Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 10(6): 13151–13165.
DOI: 10.3934/math.2025590
Received: 19 March 2025
Revised: 26 May 2025
Accepted: 04 June 2025
Published: 09 June 2025

Research article

Output statistics, equivocation, and state masking

Ligong Wang*

Department of Information Technology and Electrical Engineering, ETH Zurich, 8092 Zurich, Switzerland

* **Correspondence:** Email: ligwang@isi.ee.ethz.ch.

Abstract: Given a discrete memoryless channel and a target distribution on its output alphabet, one wishes to construct a length-n rate-R codebook such that the output distribution—computed over a codeword that is chosen uniformly at random—should be close to the n-fold tensor product of the target distribution. Here "close" means that the relative entropy between the output distribution and said n-fold product should be small. We characterize the smallest achievable relative entropy divided by n as n tends to infinity. We then demonstrate two applications of this result. The first application is an alternative proof of the achievability of the rate-equivocation region of the wiretap channel. The second application is a new capacity result for communication subject to state masking in the scenario where the decoder has access to channel-state information.

Keywords: relative entropy; soft covering; approximation of output statistics; equivocation; wiretap channel; state masking
Mathematics Subject Classification: 94A15, 94A17, 94A24, 94A40

1. Introduction

The problem of approximating a certain output distribution over a noisy channel appears in many works in Information Theory [1–9]. Here we consider a variant of this problem where the approximation error—measured in normalized relative entropy—is not required to approach zero as the number of channel uses increases. Instead, we study the trade-off between the approximation error and the rate of the codebook.

Consider a discrete memoryless channel (DMC) with finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, and channel law

$$W(y|x), \quad x \in \mathcal{X}, y \in \mathcal{Y}. \tag{1}$$

Let $\hat{Q}_Y$ be a probability mass function (PMF) on $\mathcal{Y}$ such that

$$\text{supp}(\hat{Q}_Y) = \mathcal{Y}. \tag{2}$$

We wish to approximate its n-fold product $\hat{Q}_Y^{\times n}$. A length-n rate-R codebook is given by

$$C = \left\{x^n(1), \dots, x^n(2^{nR})\right\}. \tag{3}$$

When a codeword is chosen equiprobably from C and sent over n uses of the channel, the output distribution is

$$P_{Y^n}(y^n) = 2^{-nR} \sum_{m=1}^{2^{nR}} \prod_{i=1}^{n} W(y_i \big| x_i(m)), \quad y^n \in \mathcal{Y}^n. \tag{4}$$

The approximation error that we consider is

$$\frac{1}{n} D(P_{Y^n} \big\| \hat{Q}_Y^{\times n}). \tag{5}$$

Assume that there exists an input PMF Q_X such that $\hat{Q}_Y = Q_X W$, by which notation we mean

$$(Q_X W)(y) = \sum_{x \in \mathcal{X}} Q_X(x) W(y|x), \quad y \in \mathcal{Y}. \tag{6}$$

Wyner's Soft-Covering Lemma [1] asserts that there exists a sequence of codebooks like (3) such that (5) tends to zero as $n \to \infty$ provided

$$R > I(Q_X, W). \tag{7}$$

Later, Cuff [5, 6] showed that, under (7), the unnormalized relative entropy $D(P_{Y^n} \big\| \hat{Q}_Y^{\times n})$ can also be made to approach zero, as can the total variation distance between P_{Y^n} and $\hat{Q}_Y^{\times n}$. Furthermore, the probability that a randomly generated codebook (according to Q_X) produces an output distribution that is *not* close to $\hat{Q}_Y^n$ in total variation distance is doubly-exponentially small in n.

Now consider the case where again $\hat{Q}_Y = Q_X W$, but where R approaches $I(Q_X, W)$ from below. Furthermore, assume that Q_X and $\hat{Q}_Y$ are capacity-achieving input and output distributions, respectively, so $I(Q_X, W)$ equals the capacity of the channel. Then there exists a sequence of codes such that (5) approaches zero [2, Theorem 15], [3, Theorem 2]. In fact, any sequence of "good codes"—those whose rates approach capacity and whose average error probabilities approach zero as $n \to \infty$—*must* be such that (5) tends to zero as $n \to \infty$. This result has been generalized to codes with nonvanishing error probability [8, 9].

In all results that we recalled above, the minimum of (5) approaches zero as $n \to \infty$. Here we are interested in cases where R is not large enough for (5) to approach zero. Specifically, either R is smaller than and possibly bounded away from $I(Q_X, W)$ for any Q_X such that $\hat{Q}_Y = Q_X W$, or there exists no input distribution that can induce $\hat{Q}_Y$ via W. In Section 2, we study the tradeoff between the rate R and the approximation error (5) in such cases.

The Soft-Covering Lemma and its stronger versions are useful in many problems in information-theoretic security. For example, the original version of the lemma can be used to derive the secrecy capacity of the wiretap channel, although Wyner did not observe this connection in his original work [10].

Here we demonstrate two applications of the rate-error tradeoff that we derive. The first is an alternative proof of the achievability of the rate-equivocation region of the wiretap channel; see Section 3. The second application is in *state masking* [11, 12]. Specifically, we derive capacity subject to state masking when the decoder has channel-state information (CSI); see Section 4.

2. Output statistics

2.1. An identity

The following simple identity will be used repeatedly in our proofs.

Lemma 1. *Let $P_{X^nY^n}$ be a joint distribution on $\mathcal{X}^n \times \mathcal{Y}^n$ such that $P_{Y^n|X^n} = W^{\times n}$, and let $\bar{P}_{XY}$ be its average distribution on $\mathcal{X} \times \mathcal{Y}$. Further, let $\hat{Q}_Y$ be a distribution on $\mathcal{Y}$. The following holds:*

$$D(P_{Y^n} \| \hat{Q}_Y^{\times n}) = nI(\bar{P}_X, W) + nD(\bar{P}_Y \| \hat{Q}_Y) - I(X^n; Y^n). \tag{8}$$

Proof. We will use the following well-known identity (which is easily verified by expanding its both sides): For $Q_Y = Q_X W$,

$$I(Q_X, W) + D(Q_Y \| \hat{Q}_Y) = \mathsf{E}_{Q_X}\left[D(W(\cdot|X) \| \hat{Q}_Y)\right]. \tag{9}$$

Replacing Q_X, W, and $\hat{Q}_Y$ in the above respectively by P_{X^n}, $W^{\times n}$, and $\hat{Q}_Y^{\times n}$, we obtain

$$\begin{aligned} I(X^n; Y^n) + D(P_{Y^n} \| \hat{Q}_Y^{\times n}) &= \mathsf{E}_{P_{X^n}}\left[D(W^{\times n}(\cdot|X^n) \| \hat{Q}_Y^{\times n})\right] &(10)\\ &= n\, \mathsf{E}_{\bar{P}_X}\left[D(W(\cdot|X) \| \hat{Q}_Y)\right] &(11)\\ &= nI(\bar{P}_X, W) + D(\bar{P}_Y \| \hat{Q}_Y), &(12) \end{aligned}$$

where the last step follows by applying (9) again, this time with Q_X replaced by $\bar{P}_X$. □

2.2. Direct result

Theorem 2 (Direct result)**.** *Given some DMC* (1) *and $\hat{Q}_Y$ satisfying* (2)*, fix a PMF Q_X on $\mathcal{X}$ and let $Q_Y = Q_X W$. For any $R < I(Q_X, W)$, $\epsilon > 0$, and $\zeta > 0$, let C be a random codebook whose entries are independent and identically distributed (IID) according to Q_X. Then, as $n \to \infty$, the probability that C satisfies both of the following tends to one:*

(C1) The average probability of a decoding error (by an optimal decoder) is at most ϵ;

(C2) Let P_{Y^n} be given by (4)*, then* (5) *is upper- and lower-bounded as*

$$I(Q_X, W) + D(Q_Y \| \hat{Q}_Y) - R - \zeta \le \frac{1}{n} D(P_{Y^n} \| \hat{Q}_Y^{\times n}) \le I(Q_X, W) + D(Q_Y \| \hat{Q}_Y) - R + R\epsilon + \zeta. \tag{13}$$

Proof. Let $P_{X^n}^{\mathsf{C}}$ be the uniform distribution over the codewords of C, and let $\bar{P}_X^{\mathsf{C}}$ denote its average PMF on $\mathcal{X}$, i.e.,

$$\bar{P}_X^{\mathsf{C}}(x) = \frac{1}{n} \sum_{i=1}^{n} P_{X_i}^{\mathsf{C}}(x), \quad x \in \mathcal{X}. \tag{14}$$

By standard arguments (see, e.g., [13]) we can deduce that, for sufficiently large n, with high probability, the average decoding error probability of C is less than ϵ, i.e., it satisfies (C1). Furthermore, by the Law of Large Numbers, as $n \to \infty$, $\bar{P}_X^{\mathsf{C}}(x)$—which is random because C is random—converges to $Q_X(x)$ with probability one for all $x \in \operatorname{supp}(Q_X)$. Therefore, for any $\alpha > 0$, for sufficiently large n, we have, with high probability,

$$\delta_{\mathrm{TV}}\left(\bar{P}_X^{\mathsf{C}}, Q_X\right) \le \alpha, \tag{15}$$

where $\delta_{\mathrm{TV}}(\cdot,\cdot)$ denotes the total variation distance:

$$\delta_{\mathrm{TV}}(P,Q) \triangleq \frac{1}{2}\|P-Q\|_1. \tag{16}$$

By the union bound, we deduce that, with high probability, C satisfies both (C1) and (15). We fix $\mathsf{C} = C$ for any C satisfying (C1) and (15) and show that it must also satisfy (C2) (for an appropriately chosen α), which will then conclude the proof. From now on, we drop the superscript C: $P_{X^nY^n}$ denotes the joint input-output distribution on $\mathcal{X}^n \times \mathcal{Y}^n$ resulting from sending a uniformly chosen codeword from C, and $\bar{P}_{XY}$ denotes its average PMF on $\mathcal{X}\times\mathcal{Y}$. We can upper- and lower-bound $I(X^n;Y^n)$ as

$$nR(1-\epsilon)-1 \le I(X^n;Y^n) \le nR, \tag{17}$$

where the lower bound (first inequality) follows by Fano's Inequality [13], and the upper bound (second inequality) because there are only 2^{nR} codewords, so $H(X^n)\le nR$. Using (17) together with Lemma 1, we have

$$\begin{aligned}\frac{1}{n}D(P_{Y^n}\|\hat{Q}_Y^{\times n}) &\le I(\bar{P}_X,W)+D(\bar{P}_Y\|\hat{Q}_Y)-R+R\epsilon+\frac{1}{n} &(18)\\ &=\sum_{x\in\mathcal{X}}\bar{P}_X(x)D(W(\cdot|x)\|\hat{Q}_Y)-R+R\epsilon+\frac{1}{n}, &(19)\end{aligned}$$

and also

$$\frac{1}{n}D(P_{Y^n}\|\hat{Q}_Y^{\times n}) \ge \sum_{x\in\mathcal{X}}\bar{P}_X(x)D(W(\cdot|x)\|\hat{Q}_Y)-R. \tag{20}$$

Denote

$$d \triangleq \max_{x\in\mathcal{X}} D(W(\cdot|x)\|\hat{Q}_Y), \tag{21}$$

which is finite due to (2). Then

$$\left|\sum_{x\in\mathcal{X}}\bar{P}_X(x)D(W(\cdot|x)\|\hat{Q}_Y)-\sum_{x\in\mathcal{X}}Q_X(x)D(W(\cdot|x)\|\hat{Q}_Y)\right| \le \delta_{\mathrm{TV}}(\bar{P}_X,Q_X)\cdot d \le \alpha d. \tag{22}$$

We combine (19) and (22) to obtain that, for sufficiently large n,

$$\begin{aligned}\frac{1}{n}D(P_{Y^n}\|\hat{Q}_Y^{\times n}) &\le \sum_{x\in\mathcal{X}}Q_X(x)D(W(\cdot|x)\|\hat{Q}_Y)-R+R\epsilon+\alpha d+\frac{1}{n} &(23)\\ &= I(Q_X,W)+D(Q_Y\|\hat{Q}_Y)-R+R\epsilon+\alpha d+\frac{1}{n}. &(24)\end{aligned}$$

Similarly, by (20) and (22),

$$\frac{1}{n}D(P_{Y^n}\|\hat{Q}_Y^{\times n}) \ge I(Q_X,W)+D(Q_Y\|\hat{Q}_Y)-R-\alpha d. \tag{25}$$

Combining (24) and (25) and choosing $\alpha < \zeta/d$ (strictly) proves (13). □

2.3. Converse results

We first present a lower bound on (5) for any codebook.

Theorem 3 (Converse: General lower bound)**.** *For all n, any code C satisfies the following: Let P_{Y^n} be given by* (4)*. Then*

$$\frac{1}{n}D(P_{Y^n}\|\hat{Q}_Y^{\times n}) \geq \min_{Q_X}\left\{(I(Q_X, W) - R)^+ + D(Q_Y\|\hat{Q}_Y)\right\}, \tag{26}$$

where $Q_Y = Q_X W$.

Proof. We have the following standard upper bound on $I(X^n; Y^n)$ that follows by the chain rule, the channel being memoryless, and concavity of mutual information in the input distribution:

$$I(X^n; Y^n) \leq nI(\bar{P}_X, W), \tag{27}$$

where $\bar{P}_X$ is the average PMF on $\mathcal{X}$ induced by the codebook. Also recall the upper bound (second inequality) in (17). Combining these two bounds with Lemma 1 immediately yields

$$\frac{1}{n}D(P_{Y^n}\|\hat{Q}_Y^{\times n}) \geq (I(\bar{P}_X, W) - R)^+ + D(\bar{P}_Y\|\hat{Q}_Y), \tag{28}$$

which continues to hold when its right-hand side (RHS) is minimized over $\bar{P}_X$. □

Maximizing (5) over all codebooks is trivial: one should choose all codewords to be identical and to consist only of the input symbol x^* that maximizes $D(W(\cdot|x^*)\|\hat{Q}_Y)$. The problem becomes meaningful when we require the codebook to have a small decoding error probability, in which case we have the following upper bound.

Theorem 4 (Converse: Upper bound for decodable codes)**.** *Any code C with decoding error probability less than or equal to ϵ must satisfy: There exists some Q_X such that, for $Q_Y = Q_X W$,*

$$I(Q_X, W) \geq R - R\epsilon - \frac{1}{n} \tag{29}$$

$$\frac{1}{n}D(P_{Y^n}\|\hat{Q}_Y^{\times n}) \leq I(Q_X, W) + D(Q_Y\|\hat{Q}_Y) - R + R\epsilon + \frac{1}{n}. \tag{30}$$

Proof. Choose $Q_X = \bar{P}_X$, the average input PMF induced by the codebook, then (29) follows by Fano's Inequality together with (27), and (30) follows by Fano's Inequality together with Lemma 1. □

2.4. Asymptotic expressions

The above results yield the following asymptotics on (5). First, combining Theorems 2 and 3, we have

$$\lim_{n\to\infty}\min_{C}\frac{1}{n}D(P_{Y^n}\|\hat{Q}_Y^{\times n}) = \min_{Q_X}\left\{(I(Q_X, W) - R)^+ + D(Q_Y\|\hat{Q}_Y)\right\}, \tag{31}$$

where the minimum on the left-hand side is over all length-n rate-R codebooks. If we only allow codes with vanishing decoding error probabilities—which, with some abuse of notation, we denote as C_{cor}—then, by Theorems 2 and 4,

$$\lim_{n\to\infty}\min_{C_{\text{cor}}}\frac{1}{n}D(P_{Y^n}\|\hat{Q}_Y^{\times n}) = \min_{Q_X:\, I(Q_X,W)\geq R}\left\{I(Q_X, W) - R + D(Q_Y\|\hat{Q}_Y)\right\}. \tag{32}$$

Note that (31) and (32) coincide when

$$R \leq \min_{Q_X:\, Q_X W = \hat{Q}_Y} I(Q_X, W). \tag{33}$$

Beyond this threshold, (31) remains zero, whereas (32) increases with R.

Finally, again for codes with vanishing decoding error probabilities, Theorems 2 and 4 together imply

$$\lim_{n\to\infty} \max_{\mathcal{C}_{\mathrm{cor}}} \frac{1}{n} D(P_{Y^n} \| \hat{Q}_Y^{\times n}) = \max_{Q_X:\, I(Q_X,W)\geq R} \left\{ I(Q_X, W) - R + D(Q_Y \| \hat{Q}_Y) \right\}. \tag{34}$$

Note that (32) is meaningful only when R is below the capacity of the channel; else correct decoding would not be possible, and the minimization on both sides would be over empty sets; similarly for (34).

Due to the identity (9), the minimization on the RHS of (31) (when restricted to Q_X such that $I(Q_X, W) \geq R$) and (32) is linear, and so is the maximization on the RHS of (34). Therefore they are achieved on the boundary, as we demonstrate in the following example.

Example 5. *Let W be a binary symmetric channel with crossover probability $p \in (0, 0.5)$, and let $\hat{Q}_Y$ be Bernoulli(q) with $q \in (p, 0.5)$, so $D(W(\cdot|0) \| \hat{Q}_Y) < D(W(\cdot|1) \| \hat{Q}_Y)$. It then follows that the minimum on the RHS of of* (32) *is achieved by using* 0 *as frequently as possible, i.e., it is achieved by Q_X^* being Bernoulli(a^*) such that $I(Q_X^*, W) = R$ and $a^* < 0.5$. Similarly, the maximum on the RHS of* (34) *is achieved by Bernoulli$(1 - a^*)$. We plot* (31), (32), *and* (34) *for this example in Figure 1.*

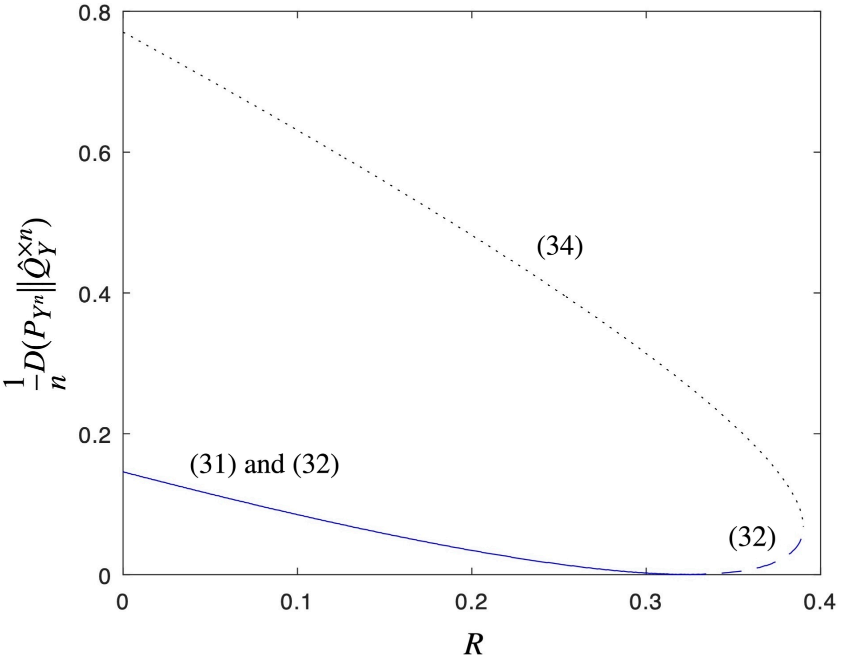

Figure 1. Example 5 with $p = 0.15$ and $q = 0.35$. The solid line depicts (31) and (32) when R satisfies (33); the dashed line depicts (32) when R exceeds the RHS of (33); and the dotted line depicts (34).

2.5. Relation to soft covering

Our results above recover the Soft-Covering Lemma [1] as well as its converse, and the proofs are considerably simpler than the ones that are usually seen in the literature. We prove both direct and converse results with the approximation error measured by the normalized relative entropy (5). As shown in [5], (5) tending to zero is a weaker requirement than $\delta_{\mathrm{TV}}(P_{Y^n}, \hat{Q}_Y^{\times n})$ tending to zero, so the direct result that we recover (which is the same as Wyner's original version [1]) is weaker than those in [5, 6], but the converse we prove is stronger than a converse stated in terms of $\delta_{\mathrm{TV}}(P_{Y^n}, \hat{Q}_Y^{\times n})$.

The converse to the Soft-Covering Lemma asserts that, in order for (5) to approach zero as $n \to \infty$, R must satisfy

$$R \geq \min_{Q_X:\, Q_X W = \hat{Q}_Y} I(Q_X, W). \tag{35}$$

This follows immediately from Theorem 3: For the RHS of (26) to approach zero, both summands inside the minimization must tend to zero. This means $Q_Y = Q_X W$ must tend to $\hat{Q}_Y$, and R must approach or exceed $I(Q_X, W)$.

For the direct Soft-Covering Lemma, we apply Theorem 2, choose Q_X to be such that $Q_Y = \hat{Q}_Y$, and let R approach $I(Q_X, W)$ from below. Then the second inequality in (13) implies that the normalized relative entropy (5) can be made arbitrarily close to zero.[*]

3. Equivocation in the wiretap channel

Consider a wiretap channel characterized by input alphabet $\mathcal{X}$, two output alphabets $\mathcal{Y}$ and $\mathcal{Z}$, and channel law $W(y, z|x)$, $(x, y, z) \in \mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$. A length-$n$ rate-R code consists of a *stochastic* encoder that maps a message $m \in \{1, \ldots, 2^{nR}\}$, possibly together with some local randomness, to x^n, and a decoder that maps y^n to $\hat{m} \in \{1, \ldots, 2^{nR}\}$; an error occurs whenever $\hat{m} \neq m$. A rate-equivocation pair $(R, \Delta_{\mathrm{eq}})$ is said to be achievable if there exists a sequence of length-n rate-R codes (indexed by n) whose error probability tends to zero as $n \to \infty$, and

$$\liminf_{n\to\infty} \frac{1}{n} H(M|Z^n) \geq \Delta_{\mathrm{eq}}. \tag{36}$$

The following result is well known [10, 14]; we provide a simple proof via Theorem 2.[†]

Theorem 6. *For any input distribution P_X such that $I(X; Y) > I(X; Z)$, any rate-equivocation pair $(R, \Delta_{\mathrm{eq}})$ satisfying the following is achievable:*

$$\Delta_{\mathrm{eq}} < R < I(X; Y) \tag{37}$$

$$\Delta_{\mathrm{eq}} < I(X; Y) - I(X; Z). \tag{38}$$

Proof. Fix some $R' > 0$. Generate a codebook

$$\{x^n(m, \ell),\ m \in \{1, \ldots, 2^{nR}\}, \ell \in \{1, \ldots, 2^{nR'}\}\} \tag{39}$$

[*] Note that our proof of the direct Soft-Covering Lemma builds upon the classic proof that, with high probability, the random codebook C admits a small decoding error probability. The latter is standard, albeit nontrivial.

[†] The optimal rate-equivocation region also involves an auxiliary random variable U satisfying the Markov relation $U \multimap\!\!- X \multimap\!\!- (Y, Z)$. The direct part of that result can be obtained immediately from Theorem 6 by viewing U, instead of X, as the channel input. We therefore focus on proving the achievability result that does not involve U.

IID according to P_X. The encoder draws L uniformly at random over $\{1, \ldots, 2^{nR'}\}$. It then maps (m, ℓ), with m being the message, to the codeword $x^n(m, \ell)$, which it subsequently sends to the channel.

Following standard procedures [13], one can show that the probability that the randomly generated codebook has small average error probability (even for decoding both m and ℓ) will tend to one as $n \to \infty$ provided

$$R + R' < I(X; Y). \tag{40}$$

We next study equivocation. Let C denote the entire (randomly generated) codebook, let $\mathsf{C}(m)$ denote $\{X^n(m, 1), \ldots, X^n(m, 2^{nR'})\}$, the random sub-codebook for message $M = m$, and let $P_{Z^n|\mathsf{C}(m)}$ denote the distribution at the eavesdropper given $M = m$ and with a uniformly chosen L. Then

$$\begin{aligned}
I(M; Z^n|\mathsf{C}) &= 2^{-nR} \sum_{m=1}^{2^{nR}} D\left(P_{Z^n|\mathsf{C}(m)} \middle\| P_{Z^n}\right) && (41)\\
&= 2^{-nR} \sum_{m=1}^{2^{nR}} D\left(P_{Z^n|\mathsf{C}(m)} \middle\| P_Z^{\times n}\right) - D\left(P_{Z^n} \middle\| P_Z^{\times n}\right) && (42)\\
&\leq 2^{-nR} \sum_{m=1}^{2^{nR}} D\left(P_{Z^n|\mathsf{C}(m)} \middle\| P_Z^{\times n}\right). && (43)
\end{aligned}$$

By Theorem 2, for any $\alpha > 0$, for all sufficiently large n,

$$\Pr\left[\frac{1}{n} D\left(P_{Z^n|\mathsf{C}(m)} \middle\| P_Z^{\times n}\right) \leq I(P_X, W) - R' + \alpha\right] \geq 1 - \alpha. \tag{44}$$

Let $B(m)$ be the indicator function

$$B(m) = \mathbf{1}\left[\frac{1}{n} D\left(P_{Z^n|\mathsf{C}(m)} \middle\| P_Z^{\times n}\right) > I(P_X, W) - R' + \alpha\right], \tag{45}$$

then $\{B(m)\}$ are IID Bernoulli(p) for some $p \leq \alpha$. Denote

$$B = 2^{-nR} \sum_m B(m), \tag{46}$$

then

$$\mathsf{E}\left[B\right] = p \leq \alpha \tag{47}$$

$$\mathsf{Var}\left[B\right] = 2^{-nR}(p - p^2) \leq 2^{-nR}\alpha. \tag{48}$$

Now define

$$d' = \max_x D(P_{Z|X=x} || P_Z), \tag{49}$$

then, by the definition of B,

$$\frac{1}{n} I(M; Z^n|\mathsf{C}) \leq I(X; Z) - R' + \alpha + B \cdot d'. \tag{50}$$

Consequently,

$$\Pr\left[\frac{1}{n}I(M;Z^n|\mathsf{C}) > I(X;Z) - R' + \alpha + 2\alpha d'\right] \leq \Pr\left[B > 2\alpha\right] \leq 2^{-nR}. \tag{51}$$

Thus, there must be a code $\mathsf{C} = C$ such that the decoding error probability is small and

$$\frac{1}{n}I(M;Z^n|\mathsf{C} = C) \leq I(X;Z) - R' + \beta \tag{52}$$

where $\beta \triangleq \alpha + 2\alpha d'$ can be made arbitrarily close to zero. This means we can achieve equivocation

$$\Delta_{\text{eq}} = \frac{1}{n}(H(M) - I(M;Z^n)) \geq R + R' - I(X;Z) - \beta. \tag{53}$$

Combining (40) and (53) together with

$$R' > 0, \tag{54}$$

and eliminating R', we obtain the desired result. □

4. State masking with CSI at decoder

State masking was first studied in [11] in a setting where the encoder has noncausal CSI while the decoder does not. Here we are mainly interested in the setting where the encoder does not have CSI but the decoder does.

4.1. Problem setup and results

Consider a state-dependent DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, state alphabet $\mathcal{S}$, and channel law

$$W(y|x,s), \quad x \in \mathcal{X}, y \in \mathcal{Y}, s \in \mathcal{S}. \tag{55}$$

The state sequence is IID according to P_S. In a length-n rate-R code, the encoder is a *possibly stochastic* mapping

$$\text{enc}\colon \left\{1, \ldots, 2^{nR}\right\} \to \mathcal{X}^n, \quad m \mapsto x^n, \tag{56}$$

and the decoder is a mapping from the output y^n and the state sequence s^n to its guess of the message:

$$\text{dec}\colon \mathcal{Y}^n \times \mathcal{S}^n \to \left\{1, \ldots, 2^{nR}\right\}, \quad (y^n, s^n) \mapsto \hat{m}. \tag{57}$$

As usual, we say a decoding error occurs if $\hat{m} \neq m$, and we consider the average error probability, where the average is over a uniformly drawn message and the randomness of the state sequence. We impose the following state-masking constraint: For some given constant $K > 0$, the joint distribution induced by a uniformly drawn message, the encoder, the random state sequence, and the channel law must satisfy, for all n,

$$I(S^n;Y^n) \leq nK. \tag{58}$$

The supremum over all rates that are achievable—in the sense that the average error probability can be made arbitrarily small as $n \to \infty$ while (58) is satisfied—is called the capacity in this setting.

Theorem 7. *The capacity of the state-dependent DMC with CSI at the decoder under the constraint* (58) *is given by*

$$\sup I(X; Y, S) \tag{59}$$

over joint distributions of the form $P_X(x)P_S(s)W(y|x, s)$ subject to

$$I(S; Y) < K. \tag{60}$$

To approach this capacity, it is sufficient to use deterministic encoders.

Proof. See Section 4.2. □

The above result may look rather natural by itself, but perhaps less so once compared to previous results on state masking without decoder CSI. When the encoder has noncausal CSI and the decoder has no CSI, the single-letter state-masking capacity formula in [11] involves a constraint on $I(S; U, Y)$ rather than just $I(S; Y)$ as in (60), where U is the auxiliary random variable in Gel'fand-Pinsker coding [15].

For a more direct comparison with Theorem 7, let us consider the setting in which neither encoder nor decoder has CSI, so the encoder is the same as above, while (57) is replaced by

$$\text{dec}' \colon \mathcal{Y}^n \to\in \{1, \dots, 2^{nR}\}, \quad y^n \mapsto \hat{m}. \tag{61}$$

Define capacity similarly as above. When $K \downarrow 0$, the problem becomes so-called *state obfuscation*, and the capacity is given by [12, Theorem 3], of which the next theorem can be considered a generalization. Its proof does not use the results from Section 2; we include this result for comparison and its proof for completeness.

Theorem 8. *The capacity of the state-dependent DMC without CSI under the constraint* (58) *is given by*

$$\sup I(U; Y) \tag{62}$$

over joint distributions $P_U(u)P_{X|U}(x|u)P_S(s)W(y|x, s)$—where U is an auxiliary random variable taking values in some finite set $\mathcal{U}$—subject to the condition

$$I(S; U, Y) < K. \tag{63}$$

Proof. See Section 4.3. □

Note that not only is (62) different from (59), but so is (63) from (60). The difference becomes even more apparent when we only allow deterministic encoders:

Remark 9. *In the no-CSI setting, if the encoder is restricted to being a deterministic mapping, then capacity becomes*

$$\sup I(X; Y) \tag{64}$$

over joint distributions $P_X(x)P_S(s)W(y|x, s)$ subject to the condition

$$I(S; X, Y) < K. \tag{65}$$

The proof is similar to that of Theorem 8 and is omitted.

4.2. Proof of Theorem 7

We start with the converse part of the theorem, which is rather straightforward. Take any (possibly random) code satisfying (58). Let $\bar{P}_{XYS}$ denote the average distribution on $\mathcal{X} \times \mathcal{Y} \times \mathcal{S}$ induced by a uniformly drawn codeword, the random states, and the channel. We start from (58) to derive the following bound:

$$\begin{aligned} nK &\geq I(S^n; Y^n) && (66)\\ &= H(S^n) - H(S^n|Y^n) && (67)\\ &= \sum_{i=1}^{n} H(S_i) - H(S_i|Y^n, S^{i-1}) && (68)\\ &= \sum_{i=1}^{n} I(S_i; Y^n, S^{i-1}) && (69)\\ &\geq \sum_{i=1}^{n} I(S_i; Y_i) && (70)\\ &\geq nI(S;Y)\Big|_{\bar{P}_{YS}} && (71)\end{aligned}$$

where the last step follows because every S_i has the same distribution, and because $I(S;Y)$, for fixed P_S, is convex in $P_{Y|S}$. On the other hand, since the decoder knows S, we can view the pair (Y, S) as the output of the channel. Using standard arguments invoking Fano's Inequality, we can show

$$nR \leq nI(X; Y, S)\Big|_{\bar{P}_{XYS}} + n\epsilon \tag{72}$$

for some $\epsilon > 0$ that approaches zero as $n \to \infty$ and as the error probability is required to vanish. The converse part of the theorem follows from (71) and (72).

We next prove the direct part with the help of Theorem 2. Let P_X be an input distribution such that $P_{XYS}(x, y, s) \triangleq P_X(x)P_S(s)W(y|x, s)$ satisfies

$$I(S;Y) \leq K' \tag{73}$$

for some $K' < K$. All single-letter mutual informations in the following are computed according to P_{XYS}.

Once again, because the decoder knows S, the effective output of the channel is the pair (Y, S). We can apply Theorem 2 with the choice

$$\hat{Q}_{YS} = P_Y \times P_S \tag{74}$$

to obtain that, for any

$$R < I(X; Y, S) \tag{75}$$

there exists a deterministic length-n rate-R code with small error probability that induces a distribution $P_{Y^nS^n}$ satisfying, for some small ϵ',

$$\begin{aligned} \frac{1}{n}D(P_{S^nY^n} \| P_S^{\times n} \times P_Y^{\times n}) &\leq I(X; Y, S) - R + D(P_{SY} \| P_S \times P_Y) + \epsilon' && (76)\\ &= I(X; Y, S) - R + I(S;Y) + \epsilon' && (77)\end{aligned}$$

$$\leq I(X;Y,S) - R + K' + \epsilon'. \tag{78}$$

We can now bound $I(S^n;Y^n)$ as

$$I(S^n;Y^n) = D(P_{S^nY^n} \| P_{S^n} \times P_{Y^n}) \tag{79}$$
$$= D(P_{S^nY^n} \| P_S^{\times n} \times P_Y^{\times n}) - D(P_{Y^n} \| P_Y^{\times n}) \tag{80}$$
$$\leq D(P_{S^nY^n} \| P_S^{\times n} \times P_Y^{\times n}) \tag{81}$$
$$\leq nK' + n(I(X;Y,S) - R + \epsilon'), \tag{82}$$

where the last step follows from (78). Since $K' < K$, and since R can be chosen to be arbitrarily close to $I(X;Y,S)$, we conclude that $I(S^n;Y^n)$ is smaller than nK for sufficiently large n. The proof is concluded by noting that the construction works for all $K' < K$.

4.3. Proof of Theorem 8

We start with the converse part. Take a code that has a small error probability and that satisfies (58). Let $P_{X^nY^nS^n}$ be the joint distribution induced by sending a uniformly drawn codeword through the channel. Define

$$U_i \triangleq (M, Y^{i-1}, S^{i-1}), \quad i = 1, \ldots, n. \tag{83}$$

Let T be uniformly distributed over $\{1, \ldots, n\}$ and independent of everything else. With a slight abuse of notation, define $U \triangleq (U_T, T)$, $Y \triangleq Y_T$, and $S \triangleq S_T$. We first upper-bound $I(M, Y^n; S^n)$ as

$$I(M, Y^n; S^n) \leq I(Y^n; S^n) + H(M|Y^n) \tag{84}$$
$$\leq nK + n\epsilon \tag{85}$$

for some ϵ that approaches zero when the error probability approaches zero and $n \to \infty$. The last step follows by (58) and Fano's Inequality. We then lower-bound $I(M, Y^n; S^n)$ as

$$I(M, Y^n; S^n) = \sum_{i=1}^{n} I(M, Y^n; S_i | S^{i-1}) \tag{86}$$
$$= \sum_{i=1}^{n} I(M, Y^n, S^{i-1}; S_i) \tag{87}$$
$$\geq \sum_{i=1}^{n} I(M, Y^i, S^{i-1}; S_i) \tag{88}$$
$$= \sum_{i=1}^{n} I(U_i, Y_i; S_i) \tag{89}$$
$$= nI(U_T, Y_T; S_T | T) \tag{90}$$
$$= nI(U, Y; S) \tag{91}$$

where (87) follows because S^n is IID; and (91) because T is independent of S_T. Combining (85) and (91), we have

$$I(U, Y; S) \leq K + \epsilon. \tag{92}$$

On the other hand, using a standard argument invoking Fano's Inequality, we can show

$$\begin{aligned} n(R-\epsilon) &\le I(M;Y^n) && (93)\\ &\le \sum_{i=1}^{n} I(U_i;Y_i) && (94)\\ &\le nI(U;Y). && (95) \end{aligned}$$

Combining (92) and (95) proves the converse part of the theorem.

We next prove the direct part of the theorem. Fix any finite set $\mathcal{U}$ and joint distribution P_{UX} on $\mathcal{U}\times\mathcal{X}$. Generate a codebook by picking the codewords $\{u^n(1),\dots,u^n(2^{nR})\}$ IID according to P_U. To send message m, the sender first passes $u_1(m),\dots,u_n(m)$ independently through $P_{X|U}$, and then sends the resulting x-sequence to the channel. Thus, effectively, we have a channel whose input alphabet is $\mathcal{U}$ instead of $\mathcal{X}$.

It follows from standard arguments that, with high probability, the randomly generated code has small average decoding error probability provided $R < I(U;Y)$. We shall show that, again with high probability, the code will also satisfy the constraint (58). Then it will follow from the union bound that there exist codes that satisfy both. To check (58), first write

$$\begin{aligned} I(S^n;Y^n) &\le I(S^n;U^n,Y^n) && (96)\\ &= I(S^n;Y^n|U^n) && (97)\\ &= H(Y^n|U^n) - H(Y^n|U^n,S^n). && (98) \end{aligned}$$

For the first term on the RHS of (98), we have

$$H(Y^n|U^n) = \sum_{i=1}^{n} H(Y_i|U^n,Y^{i-1}) \le \sum_{i=1}^{n} H(Y_i|U_i). \tag{99}$$

Since the codebook is generated IID according to P_U, by the Law of Large Numbers, the RHS divided by n tends to $H(Y|U)$ with probability one. Therefore, for every $\epsilon > 0$, for sufficiently large n, with high probability,

$$H(Y^n|U^n) \le nH(Y|U) + n\epsilon \tag{100}$$

with the understanding that $H(Y|U)$ is computed according to the chosen joint distribution. For the second term on the RHS of (98), we have

$$H(Y^n|U^n,S^n) = \sum_{i=1}^{n} H(Y_i|U_i,S_i), \tag{101}$$

because the state-dependent channel $P_{Y|US}$ is memoryless. Again by the Law of Large Numbers, for sufficiently large n, with high probability,

$$H(Y^n|U^n,S^n) \ge nH(Y|U,S) - n\epsilon. \tag{102}$$

Summarizing (98), (100), and (102), with high probability

$$I(S^n;Y^n) \le nI(S;U,Y) + 2n\epsilon. \tag{103}$$

So (58) is indeed satisfied with high probability if $I(S;U,Y) < K$ and ϵ is sufficiently small. This completes the proof.

Use of AI tools declaration

The author declares that no Artificial Intelligence (AI) tools were utilized in creating this article.

Acknowledgments

The author thanks Stefan Moser, Sergio Verdú, and Shun Watanabe for inspiring discussions and email exchanges. He also thanks Guest Editor Igal Sason and the anonymous reviewers for their helpful comments.

Conflict of interest

The author declares no conflicts of interest.

References

1. A. D. Wyner, The common information of two independent random variables, *IEEE Trans. Inform. Theory*, **21** (1975), 163–179. https://doi.org/10.1109/TIT.1975.1055346
2. T. S. Han, S. Verdú, Approximation theory of output statistics, *IEEE Trans. Inform. Theory*, **39** (1993), 752–772. https://doi.org/10.1109/18.256486
3. S. Shamai, S. Verdú, The empirical distribution of good codes, *IEEE Trans. Inform. Theory*, **43** (1997), 836–846. https://doi.org/10.1109/18.568695
4. M. Hayashi, General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel, *IEEE Trans. Inform. Theory*, **52** (2006), 1562–1575. https://doi.org/10.1109/TIT.2006.871040
5. P. Cuff, Distributed channel synthesis, *IEEE Trans. Inform. Theory*, **59** (2013), 7071–7096. https://doi.org/10.1109/TIT.2013.2279330
6. P. Cuff, A stronger soft-covering lemma and applications, in *2015 IEEE Conf. Comm. and Network Security*, Florence, Italy, 2015. https://doi.org/10.1109/CNS.2015.7346808
7. J. Hou, G. Kramer, Effective secrecy: Reliability, confusion and stealth, in *Proc. IEEE Int. Symp. Inform. Theory*, Honolulu, HI, USA, 2014. https://doi.org/10.1109/ISIT.2014.6874903
8. M. Raginsky, I. Sason, *Concentration of measure inequalities in information theory, communications, and coding: Third edition*, Now Foundations and Trends, 2018.
9. Y. Polyanskiy, S. Verdú, Empirical distribution of good channel codes with nonvanishing error probability, *IEEE Trans. Inform. Theory*, **60** (2014), 5–21. https://doi.org/10.1109/TIT.2013.2284506
10. A. D. Wyner, The wiretap channel, *Bell Syst. Techn. J.*, **54** (1975), 1355–1387. https://doi.org/10.1002/j.1538-7305.1975.tb02040.x
11. N. Merhav, S. Shamai, Information rates subject to state masking, *IEEE Trans. Inform. Theory*, **53** (2007), 2254–2261. https://doi.org/10.1109/TIT.2007.896860

12. L. Wang, G. W. Wornell, Communication over discrete channels subject to state obfuscation, *IEEE Trans. Inform. Theory*, **70** (2024), 8455–8466. https://doi.org/10.1109/TIT.2024.3432573

13. T. M. Cover, J. A. Thomas, *Elements of information theory*, 2 Eds., John Wiley & Sons, New York, 2006.

14. I. Csiszár, J. Körner, Broadcast channels with confidential messages, *IEEE Trans. Inform. Theory*, **24** (1978), 339–348. https://doi.org/10.1109/TIT.1978.1055892

15. S. I. Gel'fand, M. S. Pinsker, Coding for channels with random parameters, *Prob. Contr. Inform. Theory*, **9** (1980), 19–31.