

ε -Capacity and Strong Converse for Channels with General State

Marco Tomamichel

Centre for Quantum Technologies,
National University of Singapore
Email: cqtmarco@nus.edu.sg

Vincent Y. F. Tan

Institute for Infocomm Research, A*STAR &
Dept. of ECE, National University of Singapore
Email: vtan@nus.edu.sg

Abstract—We consider state-dependent memoryless channels with general state available at both encoder and decoder. We establish the ε -capacity and the optimistic ε -capacity. This allows us to prove a necessary and sufficient condition for the strong converse to hold. We also provide a simpler sufficient condition on the first- and second-order statistics of the state process that ensures that the strong converse holds.

I. INTRODUCTION

We study the problem of channel coding with states [1] where the channel, viewed as a stochastic kernel from the set of inputs and states to the output, is discrete, memoryless and stationary (time-invariant), while the state is allowed to be a *general source* in the sense of Verdú-Han [2], [3]. The state is assumed to be known noncausally at both encoder and decoder. This models the scenario in which the channel is non-ergodic or having memory, where both non-ergodicity and memory are induced by the general state sequence.

We derive the ε -capacity and the optimistic ε -capacity [4] under cost constraints by using the information spectrum method [2], [3] and a tight converse bound established by us [5]. These capacities only depend on the cumulative distribution function (cdf) of the Cèsaro mean of the capacity-cost functions. This corroborates our intuition because the channel is well-behaved, thus it does not require characterization using information spectrum quantities and, in particular, probabilistic limit operations [2], [3]. The only complication that can arise is due to the generality of the state and for this, we require probabilistic limits. Thus, we observe a decoupling of the randomness induced by the channel and the state. Using the form of the ε -capacity and its optimistic, we provide a necessary and sufficient condition for the channel to have the strong converse [2, Sec. V] [3, Def. 3.7.1]. By using Chebyshev's inequality, we provide a simpler sufficient condition for the strong converse to hold. This condition is based on first- and second-order statistics of the state process and hence, is easier to verify. Finally, we provide examples to illustrate the various conditions. For more extensive results on this work including *second-order coding rates*, we refer the reader to [6].

II. PRELIMINARIES AND DEFINITIONS

A. Basic Definitions

We assume throughout that \mathcal{X} , \mathcal{Y} and \mathcal{S} are finite sets. Let $\mathcal{P}(\mathcal{X})$ be the set of probability distributions on \mathcal{X} . We also de-

note the set of channels from \mathcal{X} to \mathcal{Y} as $\mathcal{P}(\mathcal{Y}|\mathcal{X}) \cong \mathcal{P}(\mathcal{Y})^{\times|\mathcal{X}|}$. In the following, we let $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X} \times \mathcal{S})$ be a channel where \mathcal{X} denotes the input alphabet, \mathcal{S} denotes the state alphabet and \mathcal{Y} denotes the output alphabet. The set of all $x \in \mathcal{X}$ that are *admissible* for the channel in state $s \in \mathcal{S}$ is

$$\mathcal{B}_s(\Gamma) := \{x \in \mathcal{X} \mid b_s(x) \leq \Gamma\}$$

for some functions $b_s : \mathcal{X} \rightarrow \mathbb{R}^+$ and $\Gamma > 0$. We do not explicitly mention Γ if there are no cost constraints, i.e., if $\Gamma = \infty$. Furthermore, we assume that the channel state S is a random variable with probability distribution $P_S \in \mathcal{P}(\mathcal{S})$.

For any $P \in \mathcal{P}(\mathcal{X}|\mathcal{S})$, we define the conditional distribution $PW \in \mathcal{P}(\mathcal{Y}|\mathcal{S})$ as $PW(y|s) = \sum_x P(x|s)W(y|x, s)$. The following *conditional log-likelihood ratios* are also of interest:

$$i(x; y|s) := \log \frac{W(y|x, s)}{PW(y|s)}, \quad j_Q(x; y|s) := \log \frac{W(y|x, s)}{Q(y|s)},$$

where the latter definition applies for any $Q \in \mathcal{P}(\mathcal{Y}|\mathcal{S})$ with $Q(\cdot|s) \gg W(\cdot|x, s)$ for every $(x, s) \in \mathcal{X} \times \mathcal{S}$. We denote the *conditional mutual information* as $I(P, W|P_S) := E[i(X; Y|S)]$ where $(S, X, Y) \leftarrow P_S(s)P(x|s)W(y|x, s)$. Furthermore, the *capacity-cost function* of the channel $W_s := W(\cdot|*, s)$ is defined as

$$C_s(\Gamma) := \max_{P \in \mathcal{P}(\mathcal{B}_s(\Gamma))} I(P, W_s).$$

A *code* for the channel W with cost constraint Γ is defined by $\mathcal{C} := \{\mathcal{M}, e, d\}$ where \mathcal{M} is the *message set*, $e : \mathcal{M} \times \mathcal{S} \rightarrow \mathcal{X}$ is the *encoder* and $d : \mathcal{Y} \times \mathcal{S} \rightarrow \mathcal{M}$ is the *decoder*. The encoder must satisfy $e(m, s) \in \mathcal{B}_s(\Gamma)$ for all $s \in \mathcal{S}$ and $m \in \mathcal{M}$. For $S \leftarrow P_S$, the average (for uniform M) and maximum error probabilities are respectively defined as

$$p_{\text{avg}}(\mathcal{C}; W, P_S) := \Pr[M \neq M'] \quad \text{and} \\ p_{\text{max}}(\mathcal{C}; W, P_S) := \max_{m \in \mathcal{M}} \Pr[M \neq M' | M = m],$$

We let $M^*(\varepsilon, \Gamma; W, P_S)$ be the maximum *code size* $|\mathcal{M}|$ for which transmission with average error probability of at most ε is possible through the channel W when the state with distribution P_S is known at both encoder and decoder. The relation of the random variables is depicted in Figure 1.

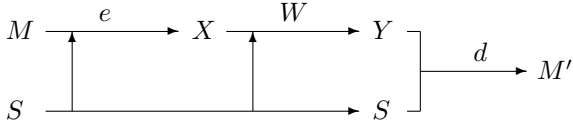


Fig. 1. Casual structure when state is known at sender and receiver.

When we consider n uses of the channel, $W^n \in \mathcal{P}(\mathcal{Y}^n | \mathcal{X}^n \times \mathcal{S}^n)$ observes the following memoryless behavior

$$W^n(\mathbf{y} | \mathbf{x}, \mathbf{s}) = \prod_{k=1}^n W(y_k | x_k, s_k), \quad (\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{S}^n.$$

and admissible inputs satisfy $\frac{1}{n} \sum_{k=1}^n b_{s_k}(x_k) \leq \Gamma$. To model general behavior of the channel, we allow the state sequence or source $\hat{S} := \{S^n = (S_1^{(n)}, \dots, S_n^{(n)})\}_{n=1}^\infty$ to evolve in an arbitrary manner in the sense of Verdú-Han [2].

We will also need the following definitions of limits in probability. For a sequence of random variables $\{A_n\}_{n=1}^\infty$, the *lim inf and lim sup in probability* [2], [3] are defined as

$$\mathfrak{p}\text{-lim inf } A_n := \sup \{r \in \mathbb{R} : \lim_{n \rightarrow \infty} \Pr[A_n < r] = 0\}, \quad \text{and}$$

$$\mathfrak{p}\text{-lim sup } A_n := \inf \{r \in \mathbb{R} : \lim_{n \rightarrow \infty} \Pr[A_n > r] = 0\}.$$

B. Definition ε -Capacity and its Dual

We say that $R \in \mathbb{R}$ is an (ε, Γ) -achievable rate if there exists a sequence of non-negative numbers $\{\varepsilon_n\}_{n=1}^\infty$ such that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log M^*(\varepsilon_n, \Gamma; W^n, P_{\hat{S}}) \geq R, \quad \limsup_{n \rightarrow \infty} \varepsilon_n \leq \varepsilon.$$

The ε -capacity-cost function $C(\varepsilon, \Gamma; W, P_{\hat{S}})$ for $\varepsilon \in [0, 1)$ is the supremum of all (ε, Γ) -achievable rates. The *capacity-cost function* $C(\Gamma; W, P_{\hat{S}}) := C(0, \Gamma; W, P_{\hat{S}})$.

Similarly, $R \in \mathbb{R}$ is a *optimistic (ε, Γ) -achievable rate* if there exists a sequence of non-negative numbers $\{\varepsilon_n\}_{n=1}^\infty$ for which

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log M^*(\varepsilon_n, \Gamma; W^n, P_{\hat{S}}) \geq R, \quad \liminf_{n \rightarrow \infty} \varepsilon_n < \varepsilon.$$

The *optimistic ε -capacity-cost function* $C^\dagger(\varepsilon, \Gamma; W, P_{\hat{S}})$ for $\varepsilon \in (0, 1]$ is the supremum of all optimistic (ε, Γ) -achievable rates. The *optimistic capacity-cost function* $C^\dagger(\Gamma; W, P_{\hat{S}}) := C^\dagger(1, \Gamma; W, P_{\hat{S}})$.

Note that $\Gamma \mapsto C(\varepsilon, \Gamma; W, P_{\hat{S}})$ and $\Gamma \mapsto C^\dagger(\varepsilon, \Gamma; W, P_{\hat{S}})$ are concave and hence continuous for $\Gamma > 0$.

A channel W with general state \hat{S} has the *strong converse property* [2, Sec. V] [3, Def. 3.7.1] if

$$C(\Gamma; W, P_{\hat{S}}) = C^\dagger(\Gamma; W, P_{\hat{S}}),$$

for all $\Gamma > 0$. This is the form of the strong converse property stated in [7]. In other words, the strong converse property holds if and only if for every sequence $\{\varepsilon_n\}_{n=1}^\infty$ for which

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log M^*(\varepsilon_n, \Gamma; W^n, P_{\hat{S}}) > C(\Gamma; W, P_{\hat{S}}),$$

we have $\lim_{n \rightarrow \infty} \varepsilon_n = 1$.

We note that even though the quantities above are defined based on the average error probability, all the results in the following also hold for the maximum error probability.

III. MAIN RESULTS

A. (ε, Γ) -Capacity and its Dual

In order to state the (ε, Γ) -capacity, it is convenient to define

$$J(R | \Gamma; W, P_{\hat{S}}) := \limsup_{n \rightarrow \infty} \Pr \left[R \geq \frac{1}{n} \sum_{k=1}^n C_{S_k}(\Gamma) \right],$$

where the probability is taken with respect to \hat{S} . Note that $\Pr[R \geq \frac{1}{n} \sum_{k=1}^n C_{S_k}(\Gamma)]$ is the cdf of the random variable $\frac{1}{n} \sum_{k=1}^n C_{S_k}(\Gamma)$ (the *information spectrum* [3]) so $J(\cdot | \Gamma; W, P_{\hat{S}})$ is the lim sup of this cdf.

Theorem 1 ((ε, Γ) -Capacity). *For every $\varepsilon \in [0, 1)$,*

$$C(\varepsilon, \Gamma; W, P_{\hat{S}}) = \sup \{R | J(R | \Gamma; W, P_{\hat{S}}) \leq \varepsilon\}. \quad (1)$$

The case of most interest is the capacity-cost function. In this case, it is easy to check from the definition of $J(R | \Gamma; W, P_{\hat{S}})$ and the \mathfrak{p} -lim inf that (1) reduces to

$$C(\Gamma; W, P_{\hat{S}}) = \mathfrak{p}\text{-lim inf}_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n C_{S_k}(\Gamma). \quad (2)$$

In the same way, in order to state our result for the optimistic (ε, Γ) -capacity, it is convenient to define the quantity:

$$J^\dagger(R | \Gamma; W, P_{\hat{S}}) := \liminf_{n \rightarrow \infty} \Pr \left[R \geq \frac{1}{n} \sum_{k=1}^n C_{S_k}(\Gamma) \right].$$

Note that $J^\dagger(\cdot | \Gamma; W, P_{\hat{S}})$ is simply the lim inf of the cdf of the random variable $\frac{1}{n} \sum_{k=1}^n C_{S_k}$.

Theorem 2 (Dual (ε, Γ) -Capacity). *For every $\varepsilon \in (0, 1]$,*

$$C^\dagger(\varepsilon, \Gamma; W, P_{\hat{S}}) = \sup \{R | J^\dagger(R | \Gamma; W, P_{\hat{S}}) < \varepsilon\}. \quad (3)$$

The case of most interest corresponds to the optimistic capacity-cost function. In this case (3) reduces to

$$C^\dagger(\Gamma; W, P_{\hat{S}}) = \mathfrak{p}\text{-lim sup}_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n C_{S_k}(\Gamma). \quad (4)$$

Note that both the ε -capacity-cost functions are expressed solely in terms of the sequence of random variables $\frac{1}{n} \sum_{k=1}^n C_{S_k}(\Gamma)$. This is because the channel W^n is well-behaved; it is memoryless and stationary and thus can effectively be characterized by $C_s(\Gamma)$ for each state $s \in \mathcal{S}$. However, the state process is general so, naturally, from the information spectrum method [3], we need to use probabilistic limit operations. It turns out that these limits are applied to the Cèsaro mean of the capacity-cost functions $\{C_s(\Gamma)\}_{s \in \mathcal{S}}$.

B. Strong Converse

Uniting (2) and (4) and recalling the definition of the strong converse property, we immediately obtain the following:

Theorem 3 (Strong Converse). *A necessary and sufficient condition for the strong converse property to hold is*

$$\mathfrak{p}\text{-lim inf}_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n C_{S_k}(\Gamma) = \mathfrak{p}\text{-lim sup}_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n C_{S_k}(\Gamma). \quad (5)$$

In other words, for the strong converse to hold, the sequence of random variables $\frac{1}{n} \sum_{k=1}^n C_{S_k}(\Gamma)$ must converge (pointwise in probability) to $C(\Gamma) \geq 0$. Furthermore, $C(\Gamma)$ is the capacity-cost function as well as the optimistic capacity-cost function of the channel W with general state \hat{S} .

While Theorem 3 provides a necessary and sufficient condition for the strong converse to hold, it requires the full statistics of \hat{S} . Thus (5) may be hard to verify in practice. We provide a simpler condition for the strong converse to hold that is based only on first- and second-order statistics.

Corollary 4 (Sufficient Condition for Strong Converse). *The strong converse holds with capacity-cost function $C(\Gamma) \geq 0$ if the following limits exist*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E}[C_{S_k}(\Gamma)] = C(\Gamma) \quad \text{and} \quad (6)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n \text{Cov}(C_{S_k}(\Gamma), C_{S_l}(\Gamma)) = 0 \quad (7)$$

Observe that if the general source \hat{S} decorrelates quickly such that $\text{Cov}(C_{S_k}(\Gamma), C_{S_l}(\Gamma))$ is small for large lags $|k-l|$, then the covariance condition in (7) is likely to be satisfied. In the following subsection, we provide some examples for which the covariance condition either holds or is violated.

C. Examples

In this section, we provide a few examples to illustrate the generality of the model and the strong converse conditions. We assume that there are no cost constraints here.

Example 1. Suppose the general source $\hat{S} := \{S^n = (S_1^{(n)}, \dots, S_n^{(n)})\}_{n=1}^{\infty}$ is such that $S_m^{(n)} = S$ for $1 \leq m \leq n$, where $S \leftarrow P_S \in \mathcal{P}(\mathcal{S})$. Then, each covariance in (7) is equal to $\text{Var}(C_S)$. If this variance is positive, then neither the sufficient condition in Corollary 4 nor the necessary condition in (5) is satisfied. For such a state sequence, which corresponds to a *mixed channel* [3, Sec. 3.3], the ε -capacity is given by

$$C(\varepsilon; W, P_{\hat{S}}) = \sup\{R \mid \Pr[R \geq C_S] \leq \varepsilon\}.$$

Example 2. Suppose the source \hat{S} is independent and identically distributed (i.i.d.) with common distribution $P_S \in \mathcal{P}(\mathcal{S})$. Then, $\text{Cov}(C_{S_k}, C_{S_l}) = 0$ for $k \neq l$ and hence the sum in (7) is simply $n \text{Var}(C_S)$. This grows linearly in n and hence, the strong converse condition holds with

$$C(W, P_{\hat{S}}) = \max_{P \in \mathcal{P}(\mathcal{X}|\mathcal{S})} I(P, W|P_S).$$

Example 3. Suppose that the source \hat{S} evolves according to a time-homogeneous, irreducible Markov chain whose states are positive recurrent. Such a Markov chain admits a unique stationary distribution $\pi \in \mathcal{P}(\mathcal{S})$. It is easy to check that $|\text{Cov}(C_{S_k}, C_{S_l})| \leq ae^{-b|k-l|}$ for some $a, b > 0$. Also, $\frac{1}{n} \sum_{k=1}^n \mathbb{E}[C_{S_k}] \rightarrow \sum_s \pi(s) C_s$. Hence, both (6) and (7) are satisfied and the channel admits a strong converse with

$$C(W, P_{\hat{S}}) = \max_{P \in \mathcal{P}(\mathcal{X}|\mathcal{S})} I(P, W|\pi).$$

A variation of the Gilbert-Elliott channel [8]–[10] with state information at the encoder and decoder is modeled in this way. In fact, since the above capacity can be achieved even without state information at the encoder (by symmetry), the above implies the strong converse for the regular Gilbert-Elliott channel.

Example 4. The covariance condition (7) is not sufficient in general. Consider the memoryless (but non-stationary) source \hat{S} given by

$$S_k = \begin{cases} S_1 & k \in \mathcal{J} \\ S_2 & k \notin \mathcal{J}, \end{cases}$$

where $\mathcal{J} := \{i \in \mathbb{N} : 2^{2k-1} \leq i < 2^{2k}, k \in \mathbb{N}\}$. Since \hat{S} is memoryless, just as in Example 2, $\text{Cov}(C_{S_k}, C_{S_l}) = 0$ for $k \neq l$. Hence, the covariance condition is satisfied. However,

$$C(W, P_{\hat{S}}) = \mathbf{p}\text{-}\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n C_{S_k} = \frac{2c}{3} + \frac{d}{3}, \quad \text{and}$$

$$C^\dagger(W, P_{\hat{S}}) = \mathbf{p}\text{-}\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n C_{S_k} = \frac{c}{3} + \frac{2d}{3},$$

where the parameters $c := \min\{\mathbb{E}[C_{S_1}], \mathbb{E}[C_{S_2}]\}$ and $d := \max\{\mathbb{E}[C_{S_1}], \mathbb{E}[C_{S_2}]\}$. If $\mathbb{E}[C_{S_1}] \neq \mathbb{E}[C_{S_2}]$, then $c < d$ and hence the necessary and sufficient condition in Theorem 3 is not satisfied and the strong converse property does not hold.

IV. PROOFS

A. One-Shot Bounds

For the direct part, we require a state-dependent generalization of Feinstein's bound [11]–[13]. The proof is omitted.

Proposition 5 (State-Dependent Feinstein Bound). *Let $\Gamma > 0$ and let $P \in \mathcal{P}(\mathcal{X}|\mathcal{S})$ be any input distribution. Then, for any $\eta > 0$, there exists a code $\mathcal{C} = \{\mathcal{M}, e, d\}$ such that*

$$p_{\max}(\mathcal{C}; W, P_S) \leq \Pr[i(X; Y|S) \leq \log |\mathcal{M}| + \eta] + \exp(-\eta) + \Pr[b_S(X) > \Gamma].$$

We prove a generalization of our one-shot converse in [5] (see also references therein).

Proposition 6 (State-Dependent Function Converse). *Let $0 \leq \varepsilon < 1$ and let $\Gamma > 0$. Then, for any $\delta > 0$,*

$$\log M^*(\varepsilon, \Gamma; W, P_S) \leq \inf_{Q \in \mathcal{P}(\mathcal{Y}|\mathcal{S})} \sup_{x: \mathcal{S} \rightarrow \mathcal{X}} \sup\{R \mid \Pr[j_Q(x(S); Y|S) \leq R] \leq \varepsilon + \delta\} - \log \delta.$$

with $x: \mathcal{S} \rightarrow \mathcal{X}$ satisfying $b_s(x(s)) \leq \Gamma$ for all $s \in \mathcal{S}$.

Proof: We consider a general code $\{\mathcal{M}, e, d\}$, where the encoder is such that $e(s, m) \subseteq \mathcal{B}_s(\Gamma)$ for all $m \in \mathcal{M}$, $s \in \mathcal{S}$. Moreover, let $Q \in \mathcal{P}(\mathcal{Y}|\mathcal{S})$ be arbitrary.

The following initial bound can be derived following the lines of [5]. See also [14]. Let $m = \log |\mathcal{M}|$ and $\delta > 0$, then

$$m \leq \sup\{R \mid \Pr[j_Q(X; Y|S) \leq R] \leq \varepsilon + \delta\} - \log \delta.$$

Here, P_{XYs} is induced by the code applied to $S \leftarrow P_S$ and a uniform M . In particular, note that $P_{X|S}(x|s) = 0$ if $x \notin \mathcal{B}_s$. We may expand the $\Pr[j_Q(X; Y|S) \leq R]$ as follows:

$$\sum_{s \in \mathcal{S}} P_S(s) \sum_{x \in \mathcal{B}_s} P_{X|S}(x|s) \Pr[j_Q(x; Y|s) \leq R | X=x, S=s].$$

Clearly, there exists a function $x_Q^* : \mathcal{S} \rightarrow \mathcal{X}$ with $x_Q^*(s) \in \arg \min_{x \in \mathcal{B}_s(\Gamma)} \Pr[j_Q(x; Y|s) \leq R | X=x, S=s]$ such that

$$\Pr[j_Q(x_Q^*(s); Y|s) \leq R | S=s] \leq \Pr[j_Q(X; Y|s) \leq R | S=s].$$

Hence, we can relax the condition on the supremum to get

$$m \leq \sup \{R \mid \Pr[j_Q(x_Q^*(S); Y|S) \leq R] \leq \varepsilon + \delta\} - \log \delta,$$

which concludes the proof after first maximizing over functions x and then minimizing over distributions Q . \blacksquare

B. ε -Capacity and its Dual

We ignore cost constraints in the remainder to make the presentation more concise. We combine the proofs of Theorems 1 and 2, but consider direct and converse bounds separately.

In the proofs, we initially fix $n \in \mathbb{N}$ and consider the n -fold channel $W^n \in \mathcal{P}(\mathcal{Y}^n | \mathcal{X}^n \times \mathcal{S}^n)$ and a state governed by $P_S \in \mathcal{P}(\mathcal{S}^n)$. We use bold letters to denote strings of length n , e.g., \mathbf{S} for S^n , and the shorthand $\Pr_*[\cdot] = \Pr[\cdot | *]$.

We use the (conditional) information variance, defined in [15]. In particular, uniting [15, Lem. 62] and [3, Rmk. 3.1.1], the following uniform bounds hold

$$V(P, W) \leq U(P, W) \leq 2.3 \log |\mathcal{Y}|. \quad (8)$$

For every $\mathbf{x} : \mathcal{S}^n \rightarrow \mathcal{X}^n$, the *joint type* [16] of $\mathbf{x}(\mathbf{s})$ and $\mathbf{s} \in \mathcal{S}^n$ is defined as

$$T_{\mathbf{x}, \mathbf{s}}(x, s) := \frac{1}{n} \sum_{k=1}^n \delta_{s_k, s} \delta_{x_k(\mathbf{s}), x} = T_{\mathbf{x}|s}(x|s) T_s(s). \quad (9)$$

Note that $\sum_x T_{\mathbf{x}}(x, s) = T_s(s)$ is the type of \mathbf{s} and $T_{\mathbf{x}, \mathbf{s}}(x|s)$ is well-defined if we set $T_{\mathbf{x}|s}(x|s)$ to uniform for all $s \in \mathcal{S}$ such that $T_s(s) = 0$. We collect all these conditional types in a set \mathcal{V}_n satisfying $|\mathcal{V}_n| \leq (n+1)^{|\mathcal{X}||\mathcal{S}|}$. We also define $T_{\mathbf{x}|s}W(y|s) := \sum_x T_{\mathbf{x}|s}(x|s)W(y|x, s)$ as the corresponding output distribution for every $T_{\mathbf{x}|s} \in \mathcal{V}_n$.

1) *Direct*: In the following, we show that for $\varepsilon \in [0, 1]$,

$$C(\varepsilon; W, P_S) \geq \sup\{R \mid J(R|W, P_S) \leq \varepsilon\}, \quad \text{and} \quad (10)$$

$$C^\dagger(\varepsilon; W, P_S) \geq \sup\{R \mid J^\dagger(R|W, P_S) < \varepsilon\}. \quad (11)$$

Proof: Consider the input distribution $P^* \in \mathcal{P}(\mathcal{X}|\mathcal{S})$ that maximizes the conditional mutual information, i.e. $P^*(\cdot|s) \in \arg \max_{P \in \mathcal{P}(\mathcal{X})} I(P, W_s)$. We now apply Proposition 5 to W^n with input distribution $P(\mathbf{x}|\mathbf{s}) = \prod_{k=1}^n P^*(x_k|s_k)$ and find that there exists a code $\mathcal{C} = \{\mathcal{M}_n, e, d\}$, with $|\mathcal{M}_n| = \exp(nR)$ for any $R > 0$, which satisfies, for any $\nu > 0$,

$$p_{\max}(\mathcal{C}; W^n, P_S) \leq \underbrace{\Pr\left[\frac{1}{n} i(\mathbf{X}; \mathbf{Y}|\mathbf{S}) \leq R + \nu\right]}_{=: p(R)} + \exp(-n\nu).$$

The probability above can be bounded as follows

$$\begin{aligned} p(R) &= \sum_{\mathbf{s} \in \mathcal{S}^n} P_S(\mathbf{s}) \Pr_{\mathbf{S}=\mathbf{s}} \left[\frac{1}{n} \sum_{k=1}^n i(X_k; Y_k | s_k) \leq R + \nu \right] \\ &\leq \sum_{\mathbf{s} \in \mathcal{S}^n} P_S(\mathbf{s}) \left(\Pr \left[\frac{1}{n} \sum_{k=1}^n C_{s_k} \leq R + 2\nu \right] \right. \\ &\quad \left. + \Pr_{\mathbf{S}=\mathbf{s}} \left[\frac{1}{n} \sum_{k=1}^n i(X_k; Y_k | s_k) < \frac{1}{n} \sum_{k=1}^n C_{s_k} - \nu \right] \right) \\ &\leq \Pr \left[\frac{1}{n} \sum_{k=1}^n C_{S_k} \leq R + 2\nu \right] + \frac{2.3 \log |\mathcal{Y}|}{n\nu^2}, \quad (12) \end{aligned}$$

where we employed Chebyshev's inequality and (8) to get

$$U(P^*(\cdot|s_k), W_{s_k}) \leq 2.3 \log |\mathcal{Y}|.$$

To show (10), set $R^* := \sup\{R \mid J(R|W, P_S) \leq \varepsilon\} - 3\nu$ and note that R^* is ε -achievable since (12) implies

$$\begin{aligned} \limsup_{n \rightarrow \infty} p_{\max}(\mathcal{C}; W^n, P_S) \\ \leq \limsup_{n \rightarrow \infty} \Pr \left[\frac{1}{n} \sum_{k=1}^n C_{S_k} \leq R + 2\nu \right] \leq \varepsilon, \quad (13) \end{aligned}$$

where the last inequality is by definition of $J(R|W, P_S)$. Since this holds for all $\nu > 0$, Eq. (10) follows. Eq. (11) follows analogously by defining $R^* := \sup\{R \mid J^\dagger(R|W, P_S) < \varepsilon\} - 3\nu$ appropriately and taking a \liminf in (13). \blacksquare

2) *Converse*: In the following, we show that for $\varepsilon \in [0, 1]$,

$$C(\varepsilon; W, P_S) \leq \sup\{R \mid J(R|W, P_S) \leq \varepsilon\}, \quad \text{and} \quad (14)$$

$$C^\dagger(\varepsilon; W, P_S) \leq \sup\{R \mid J^\dagger(R|W, P_S) < \varepsilon\}. \quad (15)$$

Proof: For n repetitions of the channel, we consider the following convex combination of conditional distributions:

$$Q^{(n)}(\mathbf{y}|\mathbf{s}) := \frac{1}{|\mathcal{V}_n|} \sum_{T_{\mathbf{x}|s} \in \mathcal{V}_n} \prod_{k=1}^n T_{\mathbf{x}|s} W(y_k | s_k).$$

Using this in Proposition 6 yields

$$\begin{aligned} \log M^*(\varepsilon; W^n, P_S) + \log \delta \\ \leq \sup_{\mathbf{x}: \mathcal{S}^n \rightarrow \mathcal{X}^n} \sup \{R \mid \Pr[j_{Q^{(n)}}(\mathbf{x}(\mathbf{S}); \mathbf{Y}|\mathbf{S}) \leq R] \leq \varepsilon + \delta\}, \end{aligned}$$

where we recall that the probability is evaluated for the distribution $\Pr[\mathbf{S}=\mathbf{s}, \mathbf{Y}=\mathbf{y}] = P_S(\mathbf{s})W^n(\mathbf{y}|\mathbf{x}(\mathbf{s}), \mathbf{s})$. Now,

$$\begin{aligned} &\Pr[j_{Q^{(n)}}(\mathbf{x}(\mathbf{S}); \mathbf{Y}|\mathbf{S}) \leq R] \\ &= \sum_{\mathbf{s} \in \mathcal{S}^n} P_S(\mathbf{s}) \Pr_{\mathbf{S}=\mathbf{s}} \left[\log \frac{\prod_{k=1}^n W(Y_k | x_k(\mathbf{s}), s_k)}{Q^{(n)}(\mathbf{Y}|\mathbf{s})} \leq R \right] \\ &\geq \sum_{\mathbf{s} \in \mathcal{S}^n} P_S(\mathbf{s}) \Pr_{\mathbf{S}=\mathbf{s}} \left[\sum_{k=1}^n \log \frac{W(Y_k | x_k(\mathbf{s}), s_k)}{T_{\mathbf{x}|s} W(Y_k | s_k)} \leq R - \log |\mathcal{V}_n| \right] \\ &= \Pr \left[\sum_{k=1}^n j_{T_{\mathbf{x}|s}}(x_k(\mathbf{S}); Y_k | s_k) \leq R - \log |\mathcal{V}_n| \right]. \end{aligned}$$

Thus, we find the following bound

$$\log M^*(\varepsilon; W^n, P_S) + \log \delta - \log |\mathcal{V}_n| \leq \underbrace{\sup_{\mathbf{x}} \sup \left\{ R \mid \Pr \left[\sum_{k=1}^n j_{T_{\mathbf{x}}|S}(x_k(\mathbf{S}); Y_k|S_k) \leq R \right] \leq \varepsilon + \delta \right\}}_{=: \text{cv}(\mathbf{x})}.$$

We next analyze $\text{cv}(\mathbf{x})$ for a fixed function $\mathbf{x} : S^n \rightarrow \mathcal{X}^n$. We first employ Markov's inequality which states for any positive-valued function f : $E[f(\mathbf{S})] \geq \gamma \Pr[f(\mathbf{S}) \geq \gamma]$. We apply this for $f(\mathbf{s}) = \Pr_{\mathbf{S}=\mathbf{s}}[\dots]$ to the above and find

$$\text{cv}(\mathbf{x}) \leq \sup \left\{ R \mid P_S[\mathcal{A}_\gamma(R)] \leq \frac{\varepsilon + \delta}{\gamma} \right\}, \quad (16)$$

where $\varepsilon + \delta < \gamma < 1$ and $\mathcal{A}_\gamma(R) \subseteq S^n$ is the set

$$\mathcal{A}_\gamma(R) := \left\{ \mathbf{s} \mid \Pr_{\mathbf{S}=\mathbf{s}} \left[\sum_{k=1}^n j_{T_{\mathbf{x}}|S}(x_k(\mathbf{s}); Y_k|S_k) \leq R \right] \geq \gamma \right\}.$$

Furthermore, consider the set

$$\mathcal{B}_\gamma(R) := \left\{ \mathbf{s} \mid nI(T_{\mathbf{x}}|S, W|T_S) + \sqrt{\frac{nV(T_{\mathbf{x}}|S, W|T_S)}{1-\gamma}} \leq R \right\},$$

where the expectation and variance are defined as

$$E \left[\sum_{k=1}^n j_{T_{\mathbf{x}}|S}(x_k(\mathbf{s}); Y_k|S_k) \mid \mathbf{S} = \mathbf{s} \right] = nI(T_{\mathbf{x}}|S, W|T_S),$$

$$\text{Var} \left[\sum_{k=1}^n j_{T_{\mathbf{x}}|S}(x_k(\mathbf{s}); Y_k|S_k) \mid \mathbf{S} = \mathbf{s} \right] = nV(T_{\mathbf{x}}|S, W|T_S),$$

and we employed the definition of the expected conditional information variance [15]. By Chebyshev's inequality, $\mathcal{B}_\gamma(R) \subseteq \mathcal{A}_\gamma(R)$ for all $R \in \mathbb{R}$. Furthermore,

$$I(T_{\mathbf{x}}|S, W|T_S) = \sum_{s' \in S} T_S(s') I(T_{\mathbf{x}}|S(\cdot|s'), W_{s'})$$

$$\leq \sum_{s' \in S} T_S(s') C_{s'} = \frac{1}{n} \sum_{k=1}^n C_{S_k}.$$

and V is bounded using (8). Substituting this into (16) yields

$$\text{cv}(\mathbf{x}) \leq \sup \left\{ R \mid \Pr \left[\sum_{k=1}^n C_{S_k} \right] \leq \frac{\varepsilon + \delta}{\gamma} \right\} + \sqrt{\frac{2.3n \log |\mathcal{Y}|}{1-\gamma}},$$

which is independent of $\mathbf{x} : S^n \rightarrow \mathcal{X}^n$.

Finally, the asymptotics (14) and (15) can be shown as follows. Due to the above, any ε -achievable rate R satisfies

$$R \leq \liminf_{n \rightarrow \infty} \frac{1}{n} M^*(\varepsilon_n; W^n, P_S)$$

$$\leq \liminf_{n \rightarrow \infty} \sup \left\{ R \mid \Pr \left[\sum_{k=1}^n C_{S_k} \right] \leq \varepsilon'_n \right\}$$

for some sequence $\hat{\varepsilon} = \{\varepsilon_n\}_{n=1}^\infty$ with $\limsup_{n \rightarrow \infty} \varepsilon_n \leq \varepsilon$ and ε'_n defined via $\varepsilon'_n = (\varepsilon_n + \delta)/\gamma$ with $\delta = 1/n$ and $\gamma = 1 - 1/\sqrt{n}$ such that the limits of the sequences coincide. Hence,

for any $\xi > 0$, there exists a constant $N_\xi \in \mathbb{N}$ such that for all $n > N_\xi$, $\Pr[\frac{1}{n} \sum_{k=1}^n C_{S_k} \leq R - \xi] \leq \varepsilon'_n$. Thus, in particular,

$$\limsup_{n \rightarrow \infty} \Pr \left[\frac{1}{n} \sum_{k=1}^n C_{S_k} \leq R - \xi \right] \leq \limsup_{n \rightarrow \infty} \varepsilon'_n \leq \varepsilon. \quad (17)$$

Hence, the supremum over all ε -achievable rates is upper bounded by the supremum of all R 's satisfying (17), concluding the proof of the first statement as $\xi \rightarrow 0$. The second statement (15) follows analogously by choosing a sequence $\hat{\varepsilon}$ with $\liminf_{n \rightarrow \infty} \varepsilon_n < \varepsilon$ and taking a \liminf in (17). ■

C. Strong Converse

Proof: It remains to show that Corollary 4 holds. If the covariance condition (7) is satisfied and we apply Chebyshev's inequality appropriately, we obtain

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n E[C_{S_k}] \leq \mathbf{p}\text{-}\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n C_{S_k}$$

$$\leq \mathbf{p}\text{-}\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n C_{S_k} \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n E[C_{S_k}].$$

Furthermore, if the limit in (6) exists, all inequalities above are equalities, which means that the condition of Theorem 3 is satisfied. Hence, the strong converse property holds. ■

REFERENCES

- [1] G. Keshet, Y. Steinberg, and N. Merhav, *Channel coding in the presence of side information*. Foundations and Trends in Communication and Information Theory, Now Publishers Inc, 2008.
- [2] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. on Inf. Th.*, vol. 40, pp. 1147–57, Apr 1994.
- [3] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer Berlin Heidelberg, Feb 2003.
- [4] P.-N. Chen and F. Alajaji, "Optimistic Shannon coding theorems for arbitrary single-user systems," *IEEE Trans. on Inf. Th.*, vol. 45, no. 7, pp. 2623–29, 1999.
- [5] M. Tomamichel and V. Y. F. Tan, "A tight upper bound for the third-order asymptotics of discrete memoryless channels," arXiv:1212.3689 [cs.IT], Nov 2012.
- [6] M. Tomamichel and V. Y. F. Tan, "ε-capacities and second-order coding rates for channels with general state," arXiv:1305.6789 [cs.IT], May 2013.
- [7] M. Hayashi and H. Nagaoka, "General Formulas for Capacity of Classical-Quantum Channels," *IEEE Trans. on Inf. Theory*, vol. 49, pp. 1753–1768, July 2003.
- [8] E. N. Gilbert, "Capacity of burst-noise channels," *Bell Syst. Tech. J.*, vol. 39, pp. 1253–65, Sep 1960.
- [9] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, pp. 1977–97, Sep 1963.
- [10] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Dispersion of the Gilbert-Elliott channel," *IEEE Trans. on Inf. Th.*, vol. 57, no. 4, pp. 1829–48, 2011.
- [11] A. Feinstein, "A new basic theorem of information theory," *IEEE Trans. on Inf. Th.*, vol. 4, no. 4, pp. 2–22, 1954.
- [12] A. J. Thomasian, "Error bounds for continuous channels," in *Fourth London Symposium on Information Theory*, pp. 46–60, C. Cherry, Ed. Butterworth, 1961.
- [13] J. N. Laneman, "On the Distribution of Mutual Information," in *Information Theory and Applications Workshop*, 2006.
- [14] L. Wang, R. Colbeck, and R. Renner, "Simple channel coding bounds," in *Intl. Symp. Inf. Th.*, (Seoul, South Korea), 2009.
- [15] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding in the finite blocklength regime," *IEEE Trans. on Inf. Th.*, vol. 56, pp. 2307–59, May 2010.
- [16] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.