Transmission Lengths that Maximize Throughput of Variable-Length Coding & ACK/NACK Feedback

R. D. Wesel, N. Wong, A. Baldauf, A. Belhouchat
University of California, Los Angeles, CA
{wesel,nsc.wong}@ucla.edu
{ambaldauf19,abelhouchat}@gmail.com

A. Heidarzadeh, J.-F. Chamberland
Texas A&M University
College Station, TX
{anoosheh,chmbrlnd}@tamu.edu

***Abstract*—To maximize the throughput achieved by variable-length codes, the length of each incremental redundancy transmission should be optimized. In previous papers, sequential differential optimization (SDO) provides transmission lengths that optimize throughput by minimizing average blocklength for a specified maximum number of feedback transmissions per message. This paper uses a Lagrangian approach to develop an alternative SDO that maximizes throughput for a specified *average* number of feedback transmissions per message. The mapping of real-valued SDO solutions to the necessarily integer transmission lengths is addressed through dithering.**

I. Introduction

This paper optimizes of the lengths of transmissions of incremental redundancy (IR) controlled by feedback of acknowledgement (ACK) or negative acknowledgement (NACK) messages from the receiver to the transmitter. ACK indicates that reliable decoding has been achieved, which can be established by a cyclic redundancy check (CRC) [1], [2] or a direct calculation of the probability of error for the selected codeword, e.g. by the reliability-output Viterbi algorithm (ROVA) [3], [4]. Feedback-based incremental redundancy is sometimes referred to as either an IR-Hybrid automatic repeat request (HARQ) or a Type-II HARQ, as described in [5], [6].

In an IR-HARQ, a variable-length (VL) code is employed; the receiver accumulates symbols through a sequence of IR transmissions until the accumulation concludes either because reliable decoding is achieved or because the available sequence of IR transmissions has been exhausted. More informative feedback messages than a simple ACK/NACK can be useful, as explored in works such as [5], [7]–[9], but in this paper the only role of feedback is to determine whether the next in a sequence of predetermined IR transmissions should be sent. This is referred to as "decision feedback" by Forney [10] or as "stop feedback" by Polyanskiy et al. [11].

The selection of each transmission length as well as the number of possible IR transmissions is central to the design of an effective IR-HARQ. As pointed out in [5], the optimization of these parameters had received relatively little attention in the literature with notable exceptions such as [8], [9] and [5] itself, which require more information than the ACK/NACK of the "stop feedback" systems that are the focus of this paper.

As explained in [5], IR-HARQs have been primarily applied to applications with fading channels. Indeed, IR-HARQs play an important role in cellular communications including 3G [12], 4G [13], and 5G [14]. However, recent theoretical analysis [11] and practical demonstrations [15], [16] show that IR-HARQ with simple ACK/NACK feedback can be useful even on non-fading channels; on these channels they allow capacity to be closely approached with short average blocklengths on the order of 200-500 symbols. The examples in this paper focus on such static channels, specifically a binary-input AWGN channel, but the analysis and techniques presented apply to any channel where an IR-HARQ is used.

Research supported by National Science Foundation (NSF) grant CCF-1618272. Any opinions, findings, and conclusions or recommendations are those of the author(s) and do not necessarily reflect the views of the NSF.

The technique of sequential differential optimization (SDO) [16]–[18] provides a sequence of transmission lengths that optimizes throughput by minimizing the average blocklength. For a a specified maximum number of feedback transmissions and a maximum probability that the decoder fails to produce an ACK even will all possible IR, SDO find the transmission lengths that minimize average blocklength. SDO requires a known probability distribution on the probability of ACK at each cumulative blocklength, but works equally well for the variety of distributions that arise from different VL codes operating on different channels [18]–[20].

The original formulation of SDO minimized the average blocklength for a fixed *maximum* number of feedback trasnsmissions. This paper re-frames the optimization problem to provide a closed form expression for the optimal transmission lengths under a constraint on the *average* number of feedback transmissions. Regardless of how feedback is constrained, SDO produces real-valued transmission lengths rather than integers. The original implementations in [16], [17] modified the lengths at each step to take on integer values, introducing a performance loss. This paper also shows how dithering can randomize integer-valued transmission lengths to provide essentially the same performance as SDO's real values.

The paper proceeds as follows: Sec. II reviews VL coding with ACK/NACK feedback, introduces the original SDO algorithm of [17] but with real-valued lengths, and presents a baseline example using convolutional codes. Sec. III introduces the average number of feedback transmissions per message as an alternative constraint and uses a Lagrangian framework to develop a new version of SDO that jointly minimizes the average blocklength and the average number of feedback transmissions. Sec. IV presents a dithering technique to produce integer transmission lengths that closely approach the performance of the real-valued transmission lengths provided by SDO. Sec. V concludes the paper.

978-1-5386-4727-1/18/$31.00 ©2018 IEEE

II. Background: SDO for VL Coding

A. Variable-Length Coding with ACK/NACK Feedback

Consider a system that communicates a k-bit message by using IR to send up to m possible transmissions in an accumulation cycle (AC). The transmissions have lengths of $\ell_1, \ldots, \ell_m$, where sending each additional transmission depends on ACK/NACK feedback. Each subsequent decoding attempt in the AC has the advantage of a successively larger cumulative blocklength of N_f, where $N_f = \sum_{j=1}^{f} \ell_j$.

The decision of whether to send an ACK to terminate the transmission or to send a NACK to request IR is based on an indicator of reliable decoding that is available at the receiver. A decoding error is only possible when the receiver decodes incorrectly and also determines erroneously that the decoding is reliable (e.g. because of passing a CRC check despite being incorrect) and sends an ACK. All such errors are undetected.

The receiver cannot know whether it has truly decoded correctly, but the mechanism for determining reliable decoding can guarantee a maximum probability of undetected error P_{UE} for a specified channel or class of channels. CRCs designed to match the structure of the encoder have well-defined upper bounds on P_{UE} that can be made quite small [1]. Similarly, the ROVA threshold is specifically designed to achieve a desired target probability of undetected error P_{UE} [4], [21].

Let $P_{\text{ACK}}^{(N_f)}$ and $P_{\text{NACK}}^{(N_f)}$ be the marginal probabilities of a decoding "success" or "failure" based on the reliability indicator when the decoder is presented with a received codeword having blocklength N_f. Note that $P_{\text{ACK}}^{(N_f)} + P_{\text{NACK}}^{(N_f)} = 1$. The SDO approach requires knowledge of $P_{\text{ACK}}^{(N_f)}$ as a function of N_f for all valid values of f. In Sec. II-C we use a Gaussian model for $P_{\text{ACK}}^{(N_f)}$. Other models or simply the empirical distribution obtained from simulation can provide the characterization of $P_{\text{ACK}}^{(N_f)}$.

Consistent with the previous literature on SDO [16]–[20], we assume a maximum transmission length N_{max} that the system will not exceed. N_{max} might be selected to achieve the desired value of ACK probability for the AC so that $P_{\text{NACK}}^{(N_{\text{max}})}$ is below a threshold such as 10^{-3}. A maximum length N_{max} might be required because there is a limited number of redundancy symbols available from the VL code. Certainly there exist approaches that have unbounded transmission length, but this paper restricts attention to the case where the transmission length of the AC is bounded by N_{max}.

While only integer values of N_f make sense in real systems, it will be useful to model $P_{\text{ACK}}^{(N_f)}$ as a cumulative density function (c.d.f.) where N_f is treated as a continuous random variable. Our SDO approach yields real-valued solutions for N_f. Sec. IV of this paper addresses how to uses integer values in an actual system and still closely approach the throughput performance of SDO with real values. This section and Sec. III focus on how real-valued SDO solutions differ in performance based on how the cost of the feedback resource is integrated into the problem formulation.

If decoding is unsuccessful after all m decoding attempts in the accumulation cycle, the associated k-bit message may have to be discarded in a delay-sensitive application or the transmission may be attempted again from scratch in applications where reliability and completeness are paramount and latency is a secondary consideration. IR-HARQs with a fixed maximum number of transmissions are referred to by Heindlmaier and Soljanin in [22] as a fixed IR scheme, and as shown in [22] the loss from the infinite IR scheme, where $m = \infty$, is small when $P_{\text{NACK}}^{(N_m)}$ is low. Even with a relatively high $P_{\text{NACK}}^{(N_m)}$ of 10^{-3}, the effect of unsuccessful decoding attempts on the throughput is negligible.

Let random variable (r.v.) I be the number of successfully transmitted information bits in an accumulation cycle. Note that I can only take on the values of k and zero. Let N be the number of symbols transmitted in an accumulation cycle. The throughput rate R_T is defined as

$$R_T = \frac{E[I]}{E[N]} = \frac{k\left(1 - P_{\text{NACK}}^{(m)} - P_{\text{UE}}\right)}{E[N]}, \tag{1}$$

where $E[N]$ is well approximated by

$$N_1 P_{\text{ACK}}^{(N_1)} + \sum_{f=2}^{m} N_f \left[P_{\text{ACK}}^{(N_f)} - P_{\text{ACK}}^{(N_{f-1})}\right] + N_m P_{\text{NACK}}^{(N_m)}. \tag{2}$$

In (2) each cumulative blocklength N_f is multiplied by the probability that reliable decoding is declared for the first time at that blocklength. The longest blocklength N_m is also multiplied by the probability that reliable decoding is never declared.

B. Sequential Differential Optimization (SDO)

The expression for $E[N]$ above is an approximation because it assumes that if an ACK was sent when decoding a message of length N_{j-1}, then certainly an ACK would also be sent when decoding the corresponding longer message with length N_f. While always true for erasure channels [20] and observed to be true for the non-binary LDPC codes explored in [16], this is not true in general and specifically not the case for the convolutional codes explored in this paper. However, events where an ACK would be followed by a NACK are relatively rare, and ignoring this effect simply leads to a slight underestimate of throughput by under-counting the ACKs that occur for the first time on the f^{th} attempt.

To further simplify optimization of R_T, note that $E[I] \approx k$ because $P_{\text{NACK}}^{(m)}$ and P_{UE} are negligibly small. Furthermore, these values are fixed by the value of N_m and the mechanism for determining ACK. Thus, maximizing R_T is essentially equivalent to minimizing $E[N]$. Thus, we seek to minimize $E[N]$ as approximated by (2).

Over a range of possible initial transmission lengths N_1, SDO [16]–[20] optimizes $\{N_2, \ldots, N_m\}$ to minimize $E[N]$ for each fixed value of N_1. For each $j \in \{2, \ldots, m\}$, the optimal value of N_f is found by setting $\frac{\partial E[N]}{\partial N_{f-1}} = 0$, yielding a sequence of relatively simple computations. In other words, SDO selects the N_f that makes the previous choice of N_{f-1} optimal in retrospect. For example, to find N_2 SDO utilizes the derivative

$$\frac{\partial E[N]}{\partial N_1} = P_{\text{ACK}}^{(N_1)} + (N_1 - N_2) P_{\text{ACK}}^{\prime(N_1)} = 0 \tag{3}$$

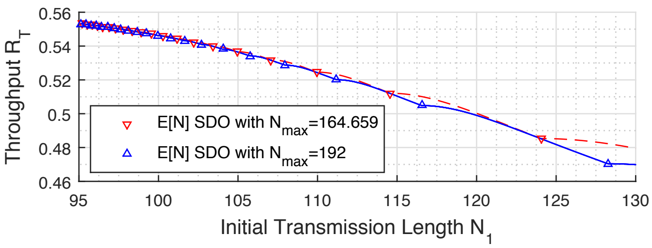

Fig. 1. Throughput R_T as a function of the initial transmission length N_1 using the cumulative transmission lengths N_f prescribed by (4) and (6) for $N_{\max} = 164.649$ and 192. The triangles show points where the final value N_m prescribed by (6) is exactly $N_{\max}$ so that truncation is not needed.

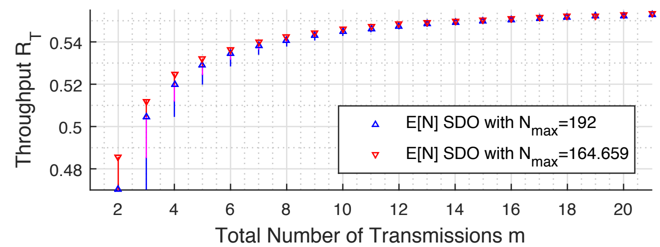

Fig. 2. Throughput R_T as a function of the maximum number of transmissions m in an AC, using the cumulative transmission lengths N_f prescribed by (4) and (6) for $N_{\max} = 164.649$ and 192. Only maximum throughput points (the triangles in Fig. 1) are shown.

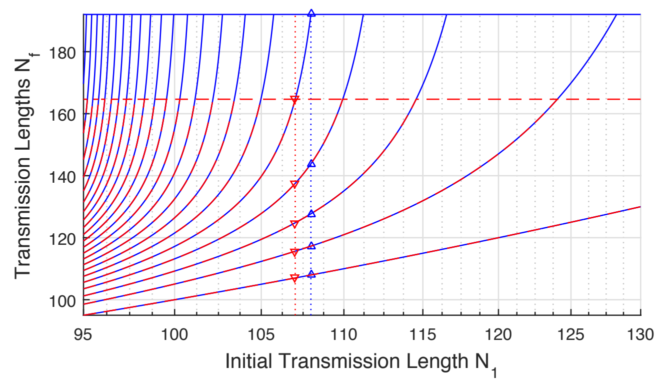

Fig. 3. Cumulative transmission lengths N_f less than 192 prescribed by (4) and (6) as a function of N_1. The blue line indicates the maximum throughput solution for $m = 5$, $\{N_f\} = \{V, W, X, Y, Z\}$.

and solves for N_2 as

$$N_2 = N_1 + \frac{P_{\text{ACK}}^{(N_1)}}{P_{\text{ACK}}^{\prime(N_1)}} . \tag{4}$$

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:

$$\frac{\partial E[N]}{\partial N_{j\text{–}1}} = P_{\text{ACK}}^{(N_{j-1})} + (N_{j-1} - N_f) P_{\text{ACK}}^{\prime(N_{j-1})} - P_{\text{ACK}}^{(N_{j-2})} . \tag{5}$$

Thus SDO solves for N_f as

$$N_j = N_{j-1} + \frac{P_{\text{ACK}}^{(N_{j-1})} - P_{\text{ACK}}^{(N_{j-2})}}{P_{\text{ACK}}^{\prime(N_{j-1})}} . \tag{6}$$

Equations (4) and (6) provide an infinite sequence of transmission lengths N_f, but many systems are constrained by a maximum transmission length $N_{\max}$. Thus, SDO computes N_f values only for indices $f \le m$ where m is the first index for which the transmission length computed by (6) exceeds $N_{\max}$. In general, the transmission length N_m is truncated to $N_{\max}$, but the most useful sequences are those for which $N_m = N_{\max}$ according to (4) so that no truncation is necessary. The transmission lengths N_f provided by (4) and (6) all increase as N_1 increases. Thus for a specified $N_{\max}$ the SDO solutions lead to a natural trade-off between throughput R_T and the maximum number of transmissions m in an AC as shown in the Figs. 1-3 for the tail-biting convolutional code described in the next subsection.

C. Example: Tail-biting Convolutional Code in AWGN

Examples are presented using the VL coding system with ACK/NACK feedback explored in [15], [18] transmitting a message of $k = 64$ bits using a tail-biting rate-1/3, 1024-state convolutional code with pseudo-random puncturing. This system has a natural $N_{\max}$ of 192 bits. For this system, reliable decoding (resulting in feedback of an ACK) is declared when the probability of error computed by the tail-biting ROVA algorithm of [21] exceeds a threshold. Our examples are for the binary-input additive white Gaussian noise (AWGN) channel with SNR of 2 dB. As shown in [18] for this system, $P_{\text{ACK}}^{(N_f)}$ is closely approximated by a Gaussian distribution on rate k/N_f with mean $\mu = 0.5666$ and standard deviation $\sigma = 0.0573$.

In the previous SDO literature [16]–[20], the real-valued solutions for $\{N_f\}$ were not considered valid solutions. These papers forced each value N_f to take an integer value (either $\lceil N_f \rceil$ or $\lfloor N_f \rfloor$) and used that integer value in the next application of (6). This paper departs from that paradigm by retaining the real-valued N_f values (including N_1) as achievable through dithering as explored in Sec. IV-B. Thus, Figs. 1-3 are novel from what has been presented in [16]–[20].

Fig. 1 shows how throughput R_T varies as a function of N_1 when $N_{\max}$ is set to 192 (the maximum number of symbols available form the convolutional code) and when $N_{\max}$ is set to 164.649, which is the value that achieves $P_{\text{NACK}}^{(N_{\max})} = 10^{-3}$ according to the Gaussian model. The triangles show points where $N_m = N_{\max}$ according to (4), which are the highest throughput points for the associated value of m. The curve connecting two triangles is the set of truncated solutions associated with one value of m. Note that the continuous throughput curve for $N_{\max} = 164.649$ dominates that for $N_{\max} = 192$, but it has a larger value of $P_{\text{NACK}}^{N_{\max}}$.

In Fig. 2, optimal points (triangles) of the R_T vs. N_1 curve of Fig. 1 provide a trade-off curve of R_T vs. m. Fig. 3 shows the evolution of the transmission-length sequence $\{N_f\}$ as a function of N_1 according to (4) and (6) for the two values of $N_{\max}$. The red and blue vertical dotted lines show the transmission lengths that maximize throughput for $m = 5$ for $N_{\max} = 164.649$ and 192 respectively.

III. Limiting Average Number of ACKs/NACKs

The previous section presented the SDO approach as practiced in [16]–[20], which seeks to maximize throughput for a fixed *maximum* number m of feedback transmissions. However, the sequence $\{N_f\}$ can be infinite and there are practical applications where $N_{\max}$ is large enough that later IR transmissions occur with low probability so that the *average* number of feedback messages required for an accumulation cycle may be a better metric for system optimization than m.

Let random variable F be the number of feedback transmissions required for an accumulation cycle. This is the number of feedback transmissions until an ACK is sent or until the maximum number of messages m has been sent. Similar to $E[N]$, the expected value of F is approximated as follows:

$$E[F] \approx P_{\text{ACK}}^{(N_1)} + \sum_{f=2}^{m} f\left[P_{\text{ACK}}^{(N_f)} - P_{\text{ACK}}^{(N_{f-1})}\right] + mP_{\text{NACK}}^{(N_m)} \quad (7)$$

$$= m - \sum_{f=1}^{m-1} P_{\text{ACK}}^{(N_f)}. \quad (8)$$

This section develops a new SDO formulation that produces transmission lengths $\{N_f\}$ that maximize throughput R_T for a specified $E[F]$.

A. Lagrangian Formulation

Consider a Lagrangian objective function that includes a cost term for the average number of feedback transmissions to enforce a constraint that $E[F] \leq \theta$:

$$\begin{aligned} J_F =& E[N] + \lambda\,(E[F] - \theta) \\ =& (N_1 + \lambda)P_{\text{ACK}}^{(N_1)} \\ &+ \sum_{f=2}^{m} (N_f + \lambda f)\left[P_{\text{ACK}}^{(N_f)} - P_{\text{ACK}}^{(N_{f-1})}\right] \\ &+ (N_m + \lambda m)\, P_{\text{NACK}}^{(N_m)} - \lambda\theta. \end{aligned}$$

Assuming a fixed k and a known distribution for $P_{\text{ACK}}^{(N_f)}$, λ is selected to minimize $E[N]$ while maintaining $E[F] \leq \theta$, i.e. such that $E[F] = \theta$. We note that this is equivalent to simply minimizing

$$J_\lambda = E[N] + \lambda E[F] \quad (9)$$

for the appropriate λ, and that minimizing (9) over an appropriate range of λ yields the trade-off of $E[N]$ vs. $E[F]$ that system designers can use to guide choices about which set of transmission lengths to adopt for the IR-HARQ given the channel conditions as described by the $P_{\text{ACK}}^{(N_f)}$ distribution. For a given λ, real-valued N_1 is varied until the solution is found that minimizes J_λ for a specified $N_{\max}$.

As with the original SDO, the optimal values of N_i for $i > 1$ can be determined for a given value of N_1 by solving a set of differential equations. Sequentially setting the derivatives of J_λ with respect to N_i to zero as described in Sec. II-B. To find N_2 we compute the derivative

$$\frac{\partial J_F}{\partial N_1} = P_{\text{ACK}}^{(N_1)} + (N_1 - N_2 - \lambda)P_{\text{ACK}}^{\prime(N_1)} = 0 \quad (10)$$

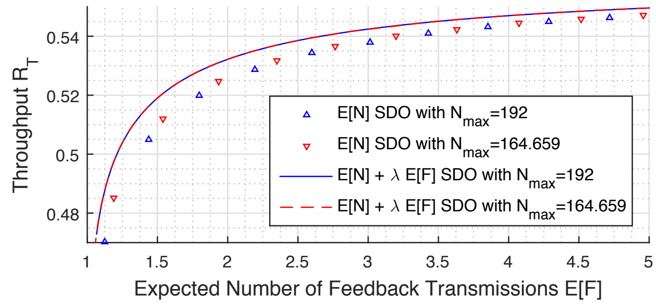

Fig. 4. Throughput R_T vs. $E[F]$ for solutions prescribed by (4)-(6) and (11)-(13) for $N_{\max} = 164.649$ and 192.

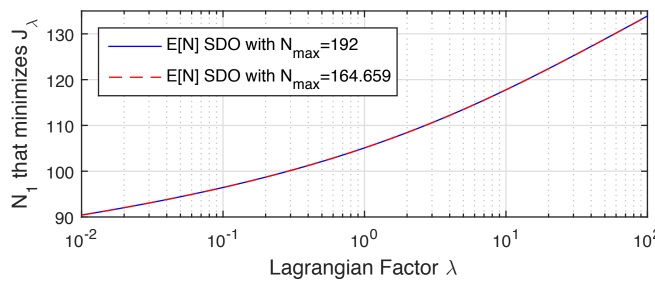

Fig. 5. The N_1 value that maximizes J_λ as a function of λ for $N_{\max} = 164.649$ and 192.

and solve for N_2 as

$$N_2 = N_1 - \lambda + \frac{P_{\text{ACK}}^{(N_1)}}{P_{\text{ACK}}^{\prime(N_1)}} \quad (11)$$

For $f > 2$, $\frac{\partial E[N]}{\partial N_{f-1}} = 0$ depends only on $\{N_{f-2}, N_{f-1}, N_f\}$:

$$\frac{\partial E[N]}{\partial N_{f\text{-}1}} = P_{\text{ACK}}^{(N_{f-1})} + (N_{f-1} - N_f - \lambda)P_{\text{ACK}}^{\prime(N_{f-1})} - P_{\text{ACK}}^{(N_{f-2})}. \quad (12)$$

Thus

$$N_f = N_{f-1} - \lambda + \frac{P_{\text{ACK}}^{(N_{f-1})} - P_{\text{ACK}}^{(N_{f-2})}}{P_{\text{ACK}}^{\prime(N_{f-1})}}. \quad (13)$$

Note from (11) and (13) that the only difference from the original SDO of (4) and (6) is that λ appears explicitly as linear term reducing the transmission lengths that would have been obtained without the cost factor. The net effect *for a fixed value of m and N_{max}* is to increase the values of $N_1, \ldots N_{m-1}$ resulting in larger values of P_{ACK} that reduce the expected number of feedback transmissions at the cost of larger expected blocklength and lower throughput. The larger λ becomes, the more $E[F]$ is reduced and $E[N]$ is increased.

Fig. 4 shows the throughput R_T as a function of the expected number of feedback messages $E[F]$ obtained using (11) and (13) with λ ranging from 10^{-2} to 10^{2} for $N_{\max} = 164.649$ and 192. For each value of λ there is a specific value of N_1^* that minimizes J_λ. Fig. 5 shows N_1^* as a function of λ for $N_{\max} = 164.649$ and 192. The values of N_1^* are slightly higher for $N_{\max} = 192$ but the difference is not visible in Fig. 5. Fig. 6 shows the evolution of the transmission-length

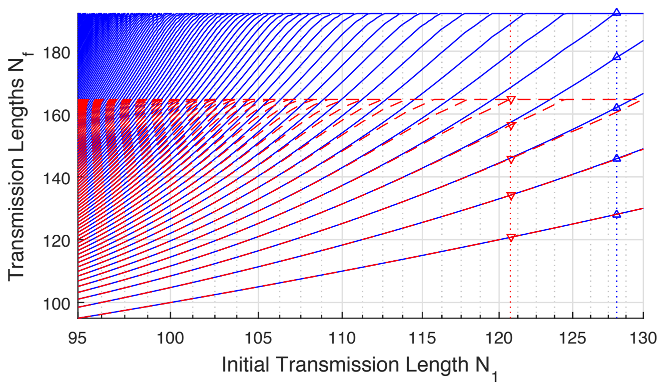

Fig. 6. Cumulative transmission lengths N_f less than 192 prescribed by (11) and (13) as a function of N_1.

sequence $\{N_f\}$ as a function of N_1 according to (11) and (13) using the λ for which that N_1 is optimal.

For a given λ, the range of N_1 values that can be considered is lower bounded; if N_1 is too small the application of (11) or (13) can produce a value of N_f that is smaller than N_{f-1}, which is not meaningful. The optimal N_1^* is very close to this bound, and for the experiments presented here we used the smallest N_1 above this bound for which a transmission length N_f was exactly equal to $N_{\max}$.

IV. Integer Transmission Lengths

The SDO equations produce real-valued transmission lengths which cannot be used in actual systems. The previous work on SDO has introduced the integer constraint as part of the SDO algorithm itself. In this paper, we have retained real values for the SDO implementation. In this section we review the previous approaches and present a dithering approach that uses integer transmission lengths but still approaches the performance of the real-valued SDO solution.

A. Integer N_1; Floor, Ceiling, and Rounding for N_f

Because the transmission lengths ultimately need to be integers, the previous work on SDO always begins with an integer value for N_1. In [20] the integer constraint is included in application of (4) and (6) by replacing the real-valued result with its ceiling and using that integer as the input for the next application of (6).

Recognizing that forcing the real-valued solution to an integer affects the later computations in unpredictable ways, in [16]–[19] a tree-search explores *all possible* sequences of transmission lengths possible by taking the floor and ceiling at each stage and considering what happens when either is used as input to the subsequent application of (6). Thus 2^{m-1} sequences of potential blocklengths $N_1, \ldots, N_m$ are computed for each candidate (integer) value of N_1 that is considered. This algorithm has exponential complexity in m (since 2^{m-1} sets $\{N_f\}$ must be checked for each potential N_1) and has been found in practice to take significant time to execute. Finally, from among these possibilities, the sequence of potential blocklengths $N_1, \ldots, N_m$ that minimizes $E[N]$ (or $E[N]+\lambda E[F]$ in the formulation of this paper) is selected. In [18], this tree search was found to achieve a higher throughput than simply rounding the real-valued solution obtained by starting with an integer N_1.

B. Dithering

As described in [23] for the application of image processing, the idea of dithering is to create the illusion of a value (or color) in between the available quantization points by randomly using quantized values above and below the desired value for the pixels where the unavailable hue is desired. We apply this principle to create the "illusion" of a real-valued N_f by randomly using the integer values just above and just below the desired N_f. Thus when SDO identifies a non-integer N_f^{SDO} the dithered implementation treats N_f as an integer-valued random variable that takes on two possible values

$$N_f = \lfloor N_f^{\text{SDO}} \rfloor = N_f^- \tag{14}$$
$$N_f = \lceil N_f^{\text{SDO}} \rceil = N_f^+ \,, \tag{15}$$

so that $E[N_f] = N_f^{\text{SDO}}$. Defining

$$\Delta_f^+ = N_f^+ - N_f^{\text{SDO}} \tag{16}$$
$$\Delta_f^- = N_f^{\text{SDO}} - N_f^- \tag{17}$$

so that

$$N_f^+ = N_f^{\text{SDO}} + \Delta_f^+ \tag{18}$$
$$N_f^- = N_f^{\text{SDO}} - \Delta_f^- \,, \tag{19}$$

N_f has probability mass function (p.m.f.)

$$P_{N_f}(N_f^-) = \Delta_f^+ \tag{20}$$
$$P_{N_f}(N_f^+) = \Delta_f^- \,. \tag{21}$$

Thus, as expected,

$$E[N_f] = N_f^- \Delta_f^+ + N_f^+ \Delta_f^- = N_f^{\text{SDO}} \tag{22}$$

The expected blocklength obtained using dithered N_f values is different from (but extremely close to) the analytical expression of (2) with the real-valued lengths N_f^{SDO}. Specifically, with the real-valued transmission lengths N_f^{SDO}, the expected blocklength $E[N_{\text{real}}]$ is

$$N_1^{\text{SDO}} P_{\text{ACK}}^{(N_1^{\text{SDO}})} + \sum_{f=2}^{m} N_f^{\text{SDO}} \left[P_{\text{ACK}}^{(N_f^{\text{SDO}})} - P_{\text{ACK}}^{(N_{f-1}^{\text{SDO}})} \right] + N_m^{\text{SDO}} P_{\text{NACK}}^{(N_m^{\text{SDO}})} . \tag{23}$$

Defining

$$P_{\text{ACK}}^{(N_f^{\text{dither}})} = \Delta_f^+ P_{\text{ACK}}^{(N_f^-)} + \Delta_f^- P_{\text{ACK}}^{(N_f^+)} \,, \tag{24}$$

with dithering, the expected blocklength $E[N_{\text{dither}}]$ is

$$\Delta_1^+ N_1^- P_{\text{ACK}}^{(N_1^-)} + \Delta_1^- N_1^+ P_{\text{ACK}}^{(N_1^+)} \tag{25}$$
$$+ \sum_{f=2}^{m} \Delta_f^+ N_f^- \left[P_{\text{ACK}}^{(N_f^-)} - P_{\text{ACK}}^{(N_{f-1}^{\text{dither}})} \right] \tag{26}$$
$$+ \sum_{f=2}^{m} \Delta_f^- N_f^+ \left[P_{\text{ACK}}^{(N_f^+)} - P_{\text{ACK}}^{(N_{f-1}^{\text{dither}})} \right] \tag{27}$$
$$+ \Delta_f^+ N_m^- P_{\text{NACK}}^{(N_m^-)} + \Delta_m^- N_m^+ P_{\text{NACK}}^{(N_m^+)} \,. \tag{28}$$

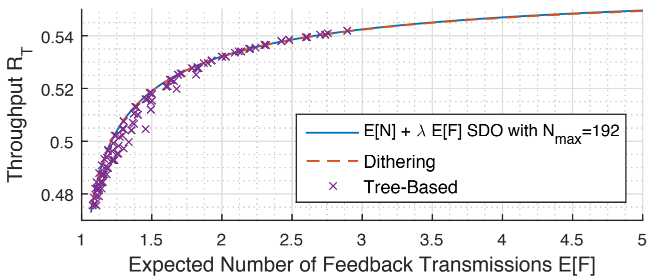

Fig. 7. Throughput R_T vs. expected number of feedback transmissions $E[F]$ for the Lagrangian SDO with real-valued $\{N_f\}$ computed using (4) and (6) and for two techniques to produce integer transmission values: dithering and the exhaustive integer tree search of [18].

For dithering, the expected number of feedback messages is

$$E[F_{\text{dither}}] = m - \sum_{f=1}^{m-1} P_{\text{ACK}}^{(N_f^{\text{dither}})}. \tag{29}$$

Fig. 7 shows that $E[N_{\text{dither}}]$ provides $E[N]$ vs. $E[F]$ performance using integer transmission lengths that is indistinguishable from that the analytical $E[N_{\text{real}}]$ computed using hypothetical real-valued transmission lengths. Moreover, for $1 < E[F] < 2$, $E[N_{\text{dither}}]$ significantly outperforms the extremely complex tree search of all floors and ceilings applied at each step of (4) and (6) starting with an integer N_1. The tree search results are only presented for $E[F] < 2$, which corresponds to $m = 30$, because after that, the compution time was prohibitive. Note that the computational complexity of computing the dithering solution using (20)-(21) is comparatively negligible regardless of $E[F]$ or m.

V. Conclusion

Vakilinia's seminal work in [16], [17] revealed sequential differential optimization (SDO) as a general approach to provide transmission lengths that optimize the throughput of IR-HARQs. This paper takes a fresh look at SDO by reconsidering the real-valued solutions provided by the basic SDO equations and introduced a new set of SDO equations that consider, as a cost, the average number of feedback messages rather than the maximum allowable number of feedback messages. Not surprisingly, explicitly considering the average number of feedback messages as a cost produces a better performance of throughput vs. that cost. Perhaps more surprising was the result that the throughput performance of the original real-valued solution of the SDO equations can be closely approached with integer valued transmission lengths through random dithering of each length, as long as all real values of the initial transmission length N_1 are considered.

References

[1] C.-Y. Lou, B. Daneshrad, and R. D. Wesel, "Convolutional-code-specific crc code design," *IEEE Transactions on Communications*, vol. 63, no. 10, pp. 3459–3470, Oct 2015. [Online]. Available: http://dx.doi.org/10.1109/TCOMM.2015.2459058

[2] P. Koopman and T. Chakravarty, "Cyclic redundancy code (crc) polynomial selection for embedded networks," in *Proc. IEEE Int. Conf. DSN*, Jun. 2004, pp. 145–154.

[3] A. R. Williamson, M. J. Marshall, and R. D. Wesel, "Reliability-output decoding of tail-biting convolutional codes," *IEEE Transactions on Communications*, vol. 62, no. 6, pp. 1768–1778, Jun 2014. [Online]. Available: http://dx.doi.org/10.1109/TCOMM.2014.2319264

[4] A. Raghavan and C. Baum, "A reliability output viterbi algorithm with applications to hybrid arq," *EEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1214–1216, May 1998.

[5] C. Lott, O. Milenkovic, and E. Soljanin, "Hybrid arq: Theory, state of the art and future directions," in *IEEE Inf. Theory Workshop (ITW)*, Bergen, Norway, Jul. 2007.

[6] D. J. Costello, J. Hagenauer, H. Imai, and S. B. Wicker, "Applications of error control coding," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2531–2560, Oct. 1998.

[7] M. Naghshvar, T. Javidi, and M. A. Wigger, "Extrinsic jensen–shannon divergence: Applications to variable-length coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 2148–2164, Apr. 2015.

[8] E. Soljanin, R. Liu, and P. Spasojević, "Hybrid arq with random transmission assignments," *Advances in Network Information Theory*, vol. 66, pp. 321–334, 2004.

[9] N. Varnica, E. Soljanin, and P. Whiting, "Ldpc code ensembles for incremental redundancy hybrid arq," *IEEE Internat. Symp. Inform. Theory*, pp. 995–999, Sep. 2005.

[10] G. Forney, "Exponential error bounds for erasure list and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. 14, no. 2, pp. 206–220, Mar. 1968.

[11] 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, August 2011.

[12] A. M. Cipriano et al., P. Gagneur, G. Vivier, and S. Sezginer, "Overview of arq and harq in beyond 3g systems," *2010 IEEE 21st Inter. Symp. on Personal, Indoor and Mobile Radio Comm. Workshops*, Sep 2010. [Online]. Available: http://dx.doi.org/10.1109/PIMRCW.2010.5671369

[13] A. Masaracchia, R. Bruno, A. Passarella, and S. Mangione, "Analysis of mac-level throughput in lte systems with link rate adaptation and harq protocols," *2015 IEEE 16th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Jun 2015. [Online]. Available: http://dx.doi.org/10.1109/WoWMoM.2015.7158162

[14] E. Cabrera, G. Fang, and R. Vesilo, "Adaptive hybrid arq (a-harq) for ultra-reliable communication in 5g," in *IEEE Vehicular Technology Conference (VTC Spring)*, Sydney, NSW, Australia, June 2017.

[15] A. R. Williamson, T.-Y. Chen, and R. D. Wesel, "Variable-length convolutional coding for short blocklengths with decision feedback," *IEEE Trans. Comm.*, vol. 63, no. 7, pp. 2389 – 2403, July 2015.

[16] K. Vakilinia, S. V. S. Ranganathan, D. Divsalar, and R. D. Wesel, "Optimizing transmission lengths for limited feedback with nonbinary ldpc examples," *IEEE Trans. Comm.*, vol. 64, no. 6, pp. 2245 – 2257, June 2016.

[17] K. Vakilinia, A. R. Williamson, S. V. S. Ranganathan, D. Divsalar, and R. D. Wesel, "Feedback systems using non-binary ldpc codes with a limited number of transmissions," *2014 IEEE Information Theory Workshop (ITW 2014)*, Nov 2014. [Online]. Available: http://dx.doi.org/10.1109/ITW.2014.6970814

[18] N. Wong, K. Vakilinia, H. Wang, S. V. S. Ranganathan, and R. D. Wesel, "Sequential differential optimization of incremental redundancy transmission lengths: An example with tail-biting convolutional codes," in *2017 Information Theory and Applications Workshop (ITA)*, San Diego, CA, Feb. 2017.

[19] H. Wang, N. Wong, A. M. Baldauf, C. K. Bachelor, S. V. S. Ranganathan, D. Divsalar, and R. D. Wesel, "An information density approach to analyzing and optimizing incremental redundancy with feedback," *2017 IEEE International Symposium on Information Theory (ISIT)*, Jun 2017. [Online]. Available: http://dx.doi.org/10.1109/ISIT.2017.8006530

[20] A. Heiderzadeh, J.-F. Chamberland, P. Parag, and R. D. Wesel, "A systematic approach to incremental redundancy over erasure channel," in *IEEE Inter. Symp. on Inform. Theory*, Vail, CO, USA, June 2018.

[21] A. R. Williamson, M. J. Marshall, and R. D. Wesel, "Reliability-output decoding of tail-biting convolutional codes," *IEEE Trans. Comm.*, vol. 62, no. 6, pp. 1768 – 1778, June 2014.

[22] M. Heindlmaier and E. Soljanin, "Isn't hybrid ARQ sufficient?" in *52nd Annu. Allerton Conf. Commun., Control, and Computing (Allerton)*, Sep. 2014, pp. 563–568.

[23] D. Ozdemir and L. Akarun, "Fuzzy algorithms for combined quantization and dithering," *IEEE Trans. Image Proc.*, vol. 10, no. 6, pp. 923–931, June 2001.