Second Order Refinements for the Classical Capacity of Quantum Channels with Separable Input States

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

Vincent Y. F. Tan
Department of Electrical and Computer Engineering,
National University of Singapore
Email: vtan@nus.edu.sg

***Abstract*—We study the non-asymptotic fundamental limits for transmitting classical information over memoryless quantum channels, i.e. we investigate the amount of information that can be transmitted when the channel is used a finite number of times and a finite average decoding error is permissible. We show that, if we restrict the encoder to use ensembles of separable states, the non-asymptotic fundamental limit admits a Gaussian approximation that illustrates the speed at which the rate of optimal codes converges to the Holevo capacity as the number of channel uses tends to infinity. To do so, several important properties of quantum information quantities, such as the capacity-achieving output state, the divergence radius, and the channel dispersion, are generalized from their classical counterparts. Further, we exploit a close relation between classical-quantum channel coding and quantum binary hypothesis testing and rely on recent progress in the non-asymptotic characterization of quantum hypothesis testing and its Gaussian approximation.**

I. Introduction

One of the landmark achievements in quantum information theory is the derivation of the coding theorem for sending classical information over a noisy quantum channel by Holevo [10], and independently by Schumacher-Westmoreland [16]: the so-called HSW theorem. These results establish that the classical capacity of a quantum channel — under the restriction that the encoder uses ensembles of product (or separable) states — is given by the Holevo capacity. This restriction of the encoder is necessary to avoid the problem of the non-additivity of the Holevo capacity [6] for general channels. In fact, without this restriction, the strong converse for the classical capacity is not known in general.[1]

Let $\mathcal{W}$ be a quantum channel (formal definitions are deferred to the next section). In order to characterize the fundamental limit of data transmission, let $M^*(\mathcal{W}^n, \varepsilon)$ denote the maximum size of a length-n block code for memoryless $\mathcal{W}^n = \mathcal{W}^{\otimes n}$ with average error probability $\varepsilon \in (0,1)$. The HSW theorem, together with the weak converse established by Holevo [11] in the 1970s (the Holevo bound), asserts that

$$\lim_{\varepsilon \to 0} \liminf_{n \to \infty} \frac{1}{n} \log M^*(\mathcal{W}^n, \varepsilon) = \chi(\mathcal{W}),$$

where $\chi(\mathcal{W}) = \sup_{\psi_A} \max_P I(P, \psi_A, \mathcal{W})$ is the Holevo capacity of the channel and $I(P, \psi_A, \mathcal{W})$ is the mutual information between the classical channel input $X \leftarrow P$ and the output quantum system B induced by $\mathcal{W}$ and the ensemble ψ_A.

The HSW theorem was strengthened significantly by Ogawa-Nagaoka [13] and Winter [25] who proved the strong converse for discrete classical-quantum (c-q) channels, namely that

$$\lim_{n \to \infty} \frac{1}{n} \log M^*(\mathcal{W}^n, \varepsilon) = \chi(\mathcal{W}), \quad \text{for all } \varepsilon \in (0,1).$$

In the work by Ogawa-Nagaoka [13], the strong converse was proved using ideas from Arimoto's strong converse proof [1] for classical channels. These ideas easily extend and can be used to show the strong converse for the classical capacity with separable input ensembles. Winter's strong converse proof [25], on the other hand, is based on the method of types [4] and strictly speaking only applies to discrete memoryless c-q channels.

We are interested in investigating the approximate behavior of $\log M^*(\mathcal{W}^n, \varepsilon)$ without taking the limit $n \to \infty$. This quantity characterizes the fundamental backoff from the Holevo capacity for finite block lengths n as a function of the error tolerance ε. In particular, we want to approximate $\log M^*(\mathcal{W}^n, \varepsilon)$ for large but finite n. Note that Winter [25] in fact showed that

$$\log M^*(\mathcal{W}^n, \varepsilon) = n\chi(\mathcal{W}) + O(\sqrt{n}), \quad \text{for all } \varepsilon \in (0,1).$$

Our present work refines the $O(\sqrt{n})$ term by identifying the implied constant in this remainder term as a function of ε and a new figure of merit for quantum channels, the *quantum channel dispersion*. The resulting Gaussian approximation generalizes results for classical channels that go back to Strassen's seminal work in 1962 [19], where he showed for most well-behaved discrete memoryless channels that

$$\log M^*(W^n, \varepsilon) = nC(W) + \sqrt{nV_\varepsilon(W)}\, \Phi^{-1}(\varepsilon) + O(\log n),$$

[1] Some of these results, although not the strong converse, can be extended to characterize the classical capacity of quantum channels without restriction on the encoder by means of a technique called regularization. However, we will not discuss this here.

where $C(W)$ is the Shannon capacity, Φ the cumulative normal Gaussian distribution, and $V_\varepsilon(W)$ is another fundamental property of the channel known as the ε-channel dispersion, a term coined by Polyanskiy *et al.* [15]. Refinements to and extensions of the expansion of $\log M^*(W^n, \varepsilon)$ were pursued by Hayashi [7], Polyanskiy *et al.* [15] as well as the present authors [21].

In this work (see [23] for the full version), we prove that an analogue of the expansion for $\log M^*(W^n, \varepsilon)$ for classical channels holds also for the classical capacity of quantum channels if we restrict to separable inputs. We show that

$$\log M^*(\mathcal{W}^n, \varepsilon) = n\chi(\mathcal{W}) + \sqrt{nV_\varepsilon(\mathcal{W})}\,\Phi^{-1}(\varepsilon) + o(\sqrt{n}),$$

for all $\varepsilon \in (0,1)$, where the ε-dispersion $V_\varepsilon(\mathcal{W})$ is, roughly speaking, the (conditional) quantum relative entropy variance between the channel and the capacity-achieving output state.

This contribution generalizes and extends a recent second order analysis by the present authors [22] that covered only discrete c-q channels. (We also take note of an alternative proof of second order achievability by Beigi and Gohari [3].) The model treated here is strictly more general in that it allows to encode into arbitrary separable states and does not assume — as does the c-q channel model — that codewords are chosen as product states *where the marginals are taken from a fixed, finite set of allowed states*. This additional generality comes at a price. Due to the restrictions on permissible codewords, our previous analysis of the converse was able to employ the method of types. Here, this is no longer possible. We thus provide a different proof of the converse that arrives at the same bounds but utilizes a net over channel output states instead. Finally, the quantum channel dispersion introduced here is a non-trivial generalization of the definition in [22].

II. Problem Setup and Definitions

Let A and B be two quantum systems modeled by Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, where we assume that $\mathcal{H}_B$ is finite-dimensional. We denote by $\mathcal{S}(A)$ and $\mathcal{S}_\circ(A)$ the set of normalized quantum states (positive semi-definite operators with unit trace) and normalized pure (rank one) quantum states on $\mathcal{H}_A$, respectively. Analogously, we define $\mathcal{S}(B)$. We choose the trace norm topology on $\mathcal{S}(A)$ and denote by $\mathcal{P}_\circ(A)$ the set of probability measures on the measurable space $(\mathcal{S}_\circ(A), \Sigma)$, where Σ is the Borel σ-algebra with regards to the trace norm on $\mathcal{S}_\circ(A)$. For a finite set $\mathcal{X}$, we denote by $\mathcal{P}(\mathcal{X})$ the set of probability mass functions on $\mathcal{X}$. For convenience of exposition, we also introduce the sets $\mathcal{X}_1 := [d^2]$ and $\mathcal{X}_2 := [d^2+1]$ that are used for indices of ensembles that achieve first and second order, respectively.

A. Channel and Codes

Let $\mathcal{W} : \mathcal{S}(A) \to \mathcal{S}(B)$ be a quantum channel, i.e. a completely positive trace preserving map from linear operators on $\mathcal{H}_A$ to linear operators on $\mathcal{H}_B$. Without loss of generality, we assume that the image of $\mathcal{W}$, denoted $\mathrm{Im}(\mathcal{W})$, has full rank on $\mathcal{H}_B$. We denote by $\mathcal{S}_* \subseteq \mathcal{S}(A)$ the set of allowed input states for $\mathcal{W}$. A *code* $\mathcal{C}$ for a quantum channel $\mathcal{W}$ with allowed input states $\mathcal{S}_*$ is defined by the triple $\{\mathcal{M}, \rho_A, \mathcal{D}\}$, where $\mathcal{M}$ is a set of messages, $e : \mathcal{M} \to \mathcal{S}_*$, $m \mapsto \rho_A^m$ an encoding function and $\mathcal{D} = \{Q_B^m\}_{m\in\mathcal{M}}$ is a positive operator valued measure (POVM). We write $|\mathcal{C}| = |\mathcal{M}|$ for the cardinality of the message set. We define the *average error probability* of a code $\mathcal{C}$ for the channel $\mathcal{W}$ as

$$p_{\text{err}}(\mathcal{C}, \mathcal{W}) := \Pr[M \neq M'] = 1 - \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \mathrm{tr}\left(\mathcal{W}(\rho_A^m) Q_B^m\right)$$

where the distribution over messages P_M is assumed to be uniform on $\mathcal{M}$, $M \xrightarrow{e} A \xrightarrow{\mathcal{W}} B \xrightarrow{\mathcal{D}} M'$ forms a Markov chain, and M' thus denotes the output of the decoder. To characterize the non-asymptotic fundamental limit of data transmission over a single use of the channel, we define

$$\begin{aligned} &M^*(\varepsilon; \mathcal{W}, \mathcal{S}_*) := \\ &\qquad \max\left\{m \in \mathbb{N} \,\middle|\, \exists\, \mathcal{C} :\ |\mathcal{C}| = m \ \wedge\ p_{\text{err}}(\mathcal{C}, \mathcal{W}) \le \varepsilon\right\}. \end{aligned}$$

We are interested in the fundamental limit for $n \ge 1$ uses of a memoryless channel, $\mathcal{W}^n = \mathcal{W}^{\otimes n}$ that takes quantum states $\mathcal{S}(A^n)$ on $\mathcal{H}_A^{\otimes n}$ to quantum states $\mathcal{S}(B^n)$ on $\mathcal{H}_B^{\otimes n}$. In this work, we consider the set of separable states $\mathcal{S}_*^n \subset \mathcal{S}(A^n)$, i.e. arbitrary convex combinations of states $\rho_{A^n} = \bigotimes_{i=1}^n \rho_{A_i}$ where $\rho_{A_i} \in \mathcal{S}(A_i)$ for all $i \in [n]$. Our goal is to find the asymptotic expansion of $M^*(\varepsilon; \mathcal{W}^n, \mathcal{S}_*^n)$.

B. First Order: Channel Capacity

The Holevo capacity of a channel $\mathcal{W}$ is most commonly defined as follows.

$$\begin{aligned} \chi(\mathcal{W}) &:= \sup_{\rho_A : \mathcal{X}_1 \to \mathcal{S}(A)} \max_{P \in \mathcal{P}(\mathcal{X}_1)} I(P, \rho_A, \mathcal{W}), \quad \text{where} \\ I(P, \rho_A, \mathcal{W}) &:= H\big(\rho_B^P\big) - \sum_{x\in\mathcal{X}_1} P(x)\, H\big(\mathcal{W}(\rho_A(x))\big) \\ &= \sum_{x\in\mathcal{X}_1} P(x)\, D\big(\mathcal{W}(\rho_A(x)) \big\| \rho_B^P\big). \qquad (1) \end{aligned}$$

Here, $H(\rho) := -\mathrm{tr}(\rho\log\rho)$ is the von Neumann entropy, $D(\rho\|\sigma) := \mathrm{tr}(\rho(\log\rho - \log\sigma))$ is the relative entropy, and $\rho_B^P := \sum_{x\in\mathcal{X}_1} P(x)\,\mathcal{W}(\rho_A(x))$ the average output state induced by P. We note that $I(P, \rho_A, \mathcal{W})$ is the mutual information between the channel input, $X \leftarrow P$, and the corresponding channel output, $\mathcal{W}\big(\rho_A(X)\big)$, for a fixed ensemble ρ_A. We find the following equivalent expressions for the Holevo capacity useful:

Proposition 1. *The following expressions for the Holevo capacity $\chi(\mathcal{W})$ are equivalent to* (1).

$$\begin{aligned} \chi(\mathcal{W}) &= \sup_{\mathbb{P}\in\mathcal{P}_\circ(A)} \int \mathrm{d}\mathbb{P}(\psi_A)\, D\big(\mathcal{W}(\psi_A) \,\big\|\, \rho_B^{\mathbb{P}}\big), \qquad (2) \\ &= \min_{\sigma_B\in\mathcal{S}(B)} \sup_{\psi_A\in\mathcal{S}_\circ(A)} D\big(\mathcal{W}(\psi_A) \big\| \sigma_B\big). \qquad (3) \end{aligned}$$

Proof Ideas: The equality (1) = (2) follows by Caratheodory's theorem, see, e.g., [17, Sec. V] and the convexity of $D(\cdot\|\cdot)$ in the first argument. The equality (2) = (3) is a consequence of Sion's minimax theorem [18]. ■

The latter expression offers a powerful geometrical interpretation of the Holevo quantity as a divergence radius.

Proposition 2. *The minimizer in* (3)*, denoted* ρ_B^**, is unique and satisfies* $\rho_B^* > 0$*. We call it the* capacity achieving output state *(CAOS) and define*

$$\Gamma := \operatorname*{arg\,max}_{\psi_A \in \mathcal{S}_\circ(A)} D\big(W(\psi_A)\big\|\rho_B^*\big) \subseteq \mathcal{S}_\circ(A),$$

the set of capacity achieving input states. *Then,* ρ_B^* *lies in the convex hull of* $\mathcal{W}(\Gamma)$*. Moreover, we have* $D(\mathcal{W}(\psi_A)\|\rho_B^*) \leq \chi(\mathcal{W})$ *with equality if and only if* $\psi_A \in \Gamma$.

Finally, the set of capacity achieving input distributions *that achieve the supremum in* (2) *is non-empty and given by*

$$\Pi := \left\{ \mathbb{P} \in \mathcal{P}_\circ(A) \,\middle|\, \mathbb{P}(\Gamma) = 1 \wedge \int \mathrm{d}\mathbb{P}(\psi)\mathcal{W}(\psi) = \rho_B^* \right\}.$$

so that we can write

$$\chi(\mathcal{W}) = \sup_{\mathbb{P}\in\Pi} \int \mathrm{d}\mathbb{P}(\psi_A)\, D\big(\mathcal{W}(\psi_A)\,\big\|\rho_B^*\big)\,. \tag{4}$$

We refer to the full version for the proof [23], which is based on ideas taken from the literature [14], [17], [22].

For later reference, let us define $\Omega_1^\nu, \Omega_2^\nu \subseteq S_*(A)^{\times n}$ for some $0 < \nu \leq 1$, which describe sets of sequences of pure states of length n that are close to capacity achieving. The first set ensures that the input states are close to Γ,

$$\Omega_1^\nu := \left\{ \phi^n \in S_*(A)^{\times n} \,\middle|\, \frac{1}{n}\sum_{i=1}^n \Delta(\phi_i, \Gamma) \leq \nu \right\},$$

where $\Delta(\phi_i, \Gamma) := \min_{\vartheta\in\Gamma} \frac{1}{2}\|\phi_i - \vartheta\|_1$. The second set ensures that the average output state is close to the CAOS, and is defined as

$$\Omega_2^\nu := \left\{ \phi^n \in S_*(A)^{\times n} \,\middle|\, \frac{1}{2}\left\| \frac{1}{n}\sum_{i=1}^n \mathcal{W}(\phi_i) - \rho_B^* \right\|_1 \leq \nu \right\}.$$

The interesting, close to capacity achieving sequences are those that are in $\Omega_1^\nu \cap \Omega_2^\nu$.

We also define the set $\Pi^\nu \subseteq \mathcal{P}_\circ(A)$ of approximately capacity achieving input probability measures on $\mathcal{S}_\circ(A)$ as follows.

$$\Pi^\nu := \left\{ \mathbb{P} \in \mathcal{P}_\circ(A) \,\middle|\, \int \mathrm{d}\mathbb{P}(\psi)\Delta(\psi, \Gamma) \leq \nu \,\wedge \right. \\ \left. \frac{1}{2}\big\|\rho_B^{\mathbb{P}} - \rho_B^*\big\|_1 \leq \nu \right\}$$

Clearly, the empirical distribution of a sequence ϕ^n, defined as $P_{\phi^n}(\psi) = \frac{1}{n}\sum_{i=1}^n 1\{\psi = \phi_i\}$, is in Π^ν if and only if $\phi^n \in \Omega_1^\nu \cup \Omega_2^\nu$. We observe that $\Pi = \bigcap_{\nu>0} \Pi^\nu$.

C. Second Order: ε-Dispersion

The *ε-channel dispersion* of a channel $\mathcal{W}$ is defined in analogy with [22] based on the relative entropy variance $V(\rho\|\sigma) := \operatorname{tr}\big(\rho(\log\rho - \log\sigma - D(\rho\|\sigma))^2\big)$ introduced in [12], [20]. We recall the following definitions, for $P \in \mathcal{P}(\mathcal{X}_2)$,

$$U(P, \psi_A, \mathcal{W}) := V\left(\sum_{x\in\mathcal{X}_2} P(x)|x\rangle\langle x| \otimes \mathcal{W}(\psi_A(x)) \,\middle\|\, \sum_{x\in\mathcal{X}_2} P(x)|x\rangle\langle x| \otimes \rho_B^P \right), \tag{5}$$

$$V(P, \psi_A, \mathcal{W}) := \sum_{x\in\mathcal{X}_2} P(x) V\big(\mathcal{W}(\psi_A(x))\big\|\,\rho_B^P\big). \tag{6}$$

Also, from [22, Lm. 3], we know that for a fixed ensemble ψ_A and $P \in \operatorname{arg\,max}_{P\in\mathcal{P}(\mathcal{X}_2)} I(P, \psi_A, \mathcal{W})$, we have $U(P, \psi_A, \mathcal{W}) = V(P, \psi_A, \mathcal{W})$.

The ε-channel dispersion is given by a generalization of the second expression, Eq. (6).

$$V_\varepsilon(\mathcal{W}) := \begin{cases} V_{\min}(\mathcal{W}) & \varepsilon < \frac{1}{2} \\ V_{\max}(\mathcal{W}) & \varepsilon \geq \frac{1}{2} \end{cases}, \quad \text{where}$$

$$V_{\min}(\mathcal{W}) := \inf_{\mathbb{P}\in\Pi} \int \mathrm{d}\mathbb{P}(\psi) V\big(\mathcal{W}(\psi)\big\|\rho_B^*\big) \quad \text{and}$$

$$V_{\max}(\mathcal{W}) := \sup_{\mathbb{P}\in\Pi} \int \mathrm{d}\mathbb{P}(\psi) V\big(\mathcal{W}(\psi)\big\|\rho_B^*\big).$$

The following lemma follows by an elementary application of Caratheodory's theorem (see, e.g., [5, Th. 18, (ii)]).

Lemma 3. *Let* $\chi = \chi(\mathcal{W})$ *and let* $\nu \in \{V_{\min}(\mathcal{W}), V_{\max}(\mathcal{W})\}$*. There exists an ensemble* $\psi_A : \mathcal{X}_2 \to \mathcal{S}_\circ(A)$ *(of size* $d^2 + 1$*) and a distribution* $P \in \mathcal{P}(\mathcal{X}_2)$ *with* $\rho_B^P = \rho_B^*$ *and*

$$I(P, \psi_A, \mathcal{W}) = \chi,\ U(P, \psi_A, \mathcal{W}) = V(P, \psi_A, \mathcal{W}) = \nu.$$

D. Hypothesis Testing Divergence

A fundamental quantity we employ in our analysis is the *ε-hypothesis testing divergence* [20]. To define this quantity, let $0 \leq \varepsilon < 1$ and $\rho \geq 0$ be a normalized state (i.e. $\operatorname{tr}\rho = 1$) and $\sigma \geq 0$ an arbitrary state. Define

$$D_h^\varepsilon(\rho\|\sigma) := -\log\frac{\beta_{1-\varepsilon}(\rho\|\sigma)}{1-\varepsilon}, \quad \text{where}$$

$$\beta_{1-\varepsilon}(\rho\|\sigma) := \min_{\substack{0\leq Q\leq \mathrm{id}: \\ \operatorname{tr}(Q\rho)\geq 1-\varepsilon}} \operatorname{tr}(Q\sigma).$$

Various properties of $D_h^\varepsilon(\rho\|\sigma)$ are summarized in [22]. In particular, we will extensively employ [22, Prop. 8], which we partially restate here for the reader's convenience.

Proposition 4. *Let* $n \geq 1$, $\{\rho^i\}_{i=1}^n$*, for* $\rho^i \in \mathcal{S}(\mathcal{A})$ *a set of states and let* $\sigma \in \mathcal{S}(\mathcal{A})$ *such that* $\sigma \gg \rho^i$ *for all* $i \in [n]$*. Moreover, let* $\varepsilon \in (0, 1)$ *and* $\delta < \min\{\varepsilon, \frac{1-\varepsilon}{4}\}$*. Define*

$$D_n := \frac{1}{n}\sum_{i=1}^n D(\rho^i\|\sigma), \quad V_n := \frac{1}{n}\sum_{i=1}^n V(\rho^i\|\sigma),$$

$$T_n := \frac{1}{n}\sum_{i=1}^n T(P^{\rho^i,\sigma}\|Q^{\rho^i,\sigma}),$$

where $T(P^{\rho^i,\sigma}\|Q^{\rho^i,\sigma})$ is defined in [22, Sec. III]. Then,

$$D_h^{\varepsilon}\Big(\bigotimes_{i=1}^{n}\rho^i\Big\|\sigma^{\otimes n}\Big) \leq nD_n + \sqrt{\frac{nV_n}{1-\varepsilon-4\delta}} + F_1(\varepsilon,\delta,\sigma)\,. \tag{7}$$

Moreover, if $V_n > 0$, we have

$$\begin{aligned}D_h^{\varepsilon}\Big(\bigotimes_{i=1}^{n}\rho^i\Big\|\sigma^{\otimes n}\Big) &\leq nD_n + \sqrt{nV_n}\,\Phi^{-1}\bigg(\varepsilon + 4\delta + \frac{6\,T_n}{\sqrt{nV_n^3}}\bigg)\\ &\quad + F_1(\varepsilon,\delta,\sigma)\end{aligned} \tag{8}$$

where $F_1(\varepsilon,\delta,\sigma) := \log\frac{n\vartheta(\sigma)(1-\varepsilon)(\varepsilon+3\delta)}{\delta^4(1-(\varepsilon+3\delta))}$ and $\vartheta(\sigma)$ is defined in [22, Sec. III.C].

III. Main Result and Proof

Theorem 5. *Let $\varepsilon \in (0,1)$ and let $\mathcal{W}$ be a quantum channel with $\chi = \chi(\mathcal{W})$ and $\nu = V_\varepsilon(\mathcal{W})$. Then, we find the following asymptotic expansion:*

$$\log M^*(\varepsilon;\mathcal{W}^n,\mathcal{S}_*^n) = n\chi + \sqrt{n\nu}\,\Phi^{-1}(\varepsilon) + o(\sqrt{n}).$$

Remark 1. *Our achievability result in fact states that*

$$\log M^*(\varepsilon;\mathcal{W}^n,\mathcal{S}_*^n) \geq n\chi + \sqrt{n\nu}\,\Phi^{-1}(\varepsilon) + O(\log n).$$

It is expected that, as in the classical case, the third order term is indeed of the form $O(\log n)$ for most channels. However, we did not investigate this issue further.

The proof is split into several parts. First we discuss single-shot bounds and then treat the direct and converse asymptotic expansions separately.

A. One-Shot Bounds

We consider the following straightforward generalization of the one-shot bounds in [22] by the present authors, which is itself based on prior results in [8] and [24].

Theorem 6. *Let $\mathcal{X}$ be a finite set and let $\varepsilon \in (0,1)$, $\eta \in (0,\varepsilon)$ and $\mu \in (0,1-\varepsilon)$. For every $\rho_A : \mathcal{X} \to \mathcal{S}_*$ and every $P \in \mathcal{P}(\mathcal{X})$, we have*

$$\begin{aligned}&D_h^{\varepsilon-\eta}\big(\rho_{XB}^{(\rho_A,P)}\big\|\rho_X^{(P)}\otimes\rho_B^{(\rho_A,P)}\big) - F_3(\varepsilon,\eta) &(9)\\ &\quad\leq \log M^*(\varepsilon;\mathcal{W},\mathcal{S}_*)\\ &\quad\leq \min_{\sigma_B\in\mathcal{S}(\mathcal{B})}\sup_{\rho_A\in\mathcal{S}_*} D_h^{\varepsilon+\mu}\big(\mathcal{W}(\rho_A)\big\|\sigma_B\big) + F_4(\varepsilon,\mu), &(10)\end{aligned}$$

where $F_3(\varepsilon,\eta) := \log\frac{4\varepsilon(1-\varepsilon+\eta)}{\eta^2}$, $F_4(\varepsilon,\mu) := \log\frac{\varepsilon+\mu}{\mu(1-\varepsilon-\mu)}$, and

$$\rho_{XB}^{(\rho_A,P)} := \sum_{x\in\mathcal{X}} P(x)|x\rangle\langle x|\otimes\mathcal{W}(\rho_A(x)).$$

B. Achievability

We refer to the full version [23] for a proof of the achievability. The essential idea is to use the achievability result in [22] for a fixed ensemble, and then apply Lemma 3 to show that a finite ensemble indeed achieves the second order asymptotics.

C. Converse

Theorem 6 applied to n uses of the channel shows that

$$\begin{aligned}\log M^*(\varepsilon;\mathcal{W}^n,\mathcal{S}_*^n) &\leq \min_{\sigma_{B^n}\in\mathcal{S}(\mathcal{B})^{\otimes n}}\ \sup_{\phi^n\in\mathcal{S}_*^{\otimes n}}\\ &\qquad D_h^{\varepsilon+\mu}\big(\mathcal{W}^n(\phi^n)\|\sigma_{B^n}\big) + F_4(\varepsilon,\mu).\end{aligned}$$

In the following we let $\mu = 1/\sqrt{n}$ ensuring that $F_4(\varepsilon,\mu) = O(\log n)$. Thus, it remains to bound the first term involving the hypothesis testing divergence. To do so, we need an appropriate choice of output state σ_{B^n} to further upper bound $\log M^*(\varepsilon;\mathcal{W}^n,\mathcal{S}_*^n)$ above. We require the following auxiliary result whose proof we omit (it is based on a construction from [9, Lm. II.4] and uses continuity results for the relative entropy [2, Thm. 2]). This result essentially states that there exists a γ-net on the set of mixed states in the output space of the channel whose cardinality can be bounded appropriately.

Lemma 7. *Let $\mathcal{S}$ be the set of quantum states on a d-dimensional Hilbert space. For every $0 < \gamma < 1$, there exists a set of states $\mathcal{G}^\gamma \subseteq \mathcal{S}$ of size $|\mathcal{G}^\gamma| \leq (5/\gamma)^{2d^2}(2d/\gamma+2)^{d-1}$ such that, for every $\rho \in \mathcal{S}$, there exists a state $\tau \in \mathcal{G}^\gamma$ with minimum eigenvalue at least $\gamma/(2d+\gamma)$ which satisfies*

$$\frac{1}{2}\|\rho-\tau\|_1 \leq \gamma \quad \textit{and} \quad D(\rho\|\tau) \leq \gamma\cdot 4(2d+1).$$

Now, we choose the output state $\sigma_{B^n} \in \mathcal{S}(\mathcal{B})^{\otimes n}$ as follows:

$$\sigma_{B^n} = \frac{1}{|\mathcal{G}^\gamma|+1}\bigg[(\rho_B^*)^{\otimes n} + \sum_{\tau_B\in\mathcal{G}^\gamma}(\tau_B)^{\otimes n}\bigg]. \tag{11}$$

Note that σ_{B^n} is normalized and is, in fact, a convex combination of the n-fold tensor product of the CAOS and the n-fold tensor product of the elements of the net, of which there are only finitely many. With this choice of σ_{B^n} we bound

$$\mathrm{cv}(\phi^n) := D_h^{\varepsilon+\mu}\big(\mathcal{W}^n(\phi^n)\|\sigma_{B^n}\big)$$

in the following. In particular, we show that it is no larger than $n\chi(\mathcal{W}) + \sqrt{nV_\varepsilon(\mathcal{W})}\Phi^{-1}(\varepsilon) + o(\sqrt{n})$ for all ϕ^n. The technique for bounding $\mathrm{cv}(\phi^n)$ differs depending on the state sequence ϕ^n. We consider the three cases: 1) $\phi^n \notin \Omega_1^\nu$, 2) $\phi^n \notin \Omega_2^\nu$ and finally, 3) $\phi^n \in \Omega_1^\nu \cap \Omega_2^\nu$ in the following.

Before we commence, we state three auxiliary results.

Lemma 8. *Let $\mathcal{I}$ be a finite index set and let $\rho, \sigma_i \in \mathcal{S}(B)$ for $i \in \mathcal{I}$ be quantum states. Let $\{\alpha_i\}_{i\in\mathcal{I}} \in \mathcal{P}(\mathcal{I})$ be non-negative numbers that sum to one. Then,*

$$D_h^{\varepsilon}\bigg(\rho\,\bigg\|\,\sum_{i\in\mathcal{I}}\alpha_i\sigma_i\bigg) \leq \min_{i\in\mathcal{I}}\big\{D_h^{\varepsilon}(\rho\|\sigma_i) - \log\alpha_i\big\}$$

Lemma 9. *Let the minimum eigenvalue of a state $\sigma \in \mathcal{S}(B)$ be $\xi > 0$. Define the function $g : \mathbb{N}\setminus\{1\} \to \mathbb{R}^+$ as $g(2) := 0.6\log^2 \mathrm{e}$ and $g(d) := \log^2 d$ for $d \geq 3$. We have*

$$\max_{\rho\in\mathcal{S}(B)} V(\rho\|\sigma) \leq g(\dim\mathcal{H}_B) + \log^2\xi.$$

Lemma 10. *Let $\phi^n \in \mathcal{S}_*(A)^{\otimes n}$ be fixed and let $0 < \nu \leq 1$. If $\phi^n \notin \Omega_1^\nu$, then there exists a set $\Xi^\nu \subseteq [n]$ of cardinality $|\Xi^\nu| > n\frac{\nu}{2}$ such that, for all $i \in \Xi^\nu$, we have $\Delta(\phi_i,\Gamma) > \frac{\nu}{2}$.*

For the proof of Lemma 8, see Properties 3 and 4 in [22, Lm. 5]. For proofs of the other lemmas, we refer the reader to the full version.

1) Sequences $\phi^n \notin \Omega_1^\nu$: Applying Lemma 8 to $\mathrm{cv}(\phi^n)$ with our choice of σ_{B^n} in (11) and picking out the CAOS $(\rho_B^*)^{\otimes n}$ yields an upper bound of the form

$$\mathrm{cv}(\phi^n) \leq D_h^{\varepsilon+\mu}\big(\mathcal{W}^n(\phi^n)\|(\rho_B^*)^{\otimes n}\big) + \log\big(|\mathcal{G}^\gamma| + 1\big).$$

Furthermore, by using the Chebyshev-type upper bound in Proposition 4, we obtain

$$\mathrm{cv}(\phi^n) \leq \sum_{i=1}^n D(\mathcal{W}(\phi_i)\|\rho_B^*) + O(\sqrt{n}),$$

where we also employed Lemma 9 since the minimum eigenvalue of the CAOS ρ_B^* is positive so $V(\mathcal{W}(\phi)\|\rho_B^*) < \infty$ uniformly in ϕ. Also note that $F_1(\varepsilon, \delta) = O(\log n)$ for $\delta = 1/\sqrt{n}$. Now, we define

$$\hat{\chi} := \sup_{\phi \in \mathcal{S}(B):\, \Delta(\phi,\Gamma) > \frac{\nu}{2}} D\big(\mathcal{W}(\phi)\big\|\rho_B^*\big) < \chi.$$

This leads us to use Lemma 10 to bound

$$\begin{aligned}\mathrm{cv}(\phi^n) &\leq \sum_{i \in \Xi^\nu} \hat{\chi} + \sum_{i \notin \Xi^\nu} \chi + O(\sqrt{n}) \\ &\leq n\chi - n(\chi - \hat{\chi})\frac{\nu}{2} + O(\sqrt{n}).\end{aligned}$$

Hence, for these sequences, we have $\mathrm{cv}(\phi^n) \leq n\chi + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + o(\sqrt{n})$ as desired.

2) Sequences $\phi^n \notin \Omega_2^\nu$: For these sequences, we extract the state $(\tau_B)^{\otimes n}$ from the convex combination that defines σ_{B^n} in (11), where $\tau_B \in \arg\min_{\tau \in \mathcal{G}^\gamma} D(\rho_B\|\tau)$ is the state closest (in the relative entropy sense) to the average output state $\rho_B = \frac{1}{n}\sum_{i=1}^n \mathcal{W}(\phi_i)$ in $\mathcal{G}^\gamma$ and the constant $\gamma > 0$ is to be chosen later. Then, by using the Chebyshev-type upper bound in Proposition 4 we have

$$\begin{aligned}\mathrm{cv}(\phi^n) \leq \sum_{i=1}^n D(\mathcal{W}(\phi_i)\|\tau_B) + \sqrt{\frac{\sum_{i=1}^n V(\mathcal{W}(\phi_i)\|\tau_B)}{1-\varepsilon-\mu-4\delta}} \\ + F_1(\varepsilon,\delta) + \log\big(|\mathcal{G}^\gamma|+1\big). \qquad (12)\end{aligned}$$

Because the states in the net have minimum eigenvalue at least $\gamma/(2d+\gamma) > 0$ (cf. Lemma 7), we have that $V(\mathcal{W}(\phi_i)\|\tau_B) \leq g(d) + \log^2\big(\gamma/(2d+\gamma)\big)$, a finite constant. Thus, the second term in (12) can be upper bounded by $O(\sqrt{n})$. The sum of the third, fourth and fifth terms is of the order $O(\log n)$. Hence, continuing the bounding of $\mathrm{cv}(\phi^n)$, we obtain

$$\begin{aligned}\mathrm{cv}(\phi^n) &\leq \sum_{i=1}^n D(\mathcal{W}(\phi_i)\|\rho_B) + nD(\rho_B\|\tau_B) + O(\sqrt{n}) \\ &\leq nI(P_{\phi^n}, \mathcal{W}) + n \cdot 4\gamma(2d+1) + O(\sqrt{n}),\end{aligned}$$

where the second inequality follows from the properties of the γ-net stated in Lemma 7. Then, we know that $\tilde{\chi} < \chi$ and

$$\mathrm{cv}(\phi^n) \leq n\chi - n(\chi - \tilde{\chi} - 4\gamma(2d+1)) + O(\sqrt{n}).$$

By choosing γ small enough such that $\chi - \tilde{\chi} - 4\gamma(2d+1) > 0$, we arrive at the desired asymptotics, i.e. that $\mathrm{cv}(\phi^n) \leq n\chi + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + o(\sqrt{n})$.

3) Sequences $\phi^n \in \Omega_1^\nu \cap \Omega_2^\nu$: These are the sequences that are close to capacity-achieving. The treatment of these is similar to the presentation in [22] and the Berry-Esseen bound (8) can be employed. However, a few additional considerations have to be made and we refer to the full version for a proof.

Acknowledgements: MT thanks Mark M. Wilde and Milán Mosonyi for helpful discussion and the Isaac Newton Institute (Cambridge) for its hospitality while part of this work was completed.

References

[1] S. Arimoto. On the Converse to the Coding Theorem for Discrete Memoryless Channels. *IEEE Trans. Inf. Theory*, 19(3):357–359, 1973.

[2] K. M. R. Audenaert and J. Eisert. Continuity bounds on the quantum relative entropy. *J. Math. Phys.*, 46(10):102104, 2005.

[3] S. Beigi and A. Gohari. Quantum Achievability Proof via Collision Relative Entropy. 2013. `arXiv: 1312.3822`.

[4] I. Csiszár. The Method of Types. *IEEE Trans. Inf. Theory*, 44(6):2505–2523, 1998.

[5] H. G. Eggleston. *Convexity*. Cambridge University Press, Cambridge, U.K., 1958.

[6] M. B. Hastings. Superadditivity of Communication Capacity Using Entangled Inputs. *Nat. Phys.*, 5(4):255–257, 2009.

[7] M. Hayashi. Information Spectrum Approach to Second-Order Coding Rate in Channel Coding. *IEEE Trans. Inf. Theory*, 55(11):4947–4966, 2009.

[8] M. Hayashi and H. Nagaoka. General Formulas for Capacity of Classical-Quantum Channels. *IEEE Trans. Inf. Theory*, 49(7):1753–1768, 2003.

[9] P. Hayden, D. Leung, P. W. Shor, and A. Winter. Randomizing Quantum States: Constructions and Applications. *Commun. Math. Phys.*, 250(2):1–21, 2004.

[10] A. Holevo. The Capacity of the Quantum Channel with General Signal States. *IEEE Trans. Inf. Theory*, 44(1):269–273, 1998.

[11] A. S. Holevo. Bounds for the Quantity of Information Transmitted by a Quantum Communication Channel. *Probl. Inform. Transm.*, 9(3):177–183, 1973.

[12] K. Li. Second-order asymptotics for quantum hypothesis testing. *Ann. Stat.*, 42(1):171–189, 2014.

[13] T. Ogawa and H. Nagaoka. Strong Converse to the Quantum Channel Coding Theorem. *IEEE Trans. Inf. Theory*, 45(7):2486–2489, 1999.

[14] M. Ohya, D. Petz, and N. Watanabe. On Capacities of Quantum Channels. *Probab. Math. Stat.*, 17(1):179–196, 1997.

[15] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel Coding Rate in the Finite Blocklength Regime. *IEEE Trans. Inf. Theory*, 56(5):2307–2359, 2010.

[16] B. Schumacher and M. Westmoreland. Sending Classical Information via Noisy Quantum Channels. *Phys. Rev. A*, 56(1):131–138, 1997.

[17] B. Schumacher and M. Westmoreland. Optimal signal ensembles. *Phys. Rev. A*, 63(2):022308, 2001.

[18] M. Sion. On General Minimax Theorems. *Pacific J. Math.*, 8:171–176, 1958.

[19] V. Strassen. Asymptotische Abschätzungen in Shannons Informationstheorie. In *Trans. Third Prague Conf. Inf. Theory*, pages 689–723, Prague, 1962.

[20] M. Tomamichel and M. Hayashi. A Hierarchy of Information Quantities for Finite Block Length Analysis of Quantum Tasks. *IEEE Trans. Inf. Theory*, 59(11):7693–7710, 2013.

[21] M. Tomamichel and V. Y. F. Tan. A Tight Upper Bound for the Third-Order Asymptotics for Most Discrete Memoryless Channels. *IEEE Trans. Inf. Theory*, 59(11):7041–7051, 2013.

[22] M. Tomamichel and V. Y. F. Tan. Second-Order Asymptotics of Classical-Quantum Channels. 2013. `arXiv: 1308.6503v1`.

[23] M. Tomamichel and V. Y. F. Tan. On the Gaussian Approximation for the Classical Capacity of Quantum Channels. 2014. `arXiv: 1308.6503`.

[24] L. Wang and R. Renner. One-Shot Classical-Quantum Capacity and Hypothesis Testing. *Phys. Rev. Lett.*, 108(20), 2012.

[25] A. Winter. Coding Theorem and Strong Converse for Quantum Channels. *IEEE Trans. Inf. Theory*, 45(7):2481–2485, 1999.