Feedback and Belief Propagation

Giuseppe Caire[1], **Shlomo Shamai**[2] and **Sergio Verdú**[3],

[1] University of Southern California, Los Angeles, California, USA `caire@usc.edu`
[2] Technion, Haifa, Israel `sshlomo@ee.technion.ac.il`
[3] Princeton University, Princeton, New Jersey, USA `verdu@princeton.edu`

Abstract

We demonstrate that feedback in discrete memoryless channels has the capability of greatly lowering the block error rate of codes designed for open-loop operation. First we show how to use full feedback of the channel output to turn any capacity achieving code into a reliability-function achieving code. Second, we propose a practical embodiment based on sparse-graph codes, belief propagation, and a variation of the closed-loop iterative doping algorithm. This scheme takes advantage of any available limited-rate feedback to bootstrap good block error rate from good bit error rate.

1 Introduction

One of the most surprising results in information theory was proven by Claude Shannon in 1956 [1]: instantaneous and noiseless feedback of the output of a discrete memoryless channel does not increase capacity. This means that causal knowledge at the encoder of the noise realization, and therefore of its effect on the decoder's belief on the transmitted codeword is useless as far as improving the maximal achievable rate with reliable communication. Open loop operation that does not try to counteract the effect of the channel is optimal.

In practice, two-way communication is common, and thus some form of feedback is often available, although rarely the return channel resources are so ample as to allow for full instantaneous noiseless feedback of the channel output (*Shannon feedback* [2]). However, in view of Shannon's result it would seem that feedback could only profitably be used to counteract the effects of the channel either when it has memory or when it does not behave ergodically. Such a conclusion is actually unwarranted because feedback is useful even for stationary memoryless channels:

1) The complexity of capacity-approaching encoding/decoding schemes may be reduced.
2) The reliability function of the channel increases with feedback.
3) The blocklength required to achieve a desired block error probability decreases with feedback.

We now pause to give a brief review of the literature in each of those three beneficial aspects of feedback.

Supported in part by the U. S. National Science Foundation under Grants CCR-0312839 and CCR-0312879, and by the U. S. - Israel Binational Science Foundation under Grant 2004140

1.1 Low complexity capacity-achieving schemes

Several sequential schemes with Shannon feedback have been proposed and analyzed by Hornstein [3] and Schwalkijk and co-workers [4], [5], [6]. In particular, Hornstein's seminal sequential scheme for the BSC [3] is very simple and achieves excellent performance versus expected decision delay tradeoff. The scheme is somewhat reminiscent of arithmetic coding for data compression: an information message of k bits is represented as a point in a subinterval of size 2^{-k} of the unit interval. Both transmitter and receiver keep track of the posterior probability cumulative distribution function of the unit interval given the channel output received up to the current time. At each round, the transmitter sends 0 if the message interval is below the current median, or 1 if it is above. Decisions are made with a fixed delay.

1.2 Reliability function with feedback

For discrete memoryless channels, the increase in the error exponent afforded by feedback has been studied in [7], [8], [9], [10] following the block coding paradigm. For the continuous-time channel, not only has it been shown that the error exponent may increase, but a double exponential behavior of the error probability has been established [11], [4], [12]. This behavior breaks down when practical constraints, such as peak power, are introduced [12].

While in the absence of feedback the reliability function is not known for all rates, Burnashev [13], [14] found the following closed form expression for the reliability function of the discrete memoryless channel with Shannon feedback, which holds for all rates from 0 to capacity:

$$E_f(R) = \overline{D}(1 - R/C) \tag{1}$$

where C is the channel capacity and

$$\overline{D} = \max_{a,b \in \mathcal{X}} D(P_{Y|X=a} \| P_{Y|X=b}) \quad (2)$$

is the maximal divergence between any two conditional output distributions. A simplified achievability scheme for (1) was given by Yamamoto-Itoh [15]. Reference [16] elaborates on a source-coding based channel coding scheme with feedback dating back to [17], and shows that it can achieve (1).

Burnashev's reliability function can be achieved even if the encoder does not initially know the discrete memoryless channel [18]. Predating [13], the infinite-bandwidth AWGN counterpart of Burnashev's reliability function is due to [19].

The reliability function in the presence of limited feedback has also been considered in the literature. Forney [20] considers block codes transmitted over a discrete memoryless channel in the presence of Ack/Nack feedback and shows that the reliability function has a slope of -1 at capacity (in contrast to $-\overline{D}/C \leq -1$ with Shannon feedback, and 0 with no feedback). A certain type of list feedback, where the order of the decoded list is conveyed back to the transmitter, has been shown not to improve exponentially over the Ack/Nack feedback [20]. The scheme in [15] attains (1) requires that the transmitter gets a noiseless replica of the message decoded by the receiver. Thus, if not full Shannon feedback in the conventional sense, the capacity of the reverse channel has to be at least as large as the capacity of the forward channel. Requiring less feedback rate, the constructive scheme in [21] bridges between error exponents achieved in [20] and (1), but without attaining necessarily the reliability function with limited feedback which is yet unknown.

1.3 Block error probability - Blocklength tradeoff

The work reviewed in subsections 1.1 and 1.2 deals with the asymptotic regime of long blocklength. However, much of the practical usefulness of feedback lies in its power to shorten the blocklength required to achieve a given desired block error probability.

In general, for a given desired block error rate there is a tradeoff between the number of feedback symbols and the required feedforward blocklength. Naturally, in the regime of nonasymptotic blocklength, the fundamental error probability - blocklength tradeoff is unknown, and the attention is focused on pragmatic schemes that can harness the potential of feedback. A simple and practically important example of such pragmatic schemes are third-generation wireless systems, which achieve good block error rate with moderate blocklengths thanks to feedback in the form of ARQ (automatic repeat request).

The use of sparse-graph channel codes for lossless data compression in conjunction with belief propagation decoders was proposed in [22], [23], [24], [25], where an algorithm called *Closed-Loop Iterative Doping* (CLID) was instrumental in achieving competitive performance with existing data compression methods. The use of belief propagation and CLID in conjunction with channel coding with limited feedback was originally proposed in [26]. An important advance was the introduction of the *noisy* CLID in [2] which achieves very significant decrease in block error rate at the expense of a modicum of feedback.

1.4 From good bit error rate to good block error rate

In this paper we champion a simple two-stage design principle: first, encode the information with a conventional codebook designed to provide good bit error rate, and second, use feedback to send further information that will enable the decoder to clean up the existing errors, to achieve good block error rate. In the absence of feedback, an established design principle is to use the concatenation of an expander code with a low bit error rate sparse-graph code to produce an overall low block error rate sparse-graph code.

In Section 2 we buttress this philosophy by a constructive proof of the achievablility of Burnashev's error exponent, which in contrast to the proofs in [13] and [15] is not based on codes that achieve exponentially vanishing block error probability. This is significant because the state-of-the-art in the design of codes that can be encoded/decoded with reasonable complexity and attain positive error exponents is still in an embryonic stage [27].

In Section 3 we design a practical embodiment of the two-stage design principle which shows the synergistic effects of belief propagation and feedback in achieving the goal of bootstrapping good block error rate from good bit error rate. In the second stage, the feedback rate available is a design parameter. The higher the feedback rate available the shorter the forward transmission is required to be in order to achieve the desired block error rate.

2 From a capacity-achieving code to a reliability-achieving code

In this section we show that it is possible to achieve the reliability function of channels with Shannon feedback for any rate between 0 and capacity using a two-stage scheme where the channel code used in the first stage (the message mode) is a "standard" open-loop code that achieves low block error rate are rates slightly below capacity.

In the Yamamoto-Itoh scheme [15], in order to convey an information message of k bits, the transmitter operates in two stages. In the first stage (message mode), it encodes the message into a codeword of length γN, using a basic code $\mathcal{C}$ of rate $R_1 = k/(\gamma N)$, where $\gamma \in (0, 1)$. At the end of the message mode, the receiver makes a decision about the message, and sends back its decision over the noiseless feedback link. In the second stage (control mode), the transmitter sends a message c if the decision is correct, or a message e if the decision is not correct. In order to do so, it makes use of a repetition code with codewords $(x, x, x, \ldots, x)$ or $(x', x', x', \ldots, x')$, where the two letters x and x' are chosen to achieve the maximum in (2) The repetition code in the control mode has length $(1-\gamma)N$.

If the decoder decides that the control message is e, then the same codeword that was transmitted in the first stage is retransmitted, otherwise, if the decoder decides that the control message is c, then the transmitter moves on to the next information message, that shall be transmitted using an identical scheme. It is clear that the detection of c resets the system: it is a *renewal event*. We define the following probabilities and events:

- P_1 is the probability of block error of the code $\mathcal{C}$ in the message mode;
- P_{ec} and P_{ce} are the probabilities of detecting c given that e was transmitted, or detecting e given that c was transmitted, in the control mode, respectively;
- One round of message + control modes will be referenced to as a "protocol round". We define $\mathcal{E}_\ell$ as the event that control message e is detected at protocol round number ℓ, and the event $\mathcal{C}_\ell$ as the event that control message c is detected at protocol round number ℓ. The event $\mathcal{A}_\ell \subset \mathcal{C}_\ell$ occurs when c is detected but e was transmitted at the ℓth protocol round.
- The random variable $\mathcal{T}$ denotes the inter-renewal time: the number of protocol rounds between consecutive occurrences of the renewal event.

The inter-renewal time is distributed as

$$P(\mathcal{T}=\ell) \stackrel{\Delta}{=} q(\ell) = P(\mathcal{E}_1, \mathcal{E}_1, \ldots, \mathcal{E}_{\ell-1}, \mathcal{C}_\ell) \quad (3)$$

$$= p(\ell-1) - p(\ell) \quad (4)$$

where $p(0) = 1$ and

$$p(\ell) \stackrel{\Delta}{=} P(\mathcal{E}_1, \mathcal{E}_1, \ldots, \mathcal{E}_\ell) \quad (5)$$

$$= \prod_{i=1}^{\ell} P(\mathcal{E}_i) \quad (6)$$

$$= \vartheta^\ell \quad (7)$$

where we have used the the fact that the events $\{\mathcal{E}_1, \mathcal{E}_1, \ldots, \mathcal{E}_\ell\}$ are independent with identical probability equal to

$$\vartheta = (1-P_1)P_{\mathsf{ce}} + P_1(1-P_{\mathsf{ec}}). \quad (8)$$

From renewal-reward theory, the coding rate (i.e., the time-average number of transmitted information bits per channel use) is given by

$$R = \frac{k}{N\mathbb{E}[\mathcal{T}]} = \frac{k}{N}\left[\sum_{\ell=0}^{\infty} p(\ell)\right]^{-1} \quad (9)$$

$$= \frac{k}{N}(1-\vartheta) \quad (10)$$

The average decision delay of the scheme is equal to the average number of rounds times the number of symbols per round, i.e.,

$$\bar{N} = \frac{N}{1-\vartheta}$$

Since a message is ultimately decoded erroneously if and only if a control message e is detected as c, The block error probability is

$$\begin{aligned} P_e &= \sum_{\ell=1}^{\infty} P(\mathcal{T}=\ell, \mathcal{A}_\ell) \\ &= \sum_{\ell=1}^{\infty} P(\mathcal{E}_1, \ldots, \mathcal{E}_{\ell-1}, \mathcal{A}_\ell) \\ &= \sum_{\ell=1}^{\infty} \vartheta^{\ell-1} P_1 P_{\mathsf{ec}} \\ &= \frac{P_1 P_{\mathsf{ec}}}{1-\vartheta} \quad (11) \end{aligned}$$

We will now show that it is possible to achieve the reliability function (1) provided that in the first stage we use a code whose block error probability can be made as small as desired for any rate below capacity for sufficiently large blocklength. To that end, fix R and an arbitrarily small $\delta > 0$.

For the first stage choose a k-to-γN code such that the rate is

$$R_1 = \frac{k}{\gamma N} = \frac{C}{1 + \frac{\delta C}{12R}} \quad (12)$$

and the block error probability satisfies

$$P_1 \leq \frac{\delta}{6} \quad (13)$$

Also, by Stein's lemma (e.g. [28]) we can find a hypothesis test that discriminates between the hypotheses c and e achieving

$$P_{\mathsf{ce}} \leq \frac{\delta}{12} \quad (14)$$

and

$$P_{\mathsf{ec}} \leq \exp\left(-(1-\gamma)N\overline{D} + N\frac{\delta}{3}\right) \quad (15)$$

for all sufficiently large N.

From (8), (14), (15) it follows that

$$\vartheta \leq \frac{\delta}{4} \quad (16)$$

The error exponent achieved by this variable length scheme (log probability divided by the average blocklength) satisfies

$$
\begin{aligned}
\frac{1}{\bar N}\log\frac{1}{P_e} &\geq \frac{1-\vartheta}{N}\log(1-\vartheta)+\frac{1-\vartheta}{N}\log\frac{1}{P_{\mathsf{ec}}}\\
&\geq -\frac{\delta}{3}+\frac{1-\vartheta}{N}\left((1-\gamma)N\overline{D}-N\frac{\delta}{3}\right)\\
&\geq -\frac{2\delta}{3}+\overline{D}(1-\vartheta)(1-\gamma) \qquad (17)\\
&= -\frac{2\delta}{3}+\overline{D}\left[1-\vartheta-\frac{R}{R_1}\right] \qquad (18)\\
&= -\frac{2\delta}{3}+\overline{D}\left[1-\vartheta-\frac{R}{C}\left(1+\frac{\delta C}{12R}\right)\right]\\
&\geq \overline{D}\left(1-\frac{R}{C}\right)-\left(\overline{D}+2\right)\frac{\delta}{3} \qquad (19)
\end{aligned}
$$

where the first inequality follows from (11); the second inequality holds for sufficiently large N because of (15); (18) follows from (10) and (12); and (19) follows from (16). Finally, since the choice of δ was arbitrary, we see that we can achieve the reliability function in (1).

Note that the proportion of symbols devoted to transmitting the codeword (as opposed to control information) is proportional to the overall rate R.

In contrast to the above analysis, [15] did not realize that open loop codes with non exponential decrease of error probabilities are sufficient to achieve the reliability function. Another benefit of our more detailed analysis is that it enables the analysis of computational complexity of coding/encoding schemes as a function of their gap to capacity to carry over to the reliability function. To see this, note that the computational effort in the solution of the binary hypothesis testing problems for the control information is, for given R, linear in the received blocklength, and furthermore we have shown in (19) that, modulo a channel-dependent proportionality constant, the gap to the reliability function at any given rate is the same as the gap to capacity of the underlying open-loop code. Thus, the results in [29] enable us to conclude that for a binary erasure channel with feedback it is possible to achieve the reliability function with bounded complexity per information bit, while for other channels, complexities that scale with the gap ε to the reliability function as $\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}$ are feasible.

3 From a good bit error rate code to a good block error rate code

In this section we report a pragmatic two-stage scheme where the open-loop code is sent in the first stage, and in the second stage the encoder and decoder maintain a dialog (aided by belief propagation and the noisy CLID algorithm) to clean up the errors resulting from the decoding of the first stage. With little added complication we present the scheme in the full generality of source-channel coding.

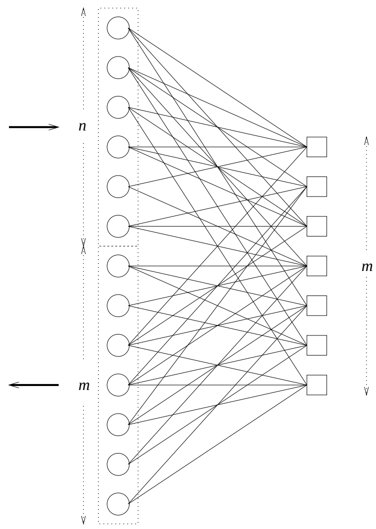

Fig. 1. LDPC for joint source-channel encoding.

3.1 Source-Channel Coding with Shannon Feedback

In [26] we proposed to do joint source-channel encoding using the sparse graph structure in Figure 1. This is a Tanner graph of a $m \times (n+m)$ parity-check matrix $\mathbf{H}$ that has the special constraint that the upper n symbols on the left are degrees of freedom, i.e. no matter how they are chosen there are m lower (parity-check) symbols that make all the equations satisfied. Thus, this is a systematic LDPC of rate $\frac{n}{n+m}$.

The systematic n bits consist of the source output if it is memoryless or of the Burrows-Wheeler transform output [25], otherwise.

We send through the channel only the nonsystematic part of the codeword. Thus, the encoding rate is $\frac{n}{m}$, which should be almost as high as the channel capacity divided by the source entropy.

In the source coding problem the approach we introduced in [22], [23] in which we encoded by multiplying the source vector with an $m \times n$ parity-check matrix $\mathbf{H}$ is a special case of this approach where the parity check matrix of the rate $n/(n+m)$ code is the matrix $[\mathbf{H} \quad -\mathbf{I}]$.

At the decoder, we run belief propagation on the Tanner graph of Figure 1, where the initial reliabilities of the systematic bits are provided by the source modeler just like in the data compression scheme in [25], and the initial reliabilities of the parity-check bits are computed from the channel outputs and the channel transition probability. Thanks to the availability of Shannon feedback, the encoder can run an exact copy of the belief propagation algorithm at the decoder and they can proceed with noisy-CLID, introduced in [2]. For convenience, we recall that in the original CLID algorithm (as used in data compression), it is possible to run an identical copy of the belief propagation algorithm at the compressor since the compressor has access

3.2 Limited Feedback

Again we consider the scheme in Figure 1 which depicts the Tanner graph of an $m \times (n+m)$ parity-check matrix $\mathbf{H}$ and assume that the n input bits are iid Bernoulli with parameter p. (In the general case these are Burrows-Wheeler transform outputs with approximately piecewise iid blocks.) The transmitted $n+m$ bits $\mathbf{x}$ pass through a memoryless binary input channel, and produce the corresponding output $\mathbf{y}$. As in Section 3.1, we use a linear code of rate $n/(n+m)$, but we transmit over the channel only the nonsystematic part of m bits. This can be viewed as all $n+m$ bits being transmitted through a memoryless time-varying channel, where the first n symbols undergo crossover probability $1/2$.

We now show how to combine in a unified manner and within a single source channel code the noisy-CLID algorithm at the whole gamut of feedback rates from Shannon feedback to zero.

Analogously to [2], the first step is to try and decode in the best possible way $\mathbf{x}$, based on the channel observations $\mathbf{y}$, using no feedback. The assumption is that if the rate is not too ambitiously close to the Shannon limit C/H (and m is sufficiently large), then the hard output of the belief propagation decoder (after a certain number of iterations) $\hat{\mathbf{x}}_1$, will be such that

$$\mathbf{e}_1 = \hat{\mathbf{x}}_1 - \mathbf{x}$$

can be approximated as a Bernoulli sequence with a rather low parameter p_e. If the reliabilities for $\hat{\mathbf{x}}_1$ are sufficiently high or if $\mathbf{H}\hat{\mathbf{x}}_1 = \mathbf{H}\hat{\mathbf{e}}_1 = \mathbf{0}$, the algorithm terminates.

In the second stage, both encoder and decoder have agreed on an $\ell \times m$ matrix $\mathbf{L}_1$ ($\ell \leq m$), and define the $\ell \times (n+m)$ matrix

$$\mathbf{H}_1 = \mathbf{L}_1\mathbf{H}$$

each of whose ℓ rows is a linear combination of rows of $\mathbf{H}$. The receiver sends back to the transmitter (Fig. 2)

$$\mathbf{H}_1\hat{\mathbf{x}}_1 = \mathbf{H}_1\mathbf{e}_1$$

over the noiseless feedback channel. Now both the transmitter and receiver know $\mathbf{H}_1\mathbf{e}_1$, and run synchronized belief propagation iterations together with noisy-CLID to try to obtain $\mathbf{e}_1$, just like in [2]. Note that this requires an agreed-upon estimate (at receiver and transmitter) of the bit error rate of the decisions emitted in Stage 1. Evidently at the receiver nothing would be accomplished without further information as it has already done its best, based on the full matrix $\mathbf{H}$. But in parallel with the belief propagation iterations synchronized with those at the transmitter, the decoder continues running its autonomous belief propagation algorithm (unsynchronized with the transmitter) which it ran at the first stage. In Figure 2 we have represented

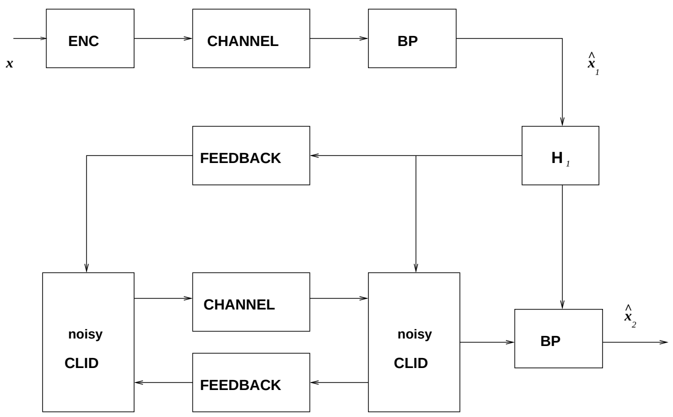

Fig. 2. Stages 1 and 2 of the noisy-CLID limited feedback scheme.

to all the information available to the decompressor. The same holds in the case of a noisy channel with Shannon feedback. In the latter case, though, the bit values are sent to the decoder via the noisy forward channel. Hence, the basic CLID algorithm is modified: instead of setting the doped bit messages to $\pm\infty$, we set them equal to $(-1)^y \log \frac{1-\delta}{\delta}$, ($\delta$ is the channel crossover probability) where y is their corresponding channel output. Notice that with Shannon feedback the transmitter knows y for each doping round, and therefore can keep its own copy of the belief propagation decoder perfectly synchronized with that of the decoder such that the position of the bits to be doped need not be communicated explicitly.

In [2] we reported the result of an experiment with a BSC with crossover probability equal to 0.1 (capacity = 0.53 bit/symbol) and an irregular LDPC code of length $n = 20,000$ and rate $r = 0.5$. With Shannon feedback, we found zero block errors in $100,000$ blocks achieving an average transmission rate equal to $R = 0.495$.

for clarity this autonomous belief propagation in two blocks labelled BP; corresponding to two stages of the same algorithm. The synchronized belief propagation algorithms run in the blocks labelled ”noisy CLID”. Every time a bit from the noisy CLID is received one or several iterations take place in both the synchronized and the autonomous belief propagation algorithms. The total number of feedforward bits and feedback bits spent in the second stage are equal to d and $d+\ell$ respectively, where d is the number of noisy-CLID bits. Whenever the reliabilities in the autonomous belief propagation algorithm have reached high enough values or its hard decisions $\hat{\mathbf{x}}_2$ satisfy $\mathbf{H}\hat{\mathbf{x}}_2 = \mathbf{0}$, the algorithm terminates outputting $\hat{\mathbf{x}}_2$. If this does not occur after a certain number of noisy-CLID iterations we repeat the second stage with a different matrix $\mathbf{L}_2$, (not necessarily the same size, likely with fewer rows) and with $\hat{\mathbf{x}}_2$ in lieu of $\hat{\mathbf{x}}_1$, thus transmitting $\mathbf{H}_2\hat{\mathbf{x}}_2$ through the feedback channel.

Unlike in the present scheme, note that in the scheme of [2] the first stage was only used to get the hard decisions $\hat{\mathbf{x}}_1$ but the reliability values computed by the belief propagation algorithm were totally wasted. Another advantage of the scheme in this section is that unlike the scheme in [2], we can encompass normalized feedback rates between $1-C$ and 1. At first sight it may seem like (aside from the fact that here we have the possibility of further uses of the feedforward channel and the concurrency of autonomous/synchronized belief propagation algorithm) we could get one scheme as a special case of the other. However, except in the extreme case, in which the inner code of [2] adds no redundancy and thus the feedback rate is $1-C$, its generator matrix is a tall matrix and thus the length of $\hat{\mathbf{x}}_1$ in [2] is strictly smaller than the (channel) blocklength. Thus the number of columns in the $\mathbf{H}_1$ in [2] is strictly smaller than the blocklength, whereas here it is equal to the blocklength.

References

[1] C.E. Shannon, “The zero error capacity of a noisy channel,” *IRE Trans. Inform. Theory*, vol. IT-2, pp. 8–19, September 1956.

[2] G. Caire, S. Shamai, and S. Verdú, “An efficient scheme for reliable error correction with limited feedback,” *Proc. 2005 IEEE Int. Symp. on Information Theory*, Sep. 2005.

[3] M. Horstein, “Sequential transmission using noiseless feedback,” *IEEE Trans. Information Theory*, vol. 9, pp. 136–143, July 1963.

[4] J. P. M. Schalkwijk and T. Kailath, “A coding scheme for additive noise channels with feedback,” *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 172–182, April 1966.

[5] J. P. M. Schalkwijk, “A class of simple and optimal strategies for block coding on the binary symmetric channel with noiseless feedback,” *IEEE Trans. Information Theory*, vol. 17, pp. 283–287, May 1971.

[6] P. Schalkwijk and K. Post, “On the error probability for a class of binary recursive feedback strategies,” *IEEE Trans. Information Theory*, vol. 19, No. 4, pp. 498–511, May 1973.

[7] R. L. Dobrushin, “Asymptotic bounds on error probability for transmission over DMC with symmetric transition probabilities,” *Theory of Probab. Applicat.*, vol. 7, pp. 283–311, 1962.

[8] K. Sh. Zigangirov, “Upper bounds on the error probability for channels with feedback,” *Prob. Peredac. Inform*, vol. 6, No. 2, pp. 87–92, 1970.

[9] A.G. Djackov, “Upper bounds on the error probability for transmission with feedback in case of memoryless discrete channel,” *Prob. Peredac. Inform*, vol. 11, No. 4, pp. 13–28, 1975.

[10] B.D. Kudryashov, “Message transmission over discrete channel with noiseless feedback,” *Prob. of Inform. Transmission*, vol. 21, No. 1, pp. 3–13, Jan-March 1979.

[11] A.J. Kramer, “Improving communication reliability by use of an intermittent feedback channel,” *IEEE Trans. Inform. Theory*, vol. IT-15, No. 1, pp. 52–60, 1969.

[12] A.D. Wyner, “On the schalkwijk-kailath coding scheme with a peak energy constraint,” *IEEE Trans. Inform. Theory*, vol. IT-14, No. 1, pp. 129–134, January 1968.

[13] M.V. Burnashev, “Data transmission over discrete channel with feedback: Random transmission time,” *Prob. Peredac. Inform*, vol. 12, No. 4, pp. 10–30, 1976.

[14] M. V. Burnashev, “Sequential discrimination of hypotheses with control observations,” *Math. USSR, Izvestia*, vol. 15, no. 3, pp. 419–440, 1980.

[15] H. Yamamoto and K. Itoh, “Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback,” *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 729–733, Nov. 1979.

[16] J.M. Ooi and G. W. Wornell, “Fast iterative coding techniques for feedback channels,” *IEEE Trans. Information Theory*, vol. 44 no. 7, pp. 2960–2976, Nov. 1998.

[17] R. Ahlswede, “A constructive proof of the coding theorem for discrete memoryless channels with feedback,” *Proc. 6th Prague Conf. Information Theory, Statistical Decision Functions, and Random Processes*, p. 3950, 1971.

[18] A. Tchamkerten and E. Telatar, “On the universality of Burnashev’s error exponent,” *IEEE Transactions on Information Theory*, vol. 51, Issue 8, pp. 2940 – 2944, Aug. 2005.

[19] J. P. M. Schalkwijk and M. Barron, “Sequential signaling under a peak power constraint,” *IEEE Trans. Information Theory*, vol. 17, pp. 278–282, May 1971.

[20] G.D. Forney, “Exponential error bounds for erasure, list and decision feedback schemes,” *IEEE Trans. Inform. Theory*, vol. IT-14, No. 2, pp. 206–220, March 1968.

[21] S. C. Draper, K. Ramchandran, B. Rimoldi, A. Sahai, and D. N. C. Tse, “Attaining maximal reliability with minimal feedback via joint channel-code and hash-function design,” *2005 Allerton Conf. on Communications, Control and Computing*, Oct. 2005.

[22] G. Caire, S. Shamai, and S. Verdú, “A new data compression algorithm for sources with memory based on error correcting codes,” *2003 IEEE Workshop on Information Theory*, pp. 291–295, Mar. 30- Apr. 4, 2003.

[23] G. Caire, S. Shamai, and S. Verdú, “Lossless data compression with error correction codes,” *2003 IEEE Int. Symp. on Information Theory*, p. 22, June 29- July 4, 2003.

[24] G. Caire, S. Shamai, and S. Verdú, “Universal data compression with LDPC codes,” *Third International Symposium On Turbo Codes and Related Topics*, pp. 55–58, Brest, France, September 1-5, 2003.

[25] G. Caire, S. Shamai, and S. Verdú, “Noiseless data compression with low density parity check codes,” in *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, P. Gupta and G. Kramer, Eds., pp. vol. 66, pp. 263–284. American Mathematical Society, 2004.

[26] G. Caire, S. Shamai, and S. Verdú, “Almost-noiseless joint source-channel coding-decoding of sources with memory,” *Proc. Fifth International ITG Conference on Source and Channel Coding (SCC)*, pp. 295–304, Jan 14-16, 2004.

[27] A. Barg and G. Zemor, “Error exponents of expander codes,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 1725–1729, June 2002.

[28] R. E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley, Reading, Mass., 1987.

[29] H. D. Pfister, I. Sason, and R. Urbanke, “Capacity-achieving ensembles for the binary erasure channel with bounded complexity,” *IEEE Trans. Information Theory*, vol. 51, no. 7, pp. 2352 – 2379, July 2005.