Reconstruction of Multiple Descriptions by MMSE Estimation

Laurent Schmalen and Peter Vary

Institute of Communication System and Data Processing, RWTH Aachen University, 52074 Aachen
E-Mail: {schmalen,vary}@ind.rwth-aachen.de
Web: www.ind.rwth-aachen.de

Abstract

This paper shows the application of *soft-decision source decoding* (SDSD) to correlated speech codec parameters given by multiple description scalar quantization. In the presence of additive white Gaussian noise on the transmission link, the reconstruction quality can be largely improved by the application of SDSD. Furthermore, if the transmission link is characterized by packet losses only, the presented approach improves the reconstruction quality whenever packet losses occur. The proposed approach is compared to a well-known soft-decision approach based on cross-decoding.

1 Introduction

Multiple Description Coding (MDC) [1, 2] is a tool to generate two (or more) descriptions of a signal which are then independently transmitted over a network with possible packet losses. If all descriptions are correctly received, the signal can be reconstructed with the best possible quality. If one or more descriptions of the signal are missing due to packet losses, the signal can still be reconstructed, however, with degraded overall quality.

Multiple description coding can also be used for a more general kind of hierarchical coding: due to bottlenecks in the network, parts of the packets may be rejected, thus allowing a flexible rate adaptation. One example of a speech and audio codec employing MDC is the *FlexCode* source coder [3, 4]. Multiple description codes are generally quantified by their *index assignment* [1] which maps a central code book index to two or more side code book indices. The set of side code book indices form the individual descriptions.

Residual redundancy of source codec parameters such as scale factors or predictor coefficients for speech, audio, and video signals, occurs due to imperfect source encoding resulting for instance from delay and complexity constraints. This redundancy can be utilized by a *soft decision source decoder* (SDSD) [5] at the receiver to improve the reconstruction quality. *Iterative source-channel decoding* (ISCD) [6, 7] is an extension of SDSD and exchanges in an iterative process so-called *extrinsic* reliabilities between a SDSD and a channel decoder.

Approaches to utilize soft information in the decoding of multiple descriptions can be found in, e.g., [8], [9]. The concept presented in [8] uses the inherent redundancy of multiple descriptions to improve the decoding performance. This approach is then also extended to realize a Turbo-like transmission scheme with iterative decoding in [8].

In this paper, we apply SDSD with *minimum mean square error* (MMSE) estimation to multiple descriptions. With SDSD, we can inherently exploit the redundancy in the quantized parameters as well as the redundancy contained in the multiple description coding scheme. Besides improving the quality in the pure-noise case (i.e., if no packets are lost), we show that our presented approach is also able to improve the quality in the packet-loss case by exploiting the correlation between consecutive frames.

2 System Model

In Fig. 1 the baseband model of the considered transmission system with non-iterative decoding is depicted. A frame $\mathbf{u}_t$ consisting of K_S unquantized codec parameters $u_{t,k}$ is quantized using a Q-level scalar quantizer $\mathcal{Q}$ to $v_{t,k} = \mathcal{Q}(u_{t,k})$ with $v_{t,k} \in \mathbb{V} = \{\bar{v}_1;\ldots;\bar{v}_Q\}$. The set $\mathbb{V}$ denotes the central quantizer code book. To each quantized parameter $v_{t,k}$, a quantizer index $i_{t,k} = \mathcal{I}(v_{t,k})$ is assigned with $i_{t,k} \in \{1;\ldots;Q\}$. It holds that $v_{t,k} = \bar{v}_{i_{t,k}}$. The *multiple description index assignment* (MDIA) generates two descriptions of the quantizer indices $i_{t,k}$ according to the method proposed in [1]. The resulting indices are denoted by $i^{[j]}_{t,k} = \mathcal{I}^{[j]}(i_{t,k})$, with $j \in \{0;1\}$ indicating the description. Furthermore, we introduce the side code books $\mathbb{V}^{[j]}$ of size $Q^{[j]}$ defined according to [1].

To each $i^{[j]}_{t,k}$, a bit pattern $\mathbf{b}^{[j]}_{t,k} = \mathcal{B}^{[j]}(i^{[j]}_{t,k})$, consisting of $M^{[j]}$ bits, is assigned. The bit patterns are selected from a set $\mathbb{B}^{[j]} = \{\bar{\mathbf{b}}^{[j]}_1;\ldots;\bar{\mathbf{b}}^{[j]}_{B^{[j]}}\}$ with $B^{[j]} = Q^{[j]}$ being the number of distinct bit patterns per description j. All bit patterns of a description are grouped to the bit stream $\mathbf{x}^{[j]}_t = (\mathbf{b}^{[j]}_{t,1},\ldots,\mathbf{b}^{[j]}_{t,K_S})$. The single descriptions are interleaved (interleaver π) and optionally channel encoded prior to transmission over the channel. Each description is independently transmitted over a packet-erasure AWGN channel. The single bipolar values of the encoded vector $\mathbf{y}^{[j]}$ (each with symbol energy $E_\mathrm{s} = 1$) may be subject to AWGN with power spectral density $\sigma^2_n = N_0/2$. Additionally, each complete packet $\mathbf{y}^{[j]}$ may be erased with probability ϵ (indicated by the switches in Fig. 1). This

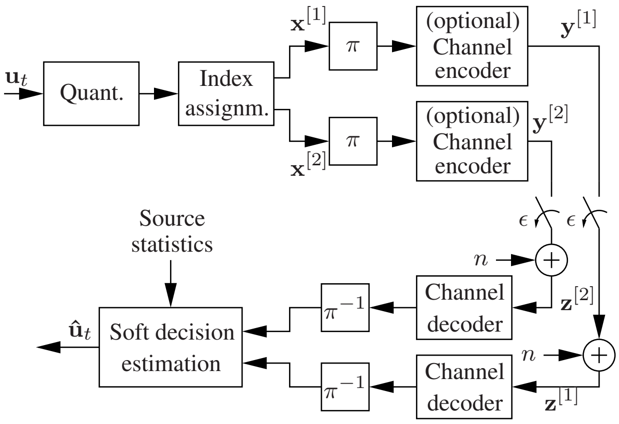

Figure 1: System model of the considered multiple description transmission with non-iterative decoding

ISBN 978-3-8007-3300-2 © VDE VERLAG GMBH · Berlin · Offenbach

channel models, e.g., the packet transmission over a wireless link: due to network congestion or synchronization issues, the complete packet may be lost. If the packet is received, the receiver noise is modelled by AWGN. At the receiver, channel decoding (if necessary) is performed using a *soft-input, soft-output* (SISO) channel decoder (e.g., the BCJR algorithm [10] for block or convolutional codes), de-interleaved and fed to the joint soft decision estimator. Note that if a description is lost, no channel-related information for the description is available.

Usually, due to numerical reasons, the channel decoder is described using L-values [11]. After transmission over an AWGN channel, the received L-value is given by $L(z_\kappa^{[j]}|y_\kappa^{[j]}) = \frac{2}{\sigma_n^2} z_\kappa^{[j]}$. If an erasure has occurred, all L-values in the description are set to zero, i.e, $L(z_\kappa^{[j]}|y_\kappa^{[j]}) = 0$. The channel decoder generates *a posteriori* L-values according to [10, 11] which, after deinterleaving, are denoted as $L_{\text{CD}}^{[\text{ap}]}(x_k^{[j]})$, with $x_k^{[j]}$ denoting the k'th bit of the j'th description $\mathbf{x}^{[j]}$. If the bit stream is described using individual bit patterns, the notation $L_{\text{CD}}^{[\text{ap}]}(b_{t,k,m}^{[j]})$ denotes the L-value of the m'th bit of bit pattern $\mathbf{b}_{t,k}^{[j]}$ at the channel decoding output.

3 MMSE Estimation of Multiple Descriptions

The multiple description quantizer consists of a single fine-resolution scalar quantizer, called central quantizer, and a mapping of a quantization index obtained by the central quantizer to a set of side indices, which are called descriptions. Each of these descriptions is transmitted independently. As there exists a one-to-one mapping between the central code book index and bit patterns in each description, we can almost directly incorporate the framework of [5] to MDC.

The first step of the soft decision source decoder (SDSD) consists in generating the channel-related probabilities

$$\gamma_{t,k}^{[j]}(\ell) = \prod_{m=1}^{M^{[j]}} \frac{1}{1+\exp\left(-(1-\bar{b}_{\ell,m}^{[j]})L_{\text{CD}}^{[\text{ap}]}(b_{t,k,m}^{[j]})\right)} \quad (1)$$

with $\bar{b}_{\ell,m}^{[j]}$ denoting the m'th bit of the bit pattern $\bar{\mathbf{b}}_\ell^{[j]} \in \mathbb{B}^{[j]}$. Equation (1) has to be evaluated for each $\ell \in \{1;\ldots;B^{[j]}\}$ and each description $j \in \{0;1\}$. If a description j has been lost due to a packet erasure, the factors $\gamma_{t,k}^{[j]}(\ell)$ are determined as $\gamma_{t,k}^{[j]}(\ell) = \frac{1}{B^{[j]}}$, $\forall \ell \in \{1;\ldots;B^{[j]}\}$. If no AWGN is present on the channel and the considered description is not lost, then $\gamma_{t,k}^{[j]}(\ell) \in \{0;1\}$ with $\gamma_{t,k}^{[j]}(\ell) = 1$, if the ℓ'th quantizer index has been transmitted (and thus received) and $\gamma_{t,k}^{[j]}(\ell) = 0$ otherwise.

The factors $\gamma_{t,k}^{[j]}(\ell)$ form the basis of all considered soft decoding algorithms for multiple descriptions. The approach of [8] for example refines (1) with cross-information from the description $1-j$. For a detailed description of the algorithm, we refer to [8].

The easiest estimator is the so-called NAK estimator which does not make use of any *a priori* knowledge of the source codec parameters. The NAK estimator is given by

$$\hat{u}_{t,k}^{[\text{NAK}]} = C_1 \sum_{q=1}^{Q} \bar{v}_q \cdot \gamma_{t,k}^{[1]}(\mathcal{I}^{[1]}(q))\gamma_{t,k}^{[2]}(\mathcal{I}^{[2]}(q)) \quad (2)$$

with C_1 being a normalization constant such that $C_1 \sum_{q=1}^{Q} \gamma_{t,k}^{[1]}(\mathcal{I}^{[1]}(q))\gamma_{t,k}^{[2]}(\mathcal{I}^{[2]}(q)) = 1$. Equation (2) makes use of the fact that there is a one-to-one correspondence between $i_{t,k}$ and $(i_{t,k}^{[1]}, i_{t,k}^{[2]})$ and that both descriptions are transmitted independently. Note that the factors $\gamma_{t,k}^{[1]}$ and $\gamma_{t,k}^{[2]}$ can be multiplied as both descriptions are transmitted over independent channels. If only a single description j is received, the estimator (2) changes to

$$\hat{u}_{t,k}^{[\text{NAK}]}\Big|_{\substack{\text{only desc. } j \\ \text{received}}} = C_2 \sum_{q=1}^{Q^{[j]}} \bar{v}_q^{[j]} \cdot \gamma_{t,k}^{[j]}(q) \,. \quad (3)$$

If both packets are erased, $\hat{u}_{t,k}^{[\text{NAK}]} = \frac{1}{Q}\sum_{q=1}^{Q} \bar{v}_q$, i.e., the mean of the central code book is utilized.

If *a priori* knowledge on parameter level is available, it can be incorporated into the estimation process, similar as in [5]. For instance, if *a priori* knowledge of zeroth order (AK0) is available, i.e., probabilities $P_V(v_{t,k})$ or $P_I(i_{t,k})$ equivalently, the estimator modifies to

$$\hat{u}_{t,k}^{[\text{AK0}]} = C_3 \sum_{q=1}^{Q} \bar{v}_q \cdot P_I(q)\gamma_{t,k}^{[1]}(\mathcal{I}^{[1]}(q))\gamma_{t,k}^{[2]}(\mathcal{I}^{[2]}(q)) \,. \quad (4)$$

Again, the normalization constant C_3 ensures that $C_3 \sum_{q=1}^{Q} P_I(q)\gamma_{t,k}^{[1]}(\mathcal{I}^{[1]}(q))\gamma_{t,k}^{[2]}(\mathcal{I}^{[2]}(q)) = 1$. If a packet loss occurs, i.e., only description j is available at the receiver, the estimator using the central code book is still utilized, with $\gamma_{t,k}^{[1-j]}(\ell) = \frac{1}{B^{[1-j]}}$. If both descriptions are lost, the estimator selects the most probable central quantizer index according to its distribution $P_I(q)$. It has been found that this estimator leads to better results in erasure situations than the estimator utilizing the side code books. If both descriptions are lost, the mean of the central code book is automatically selected using this approach.

It has been found in, e.g., [5], that tremendous gains are achievable if *a priori* knowledge of first order (AK1) is exploited by the estimator. In this contribution, we assume that the source encoder removes all intra-frame redundancies, i.e., $P_{I_k|I_{k-1}}(i_{t,k}|i_{t,k-1}) = P_I(i_{t,k})$, but there is still exploitable inter-frame correlation $P_{I_t|I_{t-1}}(i_{t,k}|i_{t-1,k})$. The estimator is given by

$$\hat{u}_{t,k}^{[\text{AK1}]} = C_4 \sum_{q=1}^{Q} \bar{v}_q \cdot \alpha_{t,k}(q) \quad (5)$$

with

$$\alpha_{t,k}(q) = \gamma_{t,k}^{[1]}(\mathcal{I}^{[1]}(q))\gamma_{t,k}^{[2]}(\mathcal{I}^{[2]}(q)) \sum_{\tilde{q}=1}^{Q} \alpha_{t-1,k}(\tilde{q}) P_{I_t|I_{t-1}}(q|\tilde{q}) \,. \quad (6)$$

ISBN 978-3-8007-3300-2 © VDE VERLAG GMBH · Berlin · Offenbach

The normalization constant C_4 in (5) ensures that $C_4 \sum_{q=1}^{Q} \alpha_{t,k}(q) = 1$. The variable $\alpha_{t,k}(q)$ (which corresponds to the *a posteriori* probability of quantizer index q, see [5]) is obtained by the forward recursion (6) with the initialization $\alpha_{0,k}(q) = P_I(q)$. Again, in the case of a packet loss of description j, the corresponding values $\gamma_{t,k}^{[j]}$ are assumed to be equiprobable. In this case, if both descriptions are lost, then although no information from the channel is available ($\gamma_{t,k}^{[1]}(q) = \frac{1}{B^{[1]}}$, $\gamma_{t,k}^{[2]}(q) = \frac{1}{B^{[2]}}$), the information from the preceding frames can be exploited and the estimator is given by

$$\hat{u}_{t,k}^{[\mathrm{AK1}]}\Big|_{\substack{\text{both}\\ \text{lost}}} = C_5 \sum_{q=1}^{Q} \bar{v}_q \cdot \sum_{\tilde{q}=1}^{Q} \alpha_{t-1,k}(\tilde{q}) P_{I_t|I_{t-1}}(q|\tilde{q}) \,.$$

The factor $\alpha_{t,k}(q)$ is updated using (6) with constant $\gamma_{t,k}^{[1]}(q)$ and $\gamma_{t,k}^{[2]}(q)$. As the estimators are similar to those in [5], extrinsic information for the use in an iterative, Turbo-like receiver can easily be generated using the approaches in [6, 7]. The iterative decoding of multiple descriptions using this receiver is however not in the scope of this paper.

4 Simulation Examples

The capabilities of the proposed receiver shall be examined by a simulation example. We use the following system setup: A source emits $K_S = 10$ i.i.d. parameters (Gaussian distribution) which are correlated over time with correlation coefficient $\rho = 0.9$. The parameters are said to possess inter-frame correlation, i.e., $P_{I_t|I_{t-1}}(i_{t,k}|i_{t-1,k}) \neq P_{I_t}(i_{t,k})$, $\forall k \in \{1;\ldots;K_S\}$. The correlation is modelled as inter-frame Markov source of first order. This block of parameters is quantized in this example using a $Q = 22$ level scalar quantizer. The multiple description index assignment is based on a 8×8 matrix with a linear 3-diagonal index assignment according to Figure 2 resulting in 6 bit per parameter (3 gray mapped bits per description). The central and side code books are generated according to the guidelines in [1]. No channel coding is used in this first example.

Example: If a parameter is quantized to the central quantizer index $i_{t,k} = 14$, then the bit pattern $\bar{\mathbf{b}}_5^{[1]} = (1\,1\,0)$ is selected for the first description while the bit pattern $\bar{\mathbf{b}}_6^{[2]} = (1\,1\,1)$ is selected for the second description.

Figure 3 depicts the parameter SNR between the original parameters u and the reconstructed parameters $\hat{u}$. It can be seen that the cross decoding method presented in [8] already considerably improves the conventional hard bit table lookup decoding. The proposed SDSD-based estimation method exploiting either no *a priori* knowledge (NAK), 0th order *a priori* knowledge (AK0, parameter distribution) or 1st order knowledge (AK1, parameter inter-frame correlation) leads to significant improvements. In the case of a packet loss probability of $\epsilon = 0.05$, it can be seen that the application of the AK1 MMSE estimator can also improve the overall signal quality in the noiseless case (for $E_s/N_0 \to \infty$) ≈ 1.5 dB. This effect is due to the inter-frame correlation of the parameters within a frame which is exploited by the AK1 MMSE estimator.

In a second experiment, $K_S = 30$ i.i.d parameters which are inter-frame correlated with correlation coefficient $\rho = 0.8$ are grouped in a block. The same multiple description index assignment as above (see Fig. 2) is utilized. Additionally, a rate $1/2$ convolutional code of constraint length 7 with (octal) generator polynomials $(133, 171)_8$ is employed (e.g. [12]). The convolutional code is decoded at the receiver using a LogMAP decoder [11]. The simulation results are depicted in Fig. 4 for a packet loss rate of $\epsilon = 0.05$. In order to account for the additional channel coding, E_b/N_0 instead of E_s/N_0 is utilized, with $E_b/N_0 = E_s/N_0 + 3.01$ dB in case of rate $1/2$ channel coding. It can be seen that all proposed MMSE estimators outperform the hard decision decoder and the cross-decoding approach of [8]. Again, the AK1-MMSE algorithm, which exploits the inter-frame correlation of the parameters, shows a superior quality than all other algorithms for all channel qualities. This is again due to the better estimation of parameters in the packet loss case.

Finally, in a third experiment, the gains by AK1 MMSE estimation in a pure packet loss scenario (i.e. no AWGN or $E_s/N_0 \to \infty$, respectively) are quantified. Again, a block consists of $K_S = 10$ i.i.d parameters which possess inter-frame correlation. It has already been shown in Figs. 3 and 4 that in a pure packet-loss case, the AK1 MMSE estimator can improve the overall reconstruction quality by

$\bar{\mathbf{b}}_1^{[1]}$ \ $\bar{\mathbf{b}}_8^{[2]}$	000	001	011	010	110	111	101	100
000	1	2						
001	3	4	6					
011		5	7	8				
010			9	10	12			
110				11	13	14		
111					15	16	18	
101						17	19	20
100							21	22

Figure 2: Multiple description index assignment and bit mappings used in the simulation example.

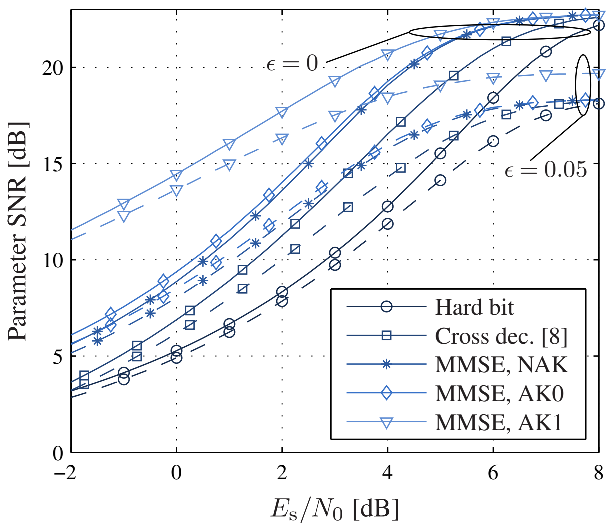

Figure 3: Comparison of the different decoding algorithms without channel coding for $\epsilon = 0$ (solid lines, ——) and for $\epsilon = 0.05$ (dashed lines, – – –), Correlation $\rho = 0.9$

ISBN 978-3-8007-3300-2 © VDE VERLAG GMBH · Berlin · Offenbach

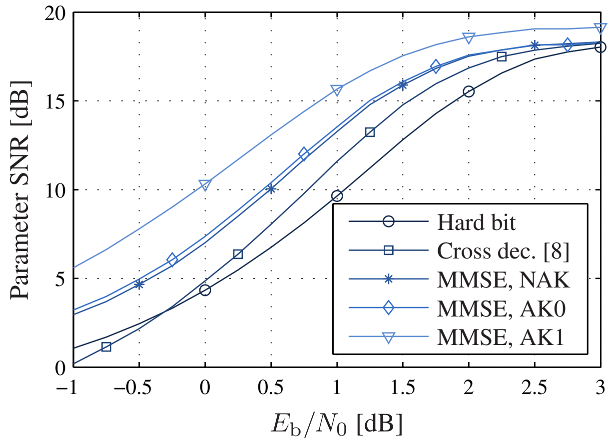

Figure 4: Comparison of the different decoding algorithms with channel coding (convolutional code) for $\epsilon = 0.05$ with correlation $\rho = 0.8$.

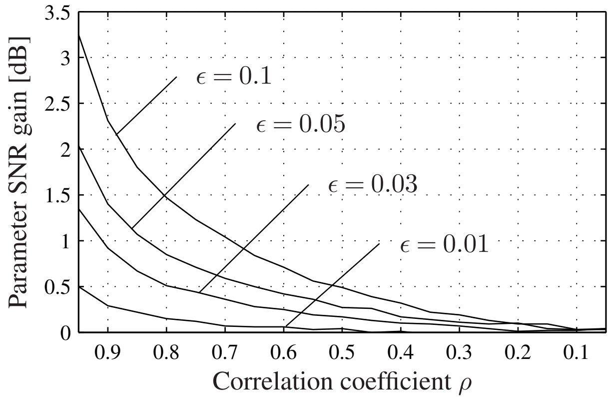

Figure 5: Parameter SNR gain of AK1 MMSE estimation compared to conventional hard bit decoding if no AWGN is present on the channel (no bit errors) and for packet loss probabilities of $\epsilon \in \{0.01; 0.03; 0.05; 0.1\}$.

exploiting the inter-frame correlation of the source codec parameters. Figure 5 depicts the achievable gain in terms of parameter SNR by employing AK1 MMSE decoding instead of conventional hard bit decoding (or any other MMSE estimator not exploiting the inter-frame correlation).

It can be seen that the improvement of the parameter SNR increases with the correlation coefficient, as can be expected. Furthermore, the gain is larger for higher packet loss rates. If no packet losses occur, the decoding result of the AK1-MMSE algorithm is identical to the result of the hard bit decoding algorithm (for the non-AWGN case). If only a small number of packet losses occur, there are only some parameters whose quality can be improved by AK1-MMSE decoding. Thus, a higher number of packet losses increases the occurrence of those situations where an improvement is obtained.

5 Conclusions

In this contribution, we have shown how conditional MMSE estimation of multiple descriptions, exploiting different amounts of residual redundancy, can lead to significant quality improvements in the case of AWGN channel noise. Additionally, we have shown that if packet losses occur and inter-frame correlation is present, the overall signal quality can be ameliorated by taking into account this correlation between the parameters of consecutive frames within the estimator. It has been shown that this gain becomes larger with increasing packet loss rates and depends, as expected, on the correlation of the parameters. Furthermore, we have shown that the proposed estimators can also successfully be applied to a system employing channel coding. An example using a convolutional code has been presented, although different channel codes, such as LDPC codes [13] can be utilized.

References

[1] V. A. Vaishampayan, "Design of Multiple Description Scalar Quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 821–834, May 1993.

[2] V. K. Goyal, "Multiple Description Coding: Compression Meets the Network," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 74–93, Sept. 2001.

[3] S. Bruhn, V. Grancharov, W. B. Kleijn, J. Klejsa, M. Li, J. Plasberg, H. Pobloth, S. Ragot, and A. Vasilache, "The FlexCode Speech and Audio Coding Approach," in *ITG Conference Speech Communication*, Aachen, Germany, Oct. 2008.

[4] FlexCode Website, http://www.flexcode.eu.

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

[6] M. Adrat, P. Vary, and J. Spittka, "Iterative Source-Channel Decoder Using Extrinsic Information from Softbit-Source Decoding," in *IEEE ICASSP*, Salt Lake City, USA, May 2001.

[7] N. Görtz, "On the Iterative Approximation of Optimal Joint Source-Channel Decoding," *IEEE J. Sel. Areas Commun.*, vol. 19, pp. 1662–1670, Sept. 2001.

[8] J. Barros, J. Hagenauer, and N. Görtz, "Turbo Cross Decoding of Multiple Descriptions," in *International Conference on Communications (ICC)*, Apr. 2002, pp. 1398–1402.

[9] T. Guionnet, C. Guillemot, and E. Fabre, "Soft Decoding of Mulitple Descriptions," in *International Conference on Multimedia and Expo (ICME)*, Lausanne, CH, Aug. 2002.

[10] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.

[11] J. Hagenauer, E. Offer, and L. Papke, "Iterative Decoding of Binary Block and Convolutional Codes," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 429–445, Mar. 1996.

[12] T. K. Moon, *Error Correction Coding - Mathematical Methods and Algorithms*. John Wiley & Sons, Inc., 2005.

[13] D. J. C. MacKay, "Good Error-Correcting Codes Based on Very Sparse Matrices," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.