Schalkwijk and Kailath Feedback Coding with Sequential Decision-Making

Christoph Jans, Meik Dörpinghaus, and Gerhard Fettweis

Vodafone Chair Mobile Communications Systems, SFB 912 HAEC, Technische Universität Dresden, Germany

Email: {christoph.jans, meik.doerpinghaus, gerhard.fettweis}@tu-dresden.de

***Abstract*—Low latency and high energy-efficiency will be key features of many future communication systems. It is known that feedback coding like the well known Schalkwijk and Kailath (SK) coding scheme outperforms feedforward coding in terms of the decoding error probability. In the present paper, the additional gain by combining the capacity-achieving feedback coding scheme by SK with sequential decision-making based on the multiple hypotheses test by Baum et al. is studied. An analytical approximation for the minimum E_b/N_0 to transmit k bits is derived under a given constraint on the decoding error probability. The theoretical results are compared to Monte Carlo simulations. With this approach a significant reduction in blocklength or latency and a further reduction of the required E_b/N_0 w.r.t. the original SK scheme can be achieved. For evaluation the results are compared to the analytical results of Polyanskiy et al. for the additive white Gaussian noise channel with and without feedback.**

I. Introduction

In future communication systems low latency and high energy-efficiency will be key. Applications in the area of cyber-physical systems, e.g., in remote health care, demand latencies below the reaction time of a human being. Moreover, in a multi-chip multi-board high performance computing cluster, researchers are aiming for replacing the wired board-to-board interconnects with wireless links [1]. To compete against wired links, ultra low latency and high energy efficiency coding schemes are required. Current communication schemes mainly adapt a feedfoward error correction (FEC) scheme to the current channel quality. With this approach they are able to increase the energy-efficiency and concurrently increase the throughput. The underlying protocol is called hybrid-ARQ, which combines FEC with automatic repeat requests (ARQ). However, hybrid-ARQ exploits only one benefit of feedback based channel coding. Namely, the feedback link is used to communicate to the transmitter when decoding was successful. Thus, it enables early stopping on favorable channel realizations. By using more extensively the feedback link, e.g., to inform the encoder about the current decoder state, various feedback coding schemes showed a significant reduction in latency, a higher decay rate of error-probability, and simplifications in the encoding and decoding process [2]–[4]. Furthermore, in [5] it is proven that feedback based channel coding outperforms any feedforward coding scheme in the non-asymptotic blocklength regime in terms of rate, latency, and energy-efficiency.

Now, the idea proposed in this paper is the combination of the well-known feedback coding scheme by Schalkwijk and Kailath (SK) [2], which uses full feedback of the channel output, with a sequential decoding strategy, i.e., combining the ideas of feeding back decoder information and using a random blocklength to stop as soon as the decoder is able to decode. Sequential decision-making allows to reduce the number of observations to decode with a given reliability. In his seminal paper Wald [6] introduced a binary probability ratio test which is optimal in the sense that it minimizes the number of observations to decide with a given probability of error. Later this test has been generalized to multiple hypotheses [7]–[9]. Hence, the proposed combination of feedback coding with sequential decision-making is promising for low latency communication and concurrently it allows to reduce the amount of transmit power in comparison to the original SK feedback coding scheme. In this paper, the performance w.r.t. the required E_b/N_0 and the required blocklength (latency) to transmit k bits is evaluated. While the studied system uses full feedback of the channel output to the transmitter, the results also give a bound on the performance gain for rate limited feedback schemes.

The rest of the paper is structured as follows. In Section II the considered system model is introduced. In Section III the basic SK coding scheme is presented, followed by an introduction into sequential decision-making in Section IV. In Section V the combination of the SK feedback coding scheme with sequential decoding is described and analyzed w.r.t. E_b/N_0 and latency. In Section VI the performance gain by combining the SK coding scheme with sequential decision-making is numerically evaluated based on the required E_b/N_0. Finally, Section VII concludes the paper.

The following notation is used in the paper. Capital letters denote random variables (e.g. Y) and lower case letters their realization (e.g. y). Analogously, bold face upper case letters (e.g. $\mathbf{Y}$) denote random vectors and bold face lower case letters (e.g. $\mathbf{y}$) denote the realization of the corresponding random vector. Moreover, a sequence, e.g., $y_1^n = (y_1, \ldots, y_n)$, is a realization of a random vector containing the random samples $\{y_k\}$ from 1 to n. The hat symbol is used to denote an estimate of a random variable. Logarithms without subscript denote natural logarithms.

II. System Model

The forward channel considered in this paper is a discrete-time additive white Gaussian noise (AWGN) channel with

$$Y_n = X_n + Z_n \tag{1}$$

where X_n, Z_n, and Y_n are the channel input, the zero-mean normal distributed noise with variance σ^2, i.e., $\mathcal{N}(0, \sigma^2)$, and the corresponding channel output at time instant n, respectively. The feedback link is modeled as a noiseless and delayless one-way link from the receiver to the transmitter without quantization or rate limitation. In general the variable length feedback (VLF) code (E, M, ϵ) for the AWGN forward channel consists of a sequence of encoding functions $\{f_n\}_{n=1}^{\infty}$ producing the channel input symbols X_n, where E is the energy constraint, M is the number of messages, and ϵ is the error constraint. The encoding function $f_n = f(W, Y_1^{n-1})$ is based on all the prior channel output symbols Y_1^{n-1} and the message W. Finally, the decoding function $g(Y_1^{n-1})$ maps the received channel output symbols to an estimate $\hat{W}$ of W, where $P[\hat{W} \neq W] \leq \epsilon$ with $\hat{W} = g(Y_1^{n-1})$.

For this specific scenario various coding strategies have been invented [2]–[4] and theoretical results on the achievable rate have been derived. Here, especially the work by Polyanskiy et al. [5], [10] needs to be mentioned, as they give theoretical results and bounds for VLF coding in the non-asymptotic regime. In [5] they studied the achievable rate for non-asymptotic blocklengths while in [10] the minimum energy to transmit k bits for an asymptotically large blocklength has been studied. These results give detailed information on the convergence of the achievable rate towards the capacity for increasing code blocklength and on the required minimum energy per bit depending on the number of information bits. This is a highly valuable reference to evaluate the performance of code designs in the short blocklength regime. The question arises, how close practical coding schemes approach these theoretical limits. A solution to get closer to the theoretical results is described in this paper.

For this purpose the fixed-blocklength capacity-achieving coding scheme by SK [2] is enhanced in this paper. The novel idea is to introduce a sequential decision-making algorithm at the decoder, which helps to further reduce the needed blocklength, as the transmission can be stopped earlier for favorable noise realizations. Before explaining the sequential decision-making algorithm, the basic concept of the SK coding scheme is summarized in the following section.

III. Schalkwijk and Kailath Feedback Coding Scheme

An intuitive explanation of the Schalkwijk and Kailath based feedback coding scheme is that the encoder tries to steer the decoder to the position of the correct codeword, such that the decoder makes a correct decision at the end of the transmission. This steering approach is realized by calculating the minimum mean square error (MMSE) estimate based on all available observations at the decoder and informing the encoder about the current estimate using the noiseless feedback link. There are two versions of the SK scheme. Both are capacity-achieving, however one fulfills for any number of feedback iterations a given bandwidth constraint [3] and for the other scheme the bandwidth scales with the number of iterations [2]. In the following no bandwidth constraint is assumed and, therefore, the presentation follows along the lines of the algorithm given in [2].

Schalkwijk and Kailath Coding Scheme: The unit interval from $(-0.5, 0.5)$ is divided into M equal-length subintervals with midpoints $(\theta_1, \dots, \theta_M)$. The midpoints represent the codewords of the feedback coding scheme. Each message out of $W \in \{1, \dots, M\}$ is mapped to one of the predefined midpoints θ_i. For encoding the following function f_n is used to generate the channel input X_n

$$f_n(W, Y_1^{n-1}) = X_n = \alpha(\hat{\theta}_n - \theta) \tag{2}$$

where θ corresponds to the codeword to be transmitted. Moreover, α can be optimally chosen based on the channel noise power, the given error constraint, and the blocklength [2]. The estimate $\hat{\theta}_n = E\left[\theta | Y_1^{n-1}\right]$ with $\hat{\theta}_1 = 0$ in (2) is calculated by the decoder and fed back to the encoder. The channel input symbols X_n are transmitted over the AWGN channel (1). An interesting remark regarding (2) is that the encoding function is only based on the current estimate as $\hat{\theta}_n$ is a sufficient statistic of all the prior channel outputs Y_1^{n-1} w.r.t. θ. Furthermore, while $\hat{\theta}_n \to \theta$ the power of the transmit symbol $E[X_n^2]$ decreases with increasing n. On the decoder side the estimate $\hat{\theta}_n$ is determined by

$$\hat{\theta}_{n+1} = E\left[\theta | Y_1^n\right] = \hat{\theta}_n - \frac{1}{\alpha n} Y_n \tag{3}$$

which can be rewritten as

$$\hat{\theta}_{n+1} = \theta - \frac{1}{\alpha n} \sum_{k=1}^{n} Z_k \tag{4}$$

where $\hat{\theta}_1 = 0$. From (4), one can easily derive that $\hat{\theta}_n$ conditioned on θ is normally distributed with $\hat{\theta}_{n+1} \sim \mathcal{N}(\theta, \frac{\sigma^2}{\alpha^2 n})$. After the predefined blocklength N the decoding function $g(Y_1^N)$ is based on the current estimate $\hat{\theta}_{N+1}$ and not explicitly on the channel output sequence Y_1^N. The decoder decides for the message $\hat{W}$ corresponding to the subinterval in which $\hat{\theta}_{N+1}$ is located. As it is known that $\hat{\theta}_{N+1}$ is normally distributed, the probability of error is bounded by

$$P\left[\hat{W} \neq W\right] = 2Q\left(\frac{\alpha\sqrt{N}}{2M\sigma}\right) \leq \epsilon. \tag{5}$$

where $Q(\cdot)$ is the tail probability of the standard normal distribution. With (5) the required blocklength N corresponding to the number of transmissions to achieve the given error probability ϵ can be calculated. Furthermore, assuming that θ is uniformly distributed over the unit interval $(-0.5, 0.5)$ with zero mean and variance $\frac{M^2-1}{12}$, the E_b/N_0 can be derived using (2)-(5):

$$\begin{aligned} \frac{E_b}{N_0} &= \frac{1}{N_0 \log_2(M)} E_{\theta,\mathbf{Z}}\left[\alpha^2\left(\hat{\theta}_1 - \theta\right)^2 + \sum_{i=2}^{N} \alpha^2 \left(\hat{\theta}_i - \theta\right)^2\right] \\ &= \frac{1}{\log_2(M)}\left(\frac{\alpha^2}{12N_0}\left(1 - \frac{1}{M^2}\right) + \frac{1}{2}\sum_{i=1}^{N-1}\frac{1}{i}\right). \end{aligned} \tag{6}$$

The input power scaling parameter α cancels out by substituting (5) with equality into (6) such that

$$\frac{E_b}{N_0} = \frac{1}{\log_2(M)}\left(\frac{1}{6N}\left(M^2 - 1\right)\left[Q^{-1}\left(\frac{\epsilon}{2}\right)\right]^2 + \frac{1}{2}H_{N-1}\right). \tag{7}$$

Here, H_{N-1} denotes the harmonic series from $i = 1$ to $N - 1$ and the E_b/N_0 is independent of α. The harmonic series in (7) can be expressed as [11, 6.3.2]

$$H_{N-1} = \sum_{i=1}^{N-1} \frac{1}{i} = \gamma + \Psi^{(0)}(N), \quad N \geq 2, \tag{8}$$

where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant and $\Psi^{(0)}(x)$ denotes the psi-function, being the n-th derivative of the digamma function. To minimize E_b/N_0 for a given ϵ one needs to take the derivative of (7) w.r.t. the blocklength N. Note that this leads to an α which differently to the choice in [2] does not maximize the rate. Substituting (8) into (7), taking the derivate w.r.t. N, and setting it equal to 0 leads to $N^2\Psi^{(1)}(N) = \frac{1}{3}\left(M^2-1\right)\left[Q^{-1}\left(\frac{\epsilon}{2}\right)\right]^2$, which can be numerically solved to get the blocklength N. One can show numerically that this blocklength N minimizes the E_b/N_0 as the second derivative is positive.

IV. Sequential Decision-Making

Sequential decision-making algorithms collect observations until they are able to stop and decide with a given reliability. Such algorithms have been studied in statistics in the context of sequential hypothesis testing.

For a binary hypothesis testing problem Wald invented the sequential probability ratio test or Wald test [6], which performs optimal in terms of minimizing the average number of observations to achieve a given decision error probability [12]. To explain the principle, a feedforward repetition coding scheme without feedback is assumed in the following. The test is based on discrete channel input symbols generated by a repetition-like encoding function $X = X_n = f_n(W) = \theta, \theta \in \{\theta_1, \theta_2\}$ for $n = 1, 2, \ldots$. A decision on the channel input symbol X should be taken based on a sequence of independent and identically distributed (i.i.d.) observations of X disturbed by AWGN. Wald proposes to accumulate the log-likelihood ratios of the binary hypotheses given by the sequence of observations and compare it against two predefined thresholds. Once one of the thresholds is reached the test decides for the corresponding hypothesis. These thresholds can be calculated solely based on the given constraints on the decision error probabilities. The accumulated log-likelihood process corresponds to a random walk and the whole test is a first passage level crossing problem with two absorbing boundaries [6].

As in the present setup the number of messages M is larger than 2, instead of a binary hypotheses test a multi hypotheses test is needed. Optimal tests on multiple hypotheses, if available at all, have larger computational complexity [9] than the binary Wald test. Hence, a lot of effort has been put on finding sub-optimal low complexity multi hypotheses tests. The works by Draglia et al. [7] and Baum et al. [8] targeted this topic and the authors came up with tests showing a good performance. However, there exist only loose bounds on the expected required number of observations. Especially, when the hypotheses are a-priori equally likely, which holds for the codewords in feedback coding, the bounds on the expected number of observations are not tight [7]–[9].

In the following, for $M \geq 2$ and the discrete channel input symbols $X = X_n = f_n(W) = \theta$ for $n = 1, 2, \ldots$ and $\theta \in \{\theta_1, \ldots, \theta_M\}$ the sequential test by Baum et al. [8] is applied. This approach works for any number of hypotheses $M \geq 2$ and reduces for $M = 2$ to the optimal Wald test. The approach by Baum et al. is based on a maximum likelihood (ML) decision given the received observations sequence $\mathbf{Y} = [Y_1, Y_2, \ldots]$. The probability of $\theta = \theta_i$ can be expressed by using Baye's rule as

$$P(\theta = \theta_i|\mathbf{Y}) = \frac{p(\mathbf{Y}|\theta = \theta_i)P(\theta = \theta_i)}{\sum_j p(\mathbf{Y}|\theta = \theta_j)P(\theta = \theta_j)} \tag{9}$$

where $p(\cdot)$ denotes a probability density function of a continuous random variable and $P(\cdot)$ denotes the probability of a discrete random variable. The probability in (9) can be further simplified if all θ_i are a-priori equally likely, i.e., $P(\theta = \theta_j) = 1/M$, $\forall j \in \{1, \ldots, M\}$, which will be assumed in the following. Now this approach is applied to the Gaussian system model (1) without feedback. The receiver collects n observations Y_1^n of the same input signal X until it stops and decides for one of the input signals. Based on the received sequence y_1^n one can write the posterior probability for each $\theta_i \in 1 \ldots M$ using (9) as

$$P\left(\theta = \theta_i|y_1^n\right) = \left(\sum_{j=1,\ldots,M} e^{\frac{\theta_j-\theta_i}{\sigma^2}\left(\sum_{k=1}^n y_k - n\frac{\theta_j+\theta_i}{2}\right)}\right)^{-1}. \tag{10}$$

Then the set of all posteriori probabilities $\{P\left(\theta = \theta_i|y_1^n\right) : i \in 1, \ldots, M\}$ is tested against a set of pre-calculated thresholds $\{A_i\}$

$$P\left(\theta = \theta_i|y_1^n\right) \geq \frac{1}{1+A_i}. \tag{11}$$

The sequential test stops at time τ with

$$\begin{aligned} \tau &= \min_i \{\tau_i\} \\ \tau_i &= \min_n \left\{ n \geq 1 : P\left(\theta = \theta_i|y_1^n\right) \geq \frac{1}{1+A_i} \right\} \end{aligned} \tag{12}$$

and decides for hypothesis θ_m, where m is

$$m = \arg\max_i P\left(\theta = \theta_i|y_1^\tau\right). \tag{13}$$

The thresholds $\{A_i\}$ can be derived from the given error constraint ϵ. Knowing that all θ_i are equally likely and the probability of an incorrect decision needs to be smaller than ϵ, the thresholds $\{A_i\}$ are the same for each hypotheses, i.e., $A_i = A$. From [8, Sect. VII.], the threshold A is given by $A = \epsilon/\gamma$, where γ can be computed as $\gamma = \frac{1}{\delta}\exp\left(-2\sum_{k=1}^{\infty}\frac{1}{k}\Phi\left(-\sqrt{\frac{\delta k}{2}}\right)\right)$, see [7], with $\Phi(\cdot)$ as cumulative distribution function of the standard normal distribution and $\delta = \min_{i\neq j} D(p(\mathbf{Y}|\theta_i)||p(\mathbf{Y}|\theta_j))$ where $D(\cdot||\cdot)$ denotes the Kullback-Leibler divergence. The corresponding expected average sample size $E[\tau]$ is given by [8, Eq. (5)-(7)]

$$E[\tau] \approx \frac{-\log A}{\delta}. \tag{14}$$

The derived $E[\tau]$ is only an approximation for the simulated sample size. The gap between simulated sample sizes and the analytically expected sample sizes $E[\tau]$ will be evaluated in the simulation results in Section VI.

V. Combination of Feedback Coding and Sequential Decision-Making

The combination of the sequential multi-hypotheses test by Baum described in Section IV and the SK coding scheme from Section III is presented in this section. The statement that the Wald test minimizes the average number of observations holds only for i.i.d. observations [12]. However, the SK coding scheme produces random variables y_n and $\hat{\theta}_n$ at the decoder which depend on the prior realizations y_{n-1} and $\hat{\theta}_{n-1}$. With the following approach an i.i.d. random variable is generated at the decoder side. Let

$$W_n = \hat{\theta}_{n+1} - \left(1 - \frac{1}{n}\right)\hat{\theta}_n \tag{15}$$

for $n = 1, 2, \ldots$, which can be calculated by the decoder without additional information. Using (2) in (1) and substitute Y_n in (3) one gets $\hat{\theta}_{n+1} = \hat{\theta}_n\left(1 - \frac{1}{n}\right) + \frac{1}{n}\theta - \frac{Z_n}{\alpha n}$ and the new random variables

$$nW_n = n\left(\hat{\theta}_{n+1} - \left(1 - \frac{1}{n}\right)\hat{\theta}_n\right) = \theta - \frac{Z_n}{\alpha} \tag{16}$$

for $n = 1, 2, \ldots$, are i.i.d. Gaussian distributed with $\mathcal{N}(\theta, \frac{\sigma^2}{\alpha^2})$. This leads to an alternative Gaussian system model based on the i.i.d. random variables nW_n, noise Z_n, and θ.

For the new system model in (16) the results on sequential decision-making from the prior section can be applied. In this regard, the sequence of nW_n defined as $U_1^n = \{1W_1, \ldots, nW_n\}$ replaces Y_1^n in (10). Moreover, U_n is Gaussian distributed with variance $\sigma_{U_n}^2 = \frac{\sigma^2}{\alpha^2}$ and mean θ, i.e. $\mathcal{N}(\theta, \frac{\sigma^2}{\alpha^2})$. For each midpoint θ_i for $i \in \{1, \ldots, M\}$ based on the sequence U_1^n the sequential test as in (12) and (13) is performed with $A = A_i = \epsilon/\gamma$. To derive the expected sample sizes, the Kullback-Leibler divergences $D\left(p\left(U_1^n|\theta_i\right)||p\left(U_1^n|\theta_j\right)\right)$ need to be calculated. For this purpose, one has to take the expectation w.r.t. $p(U_1^n|\theta_i)$ of the corresponding log-likelihood ratio L_n^{ij} between θ_i and θ_j

$$L_n^{ij} = \log\frac{p\left(U_1^n|\theta_i\right)}{p\left(U_1^n|\theta_j\right)} = \frac{\theta_i - \theta_j}{\sigma_{U_n}^2}U_1^n - \frac{\theta_i^2 - \theta_j^2}{2\sigma_{U_n}^2} \tag{17}$$

such that $D(\cdot||\cdot)$ is $E_{U_1^n|\theta_i}\left[L_n^{ij}\right] = \frac{(\theta_i - \theta_j)^2}{2\sigma_{U_n}^2}$. The minimum $\delta = \min_{i \neq j} D\left(p\left(U_1^n|\theta_i\right)||p\left(U_1^n|\theta_j\right)\right)$ is because of the symmetric distribution of θ over the real unit interval $(-0.5, 0.5)$ and the zero-mean Gaussian noise independent of i and j and given by $\delta = \frac{1}{M^2 2\sigma_{U_n}^2} = \frac{\alpha^2}{2M^2\sigma^2}$. With (14), the expected sample size which corresponds to the more meaningful expected code blocklength in this example becomes

$$E\left[\tau\right] \approx \frac{-\log A}{\delta} = \frac{-2M^2\sigma^2\log A}{\alpha^2}. \tag{18}$$

The corresponding expected E_b/N_0 for the SK feedback coding scheme with sequential decision-making can be calculated with (6) yielding

$$\frac{E_b}{N_0} = \frac{1}{N_0\log_2(M)}E_{\theta,\mathbf{Z},\tau}\left[\alpha^2\left(\hat{\theta}_1 - \theta\right)^2 + \sum_{i=2}^{\tau}\alpha^2\left(\hat{\theta}_i - \theta\right)^2\right] \tag{19}$$

$$= \frac{1}{N_0\log_2(M)}\left(\alpha^2\frac{1 - 1/M^2}{12} + E_{\mathbf{Z},\tau}\left[\sum_{i=2}^{\tau}\frac{\left(\sum_{k=1}^{i-1}Z_k\right)^2}{(i-1)^2}\right]\right) \tag{20}$$

where τ denotes the random blocklength of the sequential feedback coding scheme. For (20) equation (4) has been used. The second term of the RHS of (20) is difficult to calculate. Here, τ in the limit of the sum depends on the realization of the noise sequence $Z_1, Z_2, \ldots$. Neglecting the statistical dependency of τ and the noise sequence $Z_1, Z_2, \ldots$, one may substitute the random quantity τ in the limit of the sum by its expected value and approximate (20) with

$$\frac{E_b}{N_0} \approx \frac{1}{\log_2(M)}\left(\alpha^2\frac{1 - 1/M^2}{12N_0} + \frac{1}{2}\sum_{i=1}^{E[\tau]-1}\frac{1}{i}\right). \tag{21}$$

In Section VI, an evaluation of this approximation is presented when comparing the performance of the original SK feedback coding scheme and the SK feedback coding with sequential decision-making.

VI. Numerical Results

In the following, the performance gain of the combination of SK feedback coding and sequential decision-making for decoding in terms of the required E_b/N_0 and the required blocklength N to send k bits is evaluated. In Fig. 1a - Fig. 1b the E_b/N_0 for the original SK scheme and for the sequential SK scheme are shown. For this purpose, an error probability ϵ of 10^{-3} is considered and with (5) α is given by

$$\alpha = \frac{2M\sigma}{\sqrt{N}}Q^{-1}\left(\frac{\epsilon}{2}\right). \tag{22}$$

The derived α is used for both coding schemes, the original SK coding scheme and sequential SK coding scheme. Note that the derived α is only optimal for the standard approach without sequential decoding. In Fig. 1a the E_b/N_0 is shown over the information bits $k = \log_2 M$ and in Fig. 1b over the number of channel uses corresponding to the (average) blocklength/latency. Moreover, an analytical bound for no feedback and the result on the minimum E_b/N_0 to transmit k information bits with feedback, both for the case of $N \to \infty$, from [10] are included. The following conclusion can be drawn from the simulation results. It can be observed that the analytical expression for the E_b/N_0 (Orig. SK (Analytical, (6))) and the simulation results (Orig. SK (Simulated)) for the original SK coding scheme perfectly match. Moreover from Fig. 1a the reduction of the required E_b/N_0 can be divided into a feedback gain and an additional gain due to sequential decision-making. The original SK coding scheme easily outperforms the lower bound by Polyanskiy et al. for coding schemes with no feedback leading to the aforementioned feedback gain. Introducing a sequential decision-making algorithm at the decoder a further

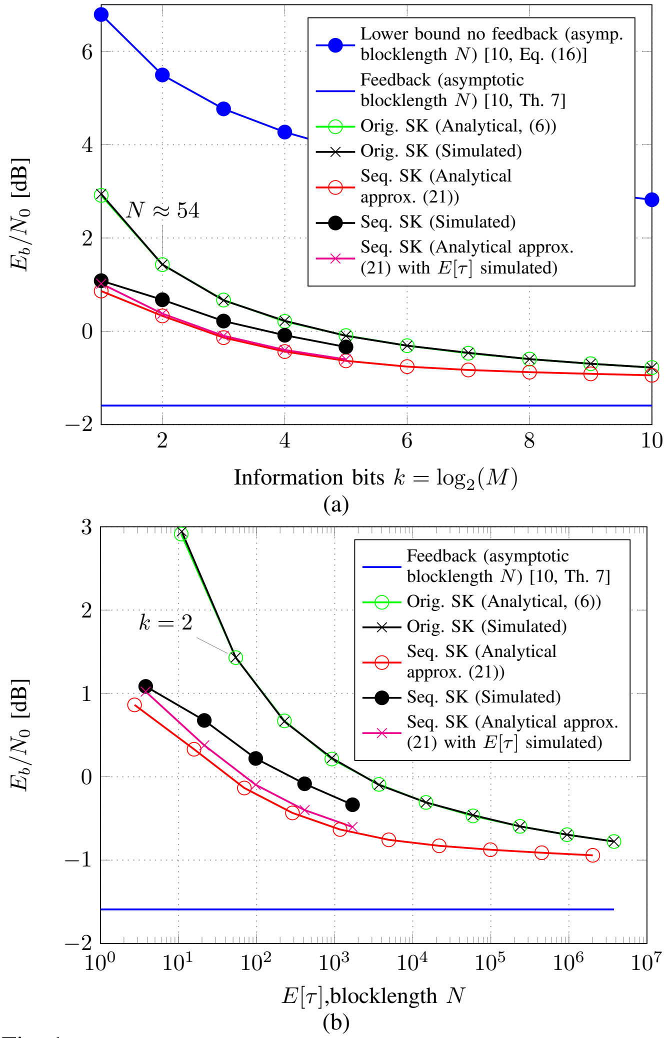

Fig. 1: Comparison of feedback vs. no feedback and original SK vs. seq. SK in terms of E_b/N_0 over the number of information bits k in (a) and over the blocklength N or $E[\tau]$ in (b); corresponding markers in (a) and (b) belong to the same operating point as exemplary highlighted for $k = 2$ and $N \approx 54$; block error rate $\epsilon = 10^{-3}$; Note that in (b) the curve 'Feedback (asymptotic blocklength N)' just act as an asymptote for $N \to \infty$.

gain can be seen. The additional sequential gain (Seq. SK (Simulated)) reduces for an increasing number of messages M. This is the case because the average transmit energy reduces with $1/n$ over the sequence of iterations. In case one aims for a minimum E_b/N_0, only for a very small number of bits k to be communicated the sequential decision-making overhead is reasonable. However, if latency is also critical one can see in Fig. 1b that a significant reduction of the average blocklength $E[\tau]$ and thus a reduction of latency is achievable with sequential decision-making. Likewise, less iterations on the forward link also means less usage of the feedback link, which will be beneficial when the feedback signaling overhead is included in the power calculations. However, this is part of further investigations as the current feedback link is assumed to be noise free and the transmit power on the feedback link is neglected.

Finally, Fig. 1a and Fig. 1b allow to evaluate the quality of the approximation of (20) by (21). The actual average E_b/N_0 is obtained by simulating (20) (Seq. SK (Simulated)). One approximation of (20) is given by (21), where the average $E[\tau]$ from (14) is used (Seq. SK (Analytical approx. (21))). Moreover, (21) is shown where the simulated sample size τ instead of $E[\tau]$ is used (Seq. SK (Analytical approx. (21) with $E[\tau]$ simulated)). Note that even here independency between τ and the realization of the noise sequence $\mathbf{Z}$ has been assumed. Fig. 1a and Fig. 1b indicate that both approaches lower-bound (20). However, an analytical proof remains for future work.

VII. Conclusion

A sequential decision-making SK feedback coding scheme was presented. The results show that the combination of the SK feedback coding and sequential decision-making for decoding leads to a further reduction in latency and required E_b/N_0 to transmit k bits compared to the original SK scheme. On the one hand, this behavior has been shown by Monte Carlo simulations and, on the other hand, by analytical approximations for the required E_b/N_0. Due to the approximation of the expected sample sizes (18) and the approximation of the expected value for the E_b/N_0 (20) a gap exits between the analytical derivation and the simulation results. Giving a rigorous analytical lower bound for the achievable E_b/N_0 with SK feedback coding in combination with sequential decoding remains for further study.

Acknowledgment

This work has been supported by the German Research Foundation (DFG) within the SFB 912 "Highly Adaptive Energy-Efficient Computing (HAEC)".

References

[1] M. Jenning et al., "Energy-efficient transceivers for ultra-highspeed computer board-to-board communication," in *Proc. IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB)*, Montreal, Canada, Oct 2015.

[2] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–I: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. 12, no. 2, pp. 172–182, Apr. 1966.

[3] J. Schalkwijk, "A coding scheme for additive noise channels with feedback–II: Band-limited signals," *IEEE Trans. Inf. Theory*, vol. 12, no. 2, pp. 183–189, Apr. 1966.

[4] P. Elias, "Channel Capacity Without Coding," *Quart. Progr. Rep., MIT Res. Lab. of Electronics*, Oct. 1956.

[5] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Feedback in the non-asymptotic regime," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4903–4925, Aug. 2011.

[6] A. Wald, *Sequential Analysis*. New York NY: John Wiley & Sons, 1947.

[7] V. P. Dragalin, A. G. Tartakovsky, and V. V. Veeravalli, "Multihypothesis sequential probability ratio tests. II. Accurate asymptotic expansions for the expected sample size," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1366–1383, Jul. 2000.

[8] C. W. Baum and V. V. Veeravalli, "A sequential procedure for multihypothesis testing," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1994–2007, Nov. 1994.

[9] A. Tartakovsky and I. Nikiforov, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*, ser. Monographs on Statistics and Applied Probability 136. Boca Raton: CRC Press, 2015.

[10] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Minimum energy to send k bits through the Gaussian channel with and without feedback," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4880–4902, Aug. 2011.

[11] M. Abramowitz, *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*. Dover Publications, Inc., 1974.

[12] A. Wald and J. Wolfowitz, "Optimum Character of the Sequential Probability Ratio Test," *The Annals of Mathematical Statistics*, vol. 19, no. 3, pp. 326–339, 1948.