Feedback Communication Over the Binary Symmetric Channel with Sparse Feedback Times

Amaael Antonini* Email: amaael@ucla.edu and Richard D. Wesel* Email: wesel@ucla.edu
*Department of Electrical and Computer Engineering, University of California, Los Angeles, CA 90095, USA

***Abstract*—Posterior matching uses variable-length encoding of the message controlled by noiseless feedback of the received symbols to achieve high rates for short average blocklengths. Traditionally, the feedback of a received symbol occurs before the next symbol is transmitted. The transmitter optimizes the next symbol transmission with full knowledge of every past received symbol.**

To move posterior matching closer to practical communication, this paper seeks to constrain how often feedback can be sent back to the transmitter. We focus on reducing the frequency of the feedback while still maintaining the high rates that posterior matching achieves with feedback after every symbol. As it turns out, the frequency of the feedback can be reduced significantly with no noticeable reduction in rate.

***Index Terms*—Posterior matching, binary symmetric channel, noiseless feedback, random coding, sparse-feedback.**

I. Introduction

Consider the problem of communicating a K-bit message Θ over a binary symmetric channel (BSC) with a noiseless feedback channel as depicted in Fig. 1. At each transmission time $t = 1, 2, \ldots, \tau$ the encoder sends binary symbols X_t through the BSC. The decoder receives symbols Y_t that are noisy versions of X_t where $\Pr(Y_t = 1 \mid X_t = 0) = \Pr(Y_t = 0 \mid X_t = 1) = p$. The receiver may choose to send the symbols Y_t to the transmitter immediately, or allow a few symbols to accumulate, sending all the accumulated symbols in a packet. The receiver needs to produce an estimate $\hat{\Theta}$ of Θ using the symbols $Y_1, Y_2, \ldots, Y_\tau$, and the process ends at the stopping time τ when the receiver is sufficiently confident of the estimate $\hat{\Theta}$. The goal is to produce the estimate $\hat{\Theta}$ with a low error probability $\Pr(\hat{\Theta} \neq \Theta)$ bounded by a small threshold and with the smallest possible average number of transmissions and average number of feedback transmission instances.

A. Background

Shannon [1] showed that feedback cannot increase the capacity of discrete memoryless channels (DMC). However, when combined with variable-length coding, Burnashev [2] showed that feedback can help increase the decay rate of the frame error rate (FER) as a function of blocklength. Horstein [3] developed one of the earliest schemes for the BSC with noiseless feedback that used sequential transmission and works well in the short blocklength regime. Shayevitz and Feder [4] later introduced a capacity achieving family of feedback schemes not limited to the BSC, which they called "posterior matching," and showed that it includes Horstein's scheme. Li and El-Gamal [5] proposed a fixed length "posterior matching" scheme that works well for block-lengths over a few thousand bits. A notable variable length "posterior matching" scheme for a general discrete memoryless channel (DMC) with feedback was proposed by Naghshvar *et. al.* [6] and used a sub-martingale analysis to prove that it achieves the channel capacity. Other "posterior matching" schemes include [7]–[12], and more variable length schemes achieving Burnashev's optimal error exponent can be found in [13]–[17].

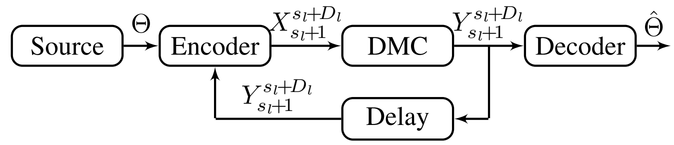

Fig. 1. System diagram of a BSC with full, noiseless feedback. At sparse times $t = s_1, s_2, \ldots, s_\eta$ transmit a block size D_l

All previous schemes are sequential schemes where every feedback symbol Y_t is made available to the transmitter before the next symbol X_{t+1} is encoded. In this paper we study the case where the receiver is allowed wait for a few transmissions before sending the accumulated feedback symbols in a single packet. In the meantime, the encoder encodes the next transmissions using only the feedback symbols received in previous feedback packets. Thus, the transmitter can also send those symbols in a single forward transmission packet. We target the short block regime and allow variable feedback transmission intervals.

B. Contributions

In our precursor journal paper [18], we introduced a new analysis for sequential transmission that simplified encoding and decoding and improved the rate bound over previous results. The contributions of the current paper include the following:

- Show that the same rate bound from [18] is achievable with packet transmissions instead of sending feedback after every symbol.
- Introduce new encoding constraints that are less restrictive and better suited for block transmissions.
- Provide the "look-ahead" encoding algorithm that enforces the new encoding constraints for a few transmissions in advance, to allow the transmission of of a packet

This work was supported by the National Science Foundation (NSF) under Grant CCF-1955660. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect views of NSF.

of symbols, and still guarantees a performance above the lower bound designed for sequential transmission.
- Provide simulation results that show the achievable feedback sparsity, with an average rate that exceeds the lower bounds developed in [18] for sequential transmissions.

The rest of the paper proceeds as follows. Sec. II describes the sparse feedback times system model, introduces the communication problem and describes the communication scheme by Naghshvar *et. al.* [6] on which our methods are based. Sec. III describes performance bounds for the non-sparse, sequential, feedback model and the encoder that achieves the bound from our previous journal paper [18], which we use to benchmark the sparse communication performance. Sec. IV introduces a new encoding constraint that guarantees the bounds in [18] for sequential feedback but allows sparseness in the feedback times, under certain conditions. Sec. IV-B introduces the "look-ahead algorithm" that implements sparse feedback times by encoding several symbols in advance, with the guarantee that the constraints in Sec. IV-A will be met for each transmission. Sec. V shows the performance of the "look-ahead algorithm" in rate, sparsity and complexity from simulations. Sec. VI concludes the paper.

Throughout the paper we denote random variables (RVs) with upper case letter and instances with lower case letters. We consider discrete times with time indexed by $t = 1, 2, \ldots$. Sequences of random variables $X_i, X_{i+1}, \ldots, X_j$ will be denoted by X_i^j, possibly dropping the sub index i if $i = 1$.

II. Posterior Matching System Model

The system model in Fig. 1 consists of a source that samples a message $\Theta \in \Omega$ from a distribution $\mathsf{U}(\Omega)$; an encoder that generates the symbols X_t at each time t; a discrete memoryless channel that transforms the transmitted symbols X_t into received symbols Y_t according to the channel transition function; a noiseless feedback channel; and a decoder that uses the channel symbols Y_t to produce an estimate $\hat{\Theta}$ of the transmitted message Θ.

A. Sparse Feedback Times Problem

The sparse feedback times model allows the receiver to wait a few time indexes between feedback transmission. The received symbols accumulated between feedback transmissions are then sent in a single packet. The time between feedback transmissions could be variable, just like the block size. Let the feedback transmissions times be at times $t = s_1, s_2, \ldots, s_\eta$, with $s_0 = 0$ and $s_\eta = \tau$. Then, at every time $t = s_{l+1}$ the receiver will send the feedback transmissions corresponding to times $s_l+1, s_l+2, \ldots, s_{l+1}$, in a block of size $D_l = s_{l+1} - s_l$, shown by the block $Y_{s_l+1}^{s_l+D_l}$ in Fig. 1.

The sparse feedback times communication problem consists of designing a variable length coding scheme to transmit a K-bit message using the smallest expected number of channel bits τ and the smallest number of feedback transmissions η that guarantees a frame error rate FER bounded by a small threshold ϵ. We note that the expectations $\mathsf{E}[\tau]$ and $\mathsf{E}[\eta]$ cannot be minimized at the same time. To see this, note that the minimum of $\mathsf{E}[\eta]$ is zero, which is achieved by any fixed length, forward error correction scheme that guarantees the FER bound. However, as shown by Burnashev [2] feedback and variable rate coding lower the error exponent, which achieves a target FER with a smaller $E[\tau]$. To formulate the communication problem we need to choose the trade-off between $\mathsf{E}[\tau]$ and $\mathsf{E}[\eta]$.

There are many ways to formulate the problem to account for the trade-off. One way could be with Lagrange multipliers, where we minimize $\mathsf{E}[\tau] + \lambda \mathsf{E}[\eta]$, for some value of λ that could represent the channel access cost, in transmission bits. However, even minimizing $\mathsf{E}[\tau]$ is an integer programming problem whose solution is not yet known to the best of our knowledge. Our approach consists of designing a scheme that aims to minimize $\mathsf{E}[\eta]$ while attaining the expected block-length $\mathsf{E}[\tau]$ that satisfies the bound from [18]. Suppose the bound on $\mathsf{E}[\tau]$ is τ_B, then we can formulate the problem as follows:

$$\text{minimize} \qquad \mathsf{E}[\eta] \tag{1}$$
$$\text{subject to:} \qquad \mathsf{E}[\tau] \le \tau_B,\ \Pr(\hat{\theta} \ne \theta) \le \epsilon \tag{2}$$
$$\text{Sparsity Constraint:} \qquad \mathbf{X}_{s_l+1}^{s_{l+1}} = \mathcal{F}(\theta, \mathbf{y}_1^{s_l})\,. \tag{3}$$

The sparsity constraint restricts the encoder to encode symbols $X_{s_l+2}, X_{s_l+3}, \ldots, X_{s_{l+1}}$ without using the feedback symbols $Y_{s_l+1}, Y_{s_l+2}, \ldots, Y_{s_{l+1}-1}$ not yet re-transmitted by the decoder. Our approach consists of finding an encoding function that guarantees that constraints (2) and (3) are satisfied and seeks to maximize sparsity in the feedback transmission times.

B. Communication Scheme by by Naghshvar et. al.

We propose a communication scheme and encoding algorithm that is based on the single phase transmission scheme proposed by Naghshvar *et. al.* [6], and combines features of the binary and the non-binary symbols. Both encoder and decoder use the channel symbol sequence up to t: $\mathbf{Y}^t = Y_1, Y_2, \ldots, Y_t$ to compute posterior probabilities $\rho_i(y^t)$ and log likelihood ratio $U_i(t)$ for each possible input message i:

$$\rho_i(y^t) \triangleq P(\theta = i \mid Y^t = y^t),\ \forall i \in \{0,1\}^K \tag{4}$$
$$U_i(t) = U_i(Y^t) \triangleq \log_2\left(\frac{\rho_i(Y^t)}{1-\rho_i(Y^t)}\right). \tag{5}$$

To encode the symbol X_{t+1} the encoder partitions the message space Ω into "bins", one for each possible input symbol, using a deterministic method known to the decoder. The encoder then transmits the symbol of the bin containing the transmitted message θ. The process terminates once a posterior crosses the threshold $1-\epsilon$ and the message with this posterior is selected as the estimate. The choice of deterministic partitioning determines the scheme's performance and thus is at the core of the scheme.

In the case of the BSC the encoder by Naghshvar *et. al.* [6], needs to construct 2 sets, S_0 and S_1, at each time t. To construct the sets, Naghshvar *et. al.* [6] proposed a deterministic algorithm, refer to as the "small enough difference" encoder (SED) in [19], because it satisfies the following constraint:

$$0 \le \sum_{i \in S_0} \rho_i(y^t) - \sum_{i \in S_1} \rho_i(y^t) < \min_{i \in S_0} \rho_i(y^t)\,. \tag{6}$$

Naghshvar *et. al.* [6] proved that the SED encoder achieves the channel capacity using extrinsic Jensen-Shannon divergence.

III. Achievable Rate for Sequential Transmission

We now describe the best rate lower bound that, to our knowledge, has been developed for sequential transmission over the BSC with noiseless feedback and a simple encoder that achieves it, from our previous work in [18]. Let ϵ be the requirement on $\Pr(\hat{\Theta} \neq \Theta)$ and let the block-length be given by the stopping time τ defined by:

$$\tau = \min_{t \in \mathrm{N}} \{\exists i \in \Omega : \rho_i(y^t) \geq 1 - \epsilon\} . \tag{7}$$

Let the rate be $K/\mathsf{E}[\tau]$, then a rate lower bound is given by an upper bound on upper bounds on expected block-length $\mathsf{E}[\tau]$. The bound on $\mathsf{E}[\tau]$ from [18] is given in terms of the channel capacity C and the constants C_1 and C_2 from [19]:

$$C \triangleq 1 + p \log_2(p) + (1-p) \log_2(1-p) \tag{8}$$

$$C_2 \triangleq \log_2 \left(\frac{1-p}{p} \right) \tag{9}$$

$$C_1 \triangleq (1-p) \log_2 \left(\frac{1-p}{p} \right) + p \log_2 \left(\frac{p}{1-p} \right) . \tag{10}$$

We proposed the "small enough absolute difference" (SEAD) encoding rule, a relaxed version of the SED encoder:

$$-\min_{i \in S_0} \rho_i(y^t) < \sum_{i \in S_0} \rho_i(y^t) - \sum_{i \in S_1} \rho_i(y^t) \leq \min_{i \in S_0} \rho_i(y^t) , \tag{11}$$

and showed that it suffices to guarantee that for all y^t and for some $a > 0$ the following inequalities hold:

$$\mathsf{E}[U_i(t+1) - U_i(t) | \mathcal{F}_t, \theta = j] \geq a \tag{12}$$

$$U_i(t+1) - U_i(t) \leq C_2 \tag{13}$$

$$E[U_\theta(t+1) - U_\theta(t) | Y^t = y^t] \geq C . \tag{14}$$

If the following singleton constraint is satisfied:

$$U_i(t) \geq 0 \implies S_0 = \{i\} \text{ or } S_1 = \{i\} , \tag{15}$$

we showed in [18] that the following inequalities also hold:

$$U_i(t) \geq 0 \implies \mathsf{E}[U_i(t+1) - U_i(t) | \mathcal{F}_t, \theta = j] = C_1 \tag{16}$$

$$U_i(t) \geq 0 \implies | U_i(t+1) - U_i(t) | = C_2 . \tag{17}$$

In [18] we used a two phase analysis, that divided the transmissions into a communication phase consisting of the times t where $U_\theta(t) \leq 0$ with total time $T \triangleq \sum_{t=1}^{\tau} \mathbb{1}_{U_\theta(t) < 0}$, and a confirmation phase with time $\tau - T$. We constructed a bound τ_{com} on $\mathsf{E}[T]$ from inequalities (12) to (14) and a bound τ_{conf} on $\mathsf{E}[\tau - T]$, with inequalities (16) and (17), given by:

$$\tau_{com} \leq \frac{\log_2(M-1)}{C} + \frac{C_2}{C} + 2^{-C_2} \frac{C_2}{C} \frac{1 - \frac{\epsilon}{1-\epsilon} 2^{-C_2}}{1 - 2^{-C_2}} \tag{18}$$

$$\tau_{conf} \leq \frac{C_2}{C_1} \left(\left\lceil \frac{\log_2(\frac{1-\epsilon}{\epsilon})}{C_2} \right\rceil - 2^{-C_2} \frac{1 - \frac{\epsilon}{1-\epsilon} 2^{-C_2}}{1 - 2^{-C_2}} \right) . \tag{19}$$

Since $\tau = (\tau - T) + T$, we can construct a bound τ_B on $\mathsf{E}[\tau]$ using (18) and (19). However, bound (18) is loose because of the terms with $\frac{C_2}{C}$ that derive from inequalities (13) and (14). In [18] the bound was tighten by constructing a strictly degraded process $U_i'(t)$ that replaced C_2 in (13) with $\frac{\log_2(2q)}{q}$, where $q = 1 - p$. The time T' of the degraded process was lower bounded by that of the original process, that is $T \leq T'$. Replacing C_2 with $\frac{\log_2(2q)}{q}$ in (13) yields an upper bound τ_{com}' on $\mathsf{E}[T']$ that applies to both $U_i(t)$ and $U_i'(t)$, given by:

$$\tau_{com}' \leq \frac{\log_2(M-1)}{C} + \frac{\log_2(2q)}{qC} \left(1 + 2^{-C_2} \frac{1 - \frac{\epsilon}{1-\epsilon} 2^{-C_2}}{1 - 2^{-C_2}} \right) \tag{20}$$

We showed in [18] that. when the source samples Θ uniformly from $\Omega = \{0,1\}^K$, systematic transmissions guarantee that all the constraints are satisfied. At time $t = K$ the posteriors produced by systematic transmissions form a binomial distribution $\mathsf{B}\{0,1\}^K$, which we used in [18] compute a bound τ_{com}^B on $\mathsf{E}[T']$ when $\Theta \sim \mathsf{B}\{0,1\}^K$. Let $\rho_K^h = p^h q^{K-h}$, then τ_{com}^B is given by:

$$\tau_{com}^B \leq \sum_{h=0}^{K} \left[\frac{\log_2 \left(\frac{1-\rho_K^h}{\rho_K^h} \right)}{C} + \frac{\log_2(2q)}{qC} \right] \binom{K}{i} \rho_K^h \mathbb{1}_{(\rho_K^h < 0.5)} + \frac{\log_2(2q)}{qC} 2^{-C_2} \frac{1 - \frac{\epsilon}{1-\epsilon} 2^{-C_2}}{1 - 2^{-C_2}} . \tag{21}$$

In [18], we used $\tau = K + (T - K) + (\tau - T)$ to obtain a tighter bound τ_B on $\mathsf{E}[\tau]$ when Θ is sampled from a uniform distribution on $\{0,1\}^K$, the bound is given by:

$$\mathsf{E}[\tau] \leq \tau_B = K + \tau_{com}^B + \tau_{conf} \tag{22}$$

Bound (22) adds K systematic transmissions to the bound τ_{com}^B on T' for binomial input and the bound τ_{conf} on $\mathsf{E}[\tau - T]$.

IV. Sparse Feedback Times Scheme

We now show that it is possible to satisfy the constraints equations (12) to (14) and (16), (17) to achieve an expected stopping time $\mathsf{E}[\tau]$ upper bounded by τ_B in (22) with some sparsity in the feedback times, i.e. where the feedback is only updated at times $s_1, s_2, \ldots$ and not after every transmission. Thus, the transmitter is restricted to encode symbols $X_{s_l+1}, X_{s_l+2}, \ldots, X_{s_{l+1}}$ using only the feedback sequence up to time s_l, given by $Y_1^{s_l}$. We will exploit systematic transmissions to make the first feedback time s_1 equal to K.

After the systematic transmissions we will use the non-binary version of the scheme proposed by Naghshvar *et al.* [6], where the number of "bins" to partition the message set Ω is the number of symbols in the channel alphabet. We consider the block of D_l bits transmitted from time $t = s_l$ to $t = s_{l+1}$ a single symbol out of an alphabet of 2^{D_l} symbols, and partitions Ω into 2^{D_l} "bins." The symbol $X_{s_l}^{s_l + D_l}$ transmitted at time s_l will be the D_l-bit label assigned to the bin that contains the transmitted message θ, which could just be the index of the bin. The binary partitions at each transmission j from time $t = s_l$ to $t = s_l + D_l$ will be given by assigning to S_0 the "bins" whose label has a 0 at the j-th entry to S_1 "bins" whose label has 1 at the j-th entry. Using this scheme, the problem reduces to finding, at each time s_l, the largest block size D_l for which we can guarantee that all constraints are met at every time $t = s_l + 1, s_l + 2, \ldots, s_l + D_l$.

A. The "Weighted Median Absolute Difference" Rule

We now introduce the "Weighted Median Absolute Difference" rule, a partitioning rule that further relaxes the tolerance in the difference of sums (11), sufficient to guarantee constraints (12) to (14). At each time t let P_0, P_1 and Δ be:

$$\Delta \triangleq \sum_{i\in S_0} \rho_i(y^t) - \sum_{i\in S_1} \rho_i(y^t) \tag{23}$$

$$P_0 \triangleq \Pr(\theta \in S_0 \mid Y^t = y^t) = \sum_{i\in S_0} \rho_i(y^t) = \frac{1+\Delta}{2} \tag{24}$$

$$P_1 \triangleq \Pr(\theta \in S_1 \mid Y^t = y^t) = \sum_{i\in S_1} \rho_i(y^t) = \frac{1-\Delta}{2} \tag{25}$$

Note that $P_0 + P_1 = 1$, and thus $P_0 = \frac{1+\Delta}{2}$ and $P_1 = \frac{1-\Delta}{2}$. Let $\{o_1, \dots, o_M\}$ be an ordering of the vector of posteriors such that $\rho_{o_1}(t) \geq \rho_{o_2}(t) \geq \cdots \geq \rho_{o_M}(t)$, and let m be the index of the "median" posterior defined by:

$$\sum_{i=1}^{m-1} \rho_{o_i}(y^t) < \frac{1}{2} \leq \sum_{i=1}^{m} \rho_{o_i}(y^t)\,. \tag{26}$$

The "Weighted Median Absolute Difference" rule is given by:

$$\Delta^2 \leq \frac{2}{5}\rho_{o_m}(y^t) \tag{27}$$

Rule (27) offers two significant advantages over SEAD and SED: the first is a larger tolerance on Δ, for most times s_l, since $\sqrt{\frac{2}{5}\rho_{o_m}(y^t)}$ is often much larger than $\rho_{o_m}(y^t)$. The second advantage is that the bound on Δ does not depend on which items are in S_0, which allows to allocate items to S_0 and S_1 to tune Δ without affecting the tolerance, unlike SED and SEAD in (6), (11) where changes in the partitioning cause changes in the tolerance.

To guarantee the bound τ_B on $\mathsf{E}[\tau]$ in equation (22) we only need to prove that rule (27) suffices to satisfy constraint (14) and enforce the singleton constraint (15). In [18] we showed that constraint (13) is satisfied by any non-empty S_0 and S_1, and that $|\Delta| \leq 1/3$ suffices to guarantee constraint (12), which can be trivially extended to the any value allowed by rule (27). The proof consists of lower bounding the left side of (14) by a function of only Δ and $\rho_{o_m}(y^t)$ and then showing that (14) holds for any Δ that satisfies (27).

A detailed proof can be found in [20]. We now provide the steps needed in the proof.

1) Use the definition to expand $E[U_\theta(t{+}1){-}U_\theta(t)|Y^t{=}y^t]$:

$$\sum_{i\in\Omega} \rho_i(y^t) E[U_i(t{+}1){-}U_i(t)|Y^t{=}y^t, \theta{=}i] \tag{28}$$

2) Write $U_i(t)$ and $U_i(t{+}1)$ in terms of $\rho_i(y^t)$ and $\rho_i(y^{t+1})$.
3) Write $\rho_i(y^{t+1})$ in terms of $\rho_i(y^t), p, \Delta$ using eq. 120 and 121 from [18].
4) Extract a term C from the sum using eq. 122 to 124 in [18]. From here the problem reduces to showing that the remaining terms combine to a non-negative value.
5) Apply Jensen's inequality over p, q to obtain a lower bound on the sum and write it as two sums over S_0 and S_1, eq: 125 and 126 in [18].
6) Write each sum over S_0 and S_1 as two sums that separate items i with $\rho_i(y^t) \geq \rho_{o_m}(y^t)$ and let R be the fraction of such items in S_0.
7) Remove the dependencies on i and $\rho_i(y^t)$, by replacing the arguments of the $\log_2(\cdot)$ with lower bounds in terms of Δ and $\rho_{o_m}(y^t)$.
8) Define the right sum in (26) as $\frac{1+\delta}{2}$ and express the weights in the four sums in terms of Δ, δ and R using eq. (24), (25) and (26).
9) Use Jensen's inequality over the four weights to lower bound $E[U_\theta(t+1) - U_\theta(t)|Y^t = y^t]$ by an expression of the form $C - \log_2(1{-}f(\Delta, \delta, R, \rho_{o_m}(y^t)))$.
10) Now it remains only to show that $f(\Delta, \delta, R, \rho_{o_m}(y^t)) < 0$. Use the worst case scenario R and δ to upper bound $f(\Delta, \delta, R, \rho_{o_m}(y^t))$ by a function $g(|\,\Delta\,|, \rho_{o_m}(y^t))$.
11) Finally show that if $|\Delta| \leq \sqrt{\frac{2}{5}\rho_{o_m}(y^t)}$ then $g(|\,\Delta\,|, \rho_{o_m}(y^t)) < 0$ for every $0 \leq \rho_{o_m}(y^t) \leq 1$.

B. The "look-ahead" Algorithm

Now we introduce the "look-ahead" algorithm, a method to design, based only on $Y_1^{s_l}$, the partitions for the next few transmissions $s_l{+}1, s_l{+}2, \dots, s_l{+}D_l$ for some D_l. The "look-ahead" algorithm needs to guarantee that constraint (27) is satisfied at each $t{=}s_l{+}1,\ s_l{+}2, \dots, s_l{+}D_l$, for the already received sequence y^{s_l} and for each future possible extension sub-sequence $Y_{s_l+1}^{s_l+j}$, $j = 1, 2, \dots, D_l - 1$. We now identify the key challenges for the "look-ahead" algorithm and the steps that we take to overcome these challenges.

First we note that at any time $t = s_l$, only a few D_l values might be feasible, thus, we need to find one such value before designing the partitions. We know from the non-sparse case that $D_l = 1$ is always feasible and know how to construct two partitions, say using Naghshvar's algorithm [6] or the thresholding algorithm 6 in [18] Sec. VII. Second, the algorithm must always converge to a solution in a finite number of steps, which we desire to be reasonably small. For this reason, we will execute a single attempt for a given D_l, and upon failure, reduce D_l by one before trying again. This procedure could fall back to the non-sparse case where $D_l = 1$. Third, if we fix S_0 and S_1 for next times $t = s_l + 1, s_l + 2, \dots s_l + D_l - 1$, then each future $\rho_{o_m}(y^t)$ and Δ is a random function of $Y_{s_l+1}, Y_{s_l+2}, \dots, Y_{s_l+D_l-1}$, the future received symbols. The "look-ahead" algorithm needs to guarantee that the pair $\rho_{o_m}(y^t)$ and Δ satisfies constraint (27) at the current time s_l any any future time up to $s_l + D_l - 1$.

To overcome these challenges, the "look-ahead" algorithm proceeds as follows: let the 2^{D_l} "bins" at time $t = s_l$, be $\mathcal{E}_k$, $k = 0, 1, \dots, 2^{D_l} - 1$ and define "bin" posteriors $P_{\mathcal{E}_k}$, δ_k, and $\delta_{\max}$ by:

$$P_{\mathcal{E}_k} \triangleq \sum_{i\in\mathcal{E}_k} \rho_i(y^{s_l}),\ \delta_k \triangleq P_{\mathcal{E}_k} - 2^{-D_l};\ \delta_{\max} \triangleq \max_k\{|\delta_k|\}, \tag{29}$$

where 2^{-D_l} is the target posterior for each bin. To overcome the uncertainty on $\rho_{o_m}(y^t)$ and guarantee that constraint (27) on Δ is satisfied at future time $t = s_l + 1, s_l + 2, \dots, s_l + D_l - 1$ the algorithm finds a lower bound $\rho_{o_m}^{\min}(y^t)$ on $\rho_{o_m}(y^t)$ that is used to compute an upper bound $\Delta_{\max}$ on Δ for each

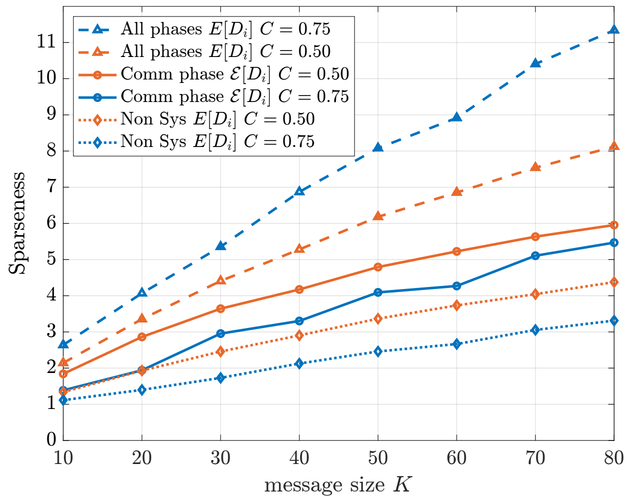

Fig. 2. Feedback sparseness vs. message size K of the "look-ahead" algorithm. The curves show average feedback packet size $\mathsf{E}[D_l]$ vs. K for channels with capacity $C = 0.50$ and $C = 0.75$. The dashed line $--\Delta$ is the overall $\mathsf{E}[D_l]$, the dotted line $\cdot\cdot\Diamond$ excludes the systematic block $D_1 = K$ and the solid line $-\circ$ is the performance for only non-systematic transmissions where $\rho_i(y^t) < 0.5\, \forall i \in \Omega$, the target region of the "look-ahead" algorithm.

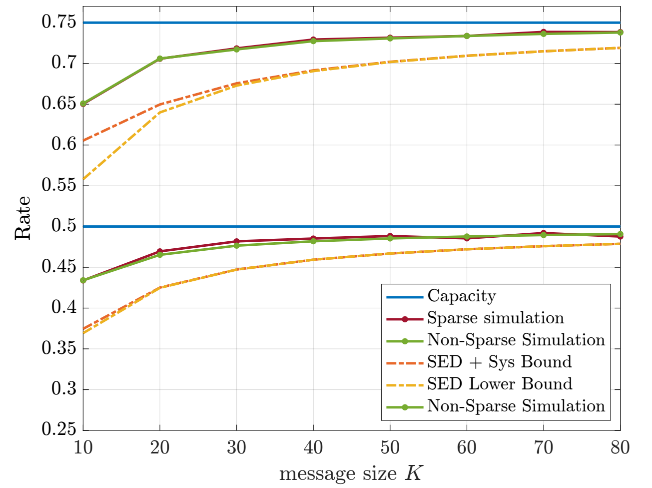

Fig. 3. Rate vs message size K for the look-ahead algorithm over two channels with capacities $C = 0.50$ and $C = 0.75$, horizontal solid blue lines. The "look-ahead algorithm" performance are the brown solid lines $-\bullet$. The green solid lines $-\bullet$ are the non-sparse algorithm from [18]. The orange dashed curves are the lower bounds $K/\mathsf{E}[\tau]$ for systematic transmission in (19) and the yellow dashed line iss are the lower bound of (20) for uniform input distribution.

future time up to $D_l - 1$. The algorithm then uses $\Delta_{\max}$ to determine $\delta_{\max}$ the largest difference δ_k between the posterior $P_{\mathcal{E}_k}$ and the target 2^{-D_l}. Note that at each time $s_l + j$, $j = 0, 1, \ldots, D_l - 1$ each set S_x, $x \in \{0, 1\}$ collects the half of "bins" whose label has x at entry j, Then, Δ at time $s_l + j$ is given by:

$$|\Delta| = |\sum_{\mathcal{E}_k \in S_0} \delta_k - \sum_{\mathcal{E}_k \in S_1} \delta_k| \leq 2^{D_l}\delta_{\max} \tag{30}$$

Since $\rho_{o_m}(y^{t+1})$ depend on Δ at time t, we use an initial $\Delta'_{\max}$ to compute $\rho^{\min}_{o_m}(y^t)$, and then compute bounds $\Delta_{\max}$ on Δ and $\delta_{\max}$ on each δ_k, $t = s_l, s_l+1, \ldots, s_l + D_l - 1$ via:

$$\Delta_{\max} \triangleq \min\{\Delta'_{\max}, \sqrt{\frac{2}{5}\rho^{\min}_{o_m}(y^t)}\},\ \delta_{\max} \triangleq \Delta_{\max} 2^{-D_l} \tag{31}$$

We now explain how to compute $\rho^{\min}_{o_m}(y^t)$. Let $x_1^{D_l}(k)$ be the label of bin $\mathcal{E}_k$, and let $Z_k \triangleq \sum_{l=1}^{j} Y_{s_l+1}^{s_l+j} \oplus x_1^{D_l}(k)$. At each time $t = s_l + j$ the posterior $\rho_i(y^t)$ for $i \in \mathcal{E}_k$ will be:

$$\begin{aligned} \rho_i(y^{s_l+j}) &= \frac{\Pr(Y_{s_l+1}^{s_l+j} = y_{s_l+1}^{s_l+j} \mid Y^{s_l} = y^{s_l}, \theta = i)\rho_i(y^{s_l})}{\sum_{k=0}^{2^{D_l-1}} \Pr(Y_{s_l+1}^{s_l+j} = y_{s_l+1}^{s_l+j} \mid Y^{s_l} = y^{s_l}, \theta \in \mathcal{E}_k) P_{\mathcal{E}_k}} \\ &\geq \frac{2^j q^{j-z_k} p^{z_k} \rho_i(y^{s_l})}{1 + \Delta_{\min}} \geq \frac{2^j q^{j-z_k} p^{z_k} \rho_i(y^{s_l})}{1 + \Delta'_{\min}}, \end{aligned} \tag{32}$$

where (32) follows since $\{Y^{s_l} = y^{s_l}\}$ determines the partitions $\mathcal{E}_k$, $k = 0, 1, \ldots, 2^{D_l} - 1$ and $\{\theta \in \mathcal{E}_k\}$ sets $X_{s_l+1}^{s_l+j} = x_1^j(k)$. A bound $\rho^{\min}_{o_m}(y^t)$ could just be the smallest item on any collection $\mathcal{C} \subset \Omega$ such that $\sum_{i \in \mathcal{C}} \rho_i(y^t) \geq 1/2$. However, we want the largest possible lower bound $\rho^{\min}_{o_m}(y^t)$, thus, we find a collection $\mathcal{C}$ using the items with largest posterior from only the "bins" U_k with larger $q^{j-z_k} p^{z_k}$. We cannot control which bins those will be, but we do know that at time $t = s_l + j$ each value of z_k will be shared by $2^{D_l - j}\binom{j}{z_k}$ "bins". Thus we choose h, a maximum z_k and find a value γ such that:

$$\gamma 2^{-D_l} \sum_{z_k}^{h} 2^{D_l - j} \binom{j}{z_k} 2^j q^{z_k} p^{j - z_k} \geq \frac{1}{2}(1 + \Delta'_{\max}). \tag{33}$$

Now suppose that the posteriors in each bin $\mathcal{E}_k$ are ordered such that $\rho_{o_i^k}(y^{s_l})$ is the i-th largest posterior in $\mathcal{E}_k$, and let $\rho_{o_\gamma^k}(y^{s_l})$ be the value of the posterior $\rho_{o_n^k}(y^{s_l})$ such that:

$$\sum_{i=1}^{n-1} \rho_{o_i^k}(y^{s_l}) < \gamma 2^{-D_l} \leq \sum_{i=1}^{n} \rho_{o_i^k}(y^{s_l}). \tag{34}$$

Then, a candidate bound $\rho^{\min}_{o_m}(y^t)$ on $\rho_{o_m}(y^t)$ at time $t = s_l + j$ is given by the smallest value of $\rho_{o_\gamma^k}(y^{s_l})$, with the worst coefficient $q^{j-h}p^h$, given by:

$$\rho^{\min}_{o_m}(y^{s_l+j}) \triangleq 2^{j-h} p^h \min_{k=0,1,\ldots,2^{D_l-1}} \{\rho_{o_\gamma^k}(y^{s_l})\} \tag{35}$$

Finally, we wish to make the smallest $\rho_{o_\gamma^k}(y^{s_l})$ as large as possible to obtain the largest possible bound $\rho^{\min}_{o_m}(y^{s_l+j})$. For this, the "look-ahead" algorithm distributes largest items $\rho_i(y^{s_l})$ across all "bins" until each crosses the $\gamma 2^{-D_l}$ threshold.

V. Simulation Results

We implemented the "look-ahead" algorithm and obtained performance results to demonstrate how "sparse" the feedback times can be while maintaining a rate above the bounds for the non-sparse case. We show sparsity by the expected size D_l of the "blocks" transmitted at each time s_l $l = 1, 2, \ldots, \eta$. The "sparsity" performance of the "look-ahead" algorithm is provided in Fig. 2 as a function of message size K for two channels with capacity 0.50 and 0.75. The solid line $-\circ$ shows the performance of the "look-ahead" in the communication phase, the target region where each $\rho_i(y^t) < 0.5$ where 2^{D_l}

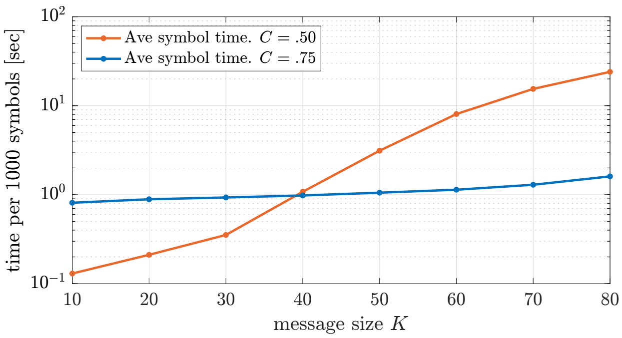

Fig. 4. Run-time complexity of the "look-ahead" algorithm vs. K, in average time per 1000 symbols for channels with capacity $C = 0.50$ and $C = 0.75$.

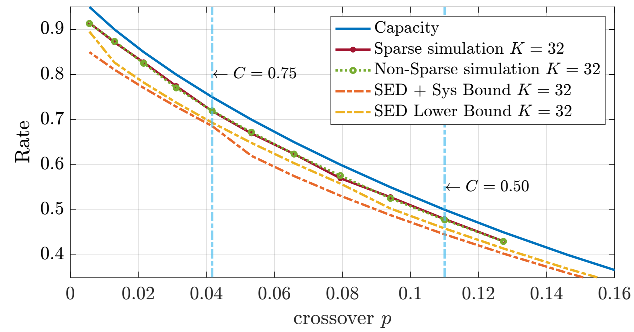

Fig. 5. Rate performance vs. channel p of the look-ahead algorithm for $K = 32$. The solid solid dark blue curve shows the channel capacity. The "look-ahead algorithm" curve is the brown solid line $-\circ$. The green solid line $-\circ$ is for the non-sparse algorithm in [18]. The orange line dash is the rate lower bound $K/\mathsf{E}[\tau]$ for systematic transmission using (19) and the yellow dash line is the lower bound from (20) for uniform input distribution.

"bins" with $D_l > 1$ could be constructed and still satisfy constraint (27). For reference we show the overall $\mathsf{E}[D_l]$ including the systematic block $D_1 = K$ and the average $\mathsf{E}[D_l \mid i \geq 2]$ that includes the times where $\exists i \in \Omega\, \rho_i(y^t) \geq 0.5$. Fig. 3 shows the rate performance of the "look-ahead" algorithm for the same simulations of Fig. 2, and the bounds (20) and (22) that validate the claim that the rate performance is above the bounds. The rate performance of the non-sparse algorithm in [18] is provided for reference, which is no better than that of the "look-ahead" algorithm. The simulations show that as K grows we can increase the sparsity, in the target region, up to an average $\mathsf{E}[D_l]$ of 5 to 6 bits per block.

The run-time complexity of the "look-ahead" algorithm as a function of channel crossover probability p for $K = 16, 32, 64, 96$ is shown in Fig. 4. The complexity curves of the algorithm increases very rapidly with p and with K. To the right the curves seem to taper down, but this is probably artifact introduced by a cap on the largest D_l, which we set at $D_l \leq 12$ because of hardware memory restrictions.

VI. Conclusion

This work explores how the frequency of feedback transmissions affects achievable rate when noiseless feedback of received symbols is used for posterior-matching communication. Although previous works usually assume that each received symbol is fed back before the next transmission, this work shows that the frequency of the feedback can be significantly reduced with no noticeable loss in achievable rate. No feedback is required until after the initial transmission of systematic bits. After that, careful partitioning allows multiple symbols to be transmitted before feedback is required for a new partitioning step.

References

[1] C. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, September 1956.

[2] M. V. Burnashev, "Data transmission over a discrete channel with feedback. random transmission time," *Problemy Peredachi Inf.*, vol. 12, no. 4, pp. 10–30, 1976.

[3] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 9, no. 3, pp. 136–143, July 1963.

[4] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1186–1222, March 2011.

[5] C. T. Li and A. E. Gamal, "An efficient feedback coding scheme with low error probability for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 2953–2963, June 2015.

[6] M. Naghshvar, T. Javidi, and M. Wigger, "Extrinsic Jensen–Shannon divergence: Applications to variable-length coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 2148–2164, April 2015.

[7] J. H. Bae and A. Anastasopoulos, "A posterior matching scheme for finite-state channels with feedback," in *2010 IEEE Int. Symp. Inf. Theory*, 2010, pp. 2338–2342.

[8] V. Kostina, Y. Polyanskiy, and S. Verd, "Joint source-channel coding with feedback," *IEEE Trans. on Inf. Theory*, vol. 63, no. 6, pp. 3502–3515, 2017.

[9] S. Kim, R. Ma, D. Mesa, and T. P. Coleman, "Efficient bayesian inference methods via convex optimization and optimal transport," in *2013 IEEE Int. Symp. on Inf. Theory*, 2013, pp. 2259–2263.

[10] O. Sabag, H. H. Permuter, and N. Kashyap, "Feedback capacity and coding for the bibo channel with a no-repeated-ones input constraint," *IEEE Tran. Inf. Theory*, vol. 64, no. 7, pp. 4940–4961, 2018.

[11] L. V. Truong, "Posterior matching scheme for gaussian multiple access channel with feedback," in *2014 IEEE Inf. Theory Workshop (ITW 2014)*, 2014, pp. 476–480.

[12] A. Anastasopoulos, "A sequential transmission scheme for unifilar finite-state channels with feedback based on posterior matching," in *2012 IEEE Int. Symp. Inf. Theory*, 2012, pp. 2914–2918.

[13] J. Schalkwijk, "A class of simple and optimal strategies for block coding on the binary symmetric channel with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 17, no. 3, pp. 283–287, May 1971.

[14] J. Schalkwijk and K. Post, "On the error probability for a class of binary recursive feedback strategies," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 498–511, July 1973.

[15] A. Tchamkerten and E. Telatar, "A feedback strategy for binary symmetric channels," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2002, pp. 362–362.

[16] A. Tchamkerten and I. E. Telatar, "Variable length coding over an unknown channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2126–2145, May 2006.

[17] M. Naghshvar, M. Wigger, and T. Javidi, "Optimal reliability over a class of binary-input channels with feedback," in *2012 IEEE Inf. Theory Workshop*, Sep. 2012, pp. 391–395.

[18] A. Antonini, R. Gimelshein, and R. Wesel, "Achievable rates for low-complexity posterior matching over the binary symmetric channel," *IEEE Trans. Inf. Theory*, pp. 1–1, 2023.

[19] H. Yang, M. Pan, A. Antonini, and R. D. Wesel, "Sequential transmission over binary asymmetric channels with feedback," *IEEE Tran. Inf. Theory*, 2021. [Online]. Available: https://arxiv.org/abs/2111.15042

[20] A. Antonini, R. Gimelshein, and R. D. Wesel, "Feedback communication over the BSC with sparse feedback times and causal encoding," 2024.