Channel Coding with Mean and Variance Cost Constraints

Adeel Mahmood
School of Electrical and Computer Engineering
Cornell University
Ithaca, NY 14853 USA
am2384@cornell.edu

Aaron B. Wagner
School of Electrical and Computer Engineering
Cornell University
Ithaca, NY 14853 USA
wagner@cornell.edu

2024 IEEE International Symposium on Information Theory (ISIT) | 979-8-3503-8284-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISIT57864.2024.10619384

***Abstract*—We consider channel coding for discrete memoryless channels (DMCs) with a novel cost constraint that constrains both the mean and the variance of the cost of the codewords. We show that the maximum (asymptotically) achievable rate under the new cost formulation is equal to the capacity-cost function; in particular, the strong converse holds. We further characterize the optimal second-order coding rate of these cost-constrained codes; in particular, the optimal second-order coding rate is finite. We then show that the second-order coding performance is strictly improved with feedback using a new variation of timid/bold coding, significantly broadening the applicability of timid/bold coding schemes from unconstrained compound-dispersion channels to all cost-constrained channels. Equivalent results on the minimum average probability of error are also given.**

I. Introduction

In practice, channel coding is subject to various cost constraints which limit the amount of resources that can be used for transmission. Such constraints may arise out of concern for interference with other terminals or, especially in the case of mobile devices, power consumption. With a cost constraint present, the role of capacity is replaced by the capacity-cost function [1, Theorem 6.11]. We focus on discrete memoryless channels (DMCs) with a cost function denoted by $c(\cdot)$. One common cost constraint called the almost-sure (a.s.) cost constraint [2], [3] bounds the time-average cost of the channel input X^n over all messages, realizations of any side randomness, channel noise (if there is feedback), etc.:

$$\frac{1}{n}\sum_{i=1}^{n} c(X_i) \leq \Gamma \quad \text{a.s.} \tag{1}$$

On the other hand, the expected cost constraint bounds the sum-cost in the average sense:

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[c(X_i)] \leq \Gamma. \tag{2}$$

These two cost constraints are also called short-term and long-term power constraints, respectively, in certain contexts [4]. With an almost-sure cost constraint, the strong converse holds. With an expected cost constraint, the strong converse ceases to hold [5, Theorem 77]. Accordingly, past work on second-order coding rates with cost constraints ([2], [6]) has focused on the almost-sure constraint. The second-order coding rate ([7], [2], [8], [9], [10]) quantifies the $O(n^{-1/2})$ convergence to the capacity-cost function (or to the capacity in the unconstrained case). Under the a.s. cost formulation, the optimal second-order coding rate (SOCR) is known [2, Theorem 3].

One of the lessons of information-theoretic studies of channel coding is that the various codewords should appear to be selected randomly and independent and identically distributed (i.i.d.) according to P^*, where P^* is a capacity-cost-achieving input distribution. The idea of generating codewords in an i.i.d. fashion is so natural and ubiquitous that it is notable that it is actually impermissible under (1).

More seriously, one incurs a performance loss by prohibiting the use of i.i.d.-generated codewords in second-order coding rate (SOCR) analyses. Consider the problem with ideal feedback from the output of the channel to the encoder. For channels without cost constraints, it is known that feedback can improve the SOCR for compound-dispersion channels [7]. Specifically, suppose a channel $W(\cdot|\cdot)$ has two capacity-achieving input distributions P_1^* and P_2^* such that

$$\mathrm{Var}_{P_1^* \circ W}\left(\log \frac{W(Y|X)}{P_1^* W(Y)}\right) < \mathrm{Var}_{P_2^* \circ W}\left(\log \frac{W(Y|X)}{P_2^* W(Y)}\right), \tag{3}$$

where $P_i^* \circ W$ denotes the joint distribution over inputs and outputs induced by the distributions P_i^* and $W(\cdot|\cdot)$, and $P_i^* W$ denotes the marginal distribution of the induced output. While codewords drawn from P_1^* and P_2^* have the same mean information-carrying ability by virtue of P_1^* and P_2^* both being capacity-achieving, codewords drawn from P_2^* are more variable as a consequence of (3). Thus, the encoder can employ codewords drawn from P_1^* ("timid") so long as the transmission is proceeding well and from P_2^* ("bold") if an error appears likely. This is referred to as *timid/bold* coding.

One limitation of the above idea is that the channel must be compound-dispersion. In particular, the capacity-achieving input distribution for $W(\cdot|\cdot)$ cannot be unique. In fact, for simple-dispersion channels for which (3) does not hold, feedback does not improve the SOCR [7, Theorem 3].

The recent work [11] studied the feedback improvement of SOCR with a cost constraint. The cost constraint in [11] was intermediate between an almost-sure cost constraint and an expected cost constraint. With the intermediate cost constraint, [11] showed that the timid/bold feedback scheme can improve the SOCR for DMCs even if the capacity-cost-achieving distribution is unique, thus broadening the scope of timid/bold coding beyond [7]. Specifically, let P^* denote a capacity-cost-achieving input distribution for the DMC $W(\cdot|\cdot)$, which might well be unique. By the law of total variance, the n-length form of the variance in (3) can be written as

$$\mathrm{Var}\left(\log\frac{W(Y^n|X^n)}{P^*W(Y^n)}\right)=\mathbb{E}\left[\mathrm{Var}\left(\log\frac{W(Y^n|X^n)}{P^*W(Y^n)}\middle|X^n\right)\right]+\mathrm{Var}\left(\mathbb{E}\left[\log\frac{W(Y^n|X^n)}{P^*W(Y^n)}\middle|X^n\right]\right). \tag{4}$$

If the channel input X^n is constant-composition, i.e., drawn uniformly from a fixed type class associated with a distribution that is close to P^*, then the quantity

$$\mathbb{E}\left[\log\frac{W(Y^n|X^n)}{P^*W(Y^n)}\middle|X^n\right] \tag{5}$$

is a.s. constant and the second term in (4) is zero. In contrast, if X^n is i.i.d. according to P^*, then the second term in (4) is order-n (see [11, Lemma 2]). The first term, in contrast, is approximately the same between the two cases. Thus both timid and bold signaling mechanisms can be created from P^* alone, depending on whether one uses constant-composition or i.i.d. codewords. Yet, this observation cannot be applied under the prevailing a.s. cost formulation for second-order rate analysis because i.i.d. codewords are impermissible under the a.s. constraint in (1).

By formulating an intermediate cost constraint that allows both i.i.d. and constant-composition channel inputs, [11] showed a strict improvement of SOCR with feedback. However, relaxing the a.s. cost constraint to permit i.i.d. codewords should be approached cautiously, as the expected cost constraint from (2), which similarly accommodates i.i.d. codewords, also admits signaling schemes characterized by a highly non-ergodic power usage, leading to the absence of a strong converse. Furthermore, the achievable SOCR with feedback in [11, Theorem 1] is only shown to exceed the optimal SOCR of the almost-sure cost constraint [2] without feedback. Hence, it is also not clear in [11] how much of the demonstrated improvement in SOCR is due to feedback and how much of it is due to relaxing the almost-sure cost constraint.

In this paper, we introduce a new (Γ, V) cost constraint which constrains both the mean and the variance of the codewords:

$$\mathbb{E}\left[\sum_{i=1}^{n}c(X_i)\right]\le n\Gamma \tag{6}$$

$$\mathrm{Var}\left(\sum_{i=1}^{n}c(X_i)\right)\le nV. \tag{7}$$

The (Γ, V) cost constraint is a natural strengthening of the expected cost constraint via a second-moment constraint. The idea is that we want power to be consumed at a limited rate but also in a predictable fashion, both to ensure a gradual consumption of energy and so that the transmitted signal is sufficiently ergodic that it can be treated as noise by other terminals (which mitigates the negative impact of interference). Note that (7) ensures that $c(X_i)$ satisfies the weak law of large numbers as $n \to \infty$.

With an additional variance constraint in (Γ, V) channel codes for DMCs, the strong converse holds. Our new cost constraint also admits a finite second-order converse. We give matching achievability and converse results characterizing the optimal SOCR in terms of a function of Γ and V, which takes the form

$$\inf_{\Pi}\mathbb{E}\left[\Phi(\Pi)\right], \tag{8}$$

where $\Phi(\cdot)$ is the standard Gaussian CDF and the infimum is over all random variables Π with an appropriately constrained expectation and variance. We characterize the solution to the optimization problem in (8) as well as the properties of a function $\mathcal{K}(r, V)$ which is equal to (8) with the expectation and variance constrained by r and V, respectively.

After establishing the optimal coding second-order coding performance under the (Γ, V) constraint, we show that this performance is strictly improved with feedback. Our feedback scheme is a new variant of timid/bold coding which requires neither multiple capacity-cost-achieving distributions as in [7] nor i.i.d. codewords as in [11]. The latter feature is useful because, although the (Γ, V) cost constraint allows i.i.d. channel inputs with a bounded variance $V > 0$, it does not admit i.i.d. P^* codewords for small values of V. Nevertheless, the (Γ, V) code with feedback shows a strict SOCR improvement for all values of $V > 0$. Therefore, a more foundational advantage of the (Γ, V) cost constraint is allowing a nonzero variance of the cost of the channel input around the cost point Γ, while still sufficiently regulating power consumption to ensure a finite second-order coding rate.

The proofs of the main lemmas and theorems are given in the longer version of this paper [12].

II. Preliminaries

Let $\mathcal{A}$ and $\mathcal{B}$ be finite input and output alphabets, respectively, of a DMC. Let $\mathcal{P}(\mathcal{A})$ be the set of probability distributions on $\mathcal{A}$. We will use W to denote the DMC. Let $\mathcal{P}_n(\mathcal{A})$ be the set of n-types on $\mathcal{A}$. For a given $t \in \mathcal{P}_n(\mathcal{A})$, $T^n_{\mathcal{A}}(t)$ denotes the type class. For a given $P \in \mathcal{P}(\mathcal{A})$, $P \circ W$ denotes the joint distribution on $\mathcal{A} \times \mathcal{B}$ induced by P and W, and PW denotes the corresponding marginal distribution on $\mathcal{B}$.

The cost function is denoted by $c(\cdot)$ where $c : \mathcal{A} \to [0, c_{\max}]$ and $c_{\max} > 0$ is a constant. Let $\Gamma_0 = \min_{a\in\mathcal{A}} c(a)$. For $\Gamma > \Gamma_0$, the capacity-cost function is defined as

$$C(\Gamma)=\max_{\substack{P\in\mathcal{P}(\mathcal{A})\\ c(P)\le\Gamma}} I(P, W), \tag{9}$$

where $c(P) := \sum_{a\in\mathcal{A}} P(a)c(a)$. Let Γ^* denote the smallest Γ such that the capacity-cost function $C(\Gamma)$ is equal to the unconstrained capacity C. We assume $\Gamma^* > \Gamma_0$ and $\Gamma \in (\Gamma_0, \Gamma^*)$ throughout the paper. We will also assume that the capacity-cost-achieving distribution for cost Γ is unique. We will use P^* to denote the unique solution to (9). For the application to feedback communication, this is the most interesting case, since if P^* is not unique, the timid/bold scheme of [7] is already applicable. This assumption also has precedent in the literature (e.g., [6]), because it affords certain technical simplifications (e.g., [12, Lemma 10]). Note that we do not assume uniqueness for costs $\Gamma' \neq \Gamma$.

We define the output distribution $Q^* := P^*W$ and dispersion $V(\Gamma) := \sum_{a\in\mathcal{A}} P^*(a)\nu_a$. A channel input $X^n \sim \text{Unif}(T^n_{\mathcal{A}}(t))$ drawn uniformly from a type class t is called a constant-composition (cc) input. We will denote by Q^{cc} the output distribution induced by the input $X^n \sim \text{Unif}(T^n_{\mathcal{A}}([P^*]_n))$ through the DMC W.

With a blocklength n and a fixed rate $R > 0$, let $\mathcal{M} = \{1, \dots, \lceil \exp(nR) \rceil\}$ denote the message set.

Definition 1: An (n, R) code for a DMC consists of an encoder f which, for each message $m \in \mathcal{M}$, chooses an input $X^n = f(m) \in \mathcal{A}^n$, and a decoder g which maps the output Y^n to $\hat{m} \in \mathcal{M}$. The code (f, g) is random if f or g is random.

Definition 2: An (n, R) code with ideal feedback for a DMC consists of an encoder f which, at each time instant k $(1 \leq k \leq n)$ and for each message $m \in \mathcal{M}$, chooses an input $x_k = f(m, x^{k-1}, y^{k-1}) \in \mathcal{A}$, and a decoder g which maps the output y^n to $\hat{m} \in \mathcal{M}$. The code (f, g) is random if f or g is random.

As noted in the introduction, we consider a cost constraint that restricts both the mean and the variance of the codewords by some $\Gamma \in (\Gamma_0, \Gamma^*)$ and $V > 0$, respectively.

Definition 3: An (n, R, Γ, V) code for a DMC is an (n, R) code such that $\mathbb{E}\left[\sum_{i=1}^n c(X_i)\right] \leq n\Gamma$ and $\text{Var}\left(\sum_{i=1}^n c(X_i)\right) \leq nV$, where the message $M \sim \text{Unif}(\mathcal{M})$ has a uniform distribution over the message set $\mathcal{M}$.

Definition 4: An (n, R, Γ, V) code with ideal feedback for a DMC is an (n, R) code with ideal feedback such that $\mathbb{E}\left[\sum_{i=1}^n c(X_i)\right] \leq n\Gamma$ and $\text{Var}\left(\sum_{i=1}^n c(X_i)\right) \leq nV$, where the message $M \sim \text{Unif}(\mathcal{M})$ has a uniform distribution over the message set $\mathcal{M}$.

We consider optimum coding performance for (n, R, Γ, V) codes defined in Definitions 3 and 4. Given $\epsilon \in (0, 1)$, define

$$M^*_{\text{fb}}(n, \epsilon, \Gamma, V) := \max\{\lceil \exp(nR) \rceil : \bar{P}_{\text{e,fb}}(n, R, \Gamma, V) \leq \epsilon\},$$

where $\bar{P}_{\text{e,fb}}(n, R, \Gamma, V)$ denotes the minimum average error probability attainable by any (n, R, Γ, V) code with feedback. Similarly, define

$$M^*(n, \epsilon, \Gamma, V) := \max\{\lceil \exp(nR) \rceil : \bar{P}_{\text{e}}(n, R, \Gamma, V) \leq \epsilon\},$$

where $\bar{P}_{\text{e}}(n, R, \Gamma, V)$ denotes the minimum average error probability attainable by any (n, R, Γ, V) code without feedback. For (n, R, Γ, V) codes for DMCs defined in Definition 3, the following (achievability) bound on the second-order coding rate can be obtained by using the coding scheme from [2, Theorem 3]:

$$\liminf_{n\to\infty} \frac{\log M^*(n, \epsilon, \Gamma, V) - nC(\Gamma)}{\sqrt{n}} \geq \sqrt{V(\Gamma)}\Phi^{-1}(\epsilon) \quad (10)$$

for $\epsilon \in (0, 1)$, where the right-hand side of (10) is the optimal SOCR associated with an almost-sure cost constraint, assuming a unique capacity-cost-achieving distribution P^* at cost Γ. It is easy to see that a constant-composition code with a fixed type class satisfies the (Γ, V) cost constraint. While constant-composition codes hit the optimal second-order coding rate with an almost-sure cost constraint, such codes are not necessarily optimal with the new (Γ, V) cost constraint.

Definition 5: A controller is a function $F : (\mathcal{A} \times \mathcal{B})^* \to \mathcal{P}(\mathcal{A})$.

We shall sometimes write $F(\cdot|x^k, y^k)$ for $F(x^k, y^k)(\cdot)$. The design of random feedback codes (f, g) can be directly related to the design of controllers [7], [11]. Specifically, any controller F satisfying

$$\mathbb{E}\left[\sum_{i=1}^n c(X_i)\right] \leq n\Gamma \text{ and } \text{Var}\left(\sum_{i=1}^n c(X_i)\right) \leq nV \quad (11)$$

can be used to construct an (n, R, Γ, V) code with ideal feedback.

Lemma 1 ([7, Lemma 14]): For any $\Gamma \in (\Gamma_0, \Gamma^*)$ and $V > 0$, a controller F satisfying (11), and any n, θ and R,

$$\begin{aligned}&\bar{P}_{\text{e,fb}}(n, R, \Gamma, V)\\ &\leq (F \circ W)\left(\frac{1}{n}\log\frac{W(Y^n|X^n)}{FW(Y^n)} \leq R + \theta\right) + e^{-n\theta}, \quad (12)\end{aligned}$$

where (X^n, Y^n) have the joint distribution specified by

$$(F \circ W)(x^n, y^n) = \prod_{k=1}^n F(x_k|x^{k-1}, y^{k-1})W(y_k|x_k),$$

and FW denotes the marginal distribution of Y^n. Furthermore, if for some α and ϵ,

$$\limsup_{n\to\infty} (F \circ W)\left(\frac{1}{n}\log\frac{W(Y^n|X^n)}{FW(Y^n)} \leq C(\Gamma) + \frac{\alpha}{\sqrt{n}}\right) < \epsilon, \quad (13)$$

then the controller F gives rise to an achievable second-order coding rate of α, i.e.,

$$\liminf_{n\to\infty} \frac{\log M^*_{\text{fb}}(n, \epsilon, \Gamma, V) - nC(\Gamma)}{\sqrt{n}} \geq \alpha. \quad (14)$$

Similar results to (12), (13) and (14) hold for $\bar{P}_{\text{e}}(n, R, \Gamma, V)$ and $M^*(n, \epsilon, \Gamma, V)$ in the non-feedback case by replacing controllers F by distributions $P \in \mathcal{P}(\mathcal{A}^n)$.

Remark 1: Lemma 1 is a starting point to prove achievability results both with feedback (Theorem 3) and without feedback (Theorem 2).

Lemma 2: Consider a channel W with cost constraint $(\Gamma, V) \in (\Gamma_0, \Gamma^*) \times (0, \infty)$, where (Γ, V) maps to some subset $\mathcal{P}_{\Gamma,V}(\mathcal{A}^n) \subset \mathcal{P}(\mathcal{A}^n)$ of distributions. Consider a random non-feedback (n, R) code with minimum average error probability at most $\epsilon \in (0, 1)$ such that the codewords are distributed according to some $\overline{P} \in \mathcal{P}_{\Gamma,V}(\mathcal{A}^n)$. Then for every $n, \rho > 0$ and $\epsilon \in (0,1)$,

$$\log\lceil \exp(nR)\rceil \leq \log\rho - \log\left[\left(1-\epsilon-\sup_{\overline{P}\in\mathcal{P}_{\Gamma,V}(\mathcal{A}^n)}\inf_{q\in\mathcal{P}(\mathcal{B}^n)}(\overline{P}\circ W)\left(\frac{W(Y^n|X^n)}{q(Y^n)}>\rho\right)\right)^+\right]. \quad (15)$$

Remark 2: Lemma 2 serves as a starting point to prove converse results. Different variants of the converse in Lemma 2 can be found in [13, Theorem 27], [14, (42)] and [7, Lemma 15]. A feedback version of Lemma 2 can be proved and stated by replacing $\overline{P}$ in (15) by controllers F such that the marginal distribution of X^n induced by $F \circ W$ lies in $\mathcal{P}_{\Gamma,V}(\mathcal{A}^n)$.

III. Main Results

Definition 6: The function $\mathcal{K} : \mathbb{R} \times (0,\infty) \to (0,1)$ is defined as

$$\mathcal{K}(r,V) := \inf_{\Pi} \mathbb{E}[\Phi(\Pi)], \quad (16)$$

where the infimum is over all random variables Π satisfying $\mathbb{E}[\Pi] \geq r$ and $\mathrm{Var}(\Pi) \leq V$.

Lemma 3: The function $\mathcal{K}(r,V)$ satisfies the following three properties:

1) The infimum in (16) is a minimum, and there exists a minimizer which is a discrete probability distribution with at most 3 point masses;
2) $\mathcal{K}(r,V)$ is a strictly increasing function w.r.t. r (for a fixed V);
3) $\mathcal{K}(r,V)$ is (jointly) continuous in (r,V).

Corollary 1: An equivalent definition of the function $\mathcal{K} : \mathbb{R} \times (0,\infty) \to (0,1)$ is

$$\mathcal{K}(r,V) = \min_{\substack{\Pi:\\ \mathbb{E}[\Pi]=r\\ \mathrm{Var}(\Pi)\leq V\\ |\mathrm{supp}(\Pi)|\leq 3}} \mathbb{E}[\Phi(\Pi)].$$

Define

$$r^* = \max\left\{r\in\mathbb{R} : \mathcal{K}\left(\frac{r}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)}\right) \leq \epsilon\right\}. \quad (17)$$

The matching converse (Theorem 1) and achievability (Theorem 2) results establish r^* as the optimal second-order coding rate of DMCs with the (Γ, V) cost constraint. The maximum on the right-hand side of (17) is well-defined. In fact, we have the following general result.

Lemma 4: For any $V > 0$ and $0 < \epsilon < 1$, the supremum,

$$\sup\{r' \in \mathbb{R} : \mathcal{K}(r', V) \leq \epsilon\}, \quad (18)$$

is achieved. Furthermore, the maximum, call it r^*, satisfies $\mathcal{K}(r^*, V) = \epsilon$.

Having denoted the optimal SOCR of (n, R, Γ, V) codes by r^* in (17), it is insightful to compare (17) with the optimal SOCR of previous cost constraints, such as almost-sure constraint and expected cost constraint.

- Almost-sure constraint ($c(X^n) \leq \Gamma$ almost surely): Recall that $r_{\text{a.s.}} := \sqrt{V(\Gamma)}\Phi^{-1}(\epsilon)$ is the optimal SOCR associated with an a.s. cost constraint[1]. We note that $r^* \geq \sqrt{V(\Gamma)}\Phi^{-1}(\epsilon)$ with equality if a minimizing probability distribution in $\mathcal{K}\left(\frac{r^*}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)}\right)$ has only one point mass. One indeed has $r^* = \sqrt{V(\Gamma)}\Phi^{-1}(\epsilon)$ for $V = 0$. While the a.s. cost constraint is not, strictly speaking, equivalent to the (Γ, V) cost constraint with $V = 0$, one nevertheless obtains $r^* = r_{\text{a.s.}}$ for $V = 0$ because in practice, optimal schemes for the case $V = 0$ satisfy the a.s. cost constraint.
- Expected cost constraint ($\mathbb{E}[c(X^n)] \leq \Gamma$): This is a special case of (Γ, V) cost constraint with $V = \infty$. As mentioned in the Introduction, the strong converse does not hold with an expected cost constraint so the optimal SOCR in this case can be considered to be infinite. One therefore expects $r^* \to \infty$ as $V \to \infty$. This is indeed the case and it suffices to show that $\mathcal{K}(r,\infty) = 0$ for all values of r. Since $\mathcal{K}(r,\infty)$ is non-decreasing in r, fix an arbitrarily large value of $r > 0$. Then for any $\epsilon > 0$, we can choose m sufficiently large so that for $p = 1 - 1/m$, $\pi_1 = -\sqrt{\log m}$ and $\pi_2 = \frac{r - p\pi_1}{1-p}$, we have $p\pi_1 + (1-p)\pi_2 = r$ and

$$\begin{aligned}\mathcal{K}(r,\infty) &\leq p\Phi(\pi_1) + (1-p)\Phi(\pi_2)\\ &\leq -\frac{p}{\pi_1}\phi(\pi_1) + 1 - p \leq \epsilon.\end{aligned}$$

A. Non-Feedback Converse

Theorem 1:

Fix an arbitrary $\epsilon \in (0,1)$. Consider a channel W with cost constraint $(\Gamma, V) \in (\Gamma_0, \Gamma^*) \times (0,\infty)$ such that P^* is unique and $V(\Gamma) > 0$. Then

$$\limsup_{n\to\infty} \frac{\log M^*(n,\epsilon,\Gamma,V) - nC(\Gamma)}{\sqrt{n}} \leq r^*.$$

Alternatively, for $R = C(\Gamma) + \frac{r}{\sqrt{n}}$,

$$\liminf_{n\to\infty} \bar{P}_{\mathrm{e}}(n, R, \Gamma, V) \geq \mathcal{K}\left(\frac{r}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)}\right).$$

Proof Outline: The starting point will be the result in Lemma 2, where $\mathcal{P}_{\Gamma,V}(\mathcal{A}^n)$ is set equal to the set of distributions $\overline{P}$ such that the channel input $X^n \sim \overline{P}$ satisfies $\mathbb{E}\left[\sum_{i=1}^n c(X_i)\right] \leq n\Gamma$ and $\mathrm{Var}\left(\sum_{i=1}^n c(X_i)\right) \leq nV$. Choosing $\rho = \exp\left(nC(\Gamma) + \sqrt{n}r\right)$ in Lemma 2 and a suitable choice of q in (15), we upper bound

$$\begin{aligned}&\sup_{\overline{P}}(\overline{P}\circ W)\left(\frac{W(Y^n|X^n)}{q(Y^n)}>\rho\right)\\ &=\sup_{\overline{P}}\sum_{x^n\in\mathcal{A}^n}\overline{P}(x^n)W\left(\log\frac{W(Y^n|x^n)}{q(Y^n)}>nC(\Gamma)+\sqrt{n}r\right)\\ &\lessapprox \sup_{\overline{P}}\mathbb{E}_{\overline{P}}\left[1-\Phi\left(\frac{\sqrt{n}C'(\Gamma)}{\sqrt{V(\Gamma)}}\left(\Gamma-\frac{1}{n}\sum_{i=1}^n c(X_i)\right)\right.\right.\end{aligned}$$

[1] assuming a unique capacity-cost-achieving distribution.

$$+ \frac{r}{\sqrt{V(\Gamma)}} \Bigg) \Bigg]$$

$$= 1 - \inf_{\Pi} \mathbb{E}\left[\Phi(\Pi)\right], \quad (19)$$

where the infimum is over random variables Π satisfying $\mathbb{E}[\Pi] \geq \frac{r}{\sqrt{V(\Gamma)}}$ and $\mathrm{Var}(\Pi) \leq \frac{C'(\Gamma)^2 V}{V(\Gamma)}$. Lemma 3 is then used in (19), giving rise to the $\mathcal{K}$-function. In view of (15), the converse is given by the maximum value of r in (19) for which the value of $\mathcal{K}$-function is at most ϵ.

B. Non-Feedback Achievability

Theorem 2: Fix an arbitrary $\epsilon \in (0,1)$. Consider a channel W with cost constraint $(\Gamma, V) \in (\Gamma_0, \Gamma^*) \times (0, \infty)$ such that P^* is unique and $V(\Gamma) > 0$. Then

$$\liminf_{n\to\infty} \frac{\log M^*(n,\epsilon,\Gamma,V) - nC(\Gamma)}{\sqrt{n}} \geq r^*.$$

Alternatively, for $R = C(\Gamma) + \frac{r}{\sqrt{n}}$,

$$\limsup_{n\to\infty} \bar{P}_{\mathrm{e}}\left(n, R, \Gamma, V\right) \leq \mathcal{K}\left(\frac{r}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)}\right).$$

Remark 3: We prove Theorem 2 under a slightly stricter cost formulation given by

$$\max_{1\leq i\leq n} \mathbb{E}\left[c(X_i)\right] \leq \Gamma \text{ and } \mathrm{Var}\left(\sum_{i=1}^{n} c(X_i)\right) \leq nV, \quad (20)$$

which trivially implies the original cost formulation,

$$\mathbb{E}\left[\sum_{i=1}^{n} c(X_i)\right] \leq n\Gamma \text{ and } \mathrm{Var}\left(\sum_{i=1}^{n} c(X_i)\right) \leq nV. \quad (21)$$

Despite the restriction, we obtain a matching lower bound in Theorem 2 to the upper bound in Theorem 1. This means that the distinction between (20) and (21) is immaterial as far as the optimal SOCR is concerned.

Proof Outline: The achievability scheme makes use of the solution (a three-point probability distribution) to the optimization problem in $\mathcal{K}\left(\frac{r^*}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)}\right)$. Specifically, the minimizing probability distribution P_Π with three point masses is mapped to three different cost values Γ_1, Γ_2 and Γ_3. The cost values are in turn mapped to three types T_1, T_2 and T_3, each type T_j being close to a capacity-cost-achieving distribution for cost Γ_j. Subsequently, we use a random coding scheme where the codewords are drawn randomly from one of the three type classes with the probability weights of P_Π.

The analysis of constant composition codes, i.e., codes that generate random codewords uniformly from a fixed type class, is complicated by the fact that the induced output distribution is Q^{cc} instead of Q^*, the former of which does not factorize into a product over the output sequence. In other words, the channel output is not i.i.d. The third part of [12, Lemma 5] bounding the ratio of Q^{cc} and Q^* is then helpful for such analysis. Specifically, it is helpful in effecting a change of measure from Q^{cc} to Q^*, although it cannot be directly applied because the induced output distribution itself is a mixture of three "Q^{cc}'s".

C. Feedback Improves the SOCR

Theorem 3: Fix an arbitrary $\epsilon \in (0,1)$. Consider a channel W with cost constraint $(\Gamma, V) \in (\Gamma_0, \Gamma^*) \times (0, \infty)$ such that P^* is unique and $V(\Gamma) > 0$. Then

$$\liminf_{n\to\infty} \frac{\log M^*_{\mathrm{fb}}(n,\epsilon,\Gamma,V) - nC(\Gamma)}{\sqrt{n}} > r^*.$$

Alternatively, for $R = C(\Gamma) + \frac{r}{\sqrt{n}}$,

$$\limsup_{n\to\infty} \bar{P}_{\mathrm{e,fb}}\left(n, R, \Gamma, V\right) < \mathcal{K}\left(\frac{r}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)}\right). \quad (22)$$

Remark 4: An explicit expression for the improvement in error probability in (22) is provided in [12, (118)] in terms of an optimizable design parameter β.

Proof Outline: We consider a random feedback code in which feedback is only used once halfway through the transmission. Recall from the discussion in the Introduction Section, specifically from (3) - (5), that timid and bold channel input distributions are associated with low and high variance, respectively, of the information density. During the first half, the distribution of the channel input is similar to the distribution used in the non-feedback achievability scheme (as described in the proof outline for Theorem 2). To reiterate, it is a mixture distribution of the three type classes emanating from a minimizing three-point probability distribution in

$$\mathcal{K}\left(\frac{r}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)}\right). \quad (23)$$

Using feedback at $t = n/2$, the encoder determines whether the information density evaluated at $(x^{n/2}, y^{n/2})$ is above some suitable threshold. If it is, then the channel input for the second half is chosen to be constant-composition over only one type class with cost Γ; otherwise, the channel input is chosen to retain the mixture distribution of the three type classes over the whole blocklength. Since the mixture distribution of three type classes has a greater spread around the cost point Γ than the constant-composition distribution over a single type class, the former can be considered a bold distribution and the latter a timid distribution. Hence, if the transmission has proceeded well, the encoder switches to a timid distribution.

One caveat in the timid/bold scheme described above is that the mixture distribution could be a constant-composition code over only one type class. This is the case when the minimizing probability distribution in (23) is a single-point mass distribution. In this case, we use a constant-composition code over one type class during the first half, which is a timid distribution. Since $V > 0$, we can construct an arbitrary mixture distribution of two type classes with costs $\Gamma_1 < \Gamma$ and $\Gamma_2 > \Gamma$ to form a bold distribution, which can be used in the second half if the information density at $t = n/2$ is below some suitable threshold. Indeed, this also leads to an SOCR improvement.

References

[1] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[2] M. Hayashi, “Information spectrum approach to second-order coding rate in channel coding,” *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 4947–4966, 2009.

[3] C. E. Shannon, “Probability of error for optimal codes in a Gaussian channel,” *The Bell System Technical Journal*, vol. 38, no. 3, pp. 611–656, 1959.

[4] W. Yang, G. Caire, G. Durisi, and Y. Polyanskiy, “Optimum power control at finite blocklength,” *IEEE Transactions on Information Theory*, vol. 61, no. 9, pp. 4598–4615, 2015.

[5] Y. Polyanskiy, “Channel coding: Non-asymptotic fundamental limits,” Ph.D. dissertation, Dept. Elect. Eng., Princeton Univ., Princeton, NJ, USA, 2010.

[6] V. Kostina and S. Verdú, “Channels with cost constraints: Strong converse and dispersion,” *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2415–2429, 2015.

[7] A. B. Wagner, N. V. Shende, and Y. Altuğ, “A new method for employing feedback to improve coding performance,” *IEEE Transactions on Information Theory*, vol. 66, no. 11, pp. 6660–6681, 2020.

[8] V. Strassen, “Asymptotische abschätzungen in Shannon’s informationstheorie,” in *Proc. Trans. 3rd Prague Conf Inf. Theory*, Prague, Czech, 1962, pp. 689–723.

[9] Y. Y. Shkel, V. Y. F. Tan, and S. C. Draper, “Second-order coding rate for m-class source-channel codes,” in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2015, pp. 620–626.

[10] M. Tomamichel and V. Y. F. Tan, “Second-order coding rates for channels with state,” *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4427–4448, 2014.

[11] A. Mahmood and A. B. Wagner, “Timid/bold coding for channels with cost constraints,” in *2023 IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 1442–1447.

[12] ——, “Channel coding with mean and variance cost constraints,” 2024. [Online]. Available: arXiv:2401.16417

[13] Y. Polyanskiy, H. V. Poor, and S. Verdu, “Channel coding rate in the finite blocklength regime,” *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[14] S. L. Fong and V. Y. F. Tan, “A tight upper bound on the second-order coding rate of the parallel Gaussian channel with feedback,” *IEEE Transactions on Information Theory*, vol. 63, no. 10, pp. 6474–6486, 2017.