2019 IEEE Information Theory Workshop (ITW)

The Additive Noise Channel with a Helper

Shraga I. Bross* and Amos Lapidoth†
*Bar Ilan University, bross@biu.ac.il
†ETH Zurich, lapidoth@isi.ee.ethz.ch

***Abstract*—The additive noise channel is studied in the presence of a helper who observes the noise and can describe it to the receiver over a rate-limited noise-free bit-pipe. It is shown that the capacity of this network is typically the sum of the capacity of the channel in the absence of the helper and the capacity of the bit-pipe from the helper to the receiver. This holds for finite-variance stationary and ergodic noises under fairly general power-like constraints on the transmitted signal. A helper that is only cognizant of the noise is thus as helpful as an omniscient helper that is cognizant of both the noise and the transmitted message. The achievability proof is based on "flash helping" and requires no binning. Extensions to additive-noise multi-access channels are also discussed.**

I. Introduction

Consider the additive noise channel with a helper that is depicted in Figure 1. The noise $\{Z_i\}$ (but not the transmitted message M) is observed by a helper who wishes to assist the decoder in recovering M. To this end, the helper uses a noise-free bit-pipe of capacity R_h to describe the noise sequence to the decoder. Our interest is in the capacity $C(R_\mathrm{h})$ of this network. In the absence of any constraints on the transmitted power, this capacity is typically infinite even without a helper. Here we will show that under fairly general cost constraints and relatively mild assumptions on the noise sequence,

$$C(R_\mathrm{h}) = C(0) + R_\mathrm{h}. \tag{1}$$

That no rate exceeding $C(0) + R_\mathrm{h}$ can be achieved, readily follows from the Cut-Set bound [1, Theorem 15.10.1] by considering the cut depicted by the dashed line in Figure 1. In fact, by this argument, no such rate would be achievable even if the helper were omniscient and cognizant not only of the noise but also of the transmitted message. Our result thus shows that on the additive noise channel, a helper that is only cognizant of the noise is as helpful as one that is also cognizant of the message.

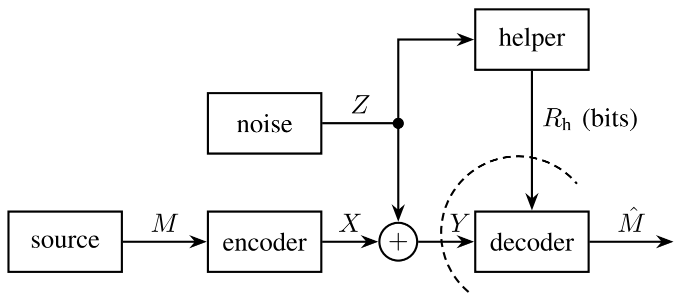

Fig. 1. The additive-noise channel with a helper.

This result is fairly general: the only assumption that we make about the noise is that it is stationary and ergodic of some finite second moment N. This guarantees an operational meaning to the mutual information. It also guarantees by the Pointwise Ergodic Theorem [2, Ch. 2, Sec. 2.2, Thm. 2.3] (and the fact that almost-sure convergence implies convergence in probability) that

$$\frac{1}{n}\sum_{i=1}^{n} Z_i^2 \xrightarrow{\mathrm{p}} N. \tag{2}$$

The constraints on the encoder are also mild. To simplify notation, we shall not state them in the fullest generality and consider only "single letter constraints": We shall assume that every channel input X_i must be in some Borel measurable "support set" $\mathcal{A} \subseteq \mathbb{R}$, so

$$x_k(m) \in \mathcal{A}, \tag{3}$$

where $x_k(m) \in \mathbb{R}$ is the symbol transmitted at Time-k to convey Message m; and each codeword $\mathbf{x}(m)$ must satisfy the constraints

$$\frac{1}{n}\sum_{k=1}^{n} g_\alpha\bigl(x_k(m)\bigr) \leq \Gamma_\alpha, \tag{4}$$

where $\{g_\alpha\}$ is a finite collection of Borel measurable mappings ("cost functions") from $\mathcal{A}$ to $\mathbb{R}^+$ (the nonnegative reals) and $\{\Gamma_\alpha\}$ is a finite collection of nonnegative numbers ("maximally-allowed average costs").

An example to keep in mind is when $\mathcal{A} \subset \mathbb{R}$ is some interval that is symmetric around the origin, and the single cost function is the quadratic $x \mapsto x^2$. This corresponds to imposing peak-power and average-power constraints. Another example arises in the study of the Exponential Noise channel [3], where $\mathcal{A}$ is the set of nonnegative reals and the cost function is the identity function $x \mapsto x$. Yet another example is the Free Space Optical Intensity channel [4] where $\mathcal{A}$ is an interval of the form $[0, A]$ and the cost function is the identity.

The final assumption we shall make rules out *inter alia* situations where $\mathcal{A}$ is finite or where the cost constraints are too restrictive. This assumption is stated formally as follows:

Assumption 1. *There exists a probability distribution P_X such that when $X \sim P_X$*

$$\Pr[X \in \mathcal{A}] = 1; \tag{5}$$

$$\mathsf{E}\bigl[g_\alpha(X)\bigr] < \Gamma_\alpha \tag{6}$$

978-1-5386-6900-6/19/$31.00 ©2019 IEEE

for each of the cost functions; and the differential entropy $h(X)$ is defined and not $-\infty$

$$h(X) > -\infty. \tag{7}$$

Here (5) and (6) guarantee that a codeword drawn at random from a random codebook whose codewords are drawn independently according to the n-fold product distribution P_X^n will satisfy the constraints with probability tending to one [1, Sec. 9.1]. Condition (7) guarantees that P_X, when used as an input distribution to an additive noise channel, gives rise to a mutual information between the channel terminals that tend to infinity as the differential entropy of the noise tends to $-\infty$.

In the example of peak and average-power constraints this assumption is satisfied whenever the peak-power and the maximally-allowed average power are positive: P_X can then be chosen as uniform over a sufficiently small (but of positive length) symmetric interval around the origin.

We can now state our main result:

Theorem 1. *Consider the additive noise channel*

$$Y_i = x_i + Z_i, \tag{8}$$

where $\{Z_i\}$ is a stationary and ergodic stochastic process of finite second moment, and where the cost constraints satisfy Assumption 1. Its capacity $C(R_\mathrm{h})$ with a rate-R_h helper is

$$C(R_\mathrm{h}) = C(0) + R_\mathrm{h}. \tag{9}$$

Our network can be viewed as a special case (albeit with cost constraints and infinite alphabets) of the problem that was studied by Ahlswede and Han [5, Sec. V] and for which they conjectured the capacity. It was solved by Kim [6] in the special case in which—as in our network—the states observed by the helper are deterministic functions of the channel inputs and outputs.[1] In fact, if one ignores the cost constraint and the infiniteness of our input and output alphabets, then Theorem 1 can be viewed as a special case of Kim's result [6]. Nevertheless, our achievability result might be of interest nonetheless because, unlike Kim's, it does not require binning. It is based on "flash helping," where help is rare but exceedingly useful.

The more general helper problem was studied by Heegard and El Gamal [7] and by Steinberg [8]. In its fullest generality it is still open [9].

Our "flash helping" approach extends to the additive noise multi-access channel (MAC):

Theorem 2. *Consider the additive noise MAC of time-i output*

$$Y_i = x_1 + x_2 + Z_i \tag{10}$$

where $\{Z_i\}$ is a stationary and ergodic stochastic process of finite second moment, and where the inputs of each of the users are restricted by a set of constraints analogous to those in the single user case. Assume that Assumption 1 holds for both sets of constraints, i.e., that for each $\nu \in \{1, 2\}$ there exists a distribution $P_X^{(\nu)}$ that assigns probability one to the subset $\mathcal{A}^{(\nu)} \subseteq \mathbb{R}$ in which X_ν must take value; that satisfies the average cost constraints on X_ν strictly; and that has differential entropy exceeding $-\infty$.

Let $\mathcal{C}(R_\mathrm{h})$ denote the capacity region of the MAC with a helper that can describe the noise to the receiver using an ideal bit-pipe of capacity R_h. Then,

$$\mathcal{C}(R_\mathrm{h}) = \mathcal{C}(0) + \bigl\{(R_1, R_2) \in \mathbb{R}^+ \times \mathbb{R}^+ \colon R_1 + R_2 \leq R_\mathrm{h}\bigr\} \tag{11}$$

where the symbol "+" denotes here Minkowski addition.

Proof: Omitted. ■

Extensions to the Broadcast Channel as well as to cases where the help is provided not to the decoder but to the encoder are discussed in [10]. For example, it is shown in [10] that when the help is provided to the encoder and the noise is Gaussian,

$$C(R_\mathrm{h}) = \frac{1}{2}\log\Bigl(1 + \frac{P}{N}\Bigr) + R_\mathrm{h}, \tag{12}$$

where P is the allowed average transmit power, and N is the variance of the Gaussian noise.

We emphasize that we assume throughout that the helper is cognizant of the exact noise sequence. Some preliminary results on the capacity when the helper only observes an approximate version of the noise can be found in [11]. The capacity in this setting is still unknown, but its low signal-to-noise ratio asymptotics are [11, Proposition 2]

II. On the Mean Squared-Error Distortion for a General Source

To prove our main results, we shall need the following variation on Sakrison's classical result [12, Section 6], [13, Theorem 3] that—among all sources of a given second moment—the memoryless Gaussian source is the most difficult to describe in mean squared-error (MSE). Unlike Sakrison, we do not require the existence of a moment higher than two. Instead, we require that the empirical average of the squares of the source symbols converge in probability.

Theorem 3. *Let $X_1, X_2, \ldots$ be a sequence of random variables whose average second moment converges to σ^2*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\bigl[X_i^2\bigr] = \sigma^2, \tag{13}$$

and assume that the empirical average of their squares converges in probability to σ^2

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{\mathrm{p}} \sigma^2. \tag{14}$$

Then, given any rate R and any $\tilde{\epsilon} > 0$, there exists for all sufficiently large blocklengths n a rate-R blocklength-n rate-distortion codebook with normalized average MSE distortion

$$\frac{1}{n}\mathsf{E}\bigl[\|\mathbf{X} - \mathbf{X}^*\|^2\bigr] \leq \sigma^2\, 2^{-2R} + \tilde{\epsilon}, \tag{15}$$

[1] The result in [6] is more general in that it pertains to the Relay Channel.

where $\mathbf{X}^*$ is the reconstruction of $\mathbf{X}$ based on its nR-bit description.

Proof: We focus on positive rates, because when the rate is zero the result follows from (13) by considering the codebook with only one codeword, namely, the all-zero codeword.

The proof relies heavily on a result of A. D. Wyner [14, Corollary to Theorem 2] that if $\theta \in (0, \pi/2)$ is arbitrary and $\rho > -\log\sin\theta$, then, for every sufficiently large n, there exist $2^{\rho n}$ points on the n-dimensional sphere such that the caps of half-angle θ around these points cover the entire n-dimensional sphere. With this result, we can proceed as follows. Let θ be the angle in the interval $(0, \pi/2)$ for which

$$R = -\log\sin\theta, \tag{16}$$

and let

$$\begin{aligned}\Delta &= \sigma^2\, 2^{-2R} \\ &= \sigma^2 \sin^2\theta.\end{aligned} \tag{17}$$

Fix some $0 < \epsilon < \cos(\theta)$ and let $\theta' \in (\theta, \pi/2)$ be such that

$$\cos(\theta) - \epsilon < \cos(\theta') < \cos(\theta). \tag{18}$$

Since $\cos(\theta') < \cos(\theta)$ with $\theta, \theta' \in (0, \pi/2)$, (16) implies that

$$R > -\log\sin\theta'. \tag{19}$$

By Wyner's result, there exist, for every n sufficiently large, 2^{nR} points on the radius-r n-dimensional sphere, where

$$r = \sqrt{n(\sigma^2 - \Delta)}, \tag{20}$$

such that the caps of half-angle θ' centered around the points cover the sphere. Our codebook will comprise these point and the all-zero point $\mathbf{0}$. Adding the latter point guarantees that, irrespective of the source sequence $\mathbf{x}$, the closest codeword $\mathbf{x}^*$ to it must satisfy

$$\|\mathbf{x} - \mathbf{x}^*\| \le \|\mathbf{x}\|, \quad \mathbf{x} \in \mathbb{R}^n. \tag{21}$$

The perfect covering of the sphere guarantees that, if $\tilde{\mathbf{x}}^*$ is the closest *nonzero* codeword to $\mathbf{x}$, then the inner product between $\mathbf{x}$ and $\tilde{\mathbf{x}}^*$ can be bounded as

$$\begin{aligned}\frac{\langle \mathbf{x}, \tilde{\mathbf{x}}^*\rangle}{\|\mathbf{x}\|\|\tilde{\mathbf{x}}^*\|} &\ge \cos(\theta') \\ &> \cos(\theta) - \epsilon,\end{aligned} \tag{22}$$

where the second inequality follows from (18). Assuming that $\mathbf{x} \neq \mathbf{0}$ and with $\hat{\theta}$ denoting the angle between $\mathbf{x}$ and $\tilde{\mathbf{x}}^*$, the above inequalities can be written as

$$\begin{aligned}\cos(\hat{\theta}) &\ge \cos(\theta') \\ &> \cos(\theta) - \epsilon.\end{aligned} \tag{23}$$

Consequently, for nonzero $\mathbf{x}$,

$$\begin{aligned}&n^{-1}\|\mathbf{x} - \mathbf{x}^*\|^2 \\ &\le n^{-1}\|\mathbf{x} - \tilde{\mathbf{x}}^*\|^2 \\ &= n^{-1}\|\mathbf{x}\|^2 + n^{-1}\|\tilde{\mathbf{x}}^*\|^2 - 2n^{-1}\|\mathbf{x}\|\|\tilde{\mathbf{x}}^*\|\cos(\hat{\theta}) \\ &< n^{-1}\|\mathbf{x}\|^2 + n^{-1}\|\tilde{\mathbf{x}}^*\|^2 - 2n^{-1}\|\mathbf{x}\|\|\tilde{\mathbf{x}}^*\|\big(\cos(\theta) - \epsilon\big) \\ &= n^{-1}\|\mathbf{x}\|^2 + n^{-1}r^2 - 2n^{-1}\|\mathbf{x}\|\, r\cos(\theta) + 2n^{-1}\|\mathbf{x}\|\, r\epsilon.\end{aligned} \tag{24}$$

As a function of $\|\mathbf{x}\|$, the RHS is monotonically increasing when $\|\mathbf{x}\|$ exceeds $r\cos(\theta)$.

To analyze $\|\mathbf{x} - \mathbf{x}^*\|^2$ we shall use (24) or (21) depending on whether or not "$\mathbf{x}$ is in the δ-shell", i.e., $n^{-1}\|\mathbf{x}\|^2$ is in the interval $(\sigma^2 - \delta, \sigma^2 + \delta)$. Here $\delta > 0$ is arbitrary but small enough to guarantee that

$$\sqrt{n(\sigma^2 - \delta)} > r\cos(\theta). \tag{25}$$

This latter condition guarantees that the RHS of (24) be monotonically increasing in $\|\mathbf{x}\|$ in the δ-shell. (It suffices, of course, that $\sqrt{n(\sigma^2 - \delta)}$ exceed r, which, in view of (20), is equivalent to δ being smaller than Δ.) Using this monotonicity we obtain that when $\mathbf{x}$ is in the δ-shell we can upper-bound the RHS of (24) by replacing $n^{-1}\|\mathbf{x}\|^2$ with $\sigma^2 + \delta$ to obtain

$$\begin{aligned}&n^{-1}\|\mathbf{x} - \mathbf{x}^*\|^2 \\ &\quad\le (\sigma^2 + \delta) + n^{-1}r^2 - 2n^{-1}\sqrt{n(\sigma^2 + \delta)}\, r\cos(\theta) \\ &\qquad + 2n^{-1}\sqrt{n(\sigma^2 + \delta)}\, r\epsilon \\ &\quad= (\sigma^2 + \delta) + (\sigma^2 - \Delta) - 2\sqrt{\sigma^2 + \delta}\,\sqrt{\sigma^2 - \Delta}\cos(\theta) \\ &\qquad + 2\sqrt{\sigma^2 + \delta}\,\sqrt{\sigma^2 - \Delta}\,\epsilon \\ &\quad= 2\sigma^2 + \delta - \Delta - 2\sqrt{\sigma^2 + \delta}\,\frac{1}{\sigma}\,(\sigma^2 - \Delta) \\ &\qquad + 2\sqrt{\sigma^2 + \delta}\,\sqrt{\sigma^2 - \Delta}\,\epsilon,\end{aligned} \tag{26}$$

where in the last equality we have used (17), which implies that

$$\cos(\theta) = \frac{1}{\sigma}\sqrt{\sigma^2 - \Delta}. \tag{27}$$

Expressing the expected unnormalized distortion as

$$\begin{aligned}\mathsf{E}\big[\|\mathbf{X} - \mathbf{X}^*\|^2\big] &= \int_{|n^{-1}\|\mathbf{x}\|^2 - \sigma^2| \le \delta} \|\mathbf{x} - \mathbf{x}^*\|^2 \,\mathrm{d}P(\mathbf{x}) \\ &\quad + \int_{|n^{-1}\|\mathbf{x}\|^2 - \sigma^2| > \delta} \|\mathbf{x} - \mathbf{x}^*\|^2 \,\mathrm{d}P(\mathbf{x}),\end{aligned} \tag{28}$$

we can upper-bound the first integral by the product of the probability of $\mathbf{X}$ being in the δ-shell and the RHS of (26). As n tends to infinity, this approaches the RHS of (26).

$$\begin{aligned}&\overline{\lim_{n\to\infty}}\,\frac{1}{n}\int_{|n^{-1}\|\mathbf{x}\|^2 - \sigma^2| \le \delta} \|\mathbf{x} - \mathbf{x}^*\|^2 \,\mathrm{d}P(\mathbf{x}) \\ &\quad\le 2\sigma^2 + \delta - \Delta - 2\sqrt{\sigma^2 + \delta}\,\frac{1}{\sigma}\,(\sigma^2 - \Delta) \\ &\qquad + 2\sqrt{\sigma^2 + \delta}\,\sqrt{\sigma^2 - \Delta}\,\epsilon.\end{aligned} \tag{29}$$

To upper-bound the second integral in (28), we upper-bound the integrand using (21) to obtain

$$\begin{aligned}\int_{\left|n^{-1}\|\mathbf{x}\|^2-\sigma^2\right|>\delta} \|\mathbf{x}-\mathbf{x}^*\|^2 \,\mathrm{d}P(\mathbf{x}) &\leq \int_{\left|n^{-1}\|\mathbf{x}\|^2-\sigma^2\right|>\delta} \|\mathbf{x}\|^2 \,\mathrm{d}P(\mathbf{x}) \\ &= \mathsf{E}\left[\|\mathbf{X}\|^2\right] - \int_{\left|n^{-1}\|\mathbf{x}\|^2-\sigma^2\right|\leq\delta} \|\mathbf{x}\|^2 \,\mathrm{d}P(\mathbf{x}) \\ &\leq \mathsf{E}\left[\|\mathbf{X}\|^2\right] - n\big(\sigma^2-\delta\big)\Pr\Big[\left|n^{-1}\|\mathbf{x}\|^2-\sigma^2\right|\leq\delta\Big]. \end{aligned} \tag{30}$$

Dividing by n and letting n tend to infinity, we obtain using (13) and (14) that

$$\varlimsup_{n\to\infty} \frac{1}{n}\int_{\left|n^{-1}\|\mathbf{x}\|^2-\sigma^2\right|>\delta} \|\mathbf{x}-\mathbf{x}^*\|^2 \,\mathrm{d}P(\mathbf{x}) \leq \delta. \tag{31}$$

Using (28), (29), and (31), we conclude that

$$\begin{aligned}\varlimsup_{n\to\infty} \frac{1}{n}\,\mathsf{E}\left[\|\mathbf{X}-\mathbf{X}^*\|^2\right] &\leq 2\sigma^2+2\delta-\Delta-2\sqrt{\sigma^2+\delta}\,\frac{1}{\sigma}\left(\sigma^2-\Delta\right) \\ &\quad + 2\sqrt{\sigma^2+\delta}\,\sqrt{\sigma^2-\Delta}\,\epsilon. \end{aligned} \tag{32}$$

Since this hold for any $\delta > 0$ for which (25) holds, we can let $\delta \downarrow 0$ to obtain

$$\varlimsup_{n\to\infty} \frac{1}{n}\,\mathsf{E}\left[\|\mathbf{X}-\mathbf{X}^*\|^2\right] \leq \Delta + 2\sigma\,\sqrt{\sigma^2-\Delta}\,\epsilon. \tag{33}$$

Since this holds for any $\epsilon > 0$, the theorem must hold. ∎

III. Proof of Theorem 1 via Flash Helping

Proof: Since the Cut-Set bound provides us with the converse, we only need to prove achievability. In the following Z^m denotes the m random variables $Z_1, \ldots, Z_m$, and $[1:m]$ denotes the set $\{1, \ldots, m\}$.

Before describing our proposed scheme, which is based on time-sharing, we begin with a calculation. Suppose that the helper describes the noise sequence Z^m using mR_{h} bits, and the decoder, based on this description, produces the estimate $\hat{Z}^m$ of Z^m with corresponding error $\tilde{Z}^m$, where

$$\tilde{Z}_i = Z_i - \hat{Z}_i, \qquad i \in [1:m]. \tag{34}$$

It then subtracts this estimate from the received sequence Y^m and obtains the sequence

$$\tilde{Y}_i = X_i + \tilde{Z}_i, \quad i \in [1:m]. \tag{35}$$

We now study $I\big(X^m;\tilde{Y}^m\big)$ when $X_1, \ldots, X_m$ are IID according to the distribution P_X whose existence is guaranteed by Assumption 1. Since the noise Z^m and its description are independent of X^m,

$$\begin{aligned} I\big(X^m;\tilde{Y}^m\big) &= h\big(\tilde{Y}^m\big) - h\big(\tilde{Z}^m\big) \\ &\geq h\big(X^m\big) - h\big(\tilde{Z}^m\big) \\ &= m\,h(P_X) - h\big(\tilde{Z}^m\big), \end{aligned} \tag{36}$$

where $h(P_X)$ is the differential entropy of a random variable that is distributed according to P_X.

To further lower-bound $I\big(X^m;\tilde{Y}^m\big)$, we shall upper-bound $h\big(\tilde{Z}^m\big)$ in terms of the estimation error. Of all multivariate distributions of a given second moment matrix, the centered multivariate Gaussian maximizes differential entropy [1, Thm. 8.6.5]. And under a constraint on the trace of the second moment matrix, the IID Gaussian distribution maximizes differential entropy. Thus, with $\tilde{\mathbf{Z}}$ denoting $\tilde{Z}^m$,

$$\begin{aligned} h\big(\tilde{\mathbf{Z}}\big) &\leq \frac{m}{2}\log\left(2\pi e\frac{\mathrm{tr}\Big(\mathsf{E}[\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^{\mathsf{T}}]\Big)}{m}\right) \\ &= \frac{m}{2}\log\left(2\pi e\frac{1}{m}\,\mathsf{E}\left[\|\tilde{\mathbf{Z}}\|^2\right]\right). \end{aligned} \tag{37}$$

From (37) and (36) we obtain upon dividing by m,

$$\begin{aligned} \frac{1}{m}I\big(X^m;\tilde{Y}^m\big) &\geq h(P_X) - \frac{1}{2}\log(2\pi e) - \frac{1}{2}\log\left(\frac{1}{m}\,\mathsf{E}\left[\|\tilde{\mathbf{Z}}\|^2\right]\right). \end{aligned} \tag{38}$$

Theorem 3 guarantees that, given any $\tilde{\delta} > 0$, we can find some sufficiently large m for which there exists a rate-R_{h} blocklength-m rate-distortion codebook that allows the helper to describe the noise sequence with average distortion that is bounded by

$$\frac{1}{m}\,\mathsf{E}\left[\|\tilde{\mathbf{Z}}\|^2\right] \leq N\,2^{-2R_{\mathrm{h}}}(1+\tilde{\delta}). \tag{39}$$

For such m, it follows from (39) and (38) that

$$\frac{1}{m}I\big(X^m;\tilde{Y}^m\big) \geq h(P_X) - \frac{1}{2}\log\big(2\pi e\,N(1+\tilde{\delta})\big) + R_{\mathrm{h}}. \tag{40}$$

Having completed our calculation, we now consider time sharing, with the channel being used without help in $(1-\alpha)$ of the time, and with rate-(R_{h}/α) help in α of the time. In the absence of help we can approach $C(0)-\epsilon'$ for any positive ϵ'. The overall achievable rate is thus lower-bounded by

$$\begin{aligned} &(1-\alpha)\big(C(0)-\epsilon'\big) \\ &\quad + \alpha\left(h(P_X) - \frac{1}{2}\log\big(2\pi e\,N(1+\tilde{\delta})\big) + \frac{R_{\mathrm{h}}}{\alpha}\right). \end{aligned} \tag{41}$$

This rate is achievable because the ergodicity of the noise process guarantees the operational meaning of the mutual information.

By considering the limit of the RHS as $\alpha \downarrow 0$ and then letting $\epsilon' \downarrow 0$ we obtain the achievability of

$$C(0) + R_{\mathrm{h}} \tag{42}$$

and hence conclude the proof. Since the helper is used only α of the time, and since α approaches zero, we refer to this form of help as “flash helping.” ∎

Acknowledgment

The authors are grateful to the anonymous reviewers for drawing their attention to [5] and [6].

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: J. Wiley & Sons, 2006.

[2] K. Petersen, *Ergodic Theory*. Cambridge University Press, 1983.

[3] S. Verdú, "The exponential distribution in information theory," *Probl. Peredachi Inf.*, vol. 32, no. 1, pp. 100–111, 1996.

[4] A. Lapidoth, S. M. Moser, and M. A. Wigger, "On the capacity of free-space optical intensity channels," *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4449–4461, Oct 2009.

[5] R. Ahlswede and T. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 396–412, May 1983.

[6] Y.-H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Trans. Inf. Theor.*, vol. 54, no. 3, pp. 1328–1329, Mar. 2008. [Online]. Available: https://doi.org/10.1109/TIT.2007.915921

[7] C. Heegard and A. E. Gamal, "On the capacity of computer memory with defects," *IEEE Trans. on Inform. Theory*, vol. 29, no. 5, pp. 731–739, Sept. 1983.

[8] Y. Steinberg, "Coding for channels with rate-limited side information at the decoder, with applications," *IEEE Trans. on Inform. Theory*, vol. 54, no. 9, pp. 4283–4295, Sep. 2008.

[9] G. Keshet, Y. Steinberg, and N. Merhav, "Channel coding in the presence of side information," *Foundations and Trends® in Communications and Information Theory*, vol. 4, no. 6, pp. 445–586, 2008.

[10] G. Marti, "Channels with a helper," Master's thesis, ETH Zurich, 2019.

[11] S. I. Bross and A. Lapidoth, "The Gaussian state-dependent channel with rate-limited decoder state-information," in *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, Nov 2016, pp. 1–5.

[12] D. J. Sakrison, "The rate distortion function for a class of sources," *Information and Control*, vol. 15, pp. 165–195, 1969.

[13] A. Lapidoth, "On the role of mismatch in rate distortion theory," *IEEE Transactions on Information Theory*, vol. 43, no. 1, pp. 38–47, Jan 1997.

[14] A. Wyner, "Random packings and coverings of the unit n-sphere," *The Bell System Technical Journal*, vol. 46, no. 9, pp. 2111–2118, 1967.