Joint source-channel coding with feedback

Victoria Kostina
California Institute of Technology
Pasadena, CA 91125
vkostina@caltech.edu

Yury Polyanskiy
Massachusetts Institute of Technology
Cambridge, MA 02139
yp@mit.edu

Sergio Verdú
Princeton University
Princeton, NJ 08544
verdu@princeton.edu

***Abstract*—This paper quantifies the fundamental limits of variable-length transmission of a general (possibly analog) source over a memoryless channel with noiseless feedback, under a distortion constraint. We consider excess distortion, average distortion and guaranteed distortion (d-semifaithful codes). In contrast to the asymptotic fundamental limit, a general conclusion is that allowing variable-length codes and feedback leads to a sizable improvement in the fundamental delay-distortion tradeoff.**

***Index Terms*—Joint source-channel coding, single-shot method, rate-distortion theory, feedback, Shannon theory.**

I. Introduction

A famous result of Shannon [1] states that feedback cannot increase the capacity of memoryless channels. Burnashev [2] showed that feedback improves the error exponent in a variable-length setting. Polyanskiy et al. [3] showed that allowing variable-length coding and non-vanishing error-probability ϵ boosts the ϵ-capacity of the discrete memoryless channel (DMC) by a factor of $1-\epsilon$. Furthermore, as shown in [3], if both feedback and variable-length coding are allowed, the asymptotic limit $\frac{C}{1-\epsilon}$ is approached at a fast speed $O\left(\frac{\log \ell}{\ell}\right)$ as the average allowable delay ℓ increases:[1]

$$(1-\epsilon)\log M^\star(\ell,\epsilon) = \ell C + O\left(\log \ell\right), \tag{1}$$

where $M^\star(\ell,\epsilon)$ is the maximum number of messages that can be distinguished with error probability ϵ at the average delay ℓ, and C is the channel capacity. This is in contrast to channel coding at fixed blocklength n where in most cases only a $O\left(\frac{1}{\sqrt{n}}\right)$ convergence rate is attainable, even when feedback is available, see [3], [4]. Thus, variable-length coding with feedback (VLF) not only boosts the ϵ-capacity of the channel, but also markedly accelerates the speed of approach to it. Moreover, zero-error communication is possible at an average rate arbitrarily close to capacity via VLFT codes, a class of codes that employs a special termination symbol to signal the end of transmission, which always recognized error-free by the receiver [3]. As discussed in [3], the availability of zero-error termination symbols models that common situation in which timing information is managed by a higher layer whose reliability is much higher than that of the payload.

In [5], we treated variable-length data compression with nonzero excess distortion probability. In particular, we showed that in fixed-to-variable-length compression of a block of k i.i.d. source outcomes, the minimum average encoded length $\ell^\star(k,d,\epsilon)$ compatible with probability ϵ of exceeding distortion threshold d satisfies, under regularity assumptions,

$$\ell^\star(k,d,\epsilon) = (1-\epsilon)kR(d) - \sqrt{\frac{k\mathcal{V}(d)}{2\pi}}e^{-\frac{(Q^{-1}(\epsilon))^2}{2}} + O\left(\log k\right) \tag{2}$$

where $R(d)$ and $\mathcal{V}(d)$ are the rate-distortion and the rate-dispersion functions, and Q is the standard normal complementary cumulative distribution function. The second term in the expansion (2) becomes more natural if one notices

$$\mathbb{E}\left[Z \cdot 1\{Z > Q^{-1}\left(\epsilon\right)\right] = \frac{1}{\sqrt{2\pi}}e^{-\frac{Q^{-1}(\epsilon)^2}{2}}\,.$$

As elaborated in [5], the expansion (2) has an unusual feature: the asymptotic fundamental limit is approached from the "wrong" side, e.g. larger dispersions and shorter blocklengths reduce the average compression rate.

In this paper, we consider variable-length transmission of a general (possibly analog) source over a DMC with feedback, under a distortion constraint. This variable-length joint source-channel coding (JSCC) setting can be viewed as a generalization of the setups in [3], [5]. Related work includes an assessment of nonasymptotic fundamental limits of fixed-length JSCC in [6]–[8], a dynamic programming formulation of zero-delay JSCC with feedback in [9], and a practical variable-length almost lossless joint compression/transmission scheme in [10]. Various feedback coding strategies are discussed in [11]–[17]. Practical feedback communication schemes in the literature that implement VLF include [18]–[21].

We treat several scenarios that differ in how the distortion is evaluated and whether a termination symbol is allowed. In all cases, we analyze the *average delay* required to achieve the objective. The summary of our results in Section III, where as before, $C, R(d), \mathcal{V}(d)$ denote the channel capacity, and the source rate-distortion and rate-dispersion functions, is:

- Under average distortion criterion, $\mathbb{E}\left[d(S^k,\hat{S}^k)\right] \le d$, the minimal average delay $\ell^\star(k,d)$ attainable via VLF codes transmitting k source symbols satisfies

$$\ell^\star(k,d)C = kR(d) + O\left(\log k\right). \tag{3}$$

- Under excess distortion criterion, $\mathbb{P}[d(S^k,\hat{S}^k) > d] \le \epsilon$, the minimal average delay attainable via VLF codes

This work was supported in part by the National Science Foundation (NSF) under Grant CCF-1016625, by the NSF CAREER award under Grant CCF-1253205, and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under Grant CCF-0939370.

[1]Unless explicitly noted, all log and exp in this paper are to arbitrary matching base, which also defines units of entropy and mutual information.

transmitting k source symbols satisfies

$$\ell^\star(k,d,\epsilon)\,C = (1-\epsilon)kR(d) - \sqrt{\frac{k\mathcal{V}(d)}{2\pi}}e^{-\frac{(Q^{-1}(\epsilon))^2}{2}} + O\left(\log k\right). \tag{4}$$

- Under guaranteed distortion criterion, $\mathbb{P}[d(S^k,\hat{S}^k) > d] = 0$, the minimal average delay attainable via VLFT codes transmitting k source symbols satisfies

$$\ell_t^\star(k,d,0)\,C = kR(d) + O\left(\log k\right). \tag{5}$$

Similar to (1), approaching the limits in (3), (4) and (5) only requires an extremely thin feedback link, namely, the decoder sends just a single acknowledgement signal once it is ready to decode (stop-feedback)[2]. Note that (4) exhibits significant similarities with (2): the asymptotic limit is approached from below, i.e. in contrast to the results in [6], [22], [23], smaller blocklengths and larger source dispersions are beneficial. Note also that the first term of the expansion in (4) can be attained with variable-length codes without feedback.

Interestingly, naive separated source/channel coding fails to attain any of the limits mentioned. For example, even the sign of the second term in (4) is not attainable. This observation led us to believe, initially, that competitive schemes in this setting should be of successive refinement and adaptation sort such as in [24], [25], or dynamic programming-like as in [9], [26]. It turns out, however, that like the fixed-length JSCC achievability schemes in [6], [7], attaining limits (3)-(5) requires a rather simple variation on the separation architecture: one only needs to allow a variable-length interface between the source coder and the channel coder. Note that typically, separation is understood in the sense that the output of the source coder (compressor) is treated as pure bits, which can be arbitrarily permuted without affecting performance of the concatenated scheme [8], [27]. Similarly, the performance of a *variable-length separated scheme* is insensitive to permutations (but not additions or deletions) of the bits at the output of the source coder. These semi-joint achievability schemes are the subject of Section II.

II. Feedback codes for non-equiprobable messages

In this section we consider joint source-channel coding assessing reliability by the probability that the (possibly non-equiprobable) message is reproduced correctly. Our key tool will be two extensions of the channel coding bounds for the DMC with feedback from [3]. VLF and VLFT codes are formally defined as follows.

Definition 1. *A variable-length feedback code (VLF) transmitting message W (taking values in $\mathcal{W}$) over the channel $\{P_{Y_i|X^iY^{i-1}}\}_{i=1}^\infty$ with input/output alphabets $\mathcal{A}$/$\mathcal{B}$ is defined by:*

1) *A random variable $U \in \mathcal{U}$ revealed to the encoder and decoder before the start of the transmission.*
2) *A sequence of encoding functions $\mathsf{f}_n\colon \mathcal{U}\times\mathcal{W}\times\mathcal{B}^{n-1}\mapsto\mathcal{A}$, defining the channel inputs*
$$X_n = \mathsf{f}_n\left(U, W, Y^{n-1}\right) \tag{6}$$
3) *A sequence of decoding functions $\mathsf{g}_n\colon \mathcal{U}\times\mathcal{B}^n\mapsto\mathcal{W}$, $n=1,2,\ldots$*
4) *A non-negative integer-valued random variable τ, a stopping time of the filtration $\mathcal{F}_n = \sigma\left\{U, Y_1,\ldots,Y_n\right\}$.*

The final decision $\widehat{W}$ is computed at the time instant τ

$$\widehat{W} = \mathsf{g}_\tau(U, Y^\tau) \tag{7}$$

The value $\mathbb{E}\left[\tau\right]$ is the average transmission length of the given code.

A very similar concept is that of an VLFT code:

Definition 2. *A variable-length feedback code with termination (VLFT) transmitting $W\in\mathcal{W}$ over the channel $\{P_{Y_i|X^iY^{i-1}}\}_{i=1}^\infty$ with input/output alphabets $\mathcal{A}$/$\mathcal{B}$ is defined similarly to VLF codes with the exception that condition 4) in the Definition 1 is replaced by*

4') *A non-negative integer-valued random variable τ, a stopping time of the filtration $\mathcal{G}_n = \sigma\{W, U, Y_1,\ldots,Y_n\}$.*

The idea of allowing τ to depend on the true message W models the practical scenarios (see [3]) where there is a highly reliable control layer operating in parallel with the data channel, which notifies the decoder when it is time to make a decision.

The two special cases of the above definitions are stop-feedback and fixed-to-variable codes:

1) *stop-feedback codes* are a special case of VLF codes where the encoder functions $\{f_n\}_{n=1}^\infty$ satisfy:
$$f_n(U,W,Y^{n-1}) = f_n(U,W)\,. \tag{8}$$
Such codes require very limited communication over feedback: only a single signal to stop the transmission once the decoder is ready to decode.
2) fixed-to-variable codes, defined in [28], are also required to satisfy (8), while the stopping time is[3]
$$\tau = \inf\{n\geq 1 : g_n(U,Y^n) = W\}\,, \tag{9}$$
and therefore, such codes are zero-error VLFT codes.

For both VLF and VLFT codes, we say that a code that satisfies $\mathbb{E}\left[\tau\right]\leq\ell$ and $\mathbb{P}\left[W\neq\widehat{W}\right]\leq\epsilon$, when averaged over U, message and channel, is an (ℓ,ϵ) code for source/channel $\left(W,\{P_{Y_i|X^iY^{i-1}}\}_{i=1}^\infty\right)$.

The random variable U serves as the common randomness shared by both transmitter and receiver, which is used to initialize the codebook. As a consequence of Caratheodory's theorem, the amount of this common randomness can always be reduced to just a few bits: as shown in [3, Theorem 19], if there exists an (ℓ,ϵ) code with $|\mathcal{U}|=\infty$, then there exists an

[2]Stop-feedback is not to be confused with the termination symbol, which is a special symbol that the encoder can transmit error-free in order to tell the decoder that the transmission has ended and it is time to decode.

[3]As explained in [28], this model encompasses fountain codes in which the decoder can get a highly reliable estimate of τ autonomously without the need for a termination symbol.

(ℓ, ϵ) code with $|\mathcal{U}| \leq 3$. Allowing for common randomness does not affect the asymptotic expansions, but leads to more concise expressions for non-asymptotic achievability bounds. Furthermore, if the channel is symmetric no common randomness is needed at all [3, Theorem 3].

Our first result generalizes the achievability result [3, (107)-(118)] to the case of non-equiprobable messages.

Theorem 1. *For every DMC with capacity C and random variable W there exists an (ℓ, ϵ) stop-feedback code for W with*

$$C\ell \leq H(W) + \log \frac{1}{\epsilon} + a_0 \tag{10}$$

where

$$a_0 \triangleq \max_{x_1, x_2, y_1, y_2} \log \frac{P_{Y|X}(y_1|x_1)}{P_{Y|X}(y_2|x_2)}. \tag{11}$$

and the maximum is taken over all pairs (x_i, y_i) with $P_{Y|X}(y_i|x_i) > 0,\ i = 1, 2$.

Proof. See extended version [29]. □

A slightly less sharp bound could be derived via a *variable-length separated scheme*: compress W losslessly into a variable-length string $\{0,1\}^*$ with average length less than $H(W)$, cf. [30], then send the length via $O(\log H(W))$ channel symbols and very small probability of error and finally send the data bits.

Next, we extend the zero-error bound in [3, Theorem 10] to the case of non-equiprobable messages:

Theorem 2. *For every DMC with capacity C there exists a constant a_1 such that for every discrete random variable W there exists an $(\ell, 0)$ VLFT code with*

$$C\ell \leq H(W) + a_1 \tag{12}$$

Proof. See extended version [29]. □

III. Asymptotic expansions of the rate-distortion tradeoff

A. Definitions

We move from the setup of Section II where a discrete message is transmitted over the feedback channel to a more general scenario of Section III, in which a possibly analog signal is transmitted over a channel with feedback, under a fidelity constraint. We will consider the following scenarios:

1) *excess distortion:* A VLF code transmitting memoryless source $(S^k, \mathcal{S}^k)$ with reproduction alphabet $\hat{\mathcal{S}}^k$ and separable distortion measure $\mathsf{d}\colon \mathcal{S}^k \times \hat{\mathcal{S}}^k \mapsto [0, +\infty]$ is called a (k, ℓ, d, ϵ) excess-distortion code if

$$\mathbb{E}[\tau] \leq \ell \tag{13}$$

$$\mathbb{P}[\mathsf{d}(S^k, \hat{S}^k) \leq d] \leq \epsilon \tag{14}$$

The corresponding fundamental limit is

$$\ell^\star(k, d, \epsilon) \triangleq \inf\{\ell\colon \exists \text{ an } (k, \ell, d, \epsilon) \text{ VLF code}\}\,. \tag{15}$$

2) *average distortion:* A VLF code satisfying, instead of (14), an average constraint

$$\mathbb{E}[\mathsf{d}(S^k, \hat{S}^k)] \leq d \tag{16}$$

is called a (k, ℓ, d) average-distortion code. The corresponding fundamental limit is

$$\ell^\star(k, d) \triangleq \inf\{\ell\colon \exists \text{ an } (k, \ell, d) \text{ VLF code}\}\,. \tag{17}$$

3) *guaranteed distortion*: A VLFT code transmitting memoryless source $(S^k, \mathcal{S}^k)$ with reproduction alphabet $\hat{\mathcal{S}}^k$ and separable distortion metric d is called a $(k, \ell, d, 0)$ guaranteed-distortion code if it achieves $\epsilon = 0$ in (14). The corresponding fundamental limit is

$$\ell_t^\star(k, d, 0) \triangleq \inf\{\ell\colon \exists \text{ an } (k, \ell, d, 0) \text{ VLFT code}\}\,. \tag{18}$$

We will use the following notation for the various minimizations of information measures:

$$\mathbb{R}_S(d) \triangleq \min_{\substack{P_{Z|S}\colon \mathcal{S} \mapsto \hat{\mathcal{S}}:\\ \mathbb{E}[\mathsf{d}(S,Z)] \leq d}} I(S;Z) \tag{19}$$

$$\mathbb{R}_S(d, \epsilon) \triangleq \min_{\substack{P_{Z|S}\colon \mathcal{S} \mapsto \hat{\mathcal{S}}:\\ \mathbb{P}[\mathsf{d}(S,Z) > d] \leq \epsilon}} I(S;Z) \tag{20}$$

$$H_{d,\epsilon}(S) \triangleq \min_{\substack{\mathsf{c}\colon \mathcal{S} \mapsto \hat{\mathcal{S}}:\\ \mathbb{P}[\mathsf{d}(S,\mathsf{c}(S)) > d] \leq \epsilon}} H(\mathsf{c}(S)). \tag{21}$$

The quantity in (21) is referred to as the (d, ϵ)-entropy of the source S [31]. The $(\epsilon, 0)$-entropy is also known as just ϵ-entropy [31].

Provided that the infimum in (19) is achieved by some transition probability kernel $P_{Z^\star|S}$, the d-tilted information in $s \in \mathcal{S}$ is defined as [23]

$$\jmath_S(s, d) \triangleq -\log \mathbb{E}\left[\exp\left(-\lambda^\star \mathsf{d}(s, Z^\star) + \lambda^\star d\right)\right] \tag{22}$$

where

$$\lambda^\star = -\mathbb{R}_S'(d). \tag{23}$$

In the almost-lossless compression, $Z^\star = S$ and

$$\jmath_S(s, d) \triangleq \imath_S(s) \tag{24}$$

$$\triangleq \log \frac{1}{P_S(s)}. \tag{25}$$

B. Regularity assumptions on the source

We assume that the source, together with its distortion measure, satisfies the following assumptions:

A1 The source $\{S_i\}$ is stationary and memoryless, $P_{S^k} = P_{\mathsf{S}} \times \ldots \times P_{\mathsf{S}}$.

A2 The distortion measure is separable, $\mathsf{d}(s^k, z^k) = \frac{1}{k} \sum_{i=1}^k \mathsf{d}(s_i, z_i)$.

A3 The distortion level satisfies $d_{\min} < d < d_{\max}$, where $d_{\min}$ is the infimum of values at which the minimal mutual information quantity $\mathbb{R}_S(d)$ is finite, and $d_{\max} = \inf_{\mathsf{z} \in \widehat{\mathcal{M}}} \mathbb{E}[\mathsf{d}(\mathsf{S}, \mathsf{z})]$, where the expectation is with respect to the unconditional distribution of S.

A4 The rate-distotortion function is achieved by a unique $P_{\mathsf{Z}^\star|\mathsf{S}}$: $\mathbb{R}_{\mathsf{S}}(d) = I(\mathsf{S};\mathsf{Z}^\star)$.

A5 $\mathbb{E}\left[\mathsf{d}^{12}(\mathsf{S},\mathsf{Z}^\star)\right] < \infty$ where the expectation is with respect to $P_{\mathsf{S}} \times P_{\mathsf{Z}^\star}$.

The rate-dispersion function of the source satisfying assumptions A1–A5 is given by [23]

$$\mathcal{V}(d) = \mathrm{Var}\left(\jmath_{\mathsf{S}}(\mathsf{S}, d)\right). \tag{26}$$

We showed in [5] that under assumptions A1–A5 for all $0 \le \epsilon \le 1$

$$\left.\begin{matrix} \mathbb{R}_{S^k}(d,\epsilon) \\ H_{d,\epsilon}(S^k) \end{matrix}\right\} = (1-\epsilon)kR(d) - \sqrt{\frac{k\mathcal{V}(d)}{2\pi}}e^{-\frac{(Q^{-1}(\epsilon))^2}{2}} + O\left(\log k\right). \tag{27}$$

C. Average distortion

Theorem 3. *Under assumptions A1–A5 we have*

$$C\ell^*(k,d) = kR(d) + O(\log k). \tag{28}$$

Proof. See extended version [29]. □

D. Excess distortion

Theorem 4. *Under assumptions A1–A5 and any $\epsilon > 0$ we have*

$$\ell^\star(k,d,\epsilon)C = (1-\epsilon)kR(d) - \sqrt{\frac{k\mathcal{V}(d)}{2\pi}}e^{-\frac{(Q^{-1}(\epsilon))^2}{2}} + O\left(\log k\right) \tag{29}$$

Proof. Achievability: Pair a lossy compressor $S^k \to W$ with excess-distortion probability $\epsilon' = \epsilon - \frac{1}{\sqrt{k}}$ and $H(W) = H_{d,\epsilon'}(S^k)$ with a VLF code from Theorem 1 transmitting W with probability of error $\frac{1}{\sqrt{k}}$. Apply (27) to (10).

Converse: Apply the data-processing inequality and [2, Lemma 1-2] to get:

$$\ell C \ge \mathbb{R}_{S^k}(d,\epsilon) \tag{30}$$

for every (k,ℓ,d,ϵ) VLF code. □

E. Guaranteed distortion

Theorem 5. *Under assumptions A1–A5, we have*

$$\ell_t^\star(k,d,0)C = kR(d) + O\left(\log k\right) \tag{31}$$

Proof. For the achievability we note that the estimate of the $H_{d,\epsilon}(S^k)$ in (27) applies with $\epsilon = 0$ and thus

$$H_{d,0}(S^k) = kR(d) + O(\log k). \tag{32}$$

Then, we can pair the mapping achieving $H_{d,0}(S^k)$ with the zero-error VLFT code from Theorem 2.

Conversely, repeating the argument of [3, Theorem 4], with the replacement of the right side of [3, (67)] by $\mathbb{R}_S(d,\epsilon)$ we conclude that any (ℓ,d,ϵ) VLFT code must satisfy

$$\mathbb{R}_S(d,\epsilon) \le C\ell + \log(\ell+1) + \log e. \tag{33}$$

□

F. Discussion

We make several remarks regarding the rate-distortion tradeoff in all three settings considered above:

1) The case $d = d_{\min}$ is special and is excluded in the assumptions of Theorems 3–5. However, in the most important special case of a distortion measure that satisfies

$$\mathsf{d}(a,b) = \begin{cases} d_{\min}, & a = b \\ > d_{\min}, & a \ne b \end{cases}, \tag{34}$$

$d = d_{\min}$ corresponds to almost-lossless transmission, and both Theorems 3 and 4 apply with $R(d)$ and $\mathcal{V}(d)$ equal to the entropy and the varentropy of the source, respectively, as long as the source is stationary and memoryless and the third moment of $\imath_{\mathsf{S}}(\mathsf{S})$ is finite.

2) For almost-lossless transmission of finite alphabet sources, the asymptotic expansion (29) can be achieved by reliably (i.e. with probability of error $\sim \frac{1}{k}$) sending through the channel the type of the source outcome first, and then reliably sending each message whose type is one the most likely types with total mass $1-\epsilon$.

3) Note that (29) is achieved by a stop-feedback code. We can further show that even without any feedback one can still achieve the optimal first order performance

$$\ell C \le (1-\epsilon)kR(d) + O(\sqrt{k\log k}), \tag{35}$$

provided variable-length channel coding is allowed. Indeed, one can first use the variable-length excess-distortion compressor from [5] on S^k to get a binary string of average length $(1-\epsilon)kR(d) + O(\sqrt{k})$, see (2). Then, truncating the length at k^2 and transmitting $2\log k$ data bits with reliability $\frac{1}{k^2}$, we can reliably inform the encoder about the total number of data bits b to be sent next. We may then use a capacity-achieving code of length $\frac{b}{C} + O(\sqrt{b\log b})$ to send the data bits with reliability $\frac{1}{k}$ [22].

4) The naive separation achieves at most:

$$\ell C \ge (1-\epsilon)kR(d) + a\sqrt{k\log k},\ a > 0. \tag{36}$$

5) The Schalkwijk-Bluestein [32] elegant linear feedback scheme for the transmission of a single Gaussian sample $S \sim \mathcal{N}(0,\sigma^2)$ over the AWGN channel achieves the mean-square error $\frac{\sigma^2}{(1+P)^n}$, after n channel uses, where P is the average transmit SNR. In other words, the minimum delay in transmitting a Gaussian sample over a Gaussian channel with feedback is given by

$$\ell^\star(1,d) = \frac{R(d)}{C}, \tag{37}$$

as long as $\frac{R(d)}{C}$ is integer.[4] Note that (37) is achieved with fixed, not variable length, and average, not maximal, power constraint. If there are k Gaussian samples to

[4] If $\frac{R(d)}{C} = 1$, no feedback is needed.

transmit, repeating the scheme for each of the samples achieves

$$\ell^\star(k, d) = k\frac{R(d)}{C}, \tag{38}$$

which implies, in particular, that in general our estimate of $O(\log k)$ in (28) is too conservative. Beyond Gaussian sources and channels, a sufficient condition for a fixed-length JSCC feedback scheme to achieve (38) is provided in [16].

IV. Conclusion

We have considered several scenarios for joint source-channel coding with feedback. Our main conclusions are:

1) The average delay vs. distortion tradeoff with feedback is governed by channel capacity, and the source rate-distortion and rate-dispersion functions. In particular, the channel dispersion plays no role.
2) In variable-length coding with feedback, the asymptotically achievable minimum average length is reduced by a factor of $1-\epsilon$, where ϵ is the excess distortion probability. This asymptotic fundamental limit is approached from below, i.e., counter-intuitively, smaller source blocklengths may lead to smaller attainable average delays.
3) Introducing a termination symbol that is always decoded error-free allows for transmission over noisy channels with guaranteed distortion.
4) Variable-length transmission without feedback still improves the asymptotic fundamental limit by a factor of $1 - \epsilon$, where ϵ is the excess distortion probability.
5) In all the cases we have analyzed the approach to the fundamental limits is very fast: $O\left(\frac{\log k}{k}\right)$, where k is the source blocklength. This behavior is attained, under average distortion, by a separated scheme with stop-feedback.
6) Variable-length separated schemes perform remarkably well in all considered scenarios.

References

[1] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, Sep. 1956.

[2] M. V. Burnashev, "Data transmission over a discrete channel with feedback: Random transmission time," *Problems of Information Transmission*, vol. 12, no. 4, pp. 10–30, 1976.

[3] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Feedback in the non-asymptotic regime," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4903–4925, Aug. 2011.

[4] S. L. Fong and V. Y. F. Tan, "Asymptotic expansions for Gaussian channels with feedback under a peak power constraint asymptotic expansions for gaussian channels with feedback under a peak power constraint," *ArXiv preprint arXiv:1410.2390*, Oct. 2014.

[5] V. Kostina, Y. Polyanskiy, and S. Verdú, "Variable-length compression allowing errors (extended)," *ArXiv preprint, arXiv:1402.0608*, Jan. 2014.

[6] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2545–2575, May 2013.

[7] A. Tauste Campo, G. Vazquez-Vilar, A. Guillén i Fàbregas, and A. Martinez, "Random-coding joint source-channel bounds," in *Proceedings 2011 IEEE International Symposium on Information Theory*, Saint-Petersburg, Russia, Aug. 2011, pp. 899–902.

[8] D. Wang, A. Ingber, and Y. Kochman, "The dispersion of joint source-channel coding," in *Proceedings 49th Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, Sep. 2011.

[9] T. Javidi and A. Goldsmith, "Dynamic joint source-channel coding with feedback," in *IEEE International Symposium on Information Theory*, Istanbul, Turkey, July 2013, pp. 16–20.

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

[11] I. A. Ovseevich, "Capacity of a randomized channel with feedback and matching of the source," *Problemy Peredachi Informatsii*, vol. 4, no. 1, pp. 52–59, 1968.

[12] R. Fang, "Optimum linear feedback scheme for matching colored Gaussian sources with white Gaussian channels," *IEEE Transactions on Information Theory*, vol. 16, no. 6, pp. 789–791, 1970.

[13] V. Chande, H. Jafarkhani, and N. Farvardin, "Joint source-channel coding of images for channels with feedback," in *Proceedings 1998 IEEE Information Theory Workshop*, Feb. 1998.

[14] V. Kafedziski, "Joint source channel coding of images over frequency selective fading channels with feedback using dct and multicarrier block pulse amplitude modulation," in *Proceedings Asilomar Conf. Signals, Systems and Computers*, vol. 1, 1998, pp. 37–41.

[15] J. Lu, A. Nosratinia, and B. Aazhang, "Progressive joint source-channel coding in feedback channels," in *Proceedings 1999 Data Compression Conference*, 1999, pp. 140–148.

[16] M. Gastpar and B. Rimoldi, "Source-channel communication with feedback," in *Proceedings 2003 IEEE Information Theory Workshop*, Paris, France, 2003, pp. 279–282.

[17] R. Manakkal and B. Rimoldi, "A source-channel coding scheme for discrete memoryless channels with feedback," in *Proceedings 2005 IEEE International Symposium on Information Theory*, Adelaide, Australia,, 2005, pp. 1511–1515.

[18] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Transactions on Information Theory*, vol. 9, no. 3, pp. 136–143, 1963.

[19] J. M. Ooi and G. W. Wornell, "Fast iterative coding techniques for feedback channels," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2960–2976, 1998.

[20] R. Ahlswede, "A constructive proof of the coding theorem for discrete memoryless channels with feedback," in *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, Prague, Sep. 1971.

[21] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1186–1222, 2011.

[22] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[23] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3309–3338, June 2012.

[24] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: An approach to combined source-channel coding," *IEEE Trans. on Information Theory*, vol. 33, no. 6, pp. 827–838, 1987.

[25] A. Amanullah and M. Salehi, "Joint source-channel coding in the presence of feedback," in *Proc. 27th Asilomar Conf. Sig. Syst. Comp.*, 1993, pp. 930–934.

[26] S. Gorantla and T. Coleman, "Information-theoretic viewpoints on optimal causal coding-decoding problems," *arXiv preprint arXiv:1102.0250*, 2011.

[27] B. Hochwald and K. Zeger, "Tradeoff between source and channel coding," *IEEE Transactions on Information Theory*, vol. 43, no. 5, pp. 1412–1424, 1997.

[28] S. Verdú and S. Shamai, "Variable-rate channel capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2651–2667, 2010.

[29] V. Kostina, Y. Polyanskiy, and S. Verdú, "Joint source-channel coding with feedback (extended)," *ArXiv preprint*, Jan. 2015.

[30] A. Wyner, "An upper bound on the entropy series," *Information and Control*, vol. 20, no. 2, pp. 176–181, 1972.

[31] E. C. Posner, E. R. Rodemich, and H. Rumsey, "Epsilon-entropy of stochastic processes," *The Annals of Mathematical Statistics*, vol. 38, no. 4, pp. 1000–1020, 1967.

[32] J. Schalkwijk and L. Bluestein, "Transmission of analog waveforms through channels with feedback," *IEEE Transactions on Information Theory*, vol. 13, no. 4, pp. 617–619, 1967.