Attaining maximal reliability with minimal feedback via joint channel-code and hash-function design

Stark C. Draper[†], Kannan Ramchandran[†], Bixio Rimoldi[‡],
Anant Sahai[†], and David N. C. Tse[†]
[†] Department of EECS, University of California, Berkeley, CA, USA
[‡] Mobile Communications Laboratory, EPFL, Lausanne, Switzerland
{sdraper,kannanr,sahai,dtse}@eecs.berkeley.edu
bixio.rimoldi@epfl.ch

Abstract

Feedback can be used to make substantial increases in the reliability of communications over discrete memoryless channels (DMCs) when coupled with a variable-length coding strategy. Burnashev has bounded the maximum achievable reliability, and shown how to achieve that reliability assuming noiseless output feedback. We explore how much feedback is required to achieve the Burnashev reliability. A first step was made by Yamamoto-Itoh who presented an asymptotically optimal scheme using only instantaneous noiseless decision feedback. We reduce the required feedback rate by feeding back only a hash of the decision. We then then further improve the scheme through joint design of the channel-code and hash-function. The result is a strategy that, depending on the feedback rate available, transitions smoothly from the Forney reliability at zero-rate to the Burnashev reliability.

1 Introduction

Feedback can be used to make great increases in the reliability of communication systems. This means that to attain a target probability of error, we can design systems with much shorter block lengths than can be achieved without feedback. In this paper we explore how much feedback is required to achieve these results. In particular, many feedback strategies rely on output feedback, i.e., feeding back the receiver's actual observations. While this is the most informative type of feedback, it is not a realistic model. Because of the availability of low-rate control-channel feedback in many communication systems, we explore what can be done when only low-bit-rate noiseless feedback is available.

Without feedback the maximum reliability (error exponent) of a DMC is upper-bounded by the sphere-packing bound. This bound $E_{sp}(R, P, W) \triangleq \min_{V:I(P,V)\leq R} D(V\|W|P)$ is defined by the code rate R, the input distribution P, the channel law W, and a worst-case channel behavior V. (See [3] for a summary of notation.) The sphere-packing bound also bounds the reliability of block-codes with feedback [4]. To realize the promised reliability gains, we relax the fixed block-length constraint, and instead consider variable-length strategies and average block-length (equivalently, average rate) constraints.

For DMCs with output feedback and an average rate constraint, we take as a point of reference a result by Burnashev [1]. Burnashev shows that the error exponent of such schemes is upper bounded by:

$$E_{burn}(\bar{R}) \leq C_1 \left(1 - \bar{R}/C\right), \quad 0 \leq \bar{R} \leq C, \tag{1}$$

where $\bar{R}$ is the expected transmission rate, C is the capacity of the forward channel, and C_1 is determined by the two "most distinguishable" input symbols as $C_1 \triangleq \max_{i,j} \sum_l p_{li} \log(p_{li}/p_{lj})$, where p_{li} is the probability of receiving output symbol l when symbol i is transmitted. We use a_{i^*} and a_{j^*} to denote the maximizing input symbols.

In [9] Yamamoto and Itoh demonstrate that the Burnashev reliability can be achieved with decision rather than observation feedback. In this setting, after transmission of an initial block-code, the decoder informs the encoder via the feedback link of its best guess or "tentative decision". The transmitter then confirms correct decisions with an `ACK` codeword and denies incorrect ones with a `NAK`. The `ACK` (`NAK`) codeword is a repetition of the a_{i^*} (a_{j^*}) symbol. The relative lengths of the initial block code and the confirm/deny phase dictate the average rate of the scheme. Errors only occur when a `NAK` is mis-detected as an `ACK`, a binary hypothesis test with exponent C_1.

In the opposite extreme when only a single bit can be fed back (i.e., zero-rate feedback), in [6] Forney describes an erasure-decoding scheme that also beats the sphere-packing bound. Roughly, if the maximum likelihood codeword isn't sufficiently more likely than the rest of the codewords, the decoder asks for a retransmission. While not achieving the Burnashev bound for most channels (the BEC is an exception), Forney's scheme improves hugely over the no-feedback case. He shows that for totally symmetric channels[1], a lower bound on the zero-rate feedback error exponent is

$$E_{forn}(\bar{R}) = E_{sp}(\bar{R}) + C\left(1 - \bar{R}/C\right), \quad 0 \leq \bar{R} \leq C. \tag{2}$$

Both these strategies make large improvements in reliability, but their underlying philosophies are markedly different. While in Forney's scheme the decision whether to retransmit is made by the decoder, in Yamamoto-Itoh it is made by the encoder. Certain error events that cannot be detected by Forney's strategy can be detected by Yamamoto-Itoh, but at the cost of a higher feedback rate. The approach we take is a hybrid of these two philosophies. Depending on the parameters of the strategy, both Forney and Yamamoto-Itoh are special cases of our scheme. More generally, our approach strikes a balance between the work done by the encoder and the decoder.

2 Hashing the decision

In this section we present an initial modification of the Yamamoto-Itoh coding strategy that reduces the feedback required to achieve the Burnashev reliability. Instead of sending back the tentative decision, we send back a decision hash. Each hash corresponds to a subset, or bin, of messages. This introduce a second source of error — when an incorrect tentative decision and the true message share the same hash – that must be balanced with the probability of a `NAK` being mis-detected as an `ACK`.

Formally, our strategy work in four stages, two feed-forward and two feedback. In the first "data-transmission" stage, a message $\mathsf{m} \in \{1, 2, \ldots M\}$ is sent over the forward channel in λn channel uses. The decoder decodes to the most likely message $\hat{\mathsf{m}}$, which is the tentative decision. In the second "decision-feedback" stage, the M messages are

[1] Forney's bound applies to other channels as well, but has a slightly more complex form as the capacity-achieving input distribution need not also maximize the sphere-packing bound at all rates.

partitioned into $M_{fb} = \exp\{nR_{fb}\}$ bins $\mathcal{B}_1 \ldots \mathcal{B}_{M_{fb}}$. The receiver feeds back the bin index k such that $\hat{m} \in \mathcal{B}_k$.[2] In the third "confirmation" stage, if $m \in \mathcal{B}_k$ the transmitter uses the forward channel $(1-\lambda)n$ times to send an ACK, else it sends a NAK. The ACK (NAK) codeword is the symbol a_{i^*} repeated $(1-\lambda)n$ times. Finally, in the fourth "synchronization" stage the receiver makes a binary decision — whether an ACK or a NAK was sent. It feeds back a single bit indicating which of the two possibilities it detected. If an ACK is detected, both transmitter and receiver start a new message. If a NAK is detected, both prepare for a repeated attempt to transmit the current message. By both abiding by the value of this bit, encoder and decoder stay synchronized. When the feedback channel is noisy, protecting this bit becomes particularly crucial, see [8].

The expected transmission rate $\bar{R}$ is determined by the number of length-n transmissions. This is geometrically distributed with mean $1/(1 - \Pr[\text{retransmission}])$ giving,

$$\begin{aligned}\Pr[\text{retransmission}] &= p_e[p_h p_{a\to n} + (1-p_h)p_{n\to n}] + (1-p_e)p_{a\to n} \\ &\leq 2p_e + p_{a\to n} \leq 2\exp\{-\lambda n E_r(\log M/\lambda n)\} + \delta,\end{aligned}$$

where the block-code has error probability p_e, and exponent $E_r(\cdot)$. The probability of a hashing collision, p_h, is the probability that $\hat{m} \neq m$ and $\hat{m} \in \mathcal{B}(m)$, where $\mathcal{B}(m)$ is the bin of the transmitted message. The probability that an ACK (NAK) is sent and is mis-detected as a NAK (ACK) is denoted $p_{a\to n}$ ($p_{n\to a}$). As we discuss below, by Stein's lemma, the constant δ can be selected as small as desired. Thus, as long as $\log M/\lambda n < C$, the average throughput can be made to approximate $\log M/n$ as closely as desired.

Communication errors result from mis-detecting NAKs or from hash collisions:

$$\Pr[\text{error}] = \begin{cases} p_e[p_h p_{a\to a} + (1-p_h)p_{n\to a}] \;\; \leq p_e(p_h + p_{n\to a}) & \text{if} \;\; R_{fb} < \log M/n, \\ p_e p_{n\to a} & \text{if} \;\; R_{fb} \geq \log M/n, \end{cases}$$

where $p_h = 0$ when feedback rate exceeds transmission rate (i.e., decision feedback).

When $R_{fb} < \log M/n$ and codewords are independently and uniformly assigned to bins, $p_h = 1/M_{fb} = \exp\{-nR_{fb}\}$. Furthermore, $p_{n\to a} = \exp\{-(1-\lambda)nC_1\}$ results from a direct application of Stein's Lemma [2] where an upper bound $\delta > 0$ is set on $p_{a\to n}$. Finally, setting the block code rate just below capacity, gives $\bar{R} \simeq \lambda C$ since the probability of retransmission can then be set arbitrarily small for large n. Hence,

$$\Pr[\text{error}] \leq p_e[\exp\{-nR_{fb}\} + \exp\{-nC_1(1-\bar{R}/C)\}]. \tag{3}$$

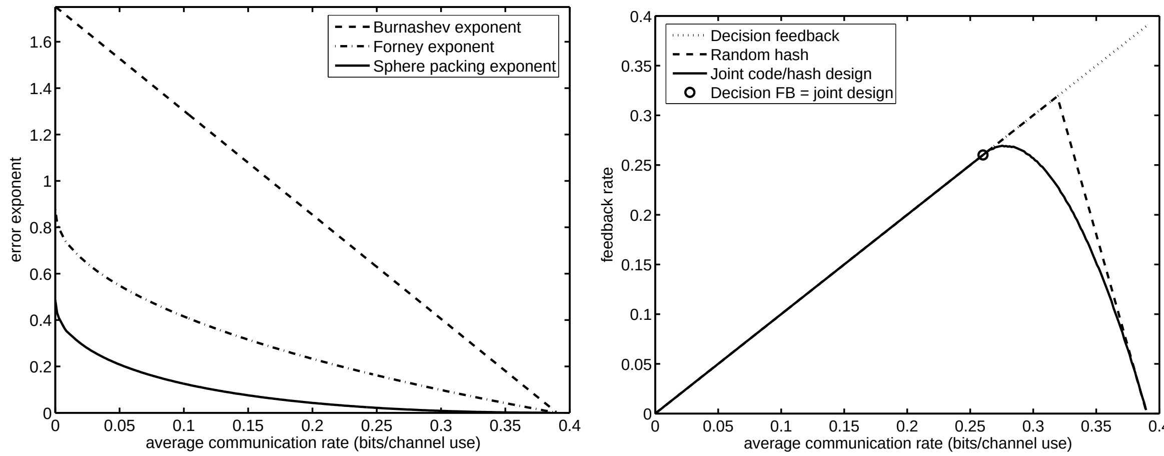
The feedback rate R_{fb} this scheme requires to attain the Burnashev exponent equals that which balances the two sources of error in (3), giving

$$R_{fb} = \min\{\bar{R}, C_1(1-\bar{R}/C)\}. \tag{4}$$

The first argument is the decision-feedback bound, which can be considered a good non-random joint channel-code and hash-function design that assigns one codeword per bin. In the next section we show how to generalize this joint design to other feedback rates.

In Fig. 1 we illustrate these results for a binary symmetric channel. We plot the Burnashev, Forney, and sphere-packing exponents in Fig. 1. In Fig. 1-(b) we compare the feedback rate required to attain the Burnashev exponent by the random hashing strategy (dashed line), by decision feedback (dotted line), and by the joint channel-code and hash-function design that we discuss next.

[2] Note that decision feedback is "bursty", occurring instantaneously at discrete decision times. Without hashing this burstiness can be smoothed out using streaming techniques as in [8]. Because such techniques chop up the block-code into a number of shorter-length codes, it is not clear yet whether similar techniques work in settings with lower feedback rates.



(a) Error exponents vs $\bar{R}$

(b) Feedback rate R_{fb} required vs $\bar{R}$

Figure 1: Figure 1-(a) compares the Burnashev, Forney, and sphere-packing exponents, while Figure 1-(b) compares the feedback rates required by the various schemes to attain the Burnashev exponent. The point where the schemes all require the same feedback rate is noted by the circle. In all cases the channel is a binary symmetric channel with cross-over probability 0.15 and capacity $C = 0.39$ bits per channel use.

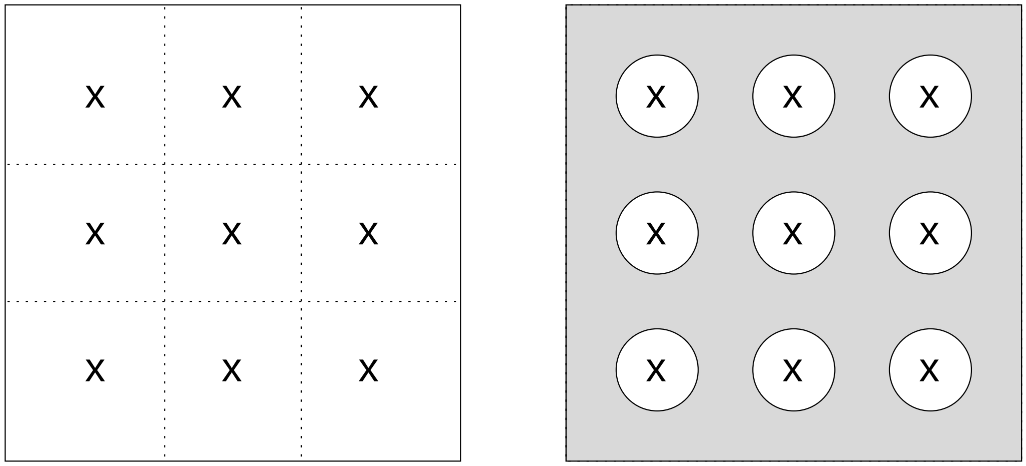
3 Joint design of channel-code and hash-function

In order to reduce the likelihood of undetected hashing errors at a given feedback rate, we design the hash functions used in the Yamamoto-Itoh plus hashing strategy to take into account the geometry of the code. Each bin of codewords – those corresponding to a single hash – themselves form a good low-rate "subcode". This reduces the likelihood that a codeword with a given hash is mis-decoded to another codeword with the same hash. We pair this code design with an erasure-type decoding rule to get a strategy that can transition smoothly between Forney reliability at zero-rate feedback and Burnashev reliability at higher rate feedback.

3.1 Why use erasure decoding?

To understand why an erasure decoding rule is appropriate, we first contrast the decision regions of maximum-likelihood (ML) decoding – plotted in Fig. 2-(a) – with those of erasure decoding – plotted in Fig. 2-(b). The decoding regions of ML fully cover the output space so a decision is always made. On the other hand, in erasure decoding, the output space is not fully covered. When the ML codeword isn't sufficiently more likely than the other codewords, the decoder instead asks for a retransmission. Errors can only occur if the noise is such that another codeword is much more likely than the transmitted one. A tunable threshold controls this comparison and allows the designer to trade off the probabilities of erasure and decoding error. In terms of Fig. 2-(b), the threshold varies the ratio of decoding volume (white) to erasure volume (gray).

One way to improve upon Forney's performance is to detect decoding errors at the encoder as in Yamamoto-Itoh. In Yamamoto-Itoh plus hashing, each bin of codewords constitute a lower-rate subcode. This is illustrated in Fig. 3-(a), where the decoding regions of all codewords in a particular subcode are shaded similarly. If we assign code-



(a) The ML decision regions cover the output space. A decision is made after each block transmission.

(b) The erasure region is shaded gray. Observations that land in this region trigger retransmissions.

Figure 2: Comparison of maximum-likelihood (ML) and erasure-decoding regions.

words to subcodes (bins) in an independent and uniform manner, we get the performance results of Section 2. To improve upon this, we make sure that the codewords that make up each subcode themselves form a good code. A lattice code and its cosets would be an example of such a code and its subcodes. Consider the error events if a codeword from the "white" bin of Fig. 3-(a) is sent. Unless the observation ends up in one of the other white decoding regions, it has landed either in the decoding region of a codeword with a different hash, or in the erasure region. In either case a `NAK` is sent. Undetectable errors only occur if the observation lies in the decoding region of another codeword in the same subcode. Such an error event is depicted in Fig. 3-(b). As can be seen by comparison with Fig. 2-(b), the probability of this event is just the probability of error in an erasure-decoding problem working on the lower-rate subcode.

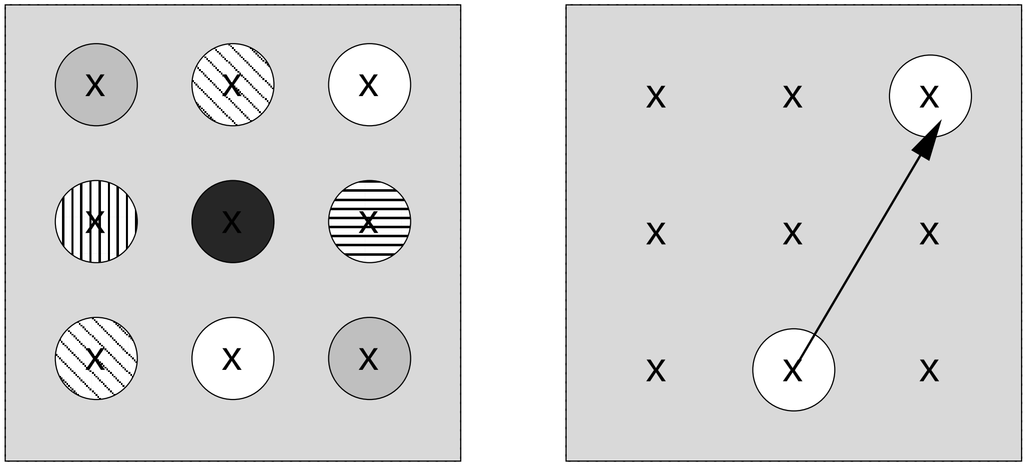
3.2 Decoding rule, feedback messages, and code design

Our decoding rule entails checking two criteria. To be the tentative decision, a codeword must satisfy one condition with respect to all other codewords in its own subcode, and a second, weaker condition, with respect to all other codewords in the mother code. If no codeword satisfies both, the decoder declares an erasure. Because of the asymmetry of the decoding rule, at most one codeword satisfies both. Each criterion is itself an erasure-type decoding comparison. We use Telatar-like [5] comparisons:

Definition 1 *The tentative decision is defined as:*

$$\phi(\mathbf{y}) \triangleq \begin{cases} i & \text{if} \quad D(V_i \| W | P) + |I(P, V_j) - (R - R_{fb})|^+ \leq T \quad \text{for all} \quad j \neq i, j \in \mathcal{B}(i) \\ & \text{and} \quad D(V_i \| W | P) + |I(P, V_j) - R|^+ \leq T \quad \text{for all} \quad j \neq i \\ e & \text{else} \end{cases}$$

where V_i is defined by $\mathbf{x}_i$ and $\mathbf{y}$ as $\mathbf{y} \in \mathcal{T}_{V_i}(\mathbf{x}_i)$, $\mathcal{B}(i)$ is the set (or bin) of codewords to which $\mathbf{x}_i$ belongs, and T is a constant threshold to be specified.



(a) The decision regions of erasure plus hashing

(b) Undetectable decoding error of erasure plus hashing.

Figure 3: In erasure plus hashing, the probability of an undetected hashing error is dictated by the erasure-decoding properties of each subcode.

We use $V_i \prec\!\!\prec V_j$ to denote a pair of codewords that satisfy the first, stronger condition, i.e., $D(V_i \| W | P) + |I(P, V_j) - (R - R_{fb})|^+ < T$. We use $V_i \prec V_j$ to denote a pair of codewords that satisfy the second condition. To denote a pair of codewords that do not satisfy these conditions, we use $\not\prec\!\!\not\prec$ and $\not\prec$, respectively.

The first condition is just Telatar's rule working on the low-rate subcode, as diagrammed in Fig. 3-(b). We use the second condition to bound the probability that the tentative decision ends up in some bin other than the one transmitted. While this probability can be shown to be small using only the first condition, the second criterion improves performance for low feedback rates so by varying the feedback rate R_{fb} and the threshold T, we can transition smoothly between the Forney and Burnashev reliabilities.

The feedback message $\psi(\mathbf{y})$ is either the index of the bin in which the tentative decision lies or an erasure message $\psi(\mathbf{y}) = e$, whenever $\phi(\mathbf{y}) = e$. Thus,

Definition 2 *(Feedback message)*

$$\psi(\mathbf{y}) \triangleq \begin{cases} k & \textit{if} \quad \phi(\mathbf{y}) = i \quad \textit{where} \quad i \in \mathcal{B}_k \\ e & \textit{else.} \end{cases} \tag{5}$$

The following lemma, an extension of Csiszár and Körner's packing lemma (lemma 2.5.1 of [3]), states that a "good" code of $M = \exp\{nR\}$ codewords exists that is the union of $M_{fb} = \exp\{nR_{fb}\}$ "good" lower-rate codes of roughly $\exp\{n(R - R_{fb})\}$ codewords each.

Lemma 1 *(Subcode packing) For every $R > R_{fb} > 0$, $\delta > 0$ and every type P of sequences in $\mathcal{X}^n$ satisfying $H(P) > R$ there exists at least $\exp\{n(R-\delta)\}$ distinct sequences $\mathbf{x}_i \in \mathcal{X}^n$ of type P, grouped into $\exp\{nR_{fb}\}$ subsets of size roughly $\exp\{n(R - R_{fb} - \delta)\}$, such that for every pair of stochastic matrices $V : \mathcal{X} \to \mathcal{Y}$, $\hat{V} : \mathcal{X} \to \mathcal{Y}$, and every i,*

$$\textit{(i)} \quad \left| T_V(\mathbf{x}_i) \cap \bigcup_{j \neq i} T_{\hat{V}}(\mathbf{x}_j) \right| \leq |T_V(\mathbf{x}_i)| \exp\{-n|I(P, \hat{V}) - R|^+\} \tag{6}$$

$$(ii)\quad \left| T_V(\mathbf{x}_i) \cap \bigcup_{j \neq i, j \in \mathcal{B}(i)} T_{\hat{V}}(\mathbf{x}_j) \right| \leq |T_V(\mathbf{x}_i)| \exp\{-n|I(P,\hat{V}) - (R - R_{fb})|^+\} \tag{7}$$

provided that $n \geq n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$*, where* $\mathcal{B}(i)$ *denotes the subset (or "bin") of messages to which message* i *belongs, and where* $|\cdot|^+ \triangleq \max\{0, \cdot\}$.

Without condition (ii) of (7), this would just be the packing lemma. Our version adds a second condition that each codeword must satisfy. We omit the proof.

3.3 Error event analysis

We now bound the various events needed to analyze the system probability of error and the expected communication rate. The derivations are similar to those given in [5].

Probability of undetected error: An undetected decoding error occurs when $\phi(\mathbf{y}) = \hat{m} \neq m$, $\hat{m} \neq e$, and $\hat{m} \in \mathcal{B}(m)$. By the decoding rule, the only way this can occur is if there is a message index $\hat{m} \neq m$ such that $\hat{m} \in \mathcal{B}(m)$ and where the observation

$$\mathbf{y} \in \mathcal{T}_V(\mathbf{x}_m) \cap \mathcal{T}_{\hat{V}}(\mathbf{x}_{\hat{m}}) \quad \text{with} \quad D(\hat{V}\|W|P) + |I(P,V) - (R - R_{fb})|^+ \leq T. \tag{8}$$

The probability of this event can be bounded in the same way as the error probability of a rate-$(R - R_{fb})$ code in erasure-decoding (e.g., in [5]).

$$\begin{aligned}
&\Pr[\text{undetected error}|\mathbf{x}_m] = W^n\left(\{\mathbf{y} : \phi(\mathbf{y}) = \hat{m}, \hat{m} \neq m, \hat{m} \neq e, \hat{m} \in \mathcal{B}(m)\}|\mathbf{x}_m\right) \\
&\leq \sum_{\substack{V,\hat{V} : \hat{V} \prec\!\!\prec V, \\ PV = P\hat{V}}} W^n\left(\mathcal{T}_V(\mathbf{x}_m) \cap \bigcup_{\hat{m} \neq m, \hat{m} \in \mathcal{B}(m)} \mathcal{T}_{\hat{V}}(\mathbf{x}_{\hat{m}}) \middle| \mathbf{x}_m \right) && (9) \\
&\leq \sum_{\substack{V,\hat{V} : \hat{V} \prec\!\!\prec V, \\ PV = P\hat{V}}} \sum_{\mathbf{y} \in \mathcal{T}_V(\mathbf{x}_m) \cap \bigcup_{\hat{m} \neq m, \hat{m} \in \mathcal{B}(m)} \mathcal{T}_{\hat{V}}(\mathbf{x}_{\hat{m}})} \exp\{-n[D(V\|W|P) + H(V|P)]\} \\
&\leq \sum_{\substack{V,\hat{V} : \hat{V} \prec\!\!\prec V, \\ PV = P\hat{V}}} \exp\{-n[D(V\|W|P) + |I(P,\hat{V}) - (R - R_{fb})|^+]\} && (10) \\
&\leq \exp\{-n[E_T(R, R_{fb}, P, W) - \gamma_n]\}. && (11)
\end{aligned}$$

In (9) we enumerate the $\mathbf{y}$ by V-shells. In (10) we apply property (ii) of Lemma 1. In (11) γ_n results from the sum over V-shells so $\lim_{n\to\infty} \gamma_n = 0$. We define the error exponent as

$$E_T(R, R_{fb}, P, W) \triangleq \min_{V,\hat{V}: PV = P\hat{V},\, D(\hat{V}\|W|P) + |I(P,V) - (R - R_{fb})|^+ \leq T} [D(V\|W|P) + |I(P,\hat{V}) - (R - R_{fb})|^+], \tag{12}$$

where the constraint on the minimization follows from (8). Since the probability in (11) does not depend on the index m of the transmitted message, it is upper bounds the probability of undetected error for every message.

Probability of erasure or detected errors: The encoder knows it should retransmit when either the tentative decision is an erasure, or the hash of the tentative decision does not match the hash of the message sent. When the former event occurs, no `NAK` is required, while one is required when the hashes do not match. Similar to the bounding of analogous events in [6, 5] we upper bound the probability of either event by the probability that the tentative decision does not equal the message sent. There are two ways this can happen: if there is some $\hat{m} \in \mathcal{B}(m)$ such that $V_m \not\prec\!\!\not\prec V_{\hat{m}}$, or some $\tilde{m} \notin \mathcal{B}(m)$ such that $V_m \not\prec V_{\tilde{m}}$.

$$
\begin{aligned}
&\Pr[\text{erasure or detected error}|\mathbf{x}_m] \le \Pr[\phi(\mathbf{y}) \neq m|\mathbf{x}_m] \\
&\le \sum_{\substack{V,\hat{V}\,:\,V \not\prec\!\!\not\prec \hat{V} \\ PV = P\hat{V}}} W^n\left(\mathcal{T}_V(\mathbf{x}_m) \cap \bigcup_{\hat{m}\neq m, \hat{m}\in\mathcal{B}(m)} \mathcal{T}_{\hat{V}}(\mathbf{x}_{\hat{m}}) \middle| \mathbf{x}_m\right) + \sum_{\substack{V,\tilde{V}\,:\,V \not\prec \tilde{V} \\ PV = P\tilde{V}}} W^n\left(\mathcal{T}_V(\mathbf{x}_m) \cap \bigcup_{\tilde{m}\notin\mathcal{B}(m)} \mathcal{T}_{\tilde{V}}(\mathbf{x}_{\tilde{m}}) \middle| \mathbf{x}_m\right) \\
&\le \sum_{\substack{V,\hat{V}\,:\,V \not\prec\!\!\not\prec \hat{V} \\ PV = P\hat{V}}} W^n\left(\mathcal{T}_V(\mathbf{x}_m) \cap \bigcup_{\hat{m}\neq m, \hat{m}\in\mathcal{B}(m)} \mathcal{T}_{\hat{V}}(\mathbf{x}_{\hat{m}}) \middle| \mathbf{x}_m\right) + \sum_{\substack{V,\tilde{V}\,:\,V \not\prec \tilde{V} \\ PV = P\tilde{V}}} W^n\left(\mathcal{T}_V(\mathbf{x}_m) \cap \bigcup_{\tilde{m}\neq m} \mathcal{T}_{\tilde{V}}(\mathbf{x}_{\tilde{m}}) \middle| \mathbf{x}_m\right) \\
&\le \sum_{\substack{V,\hat{V}\,:\,V \not\prec\!\!\not\prec \hat{V}, \\ PV = P\hat{V}}} \exp\{-n[D(V\|W|P) + |I(P,\hat{V}) - (R - R_{fb})|^+]\} \\
&\qquad + \sum_{\substack{V,\tilde{V}\,:\,V \not\prec \tilde{V}, \\ PV = P\tilde{V}}} \exp\{-n[D(V\|W|P) + |I(P,\tilde{V}) - R|^+]\} &(13) \\
&\le \sum_{V,\hat{V}} 2\exp\{-nT\} \le 2\exp\{-n[T - \gamma_n]\}. &(14)
\end{aligned}
$$

The steps leading to (13) are similar to those leading to (10) where we now use both of the code properties of Lemma 1. To get (14) we note that if $V \not\prec\!\!\not\prec \hat{V}$ or $V \not\prec \tilde{V}$ then the respective inner arguments of the exponents must each be greater than T.

Probability of a `NAK`: A `NAK` is sent when the hash of the tentative decision doesn't match the transmitted message's hash. For this to happen there must be an observation $\mathbf{y}$ and a message index from a different bin, $\hat{m} \notin \mathcal{B}(m)$, such that[3]

$$
\mathbf{y} \in \mathcal{T}_V(\mathbf{x}_m) \cap \mathcal{T}_{\hat{V}}(\mathbf{x}_{\hat{m}}) \quad \text{with} \quad D(\hat{V}\|W|P) + |I(P,V) - R|^+ \le T.
$$

We bound this event's probability in the same way we bounded the probability of undetected error in (11), but now use property (i) of Lemma 1 instead (ii).

$$
\Pr[\texttt{NAK}|\mathbf{x}_m] \le \sum_{\substack{V,\hat{V}\,:\,\hat{V} \prec V \\ PV = P\hat{V}}} W^n\left(\mathcal{T}_V(\mathbf{x}_m) \cap \bigcup_{\hat{m}\neq m} \mathcal{T}_{\hat{V}}(\mathbf{x}_{\hat{m}}) \middle| \mathbf{x}_m\right) \le \exp\{-n[E_T(R,0,P,W) - \gamma_n]\}. \quad (15)
$$

In (15) we include all non-transmitted codewords, not just those in other bins.

3.4 Expected rate and reliability

To apply the bounds developed in Sec. 3.3 we recall that in the first stage of the Yamamoto-Itoh scheme a block-code of length λn is used. Thus, when using the bounds just derived – (11), (14), and (15) – we use a code of block-length λn and rate $R_\lambda = \log M/\lambda n$. Since the feedback rate R_{fb} is defined with respect to the full block-length n (i.e., bits of feedback per forward-channel use), when calculated with respect to a block-length λn, the resulting feedback rate is R_{fb}/λ. With these substitutions we derive bounds on the expected transmission rate and the reliability of the new scheme.

[3]For our bound we do not need to use the additional more stringent constraints that $\mathbf{x}_{\hat{m}}$ must satisfy with respect to all other codewords in $\mathcal{B}(\mathbf{x}_{\hat{m}})$.

Expected transmission rate: Retransmissions occur if the fed-back message is an erasure, or if it leads to a detected error (which is then NAKed), or if a transmitted ACK is mis-detected as a NAK. We bound the probability of retransmission as

$$\Pr[\text{retransmit}] \leq \Pr[\phi(\mathbf{y}) \neq m | \mathbf{x}_m] + \Pr[\phi(\mathbf{y}) = m | \mathbf{x}_m] p_{a\to n} \leq 2\exp\{-\lambda n[T - \gamma_n]\} + \delta, \quad (16)$$

which follows from using (14) with block-length λn and $p_{a\to n} < \delta$ where δ can be picked arbitrarily small, as in Section 2. This bounds the expected transmission rate $\bar{R}$:

$$\lambda R_\lambda \geq \bar{R} = (\log M/n)[1 - \Pr[\text{retransmission}]] = \lambda R_\lambda [1 - \Pr[\text{retransmission}]]. \quad (17)$$

We will show that Pr[retransmission] can be made arbitrarily small, so $\bar{R} \simeq \lambda R_\lambda$.

Reliability: There are two ways the decoder can make an incorrect final answer. The first is when there is an undetected error that is successfully ACK-ed. The second is when there is a detected error $\hat{m} \notin \mathcal{B}(m), \hat{m} \neq e$ that is not successfully NAK-ed. Note that when $\phi(\mathbf{y}) = e$, we do not need to transmit a NAK as both encoder and decoder know that a retransmission is expected.

$$\begin{aligned} \Pr[\text{error}] = \Pr[\text{undetected error}] p_{a\to a} + \Pr[\texttt{NAK}] p_{n\to a} &\leq \exp\{-\lambda n[E_T(R_\lambda, R_{fb}/\lambda, P, W) - \gamma_n]\} \\ &+ \exp\{-\lambda n[E_T(R_\lambda, 0, P, W) - \gamma_n]\} \exp\{-(1-\lambda) n C_1\}, \end{aligned} \quad (18)$$

which follows from using (11) and (15) with block-length λn, code-rate R_λ, feedback rate R_{fb}/λ, and mis-detection probability $p_{n\to a} = \exp\{-(1-\lambda)nC_1\}$.

Picking the threshold T: To minimize the probability of undetected error (18) while keeping the probability of retransmission (16) small, we let the threshold T approach zero for n large. This regime has been studied by others (see, e.g., [6, 3, 5]). The maximum undetected error exponent can be determined by analyzing the limiting case $T = 0$. For the case of zero-rate feedback the analysis is carried out in [5]. The error exponent $E_T(R, R', P, W)$ evaluated at $T = 0$ with feedback rate R' is given by

$$E_0(R, R', W, P) \triangleq \min_{V,\hat{V}:\, PV = P\hat{V},\, D(\hat{V}\|W|P) + |I(P,V) - (R-R')|^+ \leq 0} D(V\|W|P) + |I(P,\hat{V}) - (R - R')|^+.$$

The condition $D(\hat{V}\|W|P) + |I(P,V) - (R - R')|^+ \leq 0$ and positivity of divergence and mutual information imply two things. First, that $\hat{V} = W$ (at least for those channel transitions corresponding to inputs with non-zero probability) and hence that $I(P,\hat{V}) = I(P,W)$. And, second, that $I(P,V) \leq R - R'$. Using these simplifications gives the following "feedback exponent":

$$E_T(R, R', P, W) \leq E_f(R - R', P, W) = \min_{V:\, PV = PW,\, I(P,V) \leq R - R'} D(V\|W|P) + |I(P,W) - (R - R')|^+ \quad (19)$$

Only the first term is a function of V. In [5] Telatar defines

$$E_{fsp}(\tilde{R}, P, W) \triangleq \min_{V:\, PV = PW,\, I(P,V) \leq \tilde{R}} D(V\|W|P). \quad (20)$$

Note that while akin to the sphere-packing bound $E_{fsp}(\tilde{R}, P, W) \geq E_{sp}(\tilde{R}, P, W)$ because of the extra restriction on the output distribution. The inequality can be strict, e.g., for the Z-channel. However, for totally symmetric channels, such as the binary symmetric channel, $E_{fsp}(\tilde{R}, P, W) = E_{sp}(\tilde{R}, P, W)$, where the latter is defined in Section 1.

Substituting (19) and (20) into (18) bounds the error probability for our scheme:

$$\begin{aligned} \Pr[\text{error}] \leq &\exp\{-\lambda n[E_{fsp}(R_\lambda - R_{fb}/\lambda, P, W) + |I(P,W) - R_\lambda + R_{fb}/\lambda|^+]\} \\ &+ \exp\{-\lambda n[E_{fsp}(R_\lambda, P, W) + |I(P,W) - R_\lambda|^+]\} \exp\{-n(1-\lambda)C_1\}. \end{aligned} \quad (21)$$

3.5 Minimizing feedback for a target reliability

In the first application of these results, we want to match the Burnashev reliability with the smallest feedback rate. To do so we choose P to be the capacity-achieving input distribution P^* (not necessarily a type). This means that $I(P^*, W) = C$ and we can therefore let R_λ approach C. Using this in (21) gives

$$\begin{aligned}\Pr[\text{error}] &\leq \exp\{-n[R_{fb}+\lambda E_{fsp}(C-R_{fb}/\lambda, P^*, W)]\}+\exp\{-n[\lambda E_{fsp}(C, P^*, W)+(1-\lambda)C_1]\} \\ &= \exp\{-nR_{fb}\}\exp\{-\lambda n E_{fsp}(C - R_{fb}/\lambda, P^*, W)\} + \exp\{-n(1-\lambda)C_1\},\end{aligned}$$

because $E_{fsp}(C, P^*, W) = 0$. As long as the probability of retransmission (16) is set small, the average rate $\bar{R} \simeq \lambda R_\lambda = \lambda C$ and so $\lambda \simeq \bar{R}/C$, giving:

$$\Pr[\text{error}] \leq \exp\left\{-n\left[R_{fb} + \frac{\bar{R}}{C}E_{fsp}\left(\frac{C}{\bar{R}}(\bar{R} - R_{fb}), P^*, W\right)\right]\right\} + \exp\left\{-nC_1\left(1 - \frac{\bar{R}}{C}\right)\right\}.$$

Picking the feedback rate R_{fb} to balance the two error terms gives $R_{fb} = \min\{\bar{R}, r\}$ where $r > 0$ is the smallest positive rate such that

$$r + \frac{\bar{R}}{C}E_{fsp}\left(\frac{C}{\bar{R}}(\bar{R} - r), P^*, W\right) \geq C_1\left(1 - \frac{\bar{R}}{C}\right). \tag{22}$$

Note that for low average rates $\bar{R}$ there may be no r that satisfies (22), as the probability of hashing error cannot be made as small as the probability of mis-detecting a `NAK` as an `ACK` in the confirmation stage. In these cases we must use decision feedback. Additionally, it never helps to use a feedback rate $R_{fb} > \bar{R}$, as the tentative decisions can be exactly communicated to the transmitter in those cases, which is sufficient to achieve the Burnashev exponent by Yamamoto-Itoh. By setting $r = \bar{R}$, and solving for $\bar{R}$, we find the minimum average communication rate at which our strategy requires less feedback rate than decision feedback. This rate is

$$\bar{R}^* = \frac{C_1 C}{C_1 + C + E_{fsp}(0, P^*, W)}. \tag{23}$$

In Fig. 1-(a) we plot the feedback rate R_{fb} required to achieve the Burnashev exponent on a binary symmetric channel. The general shape of the plot is common to all cross-over probabilities. The point $\bar{R}^*$ is indicated by the open circle.

We now show that the Forney scheme [6] is a special case of ours. In particular, if we choose $R_{fb} = 0$ (but still have one bit of feedback – whether to retransmit or not) and $\lambda = 1$, then $R_\lambda = \bar{R}$ and substituting these values into (21) gives

$$\Pr[\text{error}] \leq 2\exp\{-n[\sup_P E_{fsp}(\bar{R}, P, W) + |C - \bar{R}|^+]\}, \tag{24}$$

which is the the error probability derived in [5].[4]

If we want to achieve a reliability greater than E_{forn}, but below E_{burn}, we now have a methodology. We use (21) to optimize jointly over P, λ, R_λ, and R_{fb} – the input

[4]Note that we formally still need a long enough confirm/deny phase to `ACK` the correct decisions. In this context we must assign some constant number of channel uses to make sure that $p_{a\to n} < \delta$, else the probability of retransmission (16) would not be suitably small. A constant number can be used while letting $\lambda \to 1$ as $n \to \infty$. Informally, we can just get rid of the confirmation phase at zero rate. More broadly, the purpose of this section is to show that with the correct parameter settings, the current scheme can strictly dominates Forney's at any strictly positive feedback rate.

distribution, the fractional length of confirmation, the rate of the block code, and the rate of feedback. As mentioned in the introduction, our scheme allows the encoder and decoder to share the work on whether or not to make a decision. The λ parameter is the work-sharing parameter.

4 Conclusions

We have shown how to design a codebook and hash-function jointly to attain high reliabilities at reduced feedback rates. The strategy can be used to attain both Burnashev's and Forney's exponents. These initial results leave many directions to pursue.

Of primary interest is whether our approach is optimal, perhaps under the bursty feedback model we use here. Since the trade-offs between erasure and error exponents are not known to be tight, it seems hard to make general statements. However, in the limit of zero-rate, Telatar [5] shows that with the threshold $T = 0$, his error exponent $\sup_P E_f(0, P, W)$ is tight. In our context this leads to the conjecture that to obtain reliabilities greater than $\sup_P E_f(0, P, W)$, perfect decision feedback (no hashing) is required.

Related directions we are pursuing pertain to short block-length code design for channels with feedback, finding the Burnashev-like exponent for the Gel'fand-Pinkser [7] channel, and determining the Burnashev exponent for channels with input constraints.

References

[1] M. V. Burnashev. Data transmission over a discrete channel with feedback. Random transmission time. *Problems of Information Transmission*, 12(4):10–30, 1976.

[2] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, 1991.

[3] I. Csiszár and J. Körner. *Information Theory, Coding Theorems for Discrete Memoryless Systems*. Akadémiai Kiadó, 1981.

[4] R. L. Dobrushin. An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback. *Problemy Kibernetiki*, 8:161–168, 1962.

[5] İ. E. Telatar. *Multi-Access Communication with Decision Feedback Decoding*. PhD thesis, Mass. Instit. of Tech., 1992.

[6] G. D. Forney. Exponential error bounds for erasure, list, and decision feedback schemes. *IEEE Trans. Inform. Theory*, 14:206–220, March 1968.

[7] S. I. Gel'fand and M. S. Pinsker. Coding for channels with random parameters. *Problems of Control and Information Theory*, 9:19–31, 1980.

[8] A. Sahai and T. Şimşek. On the variable-delay reliability function of discrete memoryless channels with access to noisy feedback. In *IEEE Information Theory Workshop, San Antonia, Texas*, 2004.

[9] H. Yamamoto and K. Itoh. Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback. *IEEE Trans. Inform. Theory*, pages 729–733, November 1979.