Iterative AMR-WB Source and Channel-Decoding of Differential Space-Time Spreading Assisted Sphere Packing Modulation

N. S. Othman, M. El-Hajjar, O. Alamri and L. Hanzo

School of Electronics and Computer Science, University of Southampton, SO17 1BJ, UK.
Tel: +44-2380-593 125, Fax: +44-2380-593 045
Email: {nso01r,meh05r,ora02r,lh}@ecs.soton.ac.uk, http://www-mobile.ecs.soton.ac.uk

Abstract - In this paper we present a novel system that invokes jointly optimised iterative source- and channel-decoding for enhancing the error resilience of the Adaptive Multi Rate WideBand (AMR-WB) speech codec. 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. Moreover, the convergence behaviour of the proposed scheme is evaluated with the aid of both three-dimensional (3D) and two-dimensional (2D) Extrinsic Information Transfer (EXIT) charts. The proposed system exhibits an $E_b/N_0$ gain of about 1 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 using $I_{\text{system}} = 2$ system iterations and when communicating over correlated narrowband Rayleigh fading channels.

I. MOTIVATION AND BACKGROUND

The classic Shannonian source and channel coding separation theorem [1] has limited applicability in the context of finite-complexity, finite-delay lossy speech [2] and video [3] codecs, where the different encoded bits exhibit different error sensitivity. These arguments are particularly valid, when the limited-complexity, limited-delay source encoders fail to remove all the redundancy.

---

1 The financial support of the Universiti Tenaga Nasional Malaysia, of Vodafone under the auspices of the Dorothy Hodgkin Postgraduate Award, of the Ministry of Higher Education of Saudi Arabia and that of the EPSRC, UK as well as that of the European Community under Seventh Framework Programme grant agreement ICT OPTIMIX nNFSO-ICT-214625 is gratefully acknowledged.
from the correlated speech or video source signal. Fortunately, this residual redundancy may be beneficially exploited for error protection by intelligently exchanging soft information amongst the various receiver components. More explicitly, it was demonstrated in [4] 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 [5, 6], culminating in the formulation of iterative source and channel decoding (ISCD) [7]. 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 [8] the turbo principle [9] was employed for iterative soft demapping in multilevel modulation [10] 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 [11].

Explicitly, we propose and investigate the jointly optimised ISCD scheme of Figure 1 invoking the Adaptive Multirate-Wideband (AMR-WB) speech codec [12], 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 [13] over a temporally correlated narrowband 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 [13] 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 [14] that the employment of SP modulation combined with the orthogonal transmit diversity designs outperformed its conventional counterpart of [15, 16]. We will refer to this three-stage system as the DSTS-SP-RSC-AMRWB scheme.

The convergence behaviour of this iterative process is studied using Extrinsic Information Transfer
(EXIT) charts [17] by visualizing the input/output mutual information exchange of the individual constituents of the SISO decoder. Recently, the 2D EXIT-charts of two-stage concatenated systems have been extended to three-stage turbo receivers [18,19], 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) 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 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. The employment of SP modulation facilitates the joint design of several DSTS time-slots’ signal, which allows the direct minimisation of the DSTS symbol error probability.

3) A three-dimensional (3D) EXIT-chart based procedure and its two-dimensional (2D) EXIT-chart projection technique have been used for designing the optimum combination of receiver components.

The paper is structured as follows. In Section II, the AMR-WB speech codec is described briefly and the residual redundancy inherent in its parameters is quantified. In Section III, the overall system model is described, while the system’s convergence behaviour is analyzed in Section IV with the aid of 3D EXIT charts and their 2D projections. Section V quantifies the performance of our proposed three-stage scheme, while our conclusions are offered in Section VI.
II. RESIDUAL REDUNDANCY IN THE AMR-WB SPEECH CODEC

The AMR-WB speech codec is capable of supporting nine different bit rates [20], each of which may be activated in conjunction with different-rate channel codecs and different-throughput adaptive modem modes [21]. Similar near-instantaneously adaptive speech and video systems were designed in [2, 3]. In our prototype system investigated here the AMR-WB codec operated at 23.05 kbps, generating 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 code excited linear prediction (CELP) based codecs [2], it performs short-term prediction (STP), long-term prediction (LTP) and generates the excitation codebook (CB) parameters [12]. The STP coefficients produced for the 20 ms speech frame are converted to the so-called immittance spectrum pair (ISP) representation [22] and they are vector-quantised using split-multistage vector quantisation (S-MSVQ) producing 7 ISP parameters using 46 bits.

The LTP analysis [2] 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 [20].

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 [20], it leaves some residual redundancy in the encoded parameters.

Firstly, this residual redundancy manifests itself in terms of the unequal probability of occurrence
of the different values of a specific parameter in each 20 ms AMR-WB-encoded frame, which we may refer to as unequal-probability-related redundancy. This unequal probability of occurrence is exploited by our three-stage iterative detector as a priori information concerning certain AMR-WB codec parameters and hence assists the remaining two constituent decoders in their decisions. 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, although the previously decoded frames are also be exploited without any extra delay.

The second manifestation of the residual redundancy is constituted by the similarity of the parameters in the current and the immediately preceeding 20 ms AMR-WB-encoded frames, which may be referred to as inter-frame first-order correlation. This inter-frame first-order correlation may also be exploited as a source of valuable a priori information for some of the AMR-WB codec parameters. A few examples are the CB gain and the ISP parameters. 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. However, as stated before, only the previously received parameters can be used as a source of a priori information, unless the extra delay associated with postponing the processing of the correct 20 ms frame until receiving the next frame may be tolerated. In this paper we opted for avoiding any extra delay. Hence, only the inter-frame redundancy associated with the previously received 20 ms frames, but not that of the “future” frames was exploited, as we will detail in Section III. Additionally, the unequal-probability-related redundancy of the various values of those parameters, which were non-uniformly distributed, such the aforementioned LTP lag and the Fixed CB index, were capitalized on.

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 20 ms frame. 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. Similarly, the probability of occurrence was recorded for each of the AMR-WB encoded parameters which did not exhibit exploitable inter-frame
first-order correlation. The resultant residual redundancy recorded for the inter-frame correlated parameters, such as the ISP and the CB gain parameters was quantified in Table I in terms their mutual information $R_M$ between the corresponding parameters of two consecutive 20 ms-spaced speech frames. By contrast, the residual unequal-probability-related redundancy of each AMR-WB codec parameter was quantified in terms of $R_D = B - H(U)$, where $H(U)$ is the entropy of the quantized parameter $U$ and $B$ is the number of bits actually used for quantizing the parameter $U$.

As an example, we can observe from Table I that the speech-energy-related CB gain parameters have a high inter-frame first-order mutual information. 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 inter-frame first-order mutual 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, while the rest of the parameters would benefit from exploiting the non-negligible unequal-probability-based extrinsic information. Hence again, 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 MAP decoding will be briefly reviewed in Section III. It will be shown in Section V 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.

III. SYSTEM OVERVIEW

Figure 1 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 [13], while the **outer** iterative loop represents the **extrinsic** information exchange between the AMR-WB speech decoder and the RSC decoder.
A. Transmitter

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

The interleaved RSC encoded source bit sequence \( \tilde{u} \) of Figure 1 is generated by a \( \frac{1}{2} \)-rate RSC code having a code memory of \( L = 3 \) and octally represented generator polynomials of \( (G_1, G_2) = (13, 6) \). The DSTS-SP modulator of Figure 1 first maps \( B \) number of channel-coded bits \( \tilde{c} = [\tilde{c}_0, \tilde{c}_1, \ldots, \tilde{c}_{B-1}] \in \{0, 1\} \) to a sphere packing symbol \( x \in X \), using the mapping function \( x = \text{map}_{\text{sp}}(\tilde{c}) \). Furthermore, we have \( B = \log_2 L = \log_2 16 = 4 \), where \( L \) represents the set of legitimate SP constellation points, as outlined in [14]. This set of SP symbols is transmitted using DSTS in conjunction with two transmit antennas, as detailed in [14]. STS was first proposed by Marzetta et. al [23] for the sake of providing space-time-coding-style spatial transmit diversity gains for CDMA systems [24]. 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 [13] 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 considered transmissions over a temporally correlated narrowband 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 1 denote the LLRs of the bit probabilities. The notations \( \tilde{c}, c, \tilde{u} \) and \( u \) in the round brackets \( (.) \) in Figure 1 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 \( L, a, p \) and \( e \), which denote in Figure 1 \textit{a priori}, \textit{a posteriori} and \textit{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 1.
B. Receiver

**Inner Iterations:** The complex-valued received symbols, $\mathbf{z}$ are demapped to their LLR [9] representation for each of the $B$ number of RSC-encoded bits per DSTS-SP symbol. As seen in Figure 1, the *a priori* LLR values $L_{3,a}(\hat{c})$ provided by the RSC decoder are subtracted from the *a posteriori* LLR values $L_{3,p}(\hat{c})$ at the output of the SP-demapper for the sake of generating the extrinsic LLR values $L_{3,e}(\hat{c})$. Then the LLRs $L_{3,e}(\hat{c})$ are deinterleaved by a soft-bit deinterleaver. Next, the deinterleaved soft-bits $L_{2,a}(c)$ of Figure 1 are passed to the RSC decoder in order to compute the *a posteriori* LLR values $L_{2,p}(c)$ provided by the Max-log MAP algorithm [25] for all the RSC-encoded bits. The extrinsic information $L_{2,e}(c)$ seen in Figure 1 is generated by subtracting the *a priori* information $L_{2,a}(c)$ from the *a posteriori* information $L_{2,p}(c)$ according to $L_{2,p}(c) - L_{2,a}(c)$, which is then fed back to the SP demapper as the *a priori* information $L_{3,a}(\hat{c})$ after appropriately reordering them using the inner soft-value interleaver. The SP-demapper of Figure 1 exploits the *a priori* information $L_{3,a}(\hat{c})$ for the sake of providing improved *a posteriori* LLR values $L_{3,p}(\hat{c})$ 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 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 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 [7] 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 $I_{\text{system}} = 1$. The residual redundancy quantified in Section II is exploited as *a priori* information for computing the extrinsic LLR values and for estimating the speech parameters. 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 [7, 26], which are briefly reviewed below. Firstly, the
channel’s output information related to each speech parameter is given by the product of each of the constituent bits as follows:

\[ p(\hat{u}_{\kappa,\tau}|u_{\kappa,\tau}) = \prod_{m=1}^{M} p[\hat{u}_{\kappa,\tau}(m)|u_{\kappa,\tau}(m)], \]  

(1)

where \( \hat{u}_{\kappa,\tau} = [\hat{u}(1)_{\kappa,\tau}, \hat{u}(2)_{\kappa,\tau}, ..., \hat{u}(M)_{\kappa,\tau} ] \) represents the received bit sequence of the \( \kappa \)th parameter, while \( u_{\kappa,\tau} \) is the corresponding transmitted bit sequence, provided that all these bits are independent of each other.

**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 \( \kappa \)th parameter where \( \kappa = 1, ..., K_\kappa \) and \( K_\kappa = 52 \), namely from \( u_{\kappa,\tau} = [u_{\kappa,\tau}(m) \ u_{\kappa,\tau}^{[\text{ext}]}] \). In order to generate the *extrinsic* channel output information for each desired bit, \( u_{\kappa,\tau}(\lambda) \):

\[ p(\hat{u}_{\kappa,\tau}^{[\text{ext}]}|u_{\kappa,\tau}^{[\text{ext}]}) = \prod_{m\neq\lambda,m=1}^{M} p[\hat{u}_{\kappa,\tau}(m)|u_{\kappa,\tau}(m)], \]  

(2)

where the term \( u_{\kappa,\tau}^{[\text{ext}]} \) denotes all elements of the bit pattern \( u_{\kappa,\tau} \), 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 output information and the *a priori* knowledge \( p(u_{\kappa,\tau}) \) concerning the \( \kappa \)th parameter, which is given by [7, 27]:

\[ L_{S,e}(u_{\kappa,\tau}(\lambda)) = \log \frac{\sum u_{\kappa,\tau}^{[\text{ext}]}}{\sum u_{\kappa,\tau}^{[\text{ext}]}} p(u_{\kappa,\tau}^{[\text{ext}]}) u_{\kappa,\tau}(\lambda) = +1 \ p(u_{\kappa,\tau}^{[\text{ext}]}) u_{\kappa,\tau}(\lambda) = -1 \ p(u_{\kappa,\tau}^{[\text{ext}]}) u_{\kappa,\tau}(\lambda), \]  

(3)

where \( p(u_{\kappa,\tau}^{[\text{ext}]}) u_{\kappa,\tau}(\lambda) \) can also be expressed in terms of the LLR values as [27]:

\[ p(u_{\kappa,\tau}^{[\text{ext}]}) u_{\kappa,\tau}(\lambda) = \Psi_{\kappa,\tau}^{[\text{ext}]} \exp[\sum_{u_{\kappa,\tau}(l)\neq u_{\kappa,\tau}^{[\text{ext}]}} \frac{u_{\kappa,\tau}(l)}{2} (L_{CD}^{[\text{ext}]})], \]  

(4)

and \( L_{CD}^{[\text{ext}]}) \) represents the *extrinsic* LLR values generated by soft-output channel decoding, while the product \( \Psi_{\kappa,\tau} \) cancels out in Equation (3).

**Extrinsic LLR of soft speech bit generation for exploiting the inter-frame first-order correlation:** As seen in Figure 1, 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 \( p(u_{\kappa,\tau}|u_{\kappa,\tau-1}) \), yielding:

\[ \alpha_{\tau-1}(u_{\kappa,\tau-1}) = C p(\hat{u}_{\kappa,\tau-1}|u_{\kappa,\tau-1}) \sum_{u_{\kappa,\tau-2}} p(u_{\kappa,\tau-2}|u_{\kappa,\tau-2}) \alpha_{\tau-2}(u_{\kappa,\tau-2}), \]  

(5)
where $\alpha_{\tau-1}(u_{\kappa,\tau-1})$ represents a forward recursive values and $C$ represents a normalization constant.

Finally, the *a posteriori* LLR values, $L_{1,p}(u)$ generated by each bit are given by

\[
L_{S,p}(u_{\kappa,\tau}(\lambda)) = \log \frac{\sum_{u_{\kappa,\tau}(\lambda)=+1} p(\hat{u}_{\kappa,\tau} | u_{\kappa,\tau}) \sum_{u_{\kappa,\tau-1}} p(u_{\kappa,\tau} | u_{\kappa,\tau-1}) \alpha_{\tau-1}(u_{\kappa,\tau-1})}{\sum_{u_{\kappa,\tau}(\lambda)=-1} p(\hat{u}_{\kappa,\tau} | u_{\kappa,\tau}) \sum_{u_{\kappa,\tau-1}} p(u_{\kappa,\tau} | u_{\kappa,\tau-1}) \alpha_{\tau-1}(u_{\kappa,\tau-1})}.
\]

(6)

**IV. 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 mutual information amongst the constituent receiver components.

As seen from Figure 1, the RSC decoder receives inputs from and provides outputs for both the SP and the AMR-WB decoders. More explicitly, let $I_{-,A}(x)$ denote the mutual information (MI) [1] 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 $\{1,2,3\}$. 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 decoder as well as the *a priori* input, $I_{2,A}(\hat{u})$, available for the data bits $\hat{u}$, 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 1.

Correspondingly, the RSC decoder generates both the *extrinsic* output, $I_{2,E}(\hat{u})$ representing the data bits $\hat{u}$ 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 [19]:

\[
I_{2,E}(\hat{u}) = T_{\hat{u}}[I_{2,A}(\hat{u}), I_{2,A}(c)],
\]

(7)

\[
I_{2,E}(c) = T_{c}[I_{2,A}(\hat{u}), I_{2,A}(c)],
\]

(8)

which are illustrated by the 3D surfaces seen in Figures 2 and 3, respectively.

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

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

(9)
for the AMR-WB decoder and

$$I_{3,E}(\bar{c}) = T_{2}[I_{3,A}(\bar{c}), E_b/N_0],$$

(10)

for the SP decoder. Equations (9) and (10) are illustrated in Figures 2 and 3, respectively.

The EXIT chart [17] 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.

More explicitly, the intersection of the surfaces seen in Figure 2 characterizes the best possible attainable performance, when exchanging information between the RSC decoder and the AMR-WB decoder of Figure 1 for different fixed values of $I_{2,A}(c)$, which is shown as a thick solid line. For each point $[I_{2,A}(\bar{u}), I_{2,A}(c), I_{2,E}(\bar{u})]$ of this line on the 3D space of Figure 3, there is a specific value of $I_{2,E}(c)$ determined by $I_{2,A}(\bar{u})$ and $I_{2,A}(c)$ according to the EXIT function of Equation (8). Therefore the solid line on the surface of the EXIT function of the RSC decoder seen in Figure 2 is mapped to the solid line shown in Figure 3.

In order to avoid the somewhat cumbersome interpretation of 3D representation, we project the bold EXIT curve of Figure 3 onto the 2D plane at $I_{2,A}(\bar{u}) = 0$ yielding the dotted bold line in Figure 4. 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 benchmarking 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 4 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 benchmarking 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.

V. RESULTS AND DISCUSSION

In this section we characterize the attainable performance of the proposed scheme using both the BER and the Segmental Signal to Noise Ratio (SegSNR) [2] 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 $L = 16$ and a single receiver antenna, while the remaining simulation
parameters were described in Section III. In our simulations each inner iteration was followed by two outer iterations, which together formed a three-stage “system iteration”.

Figure 5 depicts the BER versus 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 correlated narrowband Rayleigh fading channels. It can be seen from Figure 5 that the DSTS-SP-RSC-AMRWB scheme outperforms the DSTS-SP-RSC benchmarker scheme by about 1 dB at BER=$4.0 \times 10^{-5}$ after $I_{\text{system}} = 2$ iterations, where again we define a system iteration $I_{\text{system}}$ as having one inner iterations followed by two outer-iteration, as mentioned in Section III. 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 4, it is expected that the DSTS-SP-RSC-AMRWB scheme outperforms 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 [28], as seen in Figure 4. More explicitly, the actual decoding trajectory of Figure 4 recorded for $I_{\text{system}} = 2$ 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.

Let us now study the speech SegSNR performance of the proposed scheme in Figure 6. It can be seen from Figure 6 that the exploitation of the residual source redundancy, which manifests itself in terms of the unequal occurrence probability of the AMR-WB parameters and the correlation of similar parameters in two consecutive 20 ms AMR-WB encoded frames during the parameter estimation in soft-bit speech decoding [5] provides valuable *a priori* information and hence 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, when allowing a maximum channel-induced SegSNR degradation of 1 dB in comparison to the maximum attainable SegSNR maintained over perfectly error-free channels. 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 $I_{\text{system}} = 2$ iterations, again when tolerating a SegSNR degradation of 1 dB.

VI. CONCLUSION

In this contribution the three-stage turbo detection aided DSTS-SP-RSC-AMRWB scheme of Figure 1 was proposed for transmission over a temporally correlated narrowband 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 performance 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.

REFERENCES


Fig. 1. Block diagram of the DSTS-SP-RSC-AMRWB scheme.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>$R_M = I(x; y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ISP_1(t); ISP_1(t - 1)$</td>
<td>1.67</td>
</tr>
<tr>
<td>$ISP_2(t); ISP_2(t - 1)$</td>
<td>1.26</td>
</tr>
<tr>
<td>$ISP_3(t); ISP_3(t - 1)$</td>
<td>0.29</td>
</tr>
<tr>
<td>$ISP_4(t); ISP_4(t - 1)$</td>
<td>0.12</td>
</tr>
<tr>
<td>$ISP_5(t); ISP_5(t - 1)$</td>
<td>0.12</td>
</tr>
<tr>
<td>$ISP_6(t); ISP_6(t - 1)$</td>
<td>0.05</td>
</tr>
<tr>
<td>$ISP_7(t); ISP_7(t - 1)$</td>
<td>0.06</td>
</tr>
<tr>
<td>$CBGain_{sub_1}(t); CBGain_{sub_1}(t - 1)$</td>
<td>0.97</td>
</tr>
<tr>
<td>$CBGain_{sub_2}(t); CBGain_{sub_2}(t - 1)$</td>
<td>1.00</td>
</tr>
<tr>
<td>$CBGain_{sub_3}(t); CBGain_{sub_3}(t - 1)$</td>
<td>1.05</td>
</tr>
<tr>
<td>$CBGain_{sub_4}(t); CBGain_{sub_4}(t - 1)$</td>
<td>1.04</td>
</tr>
</tbody>
</table>

**Inter-frame Unequal Redundancy**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$R_D = B - H(U)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ISP_1(t)$</td>
<td>0.57</td>
</tr>
<tr>
<td>$ISP_2(t)$</td>
<td>0.45</td>
</tr>
<tr>
<td>$ISP_3(t)$</td>
<td>0.10</td>
</tr>
<tr>
<td>$ISP_4(t)$</td>
<td>0.06</td>
</tr>
<tr>
<td>$ISP_5(t)$</td>
<td>0.14</td>
</tr>
<tr>
<td>$ISP_6(t)$</td>
<td>0.03</td>
</tr>
<tr>
<td>$ISP_7(t)$</td>
<td>0.07</td>
</tr>
<tr>
<td>$LTPLag_{sub_1}(t)$</td>
<td>0.23</td>
</tr>
<tr>
<td>$LTPLag_{sub_2}(t)$</td>
<td>0.40</td>
</tr>
<tr>
<td>$LTPLag_{sub_3}(t)$</td>
<td>0.20</td>
</tr>
<tr>
<td>$LTPLag_{sub_4}(t)$</td>
<td>0.39</td>
</tr>
<tr>
<td>$FixedInd_{sub_1}(t)$</td>
<td>0.33</td>
</tr>
<tr>
<td>$FixedInd_{sub_2}(t)$</td>
<td>0.31</td>
</tr>
<tr>
<td>$FixedInd_{sub_3}(t)$</td>
<td>0.31</td>
</tr>
<tr>
<td>$FixedInd_{sub_4}(t)$</td>
<td>0.31</td>
</tr>
<tr>
<td>$CBGain_{sub_1}(t)$</td>
<td>0.12</td>
</tr>
<tr>
<td>$CBGain_{sub_2}(t)$</td>
<td>0.12</td>
</tr>
<tr>
<td>$CBGain_{sub_3}(t)$</td>
<td>0.13</td>
</tr>
<tr>
<td>$CBGain_{sub_4}(t)$</td>
<td>0.14</td>
</tr>
</tbody>
</table>

**TABLE I**

Residual redundancy in the AMR-WB codec’s parameters.
Fig. 2. 3D EXIT chart of the RSC decoder and of the AMR-WB decoder.
Fig. 3. 3D EXIT chart of the RSC decoder and of the SP decoder at $E_b/N_0=5.0$ dB with projection from Figure 2.
Fig. 4. 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.
Fig. 5. BER versus $E_b/N_0$ performance of the jointly optimised DSTS-SP-RSC-AMRWB scheme of Figure 1, when communicating over correlated non-dispersive Rayleigh fading channels.
Fig. 6. Average SegSNR versus $Eb/N_0$ performance of the jointly optimised DSTS-SP-RSC-AMR-WB scheme of Figure 1, when communicating over correlated non-dispersive Rayleigh fading channels.