error Correction
GA	General Audio, refering to the MPEG-4 codec which encode any type of audio material
GSM	Global System of Mobile Communications
HCR	Huffman Codeword Reordering
HILN	Harmonic and Individual Lines plus Noise
HVXC	Harmonic Vector eXcitation Coding
IP	Internet Protocol
IRCC	Irregular Convolutional Codes

IrVLC	Irregular Variable-Length Code
ISCD	Iterative Source and Channel Decoding
ISF	Immittance Spectral Frequencies
ISI	Inter-Symbol Interference
ISP	Immittance Spectral Pairs
ISPP	Interleaved Single-Pulse Permutation
ITU	International Telecommunications Union
JSCC	Joint Source and Channel Coding
KBD	Kaiser-Bessel Derived
LDPC	Low-Density Parity-Check
LLR	Log Likelihood Ratio
LP	Linear Predictive
LPC	Linear Predictive Coding
LSF	Line Spectral Frequency
LSP	Line Spectral Pair
LTP	Long Term Prediction
M/S	Mid/Side
MA	Moving Average
MAP	Maximum A Posteriori
MDCT	Modified Discrete Cosine Transform
MELP	Mixed-Excitation Linear Predictive
MI	Mutual Information
MIMO	Multi-Input Multi-Output
ML	Maximum Likelihood

MMS	Multimedia Messaging Service
MMSE	Minimum Mean Square Error
MOS	Mean Opinion Score
MPEG	Motion Picture Expert Group
NSC	Non-Systematic Convolutional
OCM	Over-Complete source-Mapping
OFDM	Orthogonal Frequency Division Multiplexing
PDF	Probability Density Function @entrMI Mutual Information
PNS	Perceptual Noise Substitution
PQF	Polyphase Quadrature Filter
PSS	Packet-Switched Streaming
RCPC	Rate-Compatible Punctured Convolutional
RSC	Recursive Systematic Convolutional
RVLC	Reversible Variable Length Codes
S-MSVQ	Split-Multistage Vector Quantization
SA	Structured Audio
SBSD	Soft-Bit Source Decoding
SegSNR	Segmental SNR
SISO	Soft-Input-Soft-Output
SNR	Signal to Noise Ratio, noise energy compared to the signal energy
SP	Sphere-Packing
SQAM	Sound Quality Assessment Material
STBC	Space-Time Block Coding
STC	Space-Time Coding

STP	Short-Term Prediction
STS	Space-Time Spreading
STTC	Space-Time Trellis Coding
T/F	Time/Frequency
TCM	Trellis Coded Modulation
TNS	Temporal Noise Shaping
TTSI	Text-To-Speech Interface
TwinVQ	Transform-Domain Weighted Interleaved Vector Quantization
UEP	Unequal Error Protection
URC	Unity Rate Code
VAD	Voice Activity Detector
VCB11	Virtual CodeBook
VLC	Variable-Length Code
VLEC	Variable-Length Error-Correcting

Bibliography

[1] L. Hanzo, P. J. Cherriman and J. Streit, *Video Compression and Communications: H.261, H.263, H.264, MPEG4 and Proprietary Codecs*. New York: John Wiley-Sons Inc., 2007.

[2] H. Fletcher, "Auditory Patterns," *Reviews of Modern Physics*, vol. 12, pp. 47–65, January 1940.

[3] "Information Technology-Very Low Bitrate Audio-Visual Coding - Part 3: Audio, IS14496-3," in *ISO/IEC JTC1/SC29/WG11 N2503*, 1998.

[4] M. R. Schroeder and B. S. Atal, "Code-Excited Linear Prediction (CELP): High-Quality Speech at Very Low Bit Rates," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 10, pp. 937–940, April 1985.

[5] M. R. Schroeder and B. S. Atal, "Fast CELP Coding Based on Algebraic Codes," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 12, pp. 1957 – 1960, April 1987.

[6] L. Hanzo, F. C. A. Somerville and J. P. Woodard, *Voice and Audio Compression for Wireless Communications, 2nd Edition*. Chichester, UK: John Wiley-Sons Inc., 2007.

[7] S. Bruhn, E. Ekudden and K. Hellwig, "Adaptive Multi-Rate: A New Speech Service for GSM and Beyond," in *Proceedings of International ITG Conference on Source and Channel Coding (SCC)*, (Munich, Germany), pp. 319–324, January 2000.

[8] "AMR Wideband Speech Codec: Transcoding Functions," in *3GPP TS 26.090 V6 Technical Specification*, December 2004.

[9] R. Steele and L. Hanzo, *Mobile Radio Communications: Second and Third-generation Cellular and WATM Systems* . New York, USA: John Wiley and Sons, 1999.

[10] R. Salami, B. Bessette, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola and K. Jarvinen, "The Adaptive Multi-rate Wideband Codec: History and Performance," *IEEE Workshop Proceedings on Speech Coding*, pp. 144–146, October 2002.

[11] "AMR Wideband Speech Codec: Transcoding Functions," in *3GPP TS 26.190 V5*, December 2001.

[12] "Wideband Coding of Speech at Around 16 kbit/s Using Adaptive Multi-Rate Wideband (AMR-WB)," in *ITU-T Recommendation G.722.2*, July 2003.

[13] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423,623–656, July, October 1948.

[14] N. Farvardin and V. Vaishampayan, "Optimal Quantizer Design for Noisy Channels: An Approach to Combined Source - Channel Coding," *IEEE Transactions on Information Theory*, vol. 33, pp. 827–838, November 1987.

[15] N. Farvardin, "A Study of Vector Quantization for Noisy Channels," *IEEE Transactions on Information Theory*, vol. 36, pp. 799–809, July 1990.

[16] H. Jafarkhani and N. Farvardin, "Design of Channel Optimized Vector Quantizers in the Presence of Channel Mismatch," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 6, pp. 3465–3468, 12-15 May 1998.

[17] J. Hagenaruer, "Source-Controlled Channel Decoding," *IEEE Transactions on Communications*, vol. 43, pp. 2449–2457, September 1995.

[18] J. Hagenauer, N. Seshadri and C. -E. W. Sundberg, "The Performance of Rate-Compatible Punctured Convolutional Codes for Digital Mobile Radio," *IEEE Transactions on Communications*, vol. 38, pp. 966–980, July 1990.

[19] A. Nazer and F. Alajaji, "Unequal Error Protection and Source-Channel Decoding of CELP speech," *Electronics Letters*, vol. 38, pp. 347–349, 28 March 2002.

[20] C. W. Yap and K. N. Ngan, "Unequal Error Protection of Images Over Rayleigh Fading Channels," *IEEE Fifth International Symposium on Signal Processing and Its Applications*, vol. 1, pp. 19–21, 22-25 August 1999.

[21] C. Lamy-Bergot, N. Chautru and C. Bergeron, "Unequal Error Protection for H.263+ Bitstreams Over a Wireless IP Network," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. 377–380, 14-19 May 2006.

[22] R.V. Cox, J. Hagenauer, N. Seshadri, and C. -E. W. Sundberg, "Subband Speech Coding and Matched Convolutional Coding for Mobile Radio Channels," *IEEE Transaction on Signal Processing*, vol. 39, pp. 1717–1731, August 1991.

[23] J. Hagenauer, "Rate-Compatible Punctured Convolutional Codes (RCPC codes) and Their Applications," *IEEE Transactions on Communications*, vol. 36, pp. 389–400, April 1988.

[24] V. B. Balakirsky, "Joint Source-Channel Coding with Variable Length Codes," *IEEE International Symposium on Information Theory*, p. 419, 29 June-4 July 1997.

[25] R. Bauer and J. Hagenauer, "On Variable Length Codes for Iterative Source/Channel Decoding," *IEEE Data Compression Conference*, pp. 273–282, 27-29 March 2001.

[26] J. Kliewer and R. Thobaben, "Combining FEC and Optimal Soft-Input Source Decoding for the Reliable Transmission of Correlated Variable-Length Encoded Signals," *IEEE Data Compression Conference*, pp. 83–91, 2-4 April 2002.

[27] J. Kliewer and R. Thobaben, "Robust Decoding of Variable-Length Encoded Markov Sources Using a Three-Dimensional Trellis," *IEEE Communications Letters*, vol. 7, pp. 320–322, July 2003.

[28] J. Liu, G. Tu, C. Zhang and Y. Yang, "Joint Source and Channel Decoding for Variable Length Encoded Turbo Codes," *EURASIP Journal on Advances Signal Processing*, Article ID 149839, January 2008.

[29] B. Masnick and J. Wolf, "On Linear Unequal Error Protection Codes," *IEEE Transactions on Information Theory*, vol. 13, pp. 600–607, October 1967.

[30] I. Boyarinov and G. Katsman, "Linear Unequal Error Protection Codes," *IEEE Transactions on Information Theory*, vol. 27, pp. 168–175, March 1981.

[31] N. Rahnavard and F. Fekri, "Unequal Error Protection Using Low-Density Parity-Check Codes," *IEEE International Symposium on Information Theory*, p. 449, 27 June-2 July 2004.

[32] N. Rahnavard and F. Fekri, "Unequal Error Protection Using Partially Regular LDPC Codes," *IEEE Transactions on Communications*, vol. 55, pp. 387–391, March 2007.

[33] B. Vasic, A. Cvetkovic, S. Sankaranarayanan and M. Marcellin, "Adaptive Error Protection Low-Density Parity-Check Codes For Joint Source-Channel Coding Schemes," *IEEE International Symposium on Information Theory*, p. 267, 29 June-4 July 2003.

[34] C. Poulliat, D. Declercq and I. Fijalkow, "Enhancement of Unequal Error Protection Properties of LDPC Codes," *EURASIP Journal on Wireless Communications and Networking*, Article ID 92659, October 2007.

[35] H. Pishro-Nik, N. Rahnavard and F. Fekri, "Nonuniform Error Correction Using Low-Density Parity-Check Codes," *IEEE Transactions on Information Theory*, vol. 51, pp. 2702–2714, July 2005.

[36] H. Imai and S. Hirakawa, "A New Multilevel Coding Method Using Error-Correcting Codes," *IEEE Transactions on Information Theory*, vol. 23, pp. 371–377, May 1977.

[37] R. Morelos-Zaragoza, T. Kasami and S. Lin, "Multilevel Block Coded 8-PSK Modulations Using Unequal Error Protection Codes for the Rayleigh Fading Channel," *IEEE Sixth International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 2, pp. 486–490, 27-29 September 1995.

[38] R. H. Morelos-Zaragoza, M. P. C. Fossorier, S. Lin and H. Imai, "Multilevel Block Coded Modulation with Unequal Error Protection," *IEEE International Symposium on Information Theory*, p. 441, 29 June-4 July 1997.

[39] N. Seshadri and C. -E. W. Sundberg, "Multilevel Trellis Coded Modulations for the Rayleigh Fading Channel," *IEEE Transactions on Communications*, vol. 41, pp. 1300–1310, September 1993.

[40] D. -F. Yuan, Z. -W. Li and F. Zhang, "A New Unequal Error Protection Scheme for Mobile Hierarchical Image Transmission System Using Multilevel Codes," *IEEE 53rd Vehicular Technology Conference*, pp. 2036 – 2040, 6-9 May 2001.

[41] M. Aydinlik and M. Salehi, "Turbo Coded Modulation for Unequal Error Protection," *IEEE Transactions on Communications*, vol. 56, pp. 555 – 564, April 2008.

[42] L. Zhou, G. F. Bian and P. Gu, "Investigation of Multilevel Coded Modulation in the Application of the Adaptive OFDM System," *IEEE International Conference on Microwave and Millimeter Wave Technology*, vol. 4, pp. 1938 – 1941, 21-24 April 2008.

[43] T. Brüggen and P. Vary, "Unequal Error Protection by Modulation with Unequal Power Allocation," *IEEE Communications Letters*, vol. 9, pp. 484–486, June 2005.

[44] N. Gortz and E. Bresch, "Source-Adaptive Power Allocation for Digital Modulation," *IEEE Communications Letters*, vol. 7, pp. 569–571, December 2003.

[45] K. -P. Ho, "Unequal Error Protection Based on OFDM and Its Application in Digital Audio Transmission," *IEEE Global Telecommunications Conference*, vol. 3, pp. 1320–1325, 8-12 December 1998.

[46] C. Berrou, A. T. Glavieux and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo-Codes," *IEEE International Conference on Communications*, vol. 2, pp. 1064–1070, 23-26 May 1993.

[47] C. Berrou and A. T. Glavieux, "Near Optimum Error Correcting Coding and Decoding: Turbo-Codes," *IEEE Transactions on Communications*, vol. 44, pp. 1261–1271, October 1996.

[48] M. Adrat, P. Vary, and J. Spittka, "Iterative Source-Channel Decoder Using Extrinsic Information from Softbit-Source Decoding," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, pp. 2653–2656, 7-11 May 2001.

[49] F. Alajaji, N. Phamdo and T. E. Fuja, "Channel Codes that Exploit the Residual Redundancy in CELP-Encoded Speech," *IEEE Transactions on Speech and Audio Processing*, vol. 4, pp. 325–336, September 1996.

[50] T. Fazel and T. E. Fuja, "Robust Transmission of MELP-Compressed Speech: An Illustrative Example of Joint Source-Channel Decoding," *IEEE Transactions on Communications*, vol. 51, pp. 973–982, June 2003.

[51] "Federal Standard 1016 CELP," in *National Communications System (NCS) TIB 921*, January 1992.

[52] A. McCree, T. Kwan, E. B. George, T. P. Barnwell and V. Viswanathan, "A 2.4 kbit/s MELP Coder Candidate for the New U.S. Federal Standard," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. 200 – 203, 7-10 May 1996.

[53] L. R. Bahl, J. Cocke, F. Jelinek and J. Raviv, "Optimal decoding of linear codes for minimal symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, pp. 284–287, March 1974.

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

[55] A. Hedayat and A. Nosratinia, "Concatenated Error-Correcting Entropy Codes and Channel Codes," *IEEE International Conference on Communications*, vol. 5, pp. 3090–3094, 11-15 May 2003.

[56] J. Wang, L. -L. Yang and L. Hanzo, "Iterative Channel Equalization, Channel Decoding and Source Decoding," *IEEE 61st Vehicular Technology Conference*, vol. 1, pp. 518–522, 30 May-1 June 2005.

[57] S. X. Ng, F. Guo, J. Wang, L. -L Yang and L. Hanzo, "Joint Source-Coding, Channel-Coding and Modulation Schemes for AWGN and Rayleigh Fading Channels," *Electronics Letters*, vol. 39, pp. 1259–1261, August 2003.

[58] Y. Takishima, M. Wada and H. Murakami, "Reversible Variable Length Codes," *IEEE Transactions on Communications*, vol. 43, pp. 158 – 162, Feb-Mar-Apr 1995.

[59] T. Fingscheidt and P. Vary, "Speech Decoding with Error Concealment Using Residual Source Redundancy," *IEEE Workshop on Speech Coding for Telecommunication*, pp. 91–92, 7-10 Sept 1997.

[60] T. Fingscheidt and P. Vary, "Softbit Speech Decoding: A New Approach to Error Concealment," *IEEE Transactions on Speech and Audio Processing*, vol. 9, pp. 240–251, March 2001.

[61] M. Adrat, U. von Agris and P. Vary, "Convergence Behavior of Iterative Source-Channel Decoding," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 269–272, 6-10 April 2003.

[62] M. Adrat and P. Vary, "Iterative Source-Channel Decoding: Improved System Design Using EXIT Charts," *EURASIP Journal on Applied Signal Processing*, pp. 1727–1737, October 2005.

[63] T. Clevorn, M. Adrat and P. Vary, "Turbo Decodulation Using Highly Redundant Index Assignments and Multi-Dimensional Mappings," *4th Int. Symposium on Turbo Codes and Related Topics in connection with 6th Int. ITG-Conference on Source and Channel Coding*, April 2006.

[64] T. Clevorn, P. Vary and M. Adrat, "Iterative Source-Channel Decoding Using Short Block Codes," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, pp. 221–224, 14-19 May 2006.

[65] S. ten Brink, "Convergence Behaviour of Iteratively Decoded Parallel Concatenated Codes," *IEEE Transactions on Communications*, vol. 49, pp. 1727–1737, October 2001.

[66] F. Brännström, L. K. Rasmussen and A. J. Grant, "Convergence Analysis and Optimal Scheduling for Multiple Concatenated Codes," *IEEE Transactions on Information Theory*, vol. 51, pp. 3354–3364, September 2005.

[67] M. Tüchler, "Convergence Prediction for Iterative Decoding of Threefold Concatenated Systems," *IEEE Global Telecommunications Conference*, vol. 2, pp. 1358–1362, 17-21 November 2002.

[68] R. G. Maunder, J. Wang, S. X. Ng, L. -L. Yang and L. Hanzo, "Iteratively Decoded Irregular Variable Length Coding and Trellis Coded Modulation," *IEEE Workshop on Signal Processing Systems*, pp. 222–227, 17-19 October 2007.

[69] R. Perkert, M. Kaindl and T. Hindelang, "Iterative Source and Channel Decoding for GSM," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, pp. 2649 – 2652, 7-11 May 2001.

[70] 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.

[71] S. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1451–1458, October 1998.

[72] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-Time Block Codes from Orthogonal Designs," *IEEE Transactions on Information Theory*, vol. 45, pp. 1456–1467, July 1999.

[73] B. Hochwald, T.L. Marzetta and C.B. Papadias, "A Transmitter Diversity Scheme for Wideband CDMA Systems Based on Space-Time Spreading," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 48–60, September 2001.

[74] W. Su, Z. Safar and K. J. R. Liu, "Space-Time Signal Design for Time-Correlated Rayleigh Fading Channels," *IEEE International Conference on Communications*, vol. 5, pp. 3175–3179, May 2003.

[75] V. Tarokh and H. Jafarkhani, "A Differential Detection Scheme for Transmit Diversity," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1169 – 1174, July 2000.

[76] B. Hochwald and W. Sweldens, "Differential Unitary Space-Time Modulation," *IEEE Transactions on Communications*, vol. 48, pp. 2041 – 2052, December 2000.

[77] B. Hochwald and T.L. Marzetta, "Unitary Space-Time Modulation for Multiple-Antenna Communications in Rayleigh Flat Fading," *IEEE Transactions on Information Theory*, vol. 46, pp. 543–564, March 2000.

[78] A. Q. Pham, L. Hanzo and L. -L. Yang, "Joint Optimization of Iterative Source and Channel Decoding Using Over-Complete Source-Mapping," in *IEEE 66th Vehicular Technology Conference*, pp. 1072 – 1076, 30 Sept.-3 Oct. 2007.

[79] N. S. Othman, S. X. Ng and L. Hanzo, "Turbo-Detected Unequal Protection Audio and Speech Transceivers Using Serially Concantenated Convolutional Codes, trellis coded modulation and space-time trellis coding," vol. 2, pp. 1044–1048, 25-28 September 2005.

[80] N. S. Othman, S. X. Ng and L. Hanzo, "Turbo-Detected Unequal Protection MPEG-4 Audio Transceiver Using Convolutional Codes, Trellis Coded Modulation and Space-Time Trellis Coding," vol. 3, pp. 1600 – 1604, 30 May-1 June 2005.

[81] J. Wang, N. S. Othman, J. Kliewer, L. -L. Yang and L. Hanzo, "Turbo-Detected Unequal Error Protection Irregular Convolutional Codes Designed for the Wideband Advanced Multirate Speech Codec," vol. 2, pp. 927 – 931, 25-28 September 2005.

[82] M. El-Hajjar, O. Alamri, S. X. Ng and L. Hanzo, "Turbo Detection of Precoded Sphere Packing Modulation Using Four Transmit Antennas for Differential Space-Time Spreading," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 943–952, March 2008.

[83] N. S. Othman, M. El-Hajjar, O. Alamri and L. Hanzo, "Soft-Bit Assisted Iterative AMR-WB Source-Decoding and Turbo-Detection of Channel-Coded Differential Space-Time Spreading Using Sphere Packing Modulation," *IEEE 65th Vehicular Technology Conference*, pp. 2010–2014, 22-25 April 2007.

[84] N. S. Othman, M. El-Hajjar, O. Alamri and L. Hanzo, "Iterative AMR-WB Source and Channel-Decoding Using Differential Space-Time Spreading Assisted Sphere Packing Modulation," *to appear in IEEE Transactions on Vehicular Technology.*

[85] A. Q. Pham, L. -L. Yang, N. S. Othman and L. Hanzo, "EXIT-Chart Optimized Block Codes for Wireless Video Telephony," *to appear in IEEE Transactions on Circuits and Systems for Video Technology.*

[86] N. S. Othman, M. El-Hajjar, A. Q. Pham, O. Alamri, S. X. Ng and L. Hanzo, "Over-Complete Source-Mapping Aided AMR-WB MIMO Transceiver Using Three-Stage Iterative Detection," in *IEEE International Conference on Communications*, pp. 751–755, 19-23 May 2008.

[87] N. S. Othman, M. El-Hajjar, O. Alamri, S. X. Ng and L. Hanzo, "Three-Stage Iterative Detection of Precoded Soft-Bit AMR-WB for Speech MIMO Transceiver," in *To be submitted to IEEE 69th Vehicular Technology Conference*, 2009.

[88] L. Hanzo, J. Blogh and S. Ni, *3G Systems and HSDPA-Style FDD Versus TDD Networking: Smart Antennas and Adaptive Modulation, 2nd Edition.* New York, USA: John Wiley IEEE Press, 2008.

[89] "Multimedia Messaging Service, Media Formats and Codecs," in *3GPP TS 26.140*, December 2006.

[90] "Transparent End-to-end Packet-Switched Streaming Service, Protocols and Codecs," in *3GPP TS 26.234*, March 2006.

[91] P. Hoeher, J. Hagenauer, E. Offer, C. Rapp and H. Schulze, "Performance of an RCPC-Coded OFDM-Based Digital Audio Broadcasting (DAB) System," *IEEE Global Telecommunications Conference*, vol. 1, pp. 40–46, 2-5 December 1991.

[92] "Radio Broadcasting Systems: Digital Audio Broadcasting to Mobile Portable and Fixed Receivers," in *ETSI EN 300 401*, April 2001.

[93] M. Tüchler and J. Hagenauer, "EXIT charts of Irregular Codes," in *Proceedings of Conference on Information Science and Systems [CDROM]*, (Princeton University), 20-22 March 2002.

[94] M. Tüchler, "Design of Serially Concatenated Systems Depending on the Block Length," *IEEE Transactions on Communications*, vol. 52, pp. 209–218, February 2004.

[95] S. Wang, A. Sekey and A. Gersho, "An Objective Measure for Predicting Subjective Quality of Speech Coders," *IEEE Journal on Selected Areas in Communications*, vol. 10, pp. 819–829, June 1992.

[96] W. B. Kleijn and K. K. Paliwal, *Speech Coding and Synthesis*. Elsevier, 1995.

[97] F. Pereira and T. Ebrahimi, *The MPEG-4 Book*. NJ, USA : Prentice Hall, 2002.

[98] H. Purnhagen and N. Meine, "HILN-The MPEG-4 Parametric Audio Coding Tools," *IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 3, pp. 201–204, 28-31 May 2000.

[99] "Information Technology-Very Low Bitrate Audio-Visual Coding - Part 3: Audio, IS14496-3," in *ISO/IEC JTC1/SC29/WG11 N2203*, 1998.

[100] J. V. Tobias, *Foundations of Modern Auditory Theory* . New York, USA: Academic Press, 1970.

[101] J. D. Cutnell and K. W. Johnson, *Physics*. Hoboken, New Jersey: Wiley, 2007.

[102] E. Zwicker, "Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen)," *The Journal of the Acoustical Society of America*, vol. 33, p. 248, February 1961.

[103] T. Painter and A. Spanias, "Perceptual Coding of Digital Audio," *Proceedings of the IEEE*, vol. 88, pp. 451–515, April 2000.

[104] A. R. Reddy, "Design of SAW Bandpass Filters Using Window Functions," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 35, pp. 50–56, January 1988.

[105] P. Noll, "Digital Audio Coding for Visual Communications," *Proceedings of the IEEE*, vol. 83, pp. 925–943, June 1995.

[106] N. S. Jayang and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video Coding.* Englewood Cliffs, NJ: Prentice Hall, 1984.

[107] N. Iwakami, T. Moriya and S. Miki, "High-Quality Audio-Coding at Less Than 64 kbit/s by Using Transform-Domain Weighted Interleave Vector Quantization (TwinVQ)," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 3095–3098, 9-12 May 1995.

[108] K. Ikeda, T. Moriya, N. Iwakami and S. Miki, "Error-Protected TwinVQ Audio Coding at Less Than 64 kbit/s/ch," *IEEE Workshop on Speech Coding for Telecommunications*, pp. 33–34, 20-22 September 1995.

[109] K. Ikeda, T. Moriya, N. Iwakami and S. Miki, "A Design of Twin VQ Audio-Codec for Personal Communication Systems," *IEEE International Conference Universal Personal Communications*, pp. 803–807, 6-10 November 1995.

[110] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola and K. Jarvinen, "The Adaptive Multirate Wideband Speech Codec (AMR-WB)," *IEEE Transactions on Speech and Audio Processing*, vol. 10, pp. 620–636, November 2002.

[111] "Adaptive Multi-Rate Wideband Speech ANSI-C Code," in *3GPP TS 26.173 Technical Specification*, 2003.

[112] Y. Bistritz and S. Pellerm, "Immitance Spectral Pairs (ISP) for Speech Encoding," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, pp. 9–12, 27-30 April 1993.

[113] J. -P. Adoul and C. Lamblin, "A Comparison of Some Algebraic Structures for CELP Coding of Speech," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 12, pp. 1953 – 1956, April 1987.

[114] J. -P. Adoul, P. Mabilleau, M. Delprat and S. Morissette, "Fast CELP Coding Based on Algebraic Codes," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 12, pp. 1957 – 1960, April 1987.

[115] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals.* Englewood Cliffs, N.J: Prentice-Hall, 1978.

[116] K. K. Paliwal and B. S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 bits/frame," *IEEE Transactions on Speech and Audio Processing*, vol. 1, pp. 3–14, January 1993.

[117] F. Soong and B. Juang, "Line Spectrum Pair (LSP) and Speech Data Compression," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 9, pp. 37–40, March 1984.

[118] S. So and K. K. Paliwal, "A Comparative Study of LPC Parameter Representations and Quantisation Schemes for Wideband Speech Coding," *Digital Signal Processing*, vol. 17, pp. 114–137, January 2007.

[119] K. K. Paliwal and B. S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 bits/frame," *IEEE Transactions on Speech and Audio Processing*, vol. 1, pp. 3–14, January 1993.

[120] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston ; London: Kluwer Academic Publishers, 1992.

[121] "Sound Quality Assessment Material," *http://sound.media.mit.edu/mpeg4/audio/sqam/*.

[122] S. X. Ng, J. Y. Chung and L. Hanzo, "Turbo-Detected Unequal Protection MPEG-4 Wireless Video Telephony Using Multi-Level Coding, Trellis Coded Modulation and Space-Time Trellis Coding," in *IEE Communications Proceeding*, vol. 152, pp. 1116–1124, 9 December 2005.

[123] R. Koenen, "MPEG-4 Overview," in *ISO/IEC JTC1/SC29/WG11 N4668, version 21-Jeju Version, ISO/IEC*, (http://www.chiariglione.org/ mpeg/ standards/ mpeg-4/ mpeg-4.htm), March 2002.

[124] R. Koenen, "MPEG-4 Multimedia for Our Time," vol. 36, pp. 26–33, February 1999.

[125] J. Herre and B. Grill, "Overview of MPEG-4 Audio and Its Applications in Mobile Communications," *Proceedings of International Conference on Signal Processing*, vol. 1, pp. 11–20, 21-25 August 2000.

[126] R. Sperschneider, "Error Resilient Source Coding with Variable-Length Codes and Its Application to MPEG Advanced Audio Coding," *109th AES Convention*, September 2000.

[127] S. X. Ng, J. Y. Chung and L. Hanzo, "Turbo-Detected Unequal Protection MPEG-4 Wireless Video Telephony Using Trellis Coded Modulation and Space-Time Trellis Coding," in *IEE International Conference on 3G Mobile Communication Technologies (3G 2004)*, (London, UK), pp. 416–420, 18 - 20 October 2004.

[128] S. Lin and D. J. Costello Jr, *Error Control Coding, Second Edition*. Inc. Englewood Cliffs, New Jersey: Prentice-Hall, 2004.

[129] G. Ungerböck, "Channel Coding with Multilevel/Phase Signals," *IEEE Transactions on Information Theory*, vol. 28, pp. 55–67, January 1982.

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

[131] L. Hanzo, S. X. Ng, T.Keller and W. Webb, *Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised and Space-Time Coded OFDM, CDMA and MC-CDMA Systems*. New York, USA : John Wiley IEEE Press, 2004.

[132] E. Zehavi, "8-PSK Trellis Codes for a Rayleigh Fading Channel," *IEEE Transactions on Communications*, vol. 40, pp. 873–883, May 1992.

[133] S. Benedetto, D. Divsalar, G. Montorsi and F. Pollara, "Serial Concatenation of Interleaved Codes: Performance Analysis, Design, and Iterative Decoding," *IEEE Transactions on Information Theory*, vol. 44, pp. 909–926, May 1998.

[134] A. Ashikhmin, G. Kramer and S. ten Brink, "Extrinsic Information Transfer Functions: Model and Erasure Channel Properties," *IEEE Transactions on Information Theory*, vol. 50, pp. 2657– 2673, November 2004.

[135] I. Land, P. Hoeher and S. Gligorević, "Computation of Symbol-Wise Mutual Information in Transmission Systems with logAPP Decoders and Application to EXIT Charts," in *Proceedings of International ITG Conference on Source and Channel Coding (SCC)*, (Erlangen, Germany), pp. 195–202, January 2004.

[136] S. Dolinar and D. Divsalar, "Weight distributions for turbo codes using random and nonrandom permutations," *JPL-TDA Progress Report 42-122*, pp. 56–65, August 1995.

[137] T. Clevorn, J. Brauers, M. Adrat and P. Vary, "Turbo DeCodulation: Iterative Combined Demodulation and Source-Channel Decoding," *IEEE Communications Letters*, vol. 9, pp. 820–822, September 2005.

[138] L. Hanzo, C. H. Wong and M. S. Yee, *Adaptive Wireless Transceivers: Turbo-Coded, Turbo-Equalized and Space-Time Coded TDMA, CDMA, and OFDM Systems*. New York, USA : John Wiley and Sons, 2002.

[139] F. Itakura, "Line Spectrum Representation of Linear Predictive Coefficients of Speech Signals," *Journal of the Acoustical Society of America*, vol. 57, p. S35, April 1975.

[140] A. M. Kondoz, *Digital Speech Coding for Low Bit Rate Communication Systems*. Chichester, UK: John Wiley and Sons, 2004.

[141] S. ten Brink, J. Speidel and R. -H. Yan, "Iterative Demapping and Decoding for Multilevel Modulation," *IEEE Global Telecommunications Conference*, vol. 1, pp. 579–584, November 1998.

[142] L. Hanzo, L. -L. Yang, E-L. Kuan and K. Yen, *Single and Multi-Carrier DS-CDMA: Multi-User Detection, Space-Time Spreading, Synchronisation, Networking and Standards*. New York, USA : John Wiley and Sons, 2003.

[143] P. Robertson, E. Villebrun and P. Hoeher, "A Comparison of Optimal and Sub-Optimal MAP Decoding Algorithms Operating in the Log Domain," *IEEE International Conference on Communications*, vol. 2, pp. 1009–1013, 18-22 June 1995.

[144] M. Adrat, *Iterative Source-Channel Decoding for Digital Mobile Communications*. RWTH Aachen University: Ph.D Dissertation, 2003.

[145] D. Divsalar, S. Dolinar and F. Pollara, "Serial Concatenated Trellis Coded Modulation with Rate-1 Inner Code," *IEEE Global Telecommunications Conference*, vol. 2, pp. 1168–1175, 27November-1 December 2000.

[146] R. Salami, R. Lefebvre, A. Lakaniemi, K. Kontola, S. Bruhn and A. Taleb, "Extended AMR-WB for High-Quality Audio on Mobile Devices," in *IEEE Communications Magazine*, vol. 44, pp. 90–97, May 2006.

[147] J. Makinen, B. Bessette, S. Bruhn, P. Ojala, R. Salami and A. Taleb, "AMR-WB+: A New Audio Coding Standard for 3rd Generation Mobile Audio Services," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 1109–1112, 18-23 March 2005.

[148] B. Schotsch, P. Vary and T. Clevorn, "Adaptive Exploitation of Residual Redundancy in Iterative Source-Channel Decoding," in *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–5, 3-7 September 2007.

[149] J. Kliewer and A. Mertins, "Error-Resilient Transmission of Waveform Signals Using Overcomplete Expansions and Soft-Input Source Decoding," *Proceeding of 12th European Signal Processing Conference*, pp. 881–884, September 2004.

[150] J. Kliewer and A. Mertins, "Low latency Joint Source-Channel Coding Using Overcomplete Expansions and Residual Source Redundancy," *IEEE Global Telecommunications Conference*, vol. 4, pp. 2390–2394, 28 Nov.-2 Dec. 2005.

[151] J. Kliewer and A. Mertins, "Soft-Input Reconstruction of Binary Transmitted Quantized Overcomplete Expansions," *IEEE Communications Letters*, vol. 7, pp. 320–322, July 2003.

Index

Author Index

C

D

E

F

G

M

N

O

P

R

S