Iterative Source-Channel Decoding with Cross-Layer Support for Wireless VoIP

Tobias Breddermann, Helge Lüders, and Peter Vary
Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, Germany

Ismet Aktas and Florian Schmidt
Distributed Systems Group
RWTH Aachen University, Germany

International ITG Conference on Source and Channel Coding (Siegen, Germany)
January 18-21, 2010

Addendum

During the conference, it was brought to our attention that the basic concept of our proposed method is closely related to the concept developed by Ruland and Živić [1] in the area of joint channel coding and cryptography.

[1] C. Ruland and N. Živić, "Feedback in Joint Channel Coding and Cryptography," in *Proc. of International ITG Conference on Source and Channel Coding*, Ulm, Germany, January 2008.

Iterative Source-Channel Decoding with Cross-Layer Support for Wireless VoIP

Tobias Breddermann, Helge Lüders, and Peter Vary
Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, Germany

Ismet Aktas and Florian Schmidt
Distributed Systems Group
RWTH Aachen University, Germany

***Abstract*—This paper presents a cross-layer approach for iterative source-channel decoding (ISCD) in wireless VoIP networks. The novelty of the proposed method is the incorporation of both, speech bits as well as protocol header bits, into the ISCD process. The header bits take the role of pilot bits having perfect reliability. These bits are distributed over the frame as strong supporting points for the MAP decoder which results in a significant enhancement of the output speech quality compared to the benchmark scheme using ISCD for speech only. For this approach, we exploit cross-layer concepts that support the direct communication between non-adjacent layers. These concepts enable the iterative exchange of extrinsic information between the source decoder located on the application layer and the channel decoder located on the physical layer. This technique can also be applied to audio and video transmission.**

I. Introduction

Key aspects in digital mobile radio communication systems are bandwidth efficiency and bit-error robustness. It has been proven that iterative source-channel decoding (ISCD) [1], [2], [3] can perform close to the Shannon limit with reasonable computational complexity and bandwidth consumption. The improvement of decoding quality is achieved by utilizing redundancy of the source encoded signal by means of a turbo-like exchange of extrinsic information between soft-decision source decoders and channel decoders.

However, this information exchange is more difficult to realize in packet-based mobile telephone networks due to the logical separation of source coder (located on the application layer) and channel coder (located on the physical layer). Voice over Internet Protocol (VoIP) transmission is based on the layered Open System Interconnection (OSI) architecture [4] that is standardized by the International Organization for Standardization (ISO). This model prohibits the direct communication between non-adjacent layers and, consequently, the exchange of soft (extrinsic) information between source and channel decoder. In order to tap the potential of ISCD in future communication systems, cross-layer communication will be mandatory. Assuming cross-layer communication, we will further demonstrate that header bit information can be exploited to support the iterative speech decoding process.

This paper is organized as follows: Sec. II briefly reviews the fundamentals of ISCD. Sec. III describes our cross-layer VoIP system design and Sec. IV the proposed protocol-header-supported ISCD concept. We will show that this Speech-Header-ISCD (S/H-ISCD) system will result in a perceivable speech quality improvement compared to the benchmark system, Speech-ISCD (S-ISCD), which does not incorporate the header bits into the ISCD process. The achievable speech quality gains have been measured by means of simulation and are illustrated in Sec. V. Finally, conclusions are drawn in Sec. VI.

This work has been supported by the UMIC Research Centre, RWTH Aachen University.

II. Iterative Source-Channel Decoding

The concept of ISCD is one example for the efficient use of the turbo principle [5] in digital speech transmission and is illustrated in Fig. 1. By source encoding, a speech frame $\mathbf{s}$ is represented as a set of K codec-specific parameters $\mathbf{v}_\kappa$, $\kappa = 1 \dots K$, that are quantized and assigned to unique bit patterns $\mathbf{x}_\kappa = (x_1 \dots x_{M_\kappa})$ of length M_κ. Finally, these bit patterns are grouped to form the output bit stream $\mathbf{x} = (\mathbf{x_1}\,\mathbf{x_2} \dots \mathbf{x_K})$ of the source encoder. The goal of the source encoder is the reduction of natural source redundancy and irrelevancy within each speech frame for bit rate reduction while guaranteeing a certain speech quality. In conventional transmission systems, a channel encoder purposely adds redundancy to the bit stream to cope with radio-link-related bit errors at the receiver side. However, in the case of ISCD, best performance can be expected by using redundant bit mappings (RBMs) [6] concatenated with an inner channel encoder of rate $r = 1$ (see [7]). RBMs can be easily expressed by conventional bit mappings protected by parameter-individual block codes.

The ISCD decoder is based on the iterative exchange of extrinsic information between the SISO (Soft-Input/Soft-Output) channel decoder [8] and the SDSD (soft-decision source decoder) [9]. This turbo-like process can provide decoding performance close to the Shannon limit assuming a sufficient number of decoding iterations. The interleaver π within the iterative loop plays a key role in such transmission systems. It spreads the extrinsic information of each single data bit over the complete data frame. This rearranged information

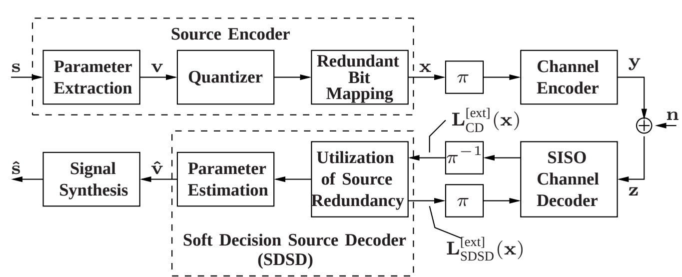


Fig. 1. Baseband model of the ISCD system.

ISBN 978-3-8007-3211-1 © VDE VERLAG -GMBH · Berlin · Offenbach

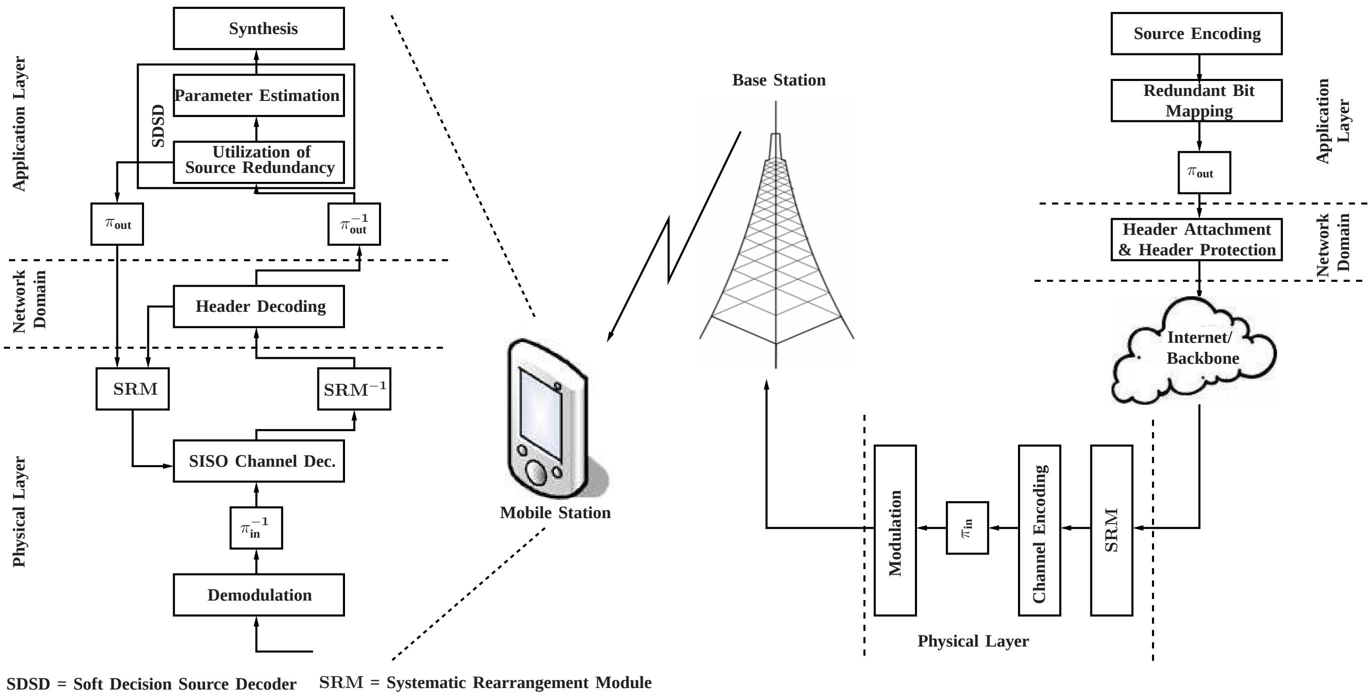


Fig. 2. Speech-Header-ISCD transceiver for the downlink in wireless VoIP systems.

can be reused in the SISO channel decoder and the SDSD, respectively, in order to refine the reliability information for each bit. The reliability information is commonly expressed in terms of log-likelihood ratios (LLRs). Assuming a bipolar representation of each bit x, i.e. $x \in \{-1, 1\}$, the *a posteriori* LLR can be stated as

$$L(x|\mathbf{z}) = \ln \frac{P(x = +1|\mathbf{z})}{P(x = -1|\mathbf{z})}, \qquad (1)$$

where $\mathbf{z} = (z(1) \dots z(n) \dots z(N))$ with $z(n) \in \mathbb{R}$, $n = 1 \dots N \in \mathbb{N}$, denotes the received data frame of length N. The sign ($\hat{x} = \text{sign}\{L(\bullet)\}$) of this LLR specifies the hard decision output bit and the magnitude $|L(\bullet)|$ represents the reliability of this decision. Applying Bayes' theorem in mixed form, (1) can be split up into the transmission related information $L(z|x)$, the bit-wise *a priori* information $L(x)$, and the terms of extrinsic information $L_{\text{CD}}^{[\text{ext}]}(x)$ (channel decoder) and $L_{\text{SDSD}}^{[\text{ext}]}(x)$ (source decoder) according to

$$L(x|\mathbf{z}) = L(z|x) + L(x) + L_{\text{CD}}^{[\text{ext}]}(x) + L_{\text{SDSD}}^{[\text{ext}]}(x). \qquad (2)$$

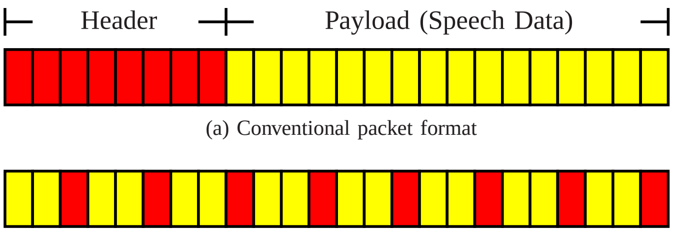
For more details, the reader is referred to [3], [10], [11].

III. Cross-Layer System Design

Let us consider the downlink scenario of our proposed S/H-ISCD transceiver system as depicted in Fig. 2. Only the downlink case has been considered in this paper, since all layers are located on the same mobile device principally enabling the utilization of ISCD. That does not hold true for the uplink scenario where the channel decoder is allocated at the base station and the source decoder is at the remote switching center. Due to delay and traffic constraints, iterations between these elements might not be possible. However, in this case parts of the kernel of the SDSD can be implemented at the base station. The parameters can be estimated and re-quantized before transmission to the remote signal synthesis module. In order to attain high data compression, a speech encoder calculates from the incoming speech frames a set of quantized codec-specific parameters, such as gains and filter coefficients, and the excitation signal for the synthesis filter. The speech data are protected against radio-link-related errors using parameter-individual block codes for the redundant bit Mappings (RBMs) before performing the outer interleaving π_{out}. Within the network domain, protocol headers are attached to the incoming speech data forming the VoIP packet. Strong forward error correction (FEC) codes are additionally applied to the protocol headers to enable error correction at the receiver while cyclic redundancy checks (CRCs) are part of the protocols to check for residual bit errors within the decoded headers. In the base station, header and speech bits are systematically rearranged within the downlink packet by a novel systematic rearrangement module (SRM) as illustrated in Figure 3b. This leads, in contrast to the conventional format, to an even distribution of header bits within the packet. The motivation for this unusual rearrangement strategy will be given in Section IV. Finally, each VoIP packet is channel encoded by a rate-1 convolutional encoder, interleaved by the inner interleaver π_{in} and modulated for transmission.

At the receiving end, the mobile device performs signal synthesis using the ISCD concept. The soft demodulated and deinterleaved packet is decoded by the SISO channel decoder that provides reliability information for each header bit and speech bit. After the separation of header and speech by the inverse SRM (SRM^{-1}), the protocol headers are decoded and checked by their CRCs for residual bit errors. While packets with faulty headers are discarded within the protocol stack, packets with correctly decoded protocol headers are forwarded

ISBN 978-3-8007-3211-1 © VDE VERLAG -GMBH · Berlin · Offenbach



(a) Conventional packet format

(b) Modified packet format for ISCD

Fig. 3. Systematic rearrangement of the header bits and the speech bits.

to the SDSD located on the application layer. This is the most important aspect of our proposed concept, since these perfectly known bits can be used in the decoding process, improving the extrinsic information and, thus, the speech quality as described in the next section.

The critical point in radio communication networks based on ISCD is the logical separation of the source decoder and the channel decoder. Therefore, we have permitted cross-layer communication in order to realize reliability feedback between the both non-adjacent layers. In this case, extrinsic information about the speech bits, calculated by the SDSD, can be passed back as additional *a priori* knowledge to the SISO channel decoder. This can be realized in the downlink, i.e., in the mobile.

IV. Exploitation of Protocol Header Information

In contrast to commonly considered ISCD systems, extrinsic header information is also generated assuming that the application layer has access to the protocol stack and that all packets reaching the application layer contain correctly decoded headers. In this case, all header bits are known and can be encoded again to determine the encoded header bits that have been incorporated by the ISCD transmitter. The related LLRs can be represented by extrinsic LLRs $L_h^{[\text{ext}]}(x)$ with an absolute value approaching infinity. Thus, $L_h^{[\text{ext}]}(x)$ can be expressed as

$$L_h^{[\text{ext}]}(x) \rightarrow x \cdot \infty \quad , \quad x \in \{-1, 1\}. \tag{3}$$

Hence, we can utilize perfect *a priori* knowledge at all header bit positions during SISO channel decoding. In the S/H-ISCD transceiver system, the interaction between the SRM and the SISO channel decoder plays a key role. While the SRM provides an even distribution of header bits within the VoIP packet, the SISO channel decoder "smears" the perfect *a priori* knowledge of the header bits over some adjacent speech bits and generates extrinsic speech LLRs with increased reliability. This successfully works due to the fact that during channel encoding header bits and speech data bits are linked together producing output bits that contain information of protocol header bits as well as speech data bits. Since each header bit supports the soft decoding of speech bits located in its surrounding vicinity, they are equally spaced within the packet by the SRM. In this case, the header bits can be interpreted as pilot bits with perfect reliability that act as strong supporting points for the maximum a posteriori (MAP) channel decoder.

To give an analytic explanation for the effectiveness of the proposed method, the trellis diagram for the convolutional IIR rate-1 code with the octal generator polynomial $\mathbf{G} = \left(\frac{2}{3}\right)_8$

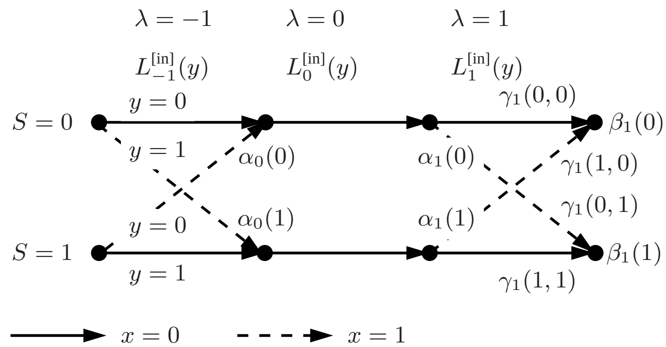


Fig. 4. Trellis-diagram for the convolutional IIR rate-1 code with the generator polynomial $\mathbf{G} = \left(\frac{2}{3}\right)_8$.

is illustrated in Fig. 4. Please note that this very simple convolutional code is not applied in our transmission system but taken as an example for the following discussion due to its simplicity. We assume w.l.o.g. that a header bit $x = 0$ is placed within the packet at time instance $\lambda = 0$ [1]. Remember that this bit is known after the first ISCD iteration and therefore only the solid transitions in Fig. 4 are valid within the trellis diagram at $\lambda = 0$. Using the well-known MAP algorithm [8], the extrinsic LLR $L_{CD}^{[\text{ext}]}(x)$ at $\lambda = 1$ (speech bit) can be calculated by means of (4) - (7). In the proposed system, the initial state reliabilities $\alpha_1(0)$ and $\alpha_1(1)$ can be refined by the channel-related LLR $L_\lambda^{[\text{in}]}(y) = L_0^{[\text{in}]}(y) = L_0(z|y)$ at the header bit position due to the reduced number of valid state transitions. Please note, as we assume a non-systematic channel code, we have to consider $L_0(z|y)$ instead of $L_0(z|x)$. The state reliabilities are given by

$$\alpha_1(0) = \alpha_0(0)\gamma_0(0,0) = \alpha_0(0) \exp\left(\frac{L_0^{[\text{in}]}(y)}{2}\right) \tag{8}$$

$$\alpha_1(1) = \alpha_0(1)\gamma_0(1,1) = \alpha_0(1) \exp\left(-\frac{L_0^{[\text{in}]}(y)}{2}\right). \tag{9}$$

Note, that the first term of (6) can be skipped since both reliabilities $\alpha_\lambda(S(\lambda))$ are equally affected by this term and the influence of $L_0^{[\text{in}]}(x) \rightarrow \infty$ has already been incorporated by skipping the impossible transitions at $\lambda = 0$. Let us further assume that $\alpha_0(0) = \alpha_0(1)$ and $\beta_1(0) = \beta_1(1)$. This implies that no extrinsic information (i.e. $L_{CD}^{[\text{ext}]}(x) = 0$) can be provided by the benchmark VoIP system that does not incorporate header bits into the ISCD process. Using (8) , (9) and

$$\tilde{\gamma_1} \doteq \gamma_1^{[\text{ext}]}(0,0) = \gamma_1^{[\text{ext}]}(1,0) = \exp\left(\frac{L_1^{[\text{in}]}(y)}{2}\right)$$

$$\hat{\gamma_1} \doteq \gamma_1^{[\text{ext}]}(1,1) = \gamma_1^{[\text{ext}]}(0,1) = \exp\left(-\frac{L_1^{[\text{in}]}(y)}{2}\right),$$

[1]The extrinsic information for $x = 1$ can be derived by setting $L_\lambda^{[\text{in}_\text{new}]}(y) = -L_\lambda^{[\text{in}]}(y)$ and $\beta_\lambda^{[\text{new}]}(S(\lambda)) = -\beta_\lambda(S(\lambda))$ for $\lambda \geq 1$. Assuming bits that are equally affected by the channel, only the signs have to be swapped to account for the state change.

ISBN 978-3-8007-3211-1

$$\alpha_\lambda(S^{(j)}(\lambda)) = \sum_{i=0}^{2^\nu-1} \gamma_\lambda(S^{(i)}(\lambda-1), S^{(j)}(\lambda)) \cdot \alpha_{\lambda-1}(S^{(i)}(\lambda-1)) \tag{4}$$

$$\beta_\lambda(S^{(i)}(\lambda)) = \sum_{j=0}^{2^\nu-1} \gamma_\lambda(S^{(i)}(\lambda), S^{(j)}(\lambda+1)) \cdot \beta_{\lambda+1}(S^{(j)}(\lambda+1)) \tag{5}$$

$$\begin{aligned}\gamma_\lambda(S(\lambda-1), S(\lambda)) &= \exp\left(\frac{x}{2}\cdot L^{[\text{in}]}(x)\right)\cdot \exp\left(\sum_{\substack{i=1\\ i\neq i_{\text{sys}}}}^{\Upsilon} \frac{y(i)}{2}\cdot L^{[\text{in}]}(y(i))\right)\\ &= \exp\left(\frac{x}{2}\cdot L^{[\text{in}]}(x)\right)\cdot \gamma_\lambda^{[\text{ext}]}(S(\lambda-1), S(\lambda)|x)\end{aligned} \tag{6}$$

$$L_{CD}^{[\text{ext}]}(x) = \ln \frac{\sum\limits_{j=0}^{2^\nu-1} \beta_\lambda(S^{(j)}(\lambda)) \cdot \sum\limits_{i=0}^{2^\nu-1} \gamma_\lambda^{[\text{ext}]}(S^{(i)}(\lambda-1), S^{(j)}(\lambda)|x=+1)\cdot \alpha_{\lambda-1}(S^{(i)}(\lambda-1))}{\sum\limits_{j=0}^{2^\nu-1} \beta_\lambda(S^{(j)}(\lambda)) \cdot \sum\limits_{i=0}^{2^\nu-1} \gamma_\lambda^{[\text{ext}]}(S^{(i)}(\lambda-1), S^{(j)}(\lambda)|x=-1)\cdot \alpha_{\lambda-1}(S^{(i)}(\lambda-1))} \tag{7}$$

Fig. 5. MAP algorithm applied for SISO channel decoding.

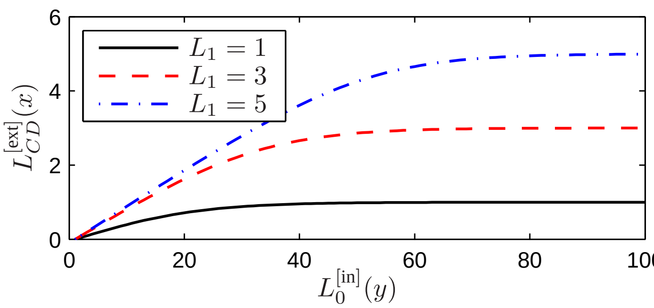


Fig. 6. Extrinsic information $L_{CD}^{[\text{ext}]}(x)$ for different $L_1^{[\text{in}]}(y) := L_1$.

we can compute $L_{CD}^{[\text{ext}]}(x)$ for our proposed system:

$$\begin{aligned}L_{CD}^{[\text{ext}]}(x) &= \ln\left(\frac{\alpha_1(0)\tilde{\gamma}_1\beta_1(0) + \alpha_1(1)\hat{\gamma}_1\beta_1(1)}{\alpha_1(0)\hat{\gamma}_1\beta_1(1) + \alpha_1(1)\tilde{\gamma}_1\beta_1(0)}\right)\\ &= \ln\left(\frac{\exp\left(\frac{L_0^{[\text{in}]}(y)+L_1^{[\text{in}]}(y)}{2}\right) + \exp\left(\frac{-L_0^{[\text{in}]}(y)-L_1^{[\text{in}]}(y)}{2}\right)}{\exp\left(\frac{L_0^{[\text{in}]}(y)-L_1^{[\text{in}]}(y)}{2}\right) + \exp\left(\frac{-L_0^{[\text{in}]}(y)+L_1^{[\text{in}]}(y)}{2}\right)}\right).\end{aligned}$$

This equation simplifies to

$$L_{CD}^{[\text{ext}]}(x) = \ln\left(\frac{\cosh\left(\frac{L_1^{[\text{in}]}(y)+L_0^{[\text{in}]}(y)}{2}\right)}{\cosh\left(\frac{L_1^{[\text{in}]}(y)-L_0^{[\text{in}]}(y)}{2}\right)}\right) \geq 0. \tag{10}$$

The extrinsic information $L_{CD}^{[\text{ext}]}(x)$ is plotted against the channel-related LLR $L_0^{[\text{in}]}(y) \geq 0$ at the header bit position in Fig. 6 for different $L_1^{[\text{in}]}(y)$. This proves that the insertion of perfectly known bits x provides additional extrinsic information $|L_{CD}^{[\text{ext}]}(x)|$ for $L_0^{[\text{in}]}(y) \neq 0$. In realistic scenarios longer headers and convolutional codes with larger constraint length are assumed, increasing the overall extrinsic information.

V. System Configuration and Simulation Results

The transmission of speech by the proposed S/H-ISCD system given in Fig. 2 is simulated. The highest mode (12.2 kbit/s) of the adaptive multi-rate narrow-band (AMR-NB) speech codec [12] is used for source encoding. Unequal error protection is realized by parameter-individual redundant bit mappings (RBMs) of rate $r_\kappa^{\text{BM}} = \frac{M_\kappa}{M_\kappa + M_\kappa^\star}$ which apply rate-flexible linear block codes for each parameter $\kappa = 1 \ldots M$ of length M_κ. $M_\kappa^\star$ is the number of used parity bits. The rate-constrained design specification can be mathematically expressed as follows:

$$\mathbf{G_1} = \left(\mathbf{I}_{M_\kappa}\ \mathbf{1}_{M_\kappa,1}\ \mathbf{P}_{M_\kappa,M_\kappa^\star-1}\right) \quad , \quad \frac{1}{2} < r_\kappa^{\text{BM}} < 1 \tag{11}$$

$$\mathbf{G_2} = \left(\mathbf{I}_{M_\kappa}\ \mathbf{I}^\star_{M_\kappa,M_\kappa^\star}\right) \quad , \quad 0 < r_\kappa^{\text{BM}} \leq \frac{1}{2}, \tag{12}$$

with $\mathbf{I}_{M_\kappa}$ denoting the $M_\kappa \times M_\kappa$ identity matrix, $\mathbf{1}_{M_\kappa,1}$ the $M_\kappa \times 1$ all-one matrix (i.e. a matrix containing only ones). The parity fractions of both block codes are given by

$$\begin{aligned}\mathbf{I}^\star_{M_\kappa,M_\kappa^\star} &= \left(\mathbf{i}_0^\star \ldots \mathbf{i}_n^\star \ldots \mathbf{i}^\star_{M_\kappa^\star-1}\right) ,\\ \mathbf{i}_n^\star &= \left(i^\star_{0,n} \ldots i^\star_{m,n} \ldots i^\star_{M_\kappa-1,n}\right) ,\\ i^\star_{m,n} &= 0 \quad \forall\ m \neq (n \bmod M_\kappa)\end{aligned} \tag{13}$$

and

$$\mathbf{P}_{M_\kappa,M_\kappa^\star-1} = \mathbf{1}_{M_\kappa,M_\kappa^\star-1} - \mathbf{I}^\star_{M_\kappa,M_\kappa^\star-1} \tag{14}$$

with $M_\kappa^\star = \frac{M_\kappa}{r_\kappa^{\text{BM}}} - M_\kappa$. For the example of $M_\kappa = 4$ and $M_\kappa^\star = 3$ the generator matrix looks like

$$\mathbf{G_1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1\\ 0 & 1 & 0 & 0 & 1 & 1 & 0\\ 0 & 0 & 1 & 0 & 1 & 1 & 1\\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

and for $M_\kappa = 4$ and $M_\kappa^\star = 7$ like

$$\mathbf{G_2} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

ISBN 978-3-8007-3211-1 © VDE VERLAG -GMBH · Berlin · Offenbach

TABLE I
REDUNDANT BIT MAPPINGS APPLIED TO THE AMR-NB CODEC AT 12.2 KBIT/S.

AMR-NB Mode	Parameter	total bits per frame uncoded / after RBM
12.2 kbit/s	5 LSFs	38 / 195
	Pitch delay	30 / 120
	Pitch gain	16 / 128
	Algebraic code	140 / 200
	Code book gain	20 / 128
	Total	**244 / 771**

Due to the special structure of this code the parity bits can easily be generated using simple binary operations and the generator matrices do not need to be explicitly stored. It has been proven in [13] that the repetition code $\mathbf{G_2}$ outperforms $\mathbf{G_1}$ for rates $r_\kappa^{\mathrm{BM}} \leq \frac{1}{2}$. However, $\mathbf{G_1}$ is more capable of performing iterative decoding for rates $r_\kappa^{\mathrm{BM}} > \frac{1}{2}$ due to the fact that the minimum hamming distance d_{min} of that code is greater than or equal to 2 for any rate. This is required for iterative decoding schemes in order to generate perfect extrinsic information in the presence of perfect a priori knowledge. The RBMs for a 20 ms speech frame is given in Table I.

The headers are error-protected by strong FECs. Three different header sizes of 8 , 24 and 40 bytes after error-protection are considered. The header bits are inserted at predefined positions into the interleaved speech bits as described in Sec. III for our proposed S/H-ISCD transceiver system. This system is benchmarked by the conventional Speech-ISCD (S-ISCD) system which does not incorporate the header bits into the ISCD process and performs separated encoding/decoding for the protocol headers. For both systems, a convolutional IIR rate-1 channel encoder with octal generator polynomial $\mathbf{G} = \left(\frac{10}{17}\right)_8$ and constraint length $K = 4$ is applied to each packet containing one speech frame. That leads to a total number of $771 + 3$ tail bits per frame and thus to a gross bit rate of 38.7 kbit/s which corresponds to the typical bit rate of the UMTS standard for 12.2 kbit/s speech before rate matching. As the purpose of this contribution is to demonstrate the effectiveness of the cross-layer concept, we use a non-fading additive white Gaussian noise (AWGN) channel for simplicity with Signal-to-Noise-Ratios (SNRs) $\frac{E_b}{N_0}$ varied between -2 dB and $+4$ dB. E_b is the energy per information bit and N_0 the noise power density. Fixing the energy per modulation symbol to $E_s = 1$, E_b can be computed for a speech frame of length N_s ($N_s = 244$ bits for the AMR-NB at 12.2 kbit/s) and a packet of length N_p to

$$E_b = E_s \cdot \frac{1}{m} \cdot \frac{N_p}{N_s}, \tag{15}$$

where m is the number of bits incorporated in one modulation symbol (i.e. 2 for QPSK). The number of ISCD iterations is 10.

There are several objective measures to quantify the subjective speech quality. In this paper, performance has been evaluated employing the perceptual evaluation of speech quality (PESQ) [14] of the received speech. PESQ is suitable to reflect the perceptual speech quality by a combination of objective instrumental measures. It rates the speech quality in terms of MOS-LQO[2] by a value between 1.02 (poor quality) and 4.56 (good quality). The maximum MOS-LQO score is limited by the output MOS-LQO of the AMR-NB codec, i.e., to a value of 3.8616 for the utilized speech sample.

The simulation results are depicted in Fig. 7: The dashed curves indicate the results achieved with the S-ISCD system that does not exploit perfect extrinsic header bit information, while the solid curves correspond to our proposed S/H-ISCD system with systematically rearranged header bits. We want to point out that in both systems within each subfigure (7a, 7b and 7c) a header of equal length is transmitted. However in all three cases the same energy for transmitting the packets (information plus header of different size) is used. If the number of header bits is increased, the energy per transmitted data bit is decreased.

It can be seen that the utilization of header bits as supporting points within the iterative speech decoding process increases the perceived speech quality. As expected, larger header sizes lead to higher gains. In all depicted cases the proposed S/H-ISCD system significantly outperforms the benchmark S-ISCD system, i.e., even for short packet headers where strong robust header compression (ROHC) [15] is applied (e.g. UMTS-LTE). This proves the high potential of our proposed concept. For 40-byte error-protected headers, a gain of approximately 1.8 dB in terms of speech quality is achieved by our S/H-ISCD transceiver system compared to the conventional S-ISCD transceiver system.

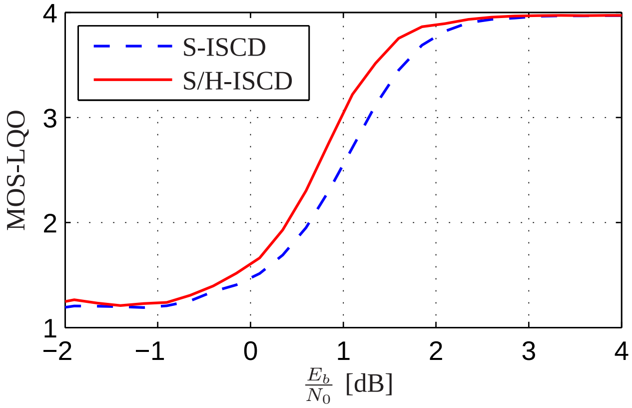
A further and rather surprising result is that, assuming a constant packet energy, our proposed system shows comparable performance for certain header sizes larger than zero, although the energy per packet bit is much less. Considering Fig. 8, we can further observe that for bad channel qualities, i.e., $\frac{E_b}{N_0} < 1.6$ dB, the performance is even 0.3 dB better in the case of 40 byte headers. Hence, the loss in decoding robustness caused by the header's energy consumption can be significantly reduced or even eliminated leading to small performance gains for certain header sizes.

VI. CONCLUSION

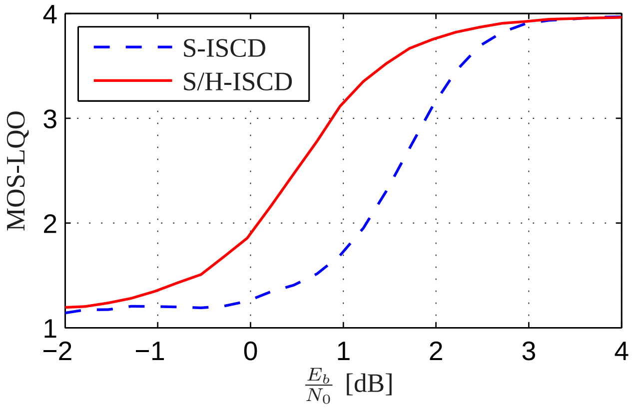
In this paper, the S/H-ISCD transceiver with cross-layer support for speech communication over packet-switched wireless networks (see Fig. 2) has been proposed. It enables, firstly, an iterative exchange of extrinsic information between the source decoder located on the application layer and the channel decoder located on the physical layer and, secondly, the incorporation of known header bits as perfect *a priori* knowledge into the iterative decoding process. This novel cross-layer concept guarantees for significantly improved quality of wireless speech transmission compared to the benchmark S-ISCD transceiver. In summary, our S/H-ISCD transceiver outperforms the benchmark S-ISCD system for any header sizes larger than zero, i.e., even if ROHC is applied to the packet headers in order to increase the goodput of the transmission system. It could be further shown that, assuming

[2]mean opinion score - listening quality objective

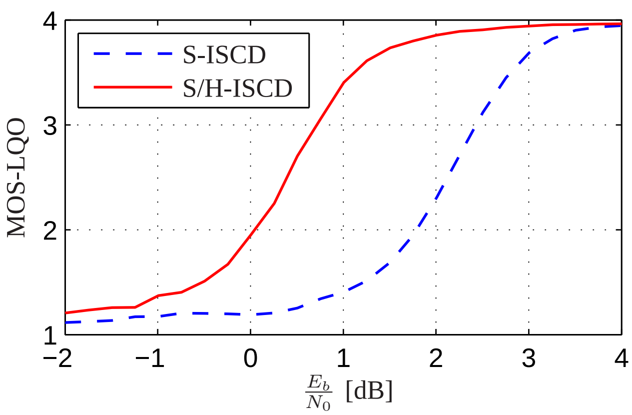
ISBN 978-3-8007-3211-1 © VDE VERLAG -GMBH · Berlin · Offenbach



(a) Header size 8 bytes (including error protection)



(b) Header size 24 bytes (including error protection)



(c) Header size 40 bytes (including error protection)

Fig. 7. MOS-LQO versus $\frac{E_b}{N_0}$ obtained using different header sizes for the benchmark Speech-ISCD (S-ISCD) system and for the proposed Speech-Header-ISCD (S/H-ISCD) system.

constant packet energy, the header size is not a critical factor as compared to state-of-the-art transmission systems, which do not exploit header information. This concept causes nearly no loss in decoding quality by the header's energy consumption for common header sizes.

ACKNOWLEDGMENT

The authors would like to thank Laurent Schmalen for providing the basic simulation software and Prof. Klaus Wehrle for fruitful discussions on the protocol aspects.

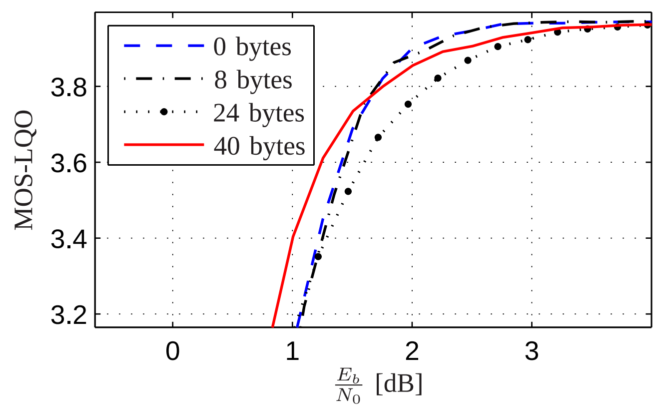


Fig. 8. MOS-LQO versus $\frac{E_b}{N_0}$ obtained using different header sizes (0, 8, 24 and 40 bytes) for the proposed Speech-Header-ISCD (S/H-ISCD) system.

REFERENCES

[1] M. Adrat, P. Vary, and J. Spittka, "Iterative Source-Channel Decoder using Extrinsic Information from Softbit-Source Decoding," in *Proc. of ICASSP*, Salt Lake City, UT, USA, 2001.

[2] N. Görtz, "Iterative Source-Channel Decoding Using Soft-In/Soft-Out Decoders," in *Proc. of ISIT*, Sorrento, Italy, 2000.

[3] M. Adrat, T. Clevorn, and L. Schmalen, "Iterative Source-Channel Decoding & Turbo DeCodulation," in *Advances in Digital Speech Transmission*, R. Martin, U. Heute, and C. Antweiler, Eds. John Wiley & Sons, Ltd, 2008, ch. 13.

[4] *ITU-T X.200: Data Networks and Open System Communications - Open System Interconnection - Model and Notation*, ITU-T, July 1994.

[5] C. Berrou and A. Glavieux, "Near Optimum Error Correcting Coding and Decoding: Turbo-Codes," *IEEE Transactions on Communications*, vol. 44, no. 10, pp. 1261–1271, October 1996.

[6] T. Clevorn, P. Vary, and M. Adrat, "Iterative Source-Channel Decoding Using Short Block Codes," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, Toulouse, France, 2006, pp. 221–224.

[7] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic Information Transfer Functions: A Model and Two Properties," in *Proc. Conference on Information Sciences and Systems (CISS)*, Princeton, New Jersey (USA), March 2002.

[8] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate (Corresp.)," *IEEE Trans. Inform. Theory*, vol. 20, no. 2, March 1974.

[9] T. Fingscheidt and P. Vary, "Softbit Speech Decoding: A New Approach to Error Concealment," *IEEE Trans. Speech Audio Processing*, vol. 9, no. 3, pp. 240–251, March 2001.

[10] T. Fingscheidt, "Parameter Models and Estimators in Soft Decision Source Decoding," in *Advances in Digital Speech Transmission*, R. Martin, U. Heute, and C. Antweiler, Eds. John Wiley & Sons, Ltd, 2008, ch. 10, pp. 281–310.

[11] S. Heinen and M. Adrat, "Optimal MMSE Estimation for Vector Sources with Spatially and Temporally Correlated Elements," in *Advances in Digital Speech Transmission*, R. Martin, U. Heute, and C. Antweiler, Eds. John Wiley & Sons, Ltd, 2008, ch. 11, pp. 311–328.

[12] ETSI EN 301 704 V7.2.1 (2000-04), *Digital Cellular Telecommunications System (Phase 2+); Adaptive Multi-Rate (AMR) Speech Transcoding (GSM 06.90 Version 7.2.1 Release 1998)*, April 2000.

[13] T. Clevorn, L. Schmalen, P. Vary, and M. Adrat, "On Redundant Index Assignments for Iterative Source-Channel Decoding," *IEEE Commun. Lett.*, vol. 12, pp. 514–516, 2008.

[14] *ITU-T P.862 Perceptual Evaluation Of Speech Quality (PESQ): An Objective Method for End-To-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs*, ITU-T, February 2001.

[15] C. Bormann et al., "Robust Header Compression (ROHC): Framework and Four Profiles," RFC 3095, Tech. Rep., 2001.

ISBN 978-3-8007-3211-1