Performance of Combined Error Correction and Error Detection for very Short Block Length Codes

Matthias Breuninger and Joachim Speidel
Institute of Telecommunications, University of Stuttgart
Pfaffenwaldring 47, 70569 Stuttgart, Germany
E-mail: {breuninger, speidel}@inue.uni-stuttgart.de

***Abstract*—A combination of error correction and error detection for channel coding under additive white Gaussian noise (AWGN) is proposed, where the error detection is based on the statistics of a metric, that is defined as a function of the received and the decoded data stream, hence, no additional redundancy is required. This approach is especially useful for applications, where short block lengths are applied, for which the channel decoding normally does not perform very well for a reasonable amount of redundancy. Simulations show that the frame error rate (FER) can be reduced significantly at the cost of a higher retransmission request rate (RRR), where both FER and RRR are subject to optimization depending on the specific application.**

I. Introduction

In factory communications systems many sensors and actuators have to send and receive digital messages. To make efficient use of the available resources they are carried out more and more as bus systems, where a group of sensors and actuators shares the same communications medium, i.e. the same transmission cable or wireless frequency band. The message blocks often carry sensitive data for that a very low undetected frame error rate (FER) is required to avoid misinterpretations of the received signals, which could harm the machinery or even humans. In addition the signal distortions in a factorial environment can be very high, so the system has to be able to operate at a low signal-to-noise ratio (SNR). Both, the very low required FER and the possibility of high interferences, lead to the demand of well performing error protection algorithms. The basic nature of these algorithms is to add redundancy at the transmitter and to exploit it at the receiver side. For this purpose a cyclic redundancy check (CRC) can be used to detect transmission errors, but if no forward error correction (FEC) code is used then the retransmission request rate (RRR) is quite high even for moderate SNR values. Hence, a FEC code is necessary to drop the tolerable RRR.

The message blocks often contain only a few binary information signals, so the lengths of these message blocks are very short. This makes the realization of an efficient error protection difficult for two reasons. First, for short block lengths FEC codes don't perform very well so the overhead relative to the transmission block length has to be chosen pretty high in order to achieve low FER for low SNR values. Second, the channel decoder may find a valid codeword different from the transmitted one, which happens especially for short block lengths quite often at low SNR values. Therefore, error detection is additionally required, which increases the relative overhead as well if conventional CRC algorithms are used as in [1]. To avoid additional redundancy in [2] a special design of relatively short Low-Density Parity-Check (LDPC) codes is proposed in a way that the decoder either finds the correct codeword or produces a decoding failure, which is typically a property of good LDPC codes of large block lengths. In [3] the same author presents solutions to process CRC codes of similar lengths with message-passing-decoding. This is a quite complex iterative soft-in decoder normally used for LDPC codes. Again, error detection is based on a code design, for which uncorrectable error patterns lead to decoding failures. Other publications focus on long codes, where the error detection is intrinsically included already. However, in industrial communications systems we are interested in very short block lengths, so the codes have only small minimum distances and therefore a high undetected error rate.

In this paper we propose a combination of error correction and error detection. Short BCH codes serve for the purpose of error correction and the error detection is based on the evaluation of the statistics of a metric that is defined as a function of both the received and decoded data stream. Hence, no additional redundancy is introduced. It is shown that the metric has to be chosen appropriately, so that its statistical properties differ significantly for correct and incorrect error correction. Therefore, depending on the FEC code, the modulation scheme, the desired residual FER and the tolerable RRR a certain upper limit for the metric can be obtained evaluating simulations. If the decoded signal yields a higher metric value than this predefined limit then it is marked to possibly contain errors. How many errors may remain undetected and how many correctly decoded words may be marked as to contain errors, and hence have to be retransmitted and contribute to the effective RRR, is subject to optimization. To find these optimal settings the cumulative density function (CDF) of the metric values is measured for both correct and incorrect error correction independently. From this the ratio of correct frames that have to be retransmitted, and the ratio of remaining frame errors can be obtained depending on the specified metric limit. Two different metric functions are proposed that can be exploited using a constant or a variable metric limit evaluation, depending on the specifically desired performance of the entire error correction and detection algorithm. Simulations show

978-1-4673-0158-9/12/$31.00 ©2012 IEEE

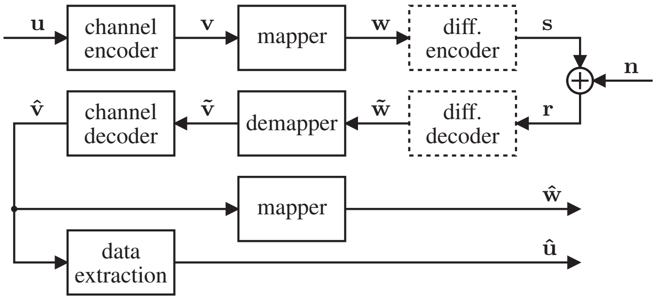

Fig. 1. Block diagram of the system model

that the FER can be reduced significantly for an appropriately set metric limit, however, the RRR is increased at the same time.

II. System model

Fig. 1 shows the basic block diagram of the system model. The binary information data sequence $\mathbf{u}$ of length K at the transmitter side is encoded to the bit sequence $\mathbf{v}$ of length N using a systematic $(N, K, d_{\min})$ linear block code with minimum distance $d_{\min}$. With M-QAM or M-PSK the coded bits are then mapped into the symbol vector $\mathbf{w}$ of length

$$S = \left\lceil \frac{N}{\log_2(M)} \right\rceil . \tag{1}$$

For coherent modulation the transmitted signal is $\mathbf{s} = \mathbf{w}$, which is corrupted by a complex additive white Gaussian noise (AWGN) vector $\mathbf{n} = \mathbf{n}_I + j\mathbf{n}_Q$. The received symbol sequences $\mathbf{r}$ and $\tilde{\mathbf{w}}$ are then given by

$$\mathbf{r}(k) = \mathbf{s}(k) + \mathbf{n}(k) \tag{2}$$

and

$$\tilde{\mathbf{w}}(k) = \mathbf{w}(k) + \mathbf{n}(k), \tag{3}$$

respectively. For $k \in \{1, \ldots, S\}$ $\mathbf{n}_I(k)$ and $\mathbf{n}_Q(k)$ are statistically independent and normally distributed with zero mean and variance σ^2. The noise power of $\mathbf{n}(k)$ is $N_0 = 2\sigma^2$ and the signal power is normalized to $E_S = 1$, hence, the SNR γ is defined as

$$\gamma = \frac{1}{2\sigma^2}. \tag{4}$$

In case of differential M-PSK modulation the differential encoder inserts an arbitrary element of the M-PSK symbol alphabet as a reference symbol $\mathbf{s}(0)$ and carries out the differential encoding for the S symbols by

$$\mathbf{s}(k) = \mathbf{w}(k)\mathbf{s}(k-1). \tag{5}$$

At the receiver side the differential decoder finds the S received noisy symbols $\tilde{\mathbf{w}}(k)$ by dividing the currenctly received symbol by the previous one according to

$$\begin{aligned} \tilde{\mathbf{w}}(k) &= \frac{\mathbf{r}(k)}{\mathbf{r}(k-1)} \\ &= \frac{\mathbf{w}(k)\mathbf{s}(k-1) + \mathbf{n}(k)}{\mathbf{s}(k-1) + \mathbf{n}(k-1)}. \end{aligned} \tag{6}$$

For both coherent and differential modulation $\tilde{\mathbf{w}}$ is demapped into the bit sequence $\tilde{\mathbf{v}}$. The channel decoder computes the estimated codeword $\hat{\mathbf{v}}$ including parity bits, which is then mapped into the estimated symbol vector $\hat{\mathbf{w}}$. The estimated information data sequence $\hat{\mathbf{u}}$ is extracted from the decoded codeword $\hat{\mathbf{v}}$ by omitting the parity bits.

A. Metric definitions

Based on this model a metric m can be defined as a function of the received noisy symbol sequence $\tilde{\mathbf{w}}$ and the estimated symbol sequence $\hat{\mathbf{w}}$, i.e.

$$m = \mathrm{f}(\tilde{\mathbf{w}}, \hat{\mathbf{w}}). \tag{7}$$

As in Generalized Minimum Distance (GMD) decoding [4] and Chase decoding [5] the squared Euclidean distance (SED) can be used to calculate such a metric, which is defined as

$$m_{\text{SED}} = \sum_{k=1}^{S} |\tilde{\mathbf{w}}(k) - \hat{\mathbf{w}}(k)|^2 . \tag{8}$$

Alternatively, for M-PSK mapping the complex valued signals can be processed using only their phase, hence, the magnitudes are not relevant and the computation of the metric can be approximated using only phase informations as well. Then calculating the SED between two points in the complex plane converts to the problem of calulating the squared circular distance (SCD), which refers to the length of the circle segment. With $E_S = 1$ the SCD metric is defined as

$$m_{\text{SCD}} = \sum_{k=1}^{S} \arg^2\left(\frac{\tilde{\mathbf{w}}(k)}{\hat{\mathbf{w}}(k)}\right) \tag{9}$$

where $\arg(\cdot) \in (-\pi, \pi]$ denotes the phase of the operand.

B. Densities of the metrics

The metrics defined above are now studied for both correct and incorrect decoding, where incorrect decoding denotes the case that the channel decoder determined a valid codeword which is different from the transmitted one. Cases where the channel decoder is not able to find a valid codeword at all are not considered for now.

For coherent modulation and correct decoding, i.e. $\hat{\mathbf{v}} = \mathbf{v}$ and thus $\hat{\mathbf{w}} = \mathbf{w}$, the SED metric given in (8) can be written in the form

$$m_{\text{coh,SED}}^{\text{correct}} = \sum_{k=1}^{S} \mathbf{n}_I^2(k) + \sum_{k=1}^{S} \mathbf{n}_Q^2(k). \tag{10}$$

For $\sigma^2 = 1$ and $k \in \{1, \ldots, S\}$ the elements $\mathbf{n}_I(k)$ and $\mathbf{n}_Q(k)$ form S independent standard normal variables each, therefore both sums of the squares follow a χ^2-distribution with S degrees of freedom. Since the sum of two independent χ^2-distributed random variables is again χ^2-distributed, for which the degrees of freedoms are summed up, $m_{\text{coh,SED}}^{\text{correct}}$ follows a χ^2-distribution with $2S$ degrees of freedom and the probability density function (PDF)

$$p_1(m) = \frac{m^{S-1}}{2^S\Gamma(S)} \mathrm{e}^{-\frac{m}{2}} \ , \ m \geq 0, \tag{11}$$

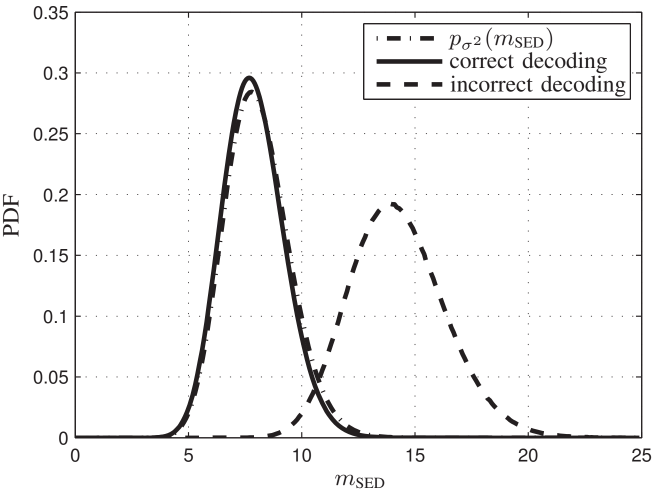

Fig. 2. PDF of m_{SED} for correct and incorrect decoding at $\gamma = 6\,\mathrm{dB}$

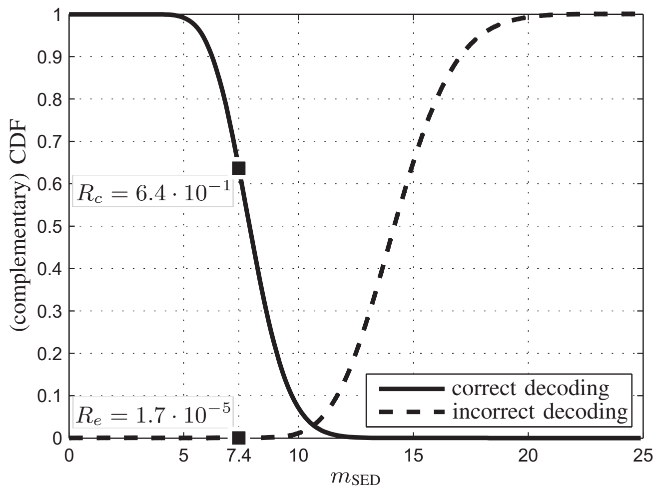

Fig. 3. CCDF of m_{SED} for correct decoding and CDF of m_{SED} for incorrect decoding at $\gamma = 6\,\mathrm{dB}$

where $\Gamma(\cdot)$ denotes the Gamma function. Then the PDF of $m^{\mathrm{correct}}_{\mathrm{coh,SED}}$ can be derived for the general case $\sigma^2 \neq 1$ by

$$\begin{aligned} p_{\sigma^2}(m) &= \frac{1}{\sigma^2} p_1\left(\frac{m}{\sigma^2}\right) \\ &= \frac{m^{S-1}}{(2\sigma^2)^S\,\Gamma(S)} \mathrm{e}^{-\frac{m}{2\sigma^2}}\ ,\ m \geq 0. \end{aligned} \tag{12}$$

For this PDF it is assumed that the decoder always finds the correct estimated codeword $\hat{\mathbf{v}} = \mathbf{v}$. However, due to high instant noise samples $\mathbf{n}(k)$ the decoder may find a codeword that is different from the transmitted one, then the corresponding metric value $m_{\mathrm{coh,SED}}$ does not contribute to the PDF $p_{\sigma^2}(m)$, which refers only to the case of correct decoding. The higher the noise variance σ^2 of the noise samples $\mathbf{n}_I(k)$ and $\mathbf{n}_Q(k)$, the smaller is the SNR γ and the more often the decoder is not able to find the transmitted codeword. Consequently, the PDF $p_{\sigma^2}(m)$ can be used as an approximation for high SNR values only.

As an example $p_{\sigma^2}(m_{\mathrm{SED}})$ according to (12) and the simulated PDF of m_{SED} are shown in Fig. 2, where the $(63, 45, 7)$-BCH code has been used with coherent 4-PSK modulation and hard decision Berlekamp-Massey decoding under AWGN for $\gamma = 6\,\mathrm{dB}$. It can be seen that the calculated PDF for correct decoding differs slightly from the simulated one, and that the densities differ significantly for correct and incorrect decoding. Hence, the calculated metric can be used as a measure for the signal quality, based on which the channel decoder can estimate if the decoded codeword $\hat{\mathbf{v}}$ is equal to the transmitted codeword $\mathbf{v}$. A straightforward approach for this estimation is to compare the calculated metric by a predefined limit $m_{\max}$ and to declare a decoded codeword as valid only if $m_{\mathrm{SED}} \leq m_{\max}$. With this procedure the undetected FER can be reduced at the cost of a higher retransmission rate RRR.

To evaluate the FER and RRR for a given metric limit $m_{\max}$ it is convenient to examine the cumulative density function (CDF) of $m^{\mathrm{incorrect}}_{\mathrm{SED}}$ and the complementary CDF (CCDF) of $m^{\mathrm{correct}}_{\mathrm{SED}}$ as shown in Fig 3. If for $\gamma = 6\,\mathrm{dB}$ the metric limit is set to $m_{\max} = 7.4$ then the ratio R_c of correct codewords that have to be retransmitted is $R_c = 6.4 \cdot 10^{-1}$, while the ratio R_e of remaining erroneous codewords is only $R_e = 1.7 \cdot 10^{-5}$, so the resulting residual FER is reduced by the factor R_e.

A first approach to detect errors is to set the metric limit dependent on the current SNR so that R_c is constant. Then R_e gets smaller for increasing SNR values, however, only at the cost that RRR approaches R_c rather than zero. If an application allows a certain retransmission rate and undetected error rate then the constant value R_c can be optimized such that both RRR and FER meet their requirements for a given minimal SNR value. Another possibility is to specify a constant upper metric limit $m_{\max}$ regardless of the SNR, which leads to an SNR dependent value for R_c. Simulations show that for low SNR values this outperforms the approach where R_c is kept constant, but for high SNR values the RRR converges to zero only at the expense of a worse FER. Results of both options will be shown in the next section in more detail.

III. Simulation Results

Some simulation results are presented for both coherent and differential modulation using the SED and SCD metrics for various FEC channel codes of very short block lengths. The metric limits have been designed for the target performances $\mathrm{FER} = 10^{-8}$ and $\mathrm{RRR} = 10^{-3}$ and all simulation results are generated under AWGN disturbance using hard decision Berlekamp-Massey decoding [6].

For both variable and constant metric limit evaluation the FER and RRR over SNR are displayed in Fig. 4 for the $(31, 16, 7)$-BCH code with coherent 4-PSK modulation. The FER without metric evaluation is shown as well, which refers to the FER at the channel decoder output, where these words are not taken into account anymore, for which the channel decoder could not find a valid codeword, since then the codeword contains an uncorrectable error pattern and the evaluation of the metric is not necessary. For the variable

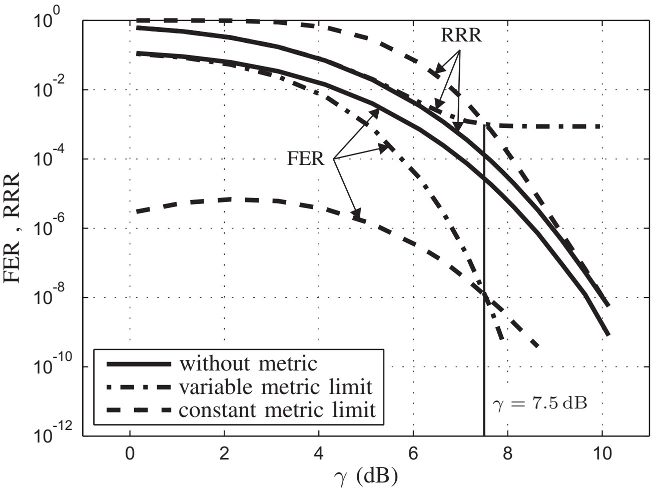

Fig. 4. Frame error rate (FER) and retransmission request rate (RRR) over SNR of the 4-PSK $(31,16,7)$-BCH code with evaluation of the SED metric

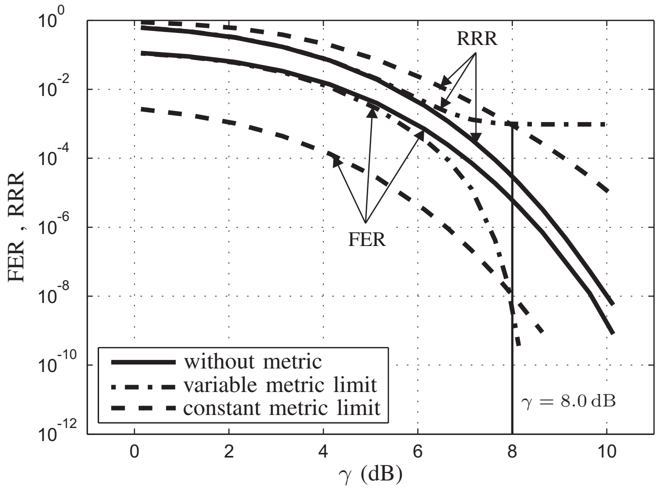

Fig. 5. Frame error rate (FER) and retransmission request rate (RRR) over SNR of the 4-PSK $(31,16,7)$-BCH code with evaluation of the SCD metric

metric limit simulations show that both the target FER and RRR can be achieved for $R_c = 8.6 \cdot 10^{-4}$ at the SNR threshold $\gamma = 7.5\,\text{dB}$, so compared to the case of no metric evaluation the SNR gain for achieving the target $\text{FER} = 10^{-8}$ is $2.2\,\text{dB}$. At this SNR threshold the variable metric limit is $m_{\max} = 5.5$, which is used for the constant metric limit curve. It can be seen that for low SNR values the undetected FER is much lower for the constant than for the variable metric limit at the expense of a higher RRR. However, beyond the threshold the variable metric limit gives a lower FER, again at the cost of a higher RRR, which approaches R_c.

Simulations with a $(32,16)$-CRC under coherent 4-PSK modulation show that it performs worse than the $(31,16,7)$-BCH code along with constant SED metric limit evaluation. For $\gamma = 7.5\,\text{dB}$ the CRC exhibits $\text{FER} = 9.0 \cdot 10^{-8}$ and $\text{RRR} = 2.4 \cdot 10^{-1}$, so using a BCH code and exploiting the statistical properties of the SED metric outperforms the conventional approach using a CRC code, while the amount of required redundancy is comparable and the RRR is far less due to the error correcting capability of the BCH code.

If now the SCD metric is used for the same channel code and modulation then the required SNR for achieving the target performances increases to $\gamma = 8.0\,\text{dB}$, so the SNR gain for the undetected error rate reduces to $1.7\,\text{dB}$, which can be seen in Fig. 5. At the same time the metric limit enhances to $m_{\max} = 5.9$ and the ratio of correct codewords that have to be retransmitted changes to $R_c = 9.6 \cdot 10^{-4}$. The performance of its RRR shows a significant difference compared to the SED metric, since the curve without metric evaluation is not approached anymore for high SNR values. Also, while for the constant SED metric limit the remaining errors are very low even for low SNR values, this does not hold for the constant SCD metric limit, where the FER monotonically approaches a constant upper bound for decreasing SNR. The reason for this behaviour is that the SCD metric is an approximation only that doesn't consider symbol magnitudes but only phases, so the SCD metric values have an upper bound given by

$$m_{\text{SCD}} \le \sum_{k=1}^{S} \pi^2. \tag{13}$$

Hence, for low SNR values m_{SCD} is saturated and does not increase anymore with increasing noise power, so its performance approaches a constant value for decreasing SNR.

Now the $(63,45,7)$-BCH code is considered, which has the same minimum distance but exhibits longer blocks and therefore has a higher code rate $R = \frac{k}{n}$. Again the metric limit curves are designed for the same target $\text{FER} = 10^{-8}$ and $\text{RRR} = 10^{-3}$. First the SED metric is used for 4-PSK mapping, for which the results for the simulated FER and RRR are displayed in Fig. 6 with and without constant metric limit evaluation. Here the performance for the variable metric limit evaluation is not shown, since it is only required for finding the metric limit $m_{\max}$, if one is interested in a low FER even for low SNR values. The required SNR is now $\gamma = 8.5\,\text{dB}$, which is $1.8\,\text{dB}$ less than for the case without metric evaluation and can be achieved for $m_{\max} = 7.4$.

The gain of the FER for the constant metric limit can be illustrated using the CDF for incorrect decoding in Fig. 3 at $\gamma = 6\,\text{dB}$ and the same metric limit $m_{\max} = 7.4$, which yields $R_e = 1.7 \cdot 10^{-5}$. Hence, the gap between the FER with and without the constant metric limit evaluation in Fig. 6 at this SNR is $R_e = 1.7 \cdot 10^{-5}$, which can be easily verified therein.

For correct decoding the CCDF of m_{SED} according to (8) converges much slower to zero for differential than for coherent modulation. This is due to the division of two consecutively received symbols in the differential decoder in (6), which can lead to a relatively high magnitude for the symbol $\tilde{\mathbf{w}}(k)$ and therefore to a high metric value m_{SED} as well. Hence, for differential modulation the SCD metric is used, since it considers only phase but not magnitude distortions.

The FER for the $(63,45,7)$-BCH code with differential 4-DPSK mapping is shown in Fig. 7. The target performances

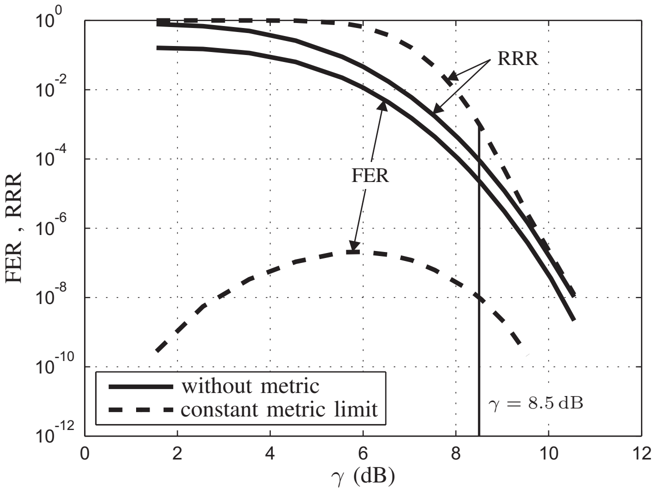

Fig. 6. Frame error rate (FER) and retransmission request rate (RRR) over SNR of the 4-PSK $(63, 45, 7)$-BCH code with evaluation of the SED metric

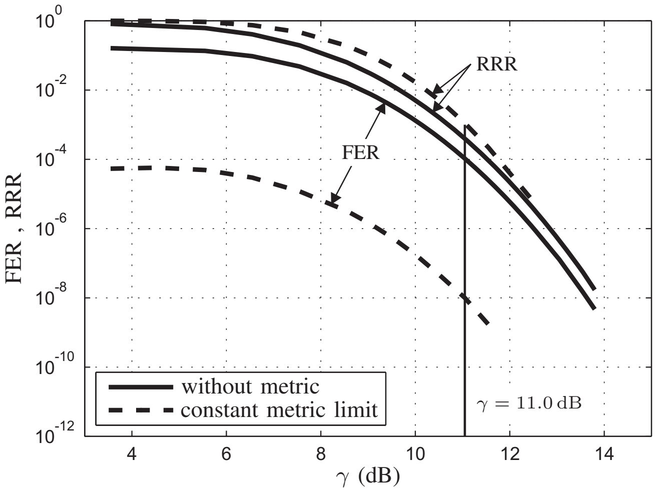

Fig. 7. Frame error rate (FER) and retransmission request rate (RRR) over SNR of the 4-DPSK $(63, 45, 7)$-BCH code with evaluation of the SCD metric

can now be achieved for $\gamma = 11.0\,\mathrm{dB}$, which is $2.5\,\mathrm{dB}$ larger than for the coherent case with the SED metric and $2.6\,\mathrm{dB}$ less than for the case without metric evaluation. For constant SCD metric limit evaluation the effect of a saturating FER for low SNR values can again be observed.

If the above shown BCH codes are processed with more powerful decoding algorithms such as the soft-in GMD or Chase decoding then the error correcting performance is enhanced but the capability of additional error detection is reduced at the same time. In general, if a maximum likelihood (ML) decoder is applied then there is no intrinsic error detection anymore, since such a decoder never results in a decoding failure, as already mentioned in [2], where only the intrinsic error detection is exploited. In contrast to that, the approach of calculating and evaluating a metric based on the received and decoded symbol sequence can still detect errors, since a codeword will only be accepted if its distance to the received word doesn't exceed the metric limit, which can still happen for a ML decoder. However, finding the ML codeword will also lead to a minimal metric value, therefore the error detection capability is reduced for better decoding algorithms. In order to degrade the error correcting performance of the Berlekamp-Massey algorithm used in the above results it can be modified to just result in a decoding success if only $t < \lfloor \frac{d_{\min}}{2} \rfloor$ bit errors were corrected [3]. So the undetected error rate can be reduced not only by choosing a low metric limit but by using a worse decoder as well, which both will increase the RRR, of course.

IV. Conclusion

To achieve a good trade-off between error detection capability and remaining data throughput we introduced the calculation and evaluation of a metric, which is based on the received and decoded data and whose statistical properties exhibit significant differences for correct and incorrect decoding. So decoded sequences can be marked to contain errors if their corresponding metrics exceed a certain limit, since then the result is considered not to be reliable enough. Using this approach redundancy is only required for the channel decoder. For the error detection no additional redundancy is introduced, which is especially useful for short block lengths. The performance of the entire system can be optimized so that the desired FER and RRR targets are both met for the same minimal SNR. Simulations show that for coherent and differential modulation the minimal SNR can be achieved if the SED and the SCD metric are applied, respectively. For low SNR values the undetected error rate for the SED metric is less than for the SCD metric. However, if the metric limit is reduced or a decoder with a worse error correcting capability is applied, then the undetected FER reduces for all SNR values at the cost of a higher RRR. So the system is very flexible and its performance can be designed according to the specific application.

Acknowledgment

This work was funded by the German Federal Ministry of Economics and Technology (BMWi 16IN0672).

References

[1] T. Mattes and F. Schiller, "Determination of the quality of error detection in communication by CRC concatenated with parity bits," in *IEEE Conference on Emerging Technology and Factory Automation*, September 2010, pp. 1 – 7.

[2] Q. Spencer, "Short-block LDPC codes for a low-frequency power-line communications system," in *IEEE International Symposium on Power Line Communications and Its Applications*, April 2005, pp. 95 – 99.

[3] ——, "Cth14-6: Error correction using a message-passing decoder to process cyclic redundancy checks," in *IEEE Global Telecommunications Conference (GLOBECOM)*, December 2006, pp. 1 – 5.

[4] J. Forney, G., "Generalized minimum distance decoding," *IEEE Transactions on Information Theory*, vol. 12, no. 2, pp. 125 – 131, April 1966.

[5] D. Chase, "Class of algorithms for decoding block codes with channel measurement information," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 170 – 182, January 1972.

[6] S. Lin and D. Costello, *Error control coding: fundamentals and applications*. Englewood Cliffs, NJ: Prentice-Hall, 1983.