International Zurich Seminar on Communications (IZS), March 2 – 4, 2016

Resource-Aware Incremental Redundancy in Feedback and Broadcast

Richard D. Wesel, Kasra Vakilinia, Sudarsan V. S. Ranganathan, Tong Mu
University of California, Los Angeles, CA 90095
Email: {wesel,vakiliniak, sudarsanvsr tongmu}@ucla.edu

Dariush Divsalar
Jet Propulsion Laboratory
Cal. Inst. of Tech., Pasadena, CA 91109
Email: Dariush.Divsalar@jpl.nasa.gov

***Abstract*—This paper reviews recent results from the UCLA Communication Systems Laboratory on the use of incremental redundancy. For channels with ACK/NACK feedback, this paper reviews how the transmission lengths used for communicating incremental redundancy should be optimized under the constraint of a limited number of incremental redundancy transmissions. For broadcast channels, this paper reviews optimization of the trade-off between packet-level erasure coding and physical-layer channel coding in the context of block fading with diversity that grows with bocklength.**

I. Introduction

This invited talk reviews two results [1], [2] optimizing the use of incremental redundancy. In systems with feedback, incremental redundancy adapts the coding rate to the the accumulated information density (the "instantaneous capacity") of the channel allowing the Shannon limit to be approached at much shorter average blocklengths than those required for the accumulated information density to concentrate around the Shannon capacity [3], [4], [5], [6], [7], [8]. In systems without feedback, incremental redundancy can provide a "fountain" of information from which a receiver need only "drink" what is needed to reliably identify the desired message [9], [10], [11]. For each of these two scenarios, this paper reviews optimization techniques that improve performance.

II. Transmission Lengths for ACK Feedback

ACK/NACK feedback is non-active in the sense that the feedback does not change what is transmitted but rather only indicates whether additional transmissions are needed. For channels with ACK/NACK feedback, the sequential differential optimization (SDO) approach of [1] optimizes the transmission lengths used to communicate incremental redundancy. This optimization maximizes throughput under the constraint of a limited number of incremental redundancy transmissions.

A. The Normal Approximation

SDO utilizes the power of the normal approximation introduced in [3] that characterizes the behavior of the rate that a channel can support at finite blocklength. Following [3], define information density $i(X,Y)$ as

$$i(X,Y) = \log_2 \frac{f_{Y|X}(y|x)}{f_Y(y)}. \tag{1}$$

The expected value of $i(X,Y)$ is the capacity of the channel. For the example of a BI-AWGN channel with noise z_k, $i(X,Y) = 1 - \log_2(1 + e^{-2(z_k+1)/\sigma^2}) = i(z_k)$. The accumulated information density I_n at the receiver after n symbols is

$$I_n = \sum_{k=1}^{n} i(z_k). \tag{2}$$

As pointed out by [3], (2) is a sum of independent random variables that will converge quickly to a normal distribution according to the central limit theorem, leading to the normal approximation of [3].

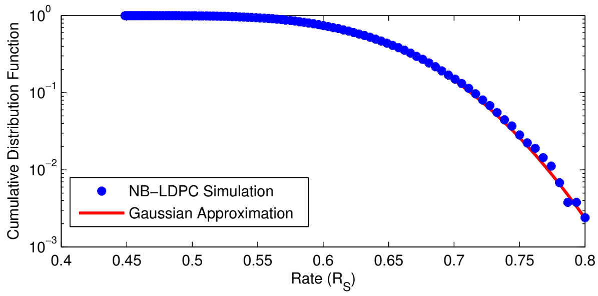

Fig. 1. Empirical complementary cumulative distribution function (c.d.f.) and the Gaussian approximation (Q-function) corresponding to the rate R_S at which the NB-LDPC code of [1] was able to decode successfully.

A key result of [1] is that a normal approximation also accurately describes the rate at which actual variable-length codes with incremental redundancy will successfully decode. Fig. 1 shows that for the NB-LDPC code used in [1] the empirical complementary cumulative distribution function on the rate at which decoding is successful is very closely approximated by a normal distribution for this example of the BI-AWGN channel with SNR of 2 dB. We have similarly confirmed the accuracy of the normal approximation to predict the rate at which decoding is successful for NB-LDPC codes in higher-SNR AWGN channels that require larger constellations and in fading channels with channel state information known at the receiver.

This material is based upon work supported by the National Science Foundation under Grant Numbers 1162501 and 1161822. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This research was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA JPL Task Plan 82-17473.

Fig. 1 shows that R_S is well-approximated by a Gaussian with mean $\mu_S = \mathrm{E}(R_S)$ and variance $\sigma_S^2 = \mathrm{Var}(R_S)$:

$$f_{R_S}(r) = \frac{1}{\sqrt{2\pi\sigma_S^2}} e^{-\frac{(r-\mu_S)^2}{2\sigma_S^2}} . \tag{3}$$

The c.d.f. of the blocklength N_S at which decoding is successful is $F_{N_S}(n) = P(N_S \leq n) = 1 - F_{R_S}(k/n)$. Taking the derivative of F_{N_S} using the Gaussian approximation of F_{R_S} produces the following "reciprocal-Gaussian" approximation for p.d.f. of N_S:

$$f_{N_S}(n) = \frac{k}{n^2\sqrt{2\pi\sigma_S^2}} e^{-\frac{\left(\frac{k}{n}-\mu_S\right)^2}{2\sigma_S^2}} . \tag{4}$$

B. Sequential Differential Optimization

SDO uses the tight Gaussian approximation discussed above to optimize the sequence of blocklengths $\{N_1, N_2, \ldots, N_m\}$ to maximize the throughput. Suppose that the number of incremental transmissions is limited to m. An accumulation cycle (AC) is a set of m or fewer transmissions and decoding attempts ending when decoding is successful or when the m^{th} decoding attempt fails. If decoding is not successful after the m^{th} decoding attempt, the accumulated transmissions are forgotten and the process starts over with a new transmission of the first block of N_1 symbols. From a strict optimality perspective, neglecting the symbols from the previous failed AC is sub-optimal. However, the probability of an AC failure is sufficiently small that the performance degradation is negligible. Neglecting these symbols greatly simplifies analysis.

The cumulative blocklength N_j at the jth stage is simply the sum of the first j increment lengths. Using the p.d.f. of N_S from (4) we can compute the probability that the decoder will need a particular incremental transmission. For $N_j < N_{j+1}$, the probability of a successful decoding attempt at blocklength N_{j+1} but not at N_j is

$$\int_{N_j}^{N_{j+1}} f_{N_S}(n)dn = \int_{N_j}^{N_{j+1}} \frac{k}{n^2\sqrt{2\pi\sigma_S^2}} e^{-\frac{\left(\frac{k}{n}-\mu_S\right)^2}{2\sigma_S^2}} dn \tag{5}$$

$$= Q\left(\frac{r_{j+1}-\mu_S}{\sigma_S}\right) - Q\left(\frac{r_j-\mu_S}{\sigma_S}\right), \tag{6}$$

where $r_j = k/N_j$.

Define the throughput as $R_T = \frac{E[K]}{E[N]}$, where $E[N]$ represents the expected number of channel uses and $E[K]$ is the effective number of information bits transferred correctly over the channel. The expression for $E[N]$ is

$$E[N] = N_1 Q\left(\frac{\frac{k}{N_1}-\mu_S}{\sigma_S}\right) \tag{7}$$

$$+ \sum_{j=2}^{m} N_j \left[Q\left(\frac{\frac{k}{N_j}-\mu_S}{\sigma_S}\right) - Q\left(\frac{\frac{k}{N_{j-1}}-\mu_S}{\sigma_S}\right)\right] \tag{8}$$

$$+ N_m \left[1 - Q\left(\frac{\frac{k}{N_m}-\mu_S}{\sigma_S}\right)\right] . \tag{9}$$

The first term (7) shows the contribution to the expected blocklength from successful decoding on the first attempt. $Q\left(\frac{\frac{k}{N_1}-\mu_S}{\sigma_S}\right)$ is the probability of decoding successfully with the initial block of N_1. Similarly, the terms in the summation of (8) are the contributions to the expected blocklength from decoding that is first successful at total blocklength N_j for $j \geq 2$ (at the j^{th} decoding attempt). Finally, the contribution to expected blocklength from not being able to decode even at N_m is (9). Even when the decoding has not been successful at N_m, the channel has been used for N_m channel symbols. The expected number of successfully transferred information bits $E[K]$ is

$$E[K] = kQ\left(\frac{\frac{k}{N_m}-\mu_S}{\sigma_S}\right) , \tag{10}$$

where $Q\left(\frac{\frac{k}{N_m}-\mu_S}{\sigma_S}\right)$ is the probability of successful decoding. Note that $E[K]$ depends only on k and N_m. In fact, $E[K] \approx k$ and is not sensitive to the specific choice of N_m for reasonably large values of N_m.

The initial blocklength is N_1 and we seek the optimal blocklengths $\{N_1, N_2, \ldots, N_m\}$ to maximize the throughput. Over a range of possible N_1 values, the SDO technique introduced in [1] selects $\{N_2, \ldots, N_m\}$ to minimize $E[N]$ for each fixed value of N_1 by setting derivatives to zero as follows:

$$\frac{\partial E[N]}{\partial N_j} = 0, \ \forall j = 1, \ldots, m-1 . \tag{11}$$

For each $j \in \{2, \ldots, m\}$, the optimal value of N_j is found by setting $\frac{\partial E[N]}{\partial N_{j-1}} = 0$, yielding a sequence of relatively simple computations. In other words, we select the N_j that makes our previous choice of N_{j-1} optimal in retrospect.

For $j > 2$, $\frac{\partial E[N]}{\partial N_{j-1}} = 0$ depends only on $\{N_{j-2}, N_{j-1}, N_j\}$ as follows:

$$\frac{\partial E[N]}{\partial N_{j-1}} = Q\left(\frac{\frac{k}{N_{j-1}}-\mu}{\sigma}\right) + (N_{j-1}-N_j)Q'\left(\frac{\frac{k}{N_{j-1}}-\mu}{\sigma}\right) - Q\left(\frac{\frac{k}{N_{j-2}}-\mu}{\sigma}\right).$$

Thus we can solve for N_j as

$$N_j = \frac{Q\left(\frac{\frac{k}{N_{j-1}}-\mu}{\sigma}\right) + N_{j-1}Q'\left(\frac{\frac{k}{N_{j-1}}-\mu}{\sigma}\right) - Q\left(\frac{\frac{k}{N_{j-2}}-\mu}{\sigma}\right)}{Q'\left(\frac{\frac{k}{N_{j-1}}-\mu}{\sigma}\right)} . \tag{12}$$

For each possible value of N_1, SDO can be used to produce an infinite sequence of N_j values that solve (11) for any choice of m. The sequence does not depend on m, only N_1. Each such sequence is an optimal sequence of increment lengths for a given density of decoding attempts on the time axis. As N_1 increases, the density of decoding attempts decreases, lowering system complexity. Using SDO to compute the optimal m decoding points is equivalent to selecting the most dense SDO-optimal sequence that when truncated to m points still meets the frame-error-rate target.

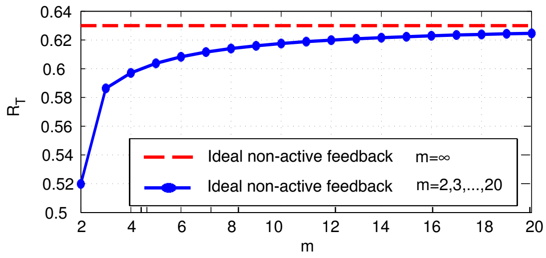

Fig. 2. Throughput as a function of the number m of incremental transmissions permitted.

C. Approaching Capacity at Short Blocklengths with Feedback

Fig. 2 shows the resulting throughputs obtained by using SDO to find the optimal increment lengths for values of m in the range of $2 < m < 20$ for the target FER of 10^{-3} for the NB-LDPC code of [1] for $k = 96$ message bits. Fig. 2 illustrates that with $m = 10$ decoding points, a system can closely approach the performance of a system that has $m = \infty$, which is the limiting case where decoding is attempted and feedback of an ACK/NACK is required after every received symbol.

Using SDO, variable-length codes with average blocklengths of around 500 symbols can closely approach capacity in theory and in practice as demonstrated in [1]. Fig. 3 illustrates the example of a binary-input (BI) additive white Gaussian noise (AWGN) channel with frame error rate (FER) required to be less than 10^{-3}. For a system transmitting k symbols at an average blocklength of λ, the throughput R_t is defined by $R_t = k/\lambda$. For reference, Fig. 3 shows the curves of possible throughput R_t as a function of λ for some values of k. The performance characterization for fixed-blocklength codes is from [3] and is based on the normal approximation, which is shown in [3] to be accurate for blocklengths as small as 100 symbols. The computation of the random coding lower bound on the performance of variable-length codes with feedback is based on the analysis in [4].

Fig. 3 shows curves from [1], [5], [6] that show simulation results that approach or exceed the performance promised by [4] in the range of average blocklengths below 500 bits. For values of $k = 16$, $k = 32$, $k = 64$, and $k = 89$ these throughput results exceed Polyanskiy's random coding lower bound. As the average blocklength becomes larger, the random coding lower bound is more predictive.

Note that variable-length codes with feedback approach capacity at very short blocklengths. In Fig. 3, the random-coding lower bound for a system with feedback is 0.27 dB from the Shannon limit for $k = 280$ with a blocklength of less than 500 bits. Looking at implemented codes for $k = 280$ in Fig. 3, the $m = \infty$ non-binary LDPC (NB-LDPC) code is 0.53 dB from Shannon limit. Using SDO, the NB-LDPC non-active feedback system in Fig. 3 that uses ten rounds of single-bit feedback to operate within 0.65 dB of the Shannon limit with an average blocklength of less than 500 bits.

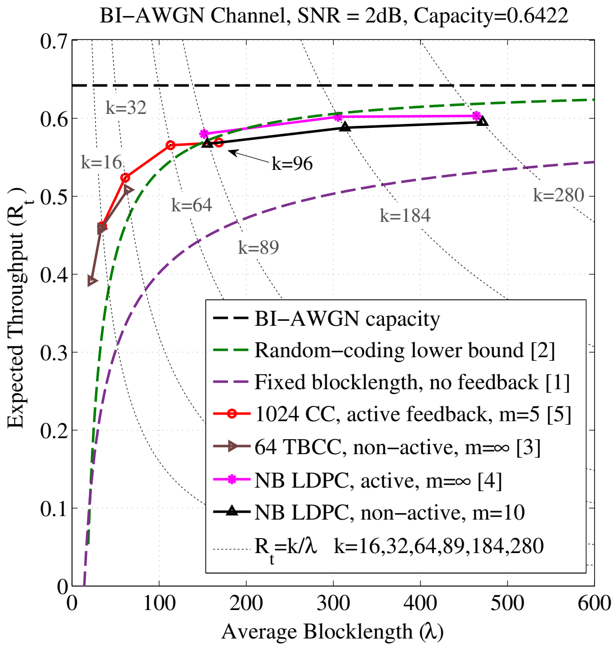

Fig. 3. Approaching capacity at short blocklengths using feedback.

III. Packet-Level vs. Physical Layer Redundancy

Consider a broadcast setting that uses a hybrid of packet-level erasure coding and physical layer coding to provide a stream of information with the goal of each receiver decoding the desired message at the earliest opportunity. There is a trade-off between using available redundancy for additional packets in a packet-level erasure code or simply for additional physical-layer code symbols.

As the amount of available redundancy grows, the work of [12] shows that in several different block fading scenarios the physical layer coding rate decreases ultimately to zero while the packet-level erasure coding rate does not. This indicates that at some point incremental redundancy should be directed to the physical layer rather than additional packet-level erasure coding. This paper reviews the recent work [2] that studies the this hybrid coding approach using a proportional diversity block fading model (in which diversity increases linearly with blocklength).

A. The Channel Model and Optimization Problem

Consider a transmitter and a receiver communicating over a fading channel [13]. The one-dimensional channel is modeled as $Y = HX + Z$ where X is the transmitted symbol, Y is the received symbol, H is the fading coefficient, and Z is i.i.d. additive white Gaussian noise (AWGN) with variance σ^2 and mean 0. We assume the channel is Rayleigh with $\mathbb{E}\left[H^2\right] = 1$, Z has unit variance, i.e. $\sigma^2 = 1$. Let the average transmit power be $\mathbb{E}\left[X^2\right] = P$. Then, the instantaneous *signal-to-noise ratio (SNR)* when $H = h$ is $h^2 P$. For this Rayleigh fading channel, SNR (denoted γ) is exponentially distributed with parameter $\frac{1}{P}$ that depends only on the average transmit power. Note that, γ has a mean of P.

A message consisting of m packets with k nats of information per packet is to be transmitted with a low *probability of message error* q; this is the probability that the receiver fails to recover all the m packets. The transmitter uses the channel for T units of time for an overall code rate of $\frac{mk}{T}$. It performs erasure coding across the m packets at a rate R_E and codes each resultant packet at a channel-coding rate R_C such that

$$\frac{mk}{T} = R_E R_C. \tag{13}$$

That is, the m packets are first coded using an erasure code at rate R_E to yield $\frac{m}{R_E}$ packets. Note that, for erasure coding, R_E has to satisfy $R_E \leq 1$. To transmit each packet, the transmitter uses a channel code at rate R_C [nats/channel-use] so that the resultant codeword block-length of each packet is $\frac{k}{R_C}$. For a fixed average transmit power, our objective is to pick the value of R_C (and thus R_E) that optimizes an objective function. The unit of channel-coding rate is "nats/channel-use" for convenience. The receiver is assumed to know the fading coefficient H while the transmitter does not.

The proportional diversity (PD) model introduce a parameter l_f, which describes the length of a fade. With the block-length being $\frac{k}{R_C}$, the number of block fades F_P in a transmitted codeword of a system with PD block fading of fade lengths l_f is

$$F_P = \left\lceil \frac{k}{R_C l_f} \right\rceil. \tag{14}$$

With PD block fading, long codewords benefit from an inherent increase in diversity. For this work, we assume that each block-fading event is independent, i.e. H assumes i.i.d. values across different block fades.

The receiver sees $\frac{m}{R_E} = R_C T k^{-1}$ packets from the channel. The number of packets that the decoder of the erasure code requires to recover the message, denoted $\hat{m} \geq m$, depends upon the erasure code. For Reed-Solomon erasure codes, $\hat{m} = m$; for fountain codes such as a Raptor code, $\hat{m} > m$ typically. Thus, the probability of message error q can be written using the binomial distribution as

$$q = \sum_{i=0}^{\hat{m}-1} \binom{R_C T k^{-1}}{i} (1-p_e)^i \, p_e^{\left(R_C T k^{-1} - i\right)}. \tag{15}$$

In the above expression, p_e denotes the probability that a packet is not decoded successfully (and declared an erasure) upon reception from the channel; this is called the *probability of packet erasure*. Owing to our assumption that the channel codes in the system operate close to capacity with zero block-error probability when the Shannon capacity exceeds the attempted rate, p_e constitutes only one event: *fading outage* [14].

The binomial sum in (15) can be computed numerically only for small values of $R_C T k^{-1}$. Hence, we approximate the random variable that denotes the number of packets successfully decoded by the channel decoder using the Central Limit Theorem (CLT), and obtain the Gaussian approximation for q [12] as

$$q \approx \Phi\left[\frac{(\hat{m}-1) - R_C T k^{-1}(1-p_e)}{\sqrt{R_C T k^{-1} p_e (1-p_e)}}\right], \tag{16}$$

where $\Phi(x)$ is the value of the c.d.f. of the standard normal random variable at $x \in \mathbb{R}$.

To summarize, the objective is to minimize the message-error probability q in (15) via (16), where p_e is also a function of R_C. Writing the minimization problem in terms of R_C, R_E can be obtained as $R_E = \frac{mk}{TR_C}$. Hence, the optimization problem is as follows:

$$\begin{aligned}
&\min_{R_C} \;\; \Phi\left[\frac{(\hat{m}-1) - R_C T k^{-1}(1-p_e)}{\sqrt{R_C T k^{-1} p_e (1-p_e)}}\right], \\
&\text{s.t.} \;\; p_e(R_C) = \mathbb{P}\left[\frac{1}{\frac{k}{R_C l_f}} \sum_{i=1}^{\left\lfloor \frac{k}{R_C l_f} \right\rfloor} C(\gamma_i) + \right. \\
&\qquad\qquad \left. \frac{\frac{k}{R_C l_f} - \left\lfloor \frac{k}{R_C l_f} \right\rfloor}{\frac{k}{R_C l_f}} C(\gamma_{\text{last}}) < (1+\epsilon) R_C \right], \\
&\qquad\qquad \frac{k\hat{m}}{T} \leq R_C \leq \frac{k}{l_f}, \;\; R_C T k^{-1} \in \mathbb{N}.
\end{aligned} \tag{17}$$

Note that, minimizing $\Phi(\cdot)$ is equivalent to minimizing its argument, and the value of q need not be explicitly computed. We have specified the dependence of p_e on R_C here for clarity.

As noted in [12], [15], and many previous works, the evaluation of p_e for the block-Rayleigh fading channel (or for its PD version) is not a straightforward task. One can use [15] or similar works for the block-Rayleigh fading channel to compute the outage probability p_e with a minuscule error. But, our fading model complicates it further as we have a sum of two random variables that are not identically distributed in the expression for p_e in (17). We first expand and rearrange the terms in p_e for our one-dimensional PD block-Rayleigh fading channel with capacity-achieving codes to obtain

$$p_e = \mathbb{P}\left[\sum_{i=1}^{\left\lfloor \frac{k}{R_C l_f} \right\rfloor} W_i + \left(\frac{k}{R_C l_f} - \left\lfloor \frac{k}{R_C l_f} \right\rfloor\right) W_{\text{last}} < \frac{ck}{l_f}\right], \tag{18}$$

where $c = 2(1+\epsilon)$, $W_i = \log(1+\gamma_i)$, $W_{\text{last}} = \log(1+\gamma_{\text{last}})$.

B. Gaussian Approximations of the Optimization Problem

Based on Gaussian approximations of p_e in (18), as inspired by [12], [2] presents four approximations to the optimization problem (17). For numerical-search based results, [2] uses a very low value of the margin, say $\epsilon = 0.05$, to obtain c.

1) Gaussian Approximation 1 (Approx. 1): Ignoring the contribution of W_{last} in (18), we get

$$p_e = \mathbb{P}\left[\sum_{i=1}^{\left\lfloor \frac{k}{R_C l_f} \right\rfloor} W_i < \frac{ck}{l_f}\right]. \tag{19}$$

The above can be approximated using the Gaussian CDF as

$$p_e = \Phi\left[\frac{\frac{ck}{l_f} - \left\lfloor \frac{k}{R_C l_f} \right\rfloor \mu(P)}{\sqrt{\left\lfloor \frac{k}{R_C l_f} \right\rfloor \mathrm{Var}(P)}}\right]. \tag{20}$$

The values of $\mu(P)$ and $\mathrm{Var}(P)$, which denote the mean and variance of $\log(1+\gamma)$ with $\gamma \sim$ Exponential $\left(\frac{1}{P}\right)$, can be computed as stated in [12]. By ignoring the flooring function, we get *Gaussian approximation 1 (Approx. 1)*, which is an adaptation of (19) in [12] to PD block-Rayleigh fading:

$$p_e = \Phi\left[\sqrt{\frac{k}{R_C l_f}}\frac{cR_C - \mu(P)}{\sqrt{\mathrm{Var}(P)}}\right]. \tag{21}$$

2) Gaussian Approximation 2 (Approx. 2): For *Approx. 2*, we evaluate (20) directly. The approximation to p_e that is being made here is imprecise in the sense that, (20) evaluates to the same value for a range of R_C values; the reason being the presence of the flooring function.

3) Gaussian Approximation 3 (Approx. 3): This approximation is the evaluation of (18) with a constrained search space that limits R_C such that both $\frac{m}{R_E}$ and $\frac{k}{R_C l_f}$ are positive integers.

4) Gaussian Approximation 4 (Approx. 4): The Gaussian approximation that we make here considers both the terms in (18), making it the most appropriate. Once we find out $\mu(P)$ and $\mathrm{Var}(P)$, we assume that $\sum_{i=1}^{\left\lfloor \frac{k}{R_C l_f} \right\rfloor} W_i$ is Gaussian and also that $\left(\frac{k}{R_C l_f} - \left\lfloor \frac{k}{R_C l_f} \right\rfloor\right) W_{\text{last}}$ is Gaussian. Thus, their linear sum is another Gaussian random variable denoted W_G, which stands for *Gaussian approximation of weighted average mutual information*, with

$$\mathrm{mean}(W_G) = \frac{k}{R_C l_f}\mu(P),$$
$$\mathrm{Var}(W_G) = \mathrm{Var}(P)\left[\left\lfloor \frac{k}{R_C l_f} \right\rfloor + \left(\frac{k}{R_C l_f} - \left\lfloor \frac{k}{R_C l_f} \right\rfloor\right)^2\right]. \tag{22}$$

Thus, p_e for this approximation *(Approx. 4)* is

$$p_e = \Phi\left[\frac{\frac{ck}{l_f} - \mathrm{mean}(W_G)}{\sqrt{\mathrm{Var}(W_G)}}\right]. \tag{23}$$

C. Results and conclusions

Fig. 4 shows an example of the optimal values of R_C and R_E obtained from Approximations 1 and 4 as the overall code rate $\frac{mk}{T}$ goes to 0. As observed by Courtade and Wesel [12] for the (fixed diversity) block-fading channel, for the PD block-fading model that the optimal channel-coding rate goes to 0. However, where the optimal value of R_E approached a non-zero constant less than 1 for fixed diversity in [12], under PD block-fading it is approaching 1. With sufficient overall redundancy, packet-level erasure coding is unnecessary in a block fading channel with proportional diversity. Note that rate-compatibility in this scenario is challenging because the rate R_E increases for a sufficiently low overall rate.

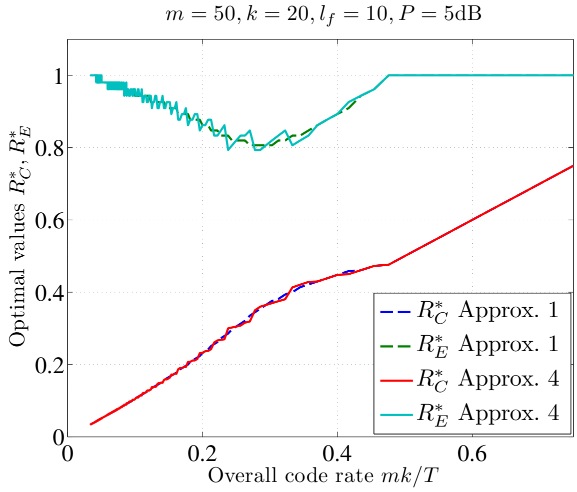

Fig. 4. Result of optimization problem (17) as a function of overall code rate for an example system with $m = 10$, $k = 20$, $l_f = 10$ and $P = 5$ dB.

References

[1] K. Vakilinia, A. R. Williamson, S. V. S. Raganathan, D. Divsalar, and R. D. Wesel. Feedback systems using non-binary LDPC codes with a limited number of transmissions. In *IEEE Information Theory Workshop*, pages 167 – 171, Hobart, Tasmania, Australia, November 2014.

[2] S. V. S. Ranganathan, T. Mu, and R. D. Wesel. Optimality and rate-compatibility for erasure-coded packet transmissions when fading channel diversity increases with packet length. In arxiv.org/abs/1602.00761, 2016.

[3] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory*, 56(5):2307–2359, May 2010.

[4] Y. Polyanskiy, H. V. Poor, and S. Verdú. Feedback in the non-asymptotic regime. *IEEE Trans. Inf. Theory*, 57(8):4903–4925, Aug. 2011.

[5] A. R. Williamson, T.-Y. Chen, and R. D. Wesel. Variable-length convolutional coding for short blocklengths with decision feedback. *IEEE Trans. Commun.*, 63(7):2389 – 2403, July 2015.

[6] A. R. Williamson, T.-Y. Chen, and R. D. Wesel. Firing the genie: Two-phase short-blocklength convolutional coding with feedback. In *IEEE Inf. Theory and Applicat. Workshop*, pages 1 – 6, San Diego, CA, February 2013.

[7] K. Vakilinia, T.-Y. Chen, S. V. S. Raganathan, A. R. Williamson, D. Divsalar, and R. D. Wesel. Short-blocklength non-binary LDPC codes with feedback-dependent incremental transmissions. In *IEEE Int. Symp. Inf. Theory*, pages 426 – 430, Honolulu, Hawaii, June 2014.

[8] A. R. Williamson, T.-Y. Chen, and R. D. Wesel. Reliability-based error detection for feedback communication with low latency. In *IEEE Int. Symp. Inf. Theory*, pages 2552 – 2556, Istanbul, Turkey, July 2013.

[9] M. G. Luby, M. Mitzenmacher, M. A. Shokrallahi, and D. A. Spielman. Efficient erasure correcting codes. *IEEE Trans. on Info. Th.*, 47(2):569–584, Feb. 2001.

[10] M. Luby. LT codes. In *The 43rd Annual IEEE Symposium on the Foundations of Computer Science*, pages 271–280, Atlanta, Georgia, Nov. 2002.

[11] A. Shokrollahi. Raptor codes. *IEEE Trans. on Info. Th.*, 52(6):2551–2567, June 2006.

[12] T.A. Courtade and R.D. Wesel. Optimal allocation of redundancy between packet-level erasure coding and physical-layer channel coding in fading channels. *IEEE Trans. on Comm.*, 59(8):2101–2109, Aug. 2011.

[13] A. Goldsmith. *Wireless Communications*. Cambridge University Press, New York, 2005.

[14] E. Biglieri. *Coding for Wireless Channels*. Springer, 2005.

[15] A. Yilmaz. Calculating outage probability of block fading channels based on moment generating functions. *IEEE Trans. Commun.*, 59(11):2945–2950, November 2011.