Second Order Refinements for the Classical Capacity of Quantum Channels

(arXiv:1308.6503)

Marco Tomamichel[1], Vincent Y. F. Tan[2]

[1]Centre for Quantum Technologies (CQT), National University of Singapore
[2]Department of Electrical Engineering and Department of Mathematics, National University of Singapore

ISIT
June 30, 2014

Information Processing with Finite Resources

- Information theory usually treats the asymptotic limit where resources are unrestricted (e.g. the number of sequential channel uses is arbitrarily large).
- This is often a good approximation as classical computers can deal with large amounts of data easily.
- However, it fails for time-sharing in wireless networks.
- Quantum computers can only cope with a comparably small amount of data coherently for the foreseeable future.

Today's Question

How much information can we transmit reliably over a noisy quantum channel when sender and receiver have access to a small-scale quantum computer?

Coding over Quantum Channels

- Let $\mathcal{W} : \mathcal{X} \to \mathcal{S}$ be a map to quantum states on a finite-dimensional Hilbert space, denoted $\mathcal{S}$.
- Here, $\mathcal{X}$ could be a discrete set (classical-quantum channel) or a continuous set of quantum states.
- The n-fold repetition of this channel, $\mathcal{W}^n : \mathcal{X}^n \to \mathcal{S}^n$, acts as

$$\mathcal{W}^n(\vec{x}) = \mathcal{W}(x_1) \otimes \mathcal{W}(x_2) \otimes \ldots \otimes \mathcal{W}(x_n)$$

  for some n-tuple $\vec{x} = (x_1, x_2, \ldots, x_n)$.
- Thus, $\mathcal{W}^n$ only accepts product states. (But our results also apply for separable input states.)
- We are interested in how much classical information we can transmit over this channel.

Coding over Quantum Channels

- A code $\mathcal{C} = (\mathcal{M}, e, d)$ for $\mathcal{W}$ consists of a set of messages, $\mathcal{M}$, an encoder $e : \mathcal{M} \to \mathcal{X}$, and a decoding POVM $d = \{D_m\}$.
- The *average* probability of error is defined as

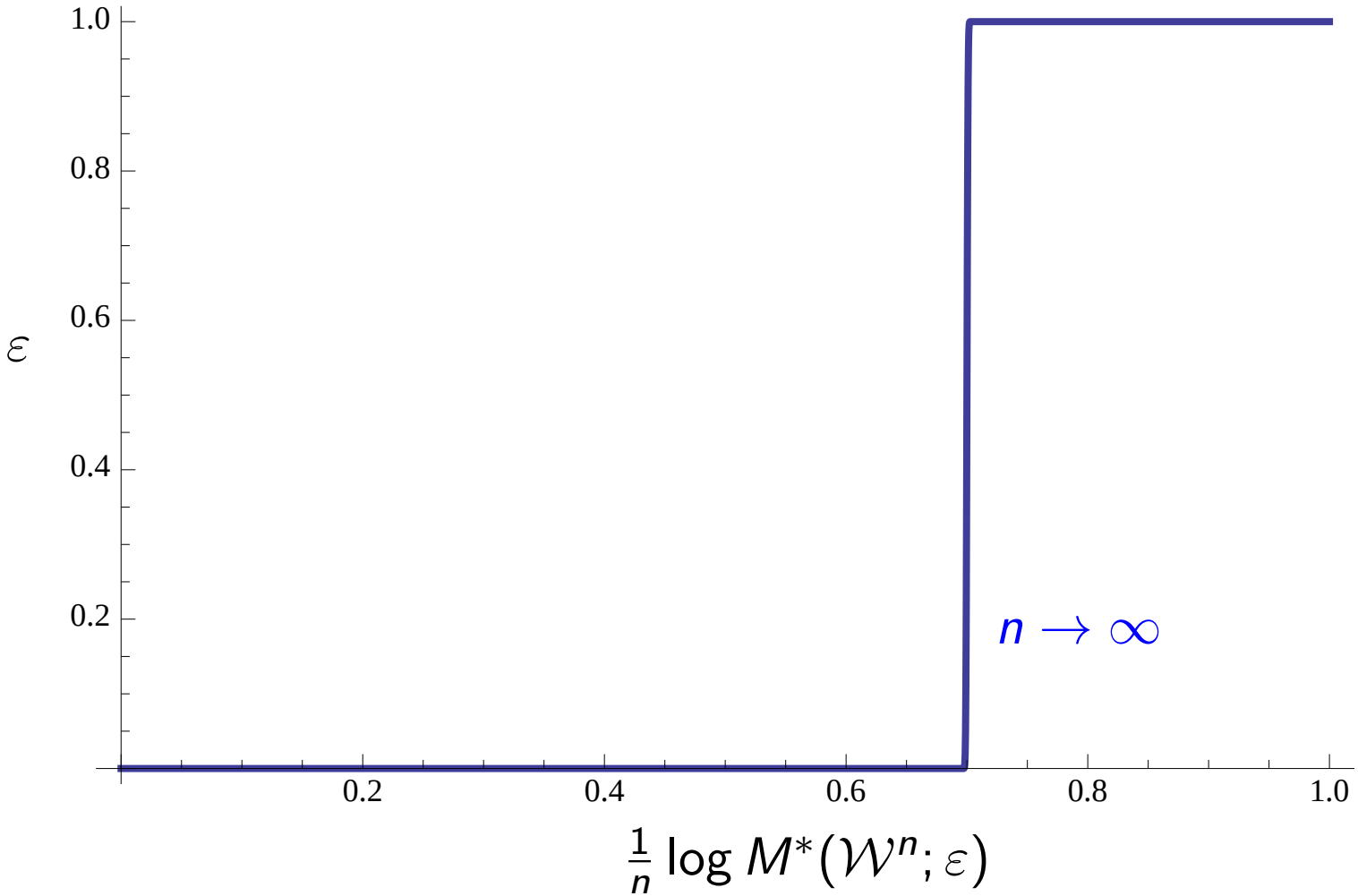
$$p_{\text{err}}(\mathcal{C}, \mathcal{W}) := 1 - \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \text{tr}(\mathcal{W}(e(m))D_m).$$

- We are interested in the maximum size of a code with error at most ε, i.e.

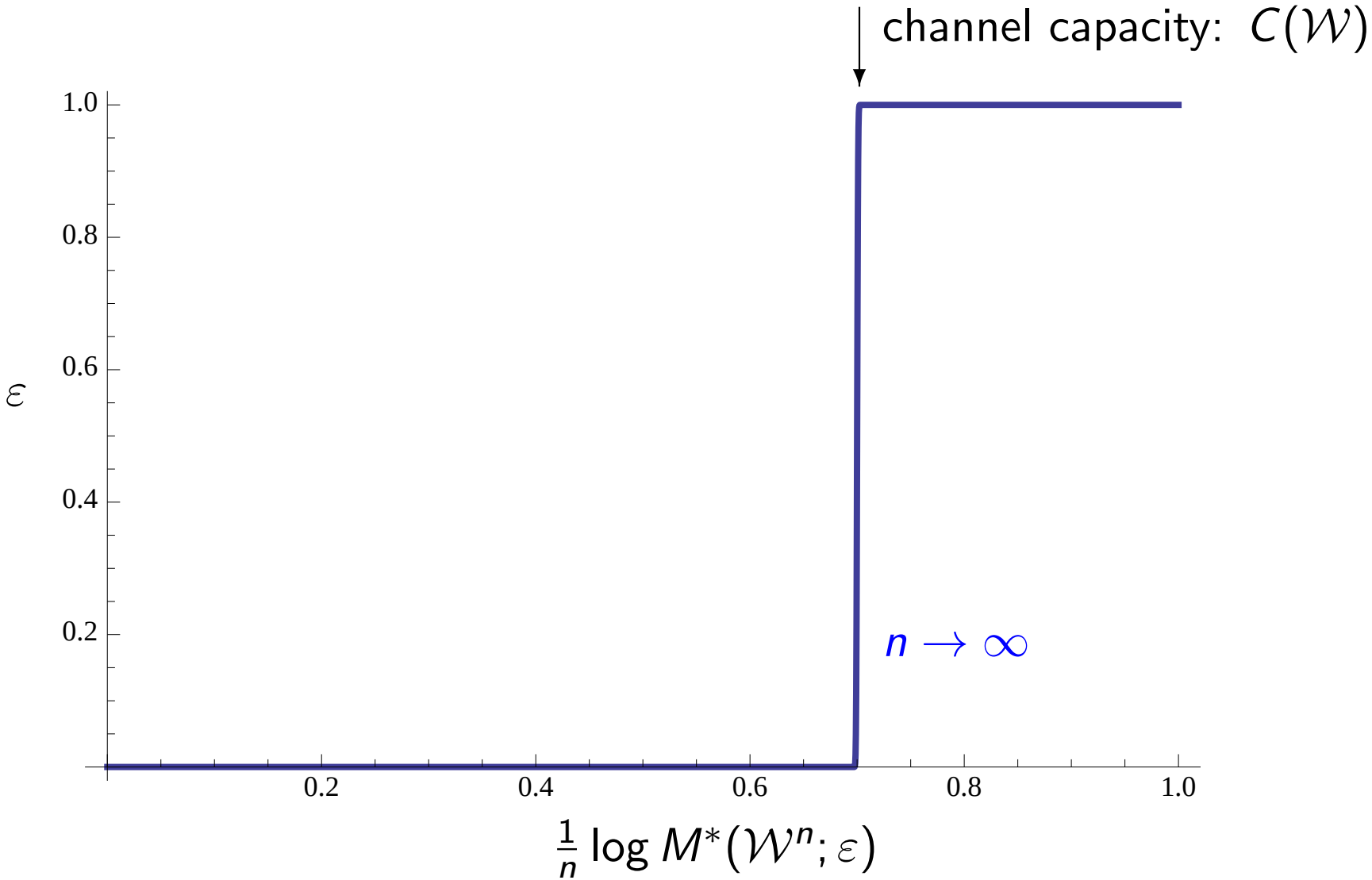
$$M^*(\mathcal{W}; \varepsilon) := \max\left\{k \in \mathbb{N} \,\middle|\, \exists \mathcal{C} : |\mathcal{M}| = k \wedge p_{\text{err}}(\mathcal{C}, \mathcal{W}) \leq \varepsilon\right\}.$$

- We want to investigate $M^*(\mathcal{W}^n; \varepsilon)$ as a function of n and ε.

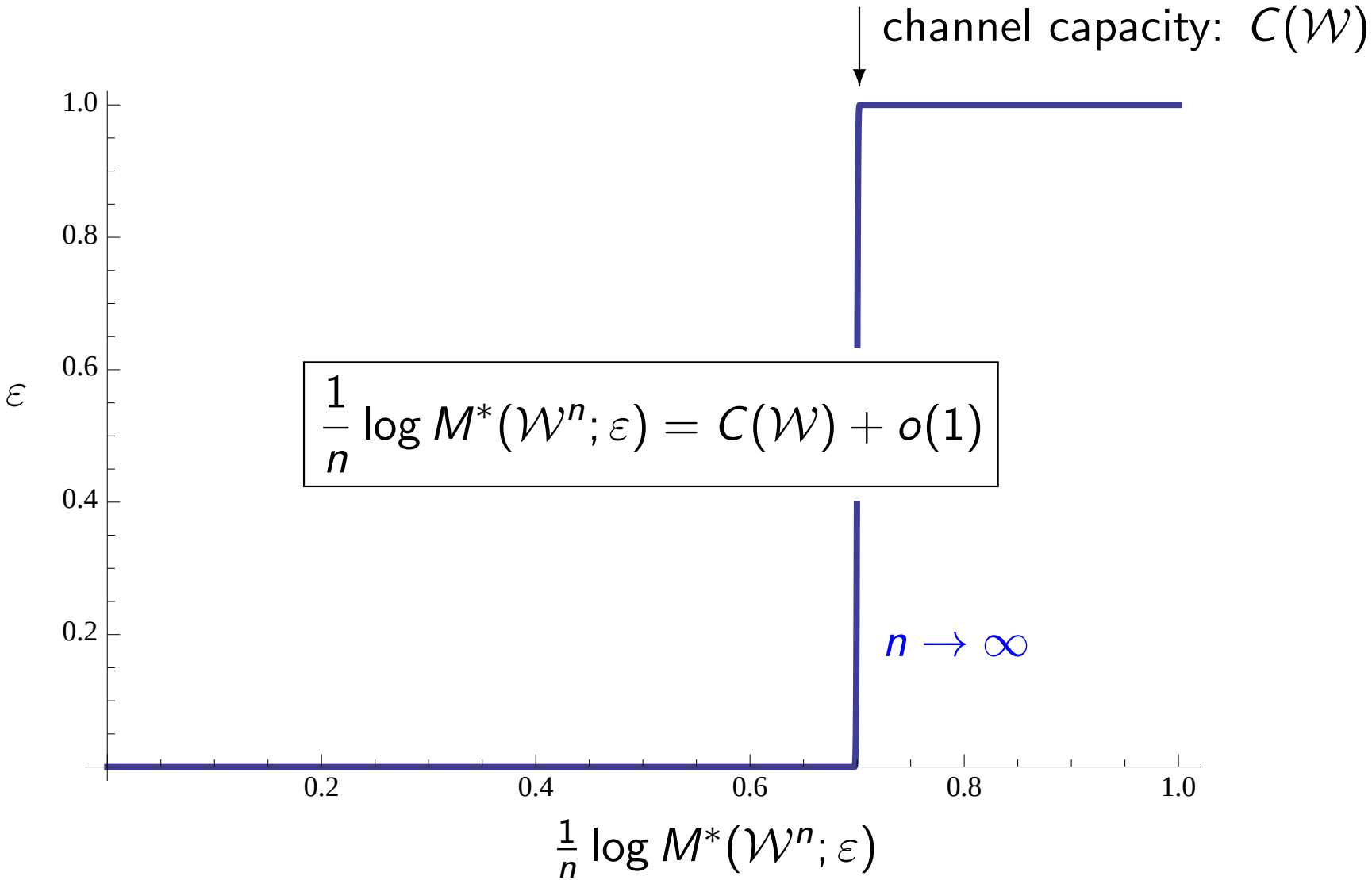
Strong Converse



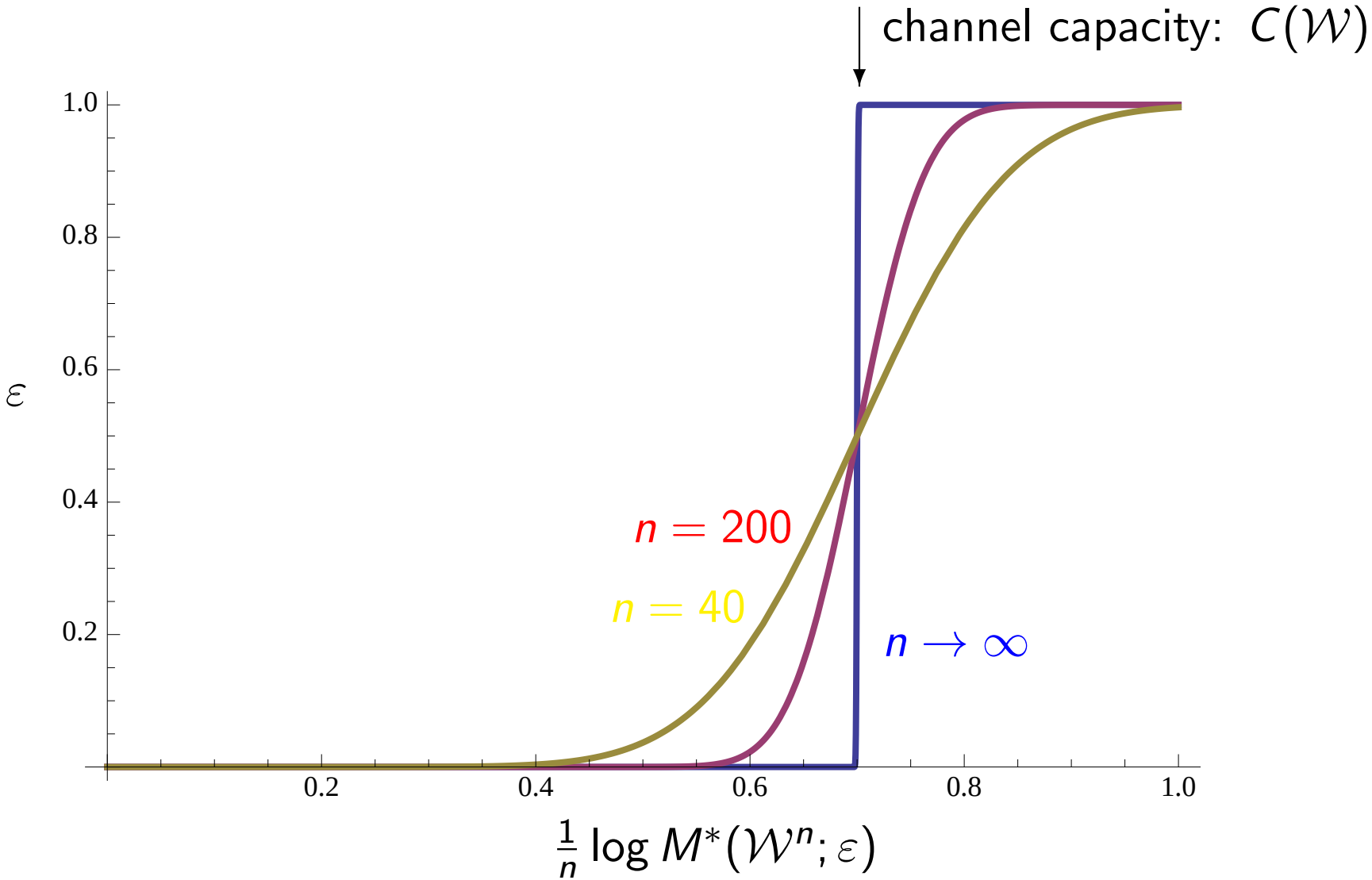
Strong Converse



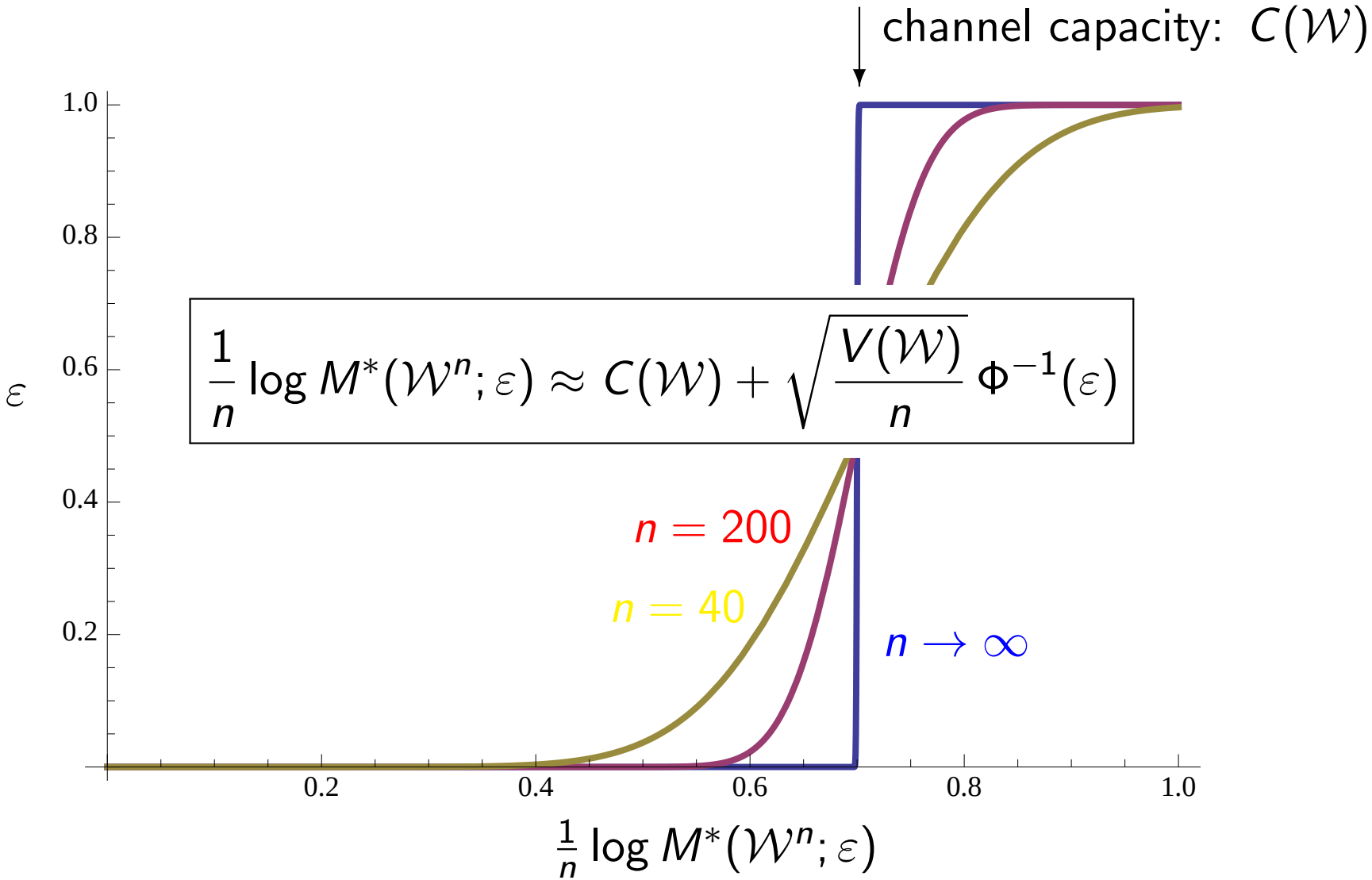
Strong Converse



Finite Resource Regime



Finite Resource Regime



Finite Resource Regime

- The main result of this work is to make this precise.

Theorem

For all channels $\mathcal{W}$ and $\varepsilon \in (0,1)$, we have

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

(Furthermore, if $\mathcal{X}$ is discrete and some regularity conditions apply, the remainder term is in fact $O(\log n)$.)

Finite Resource Regime

- The main result of this work is to make this precise.

Theorem

For all channels $\mathcal{W}$ and $\varepsilon \in (0,1)$, we have

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

(Furthermore, if $\mathcal{X}$ is discrete and some regularity conditions apply, the remainder term is in fact $O(\log n)$.)

- The *channel dispersion*, $V_\varepsilon(\mathcal{W})$ might have different values for $\varepsilon < \frac{1}{2}$ and $\varepsilon > \frac{1}{2}$, but is otherwise independent of ε.

Finite Resource Regime

- The main result of this work is to make this precise.

Theorem

For all channels $\mathcal{W}$ and $\varepsilon \in (0,1)$, we have

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

(Furthermore, if $\mathcal{X}$ is discrete and some regularity conditions apply, the remainder term is in fact $O(\log n)$.)

- The *channel dispersion*, $V_\varepsilon(\mathcal{W})$ might have different values for $\varepsilon < \frac{1}{2}$ and $\varepsilon > \frac{1}{2}$, but is otherwise independent of ε.
- We find single-letter expressions for the channel parameters, $C(\mathcal{W})$ and $V_\varepsilon(\mathcal{W})$. That is, they can actually be computed.

Finite Resource Regime

- The main result of this work is to make this precise.

Theorem

For all channels $\mathcal{W}$ and $\varepsilon \in (0,1)$, we have

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

(Furthermore, if $\mathcal{X}$ is discrete and some regularity conditions apply, the remainder term is in fact $O(\log n)$.)

- The *channel dispersion*, $V_\varepsilon(\mathcal{W})$ might have different values for $\varepsilon < \frac{1}{2}$ and $\varepsilon > \frac{1}{2}$, but is otherwise independent of ε.
- We find single-letter expressions for the channel parameters, $C(\mathcal{W})$ and $V_\varepsilon(\mathcal{W})$. That is, they can actually be computed.
- The parameters have a geometrical interpretation.

History and Previous Work

- Classical Channels:
  - Strassen'62 first introduced the Gaussian approximation for (classical) discrete memoryless channels.
  - These results were recently rediscovered and strengthened by Hayashi [IEEE TIT **55**, 2009] and Polyanskiy, Poor & Verdú [IEEE TIT **56**, 2010].
- Quantum Channels:
  - We build on one-shot bounds for $M^*(\mathcal{W};\varepsilon)$ that are due to Hayashi & Nagaoka [IEEE TIT **49**, 2003]. (See also Wang & Renner [PRL **108**, 2012].)
  - A Gaussian approximation for quantum hypothesis testing due to T & Hayashi [IEEE TIT **59**, 2013] and Li [Ann. Stat. **42**, 2014].
  - Classical proof strategies, for example in Hayashi'09, Polyanskiy *et al.*'10 as well as T & Tan [IEEE TIT **59**, 2013].

Proof Challenges

- The direct part follows from the Hayashi & Nagaoka (one-shot) bound together with the Gaussian approximation for quantum hypothesis testing. The only new ingredient here is a refined application of Carathéodory's Theorem.

Proof Challenges

- The direct part follows from the Hayashi & Nagaoka (one-shot) bound together with the Gaussian approximation for quantum hypothesis testing. The only new ingredient here is a refined application of Carathéodory's Theorem.
- For the converse, we analyze the one-shot bound

$$\log M^*(\mathcal{W}^n, \varepsilon) \lesssim \inf_{\sigma_n \in \mathcal{S}^{\otimes n}} \sup_{\rho_n \in \mathrm{im}(\mathcal{W}^n)} D_h^\varepsilon(\rho_n \| \sigma_n).$$

  by choosing an appropriate σ_n that works well for all ρ_n.
- In the classical (second order) converse proofs one usually relies on the Method of Types. This is not possible here since we allow continuous sets of channel input states.
- Instead, we span an ε-covering on the (finite-dimensional) output space to deal with codes that are far from capacity.

Some Information Quantities

- The relative entropy is given by

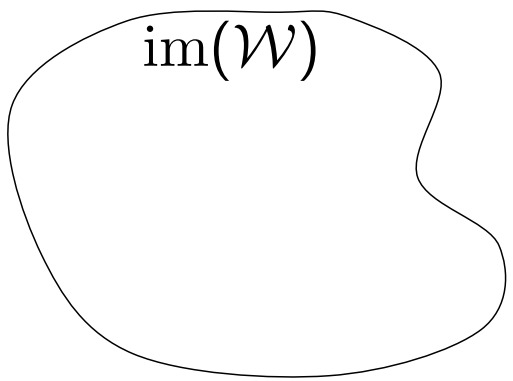
$$D(\rho\|\sigma) := \mathrm{tr}\left(\rho\big(\log\rho - \log\sigma\big)\right).$$

- The relative entropy variance is given by

$$V(\rho\|\sigma) := \mathrm{tr}\left(\rho\big(\log\rho - \log\sigma\big)^2\right) - D(\rho\|\sigma)^2.$$

Divergence Radius and Center

- Channel image: $\mathrm{im}(\mathcal{W}) := \{\rho \in \mathcal{S} \,|\, \exists x \in \mathcal{X} : \mathcal{W}(x) = \rho\}$.



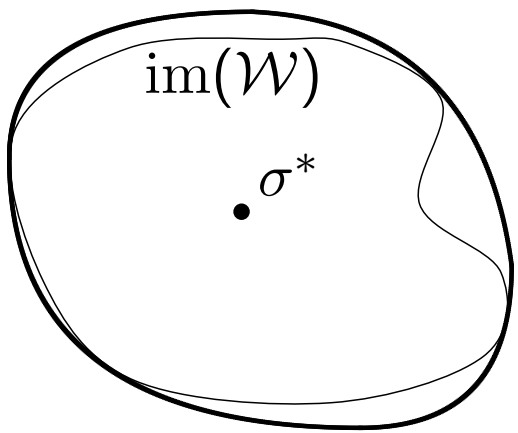
Divergence Radius and Center

- Channel image: $\mathrm{im}(\mathcal{W}) := \left\{\rho \in \mathcal{S} \,\middle|\, \exists x \in \mathcal{X} : \mathcal{W}(x) = \rho\right\}$.
- Divergence radius (of channel image):

$$\chi(\mathcal{W}) := \min_{\sigma \in \mathcal{S}} \max_{\rho \in \mathrm{im}(\mathcal{W})} D(\rho \| \sigma)$$

- Divergence center (it is unique!):

$$\sigma^*(\mathcal{W}) := \arg\min_{\sigma \in \mathcal{S}} \max_{\rho \in \mathrm{im}(\mathcal{W})} D(\rho \| \sigma)$$



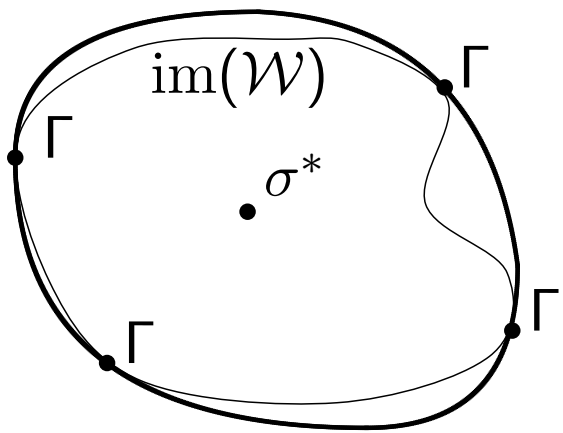
Divergence Radius and Center

- Channel image: $\mathrm{im}(\mathcal{W}) := \{\rho \in \mathcal{S} \,|\, \exists x \in \mathcal{X} : \mathcal{W}(x) = \rho\}$.
- Divergence radius (of channel image):

$$\chi(\mathcal{W}) := \min_{\sigma \in \mathcal{S}} \max_{\rho \in \mathrm{im}(\mathcal{W})} D(\rho \| \sigma)$$

- Divergence center (it is unique!):

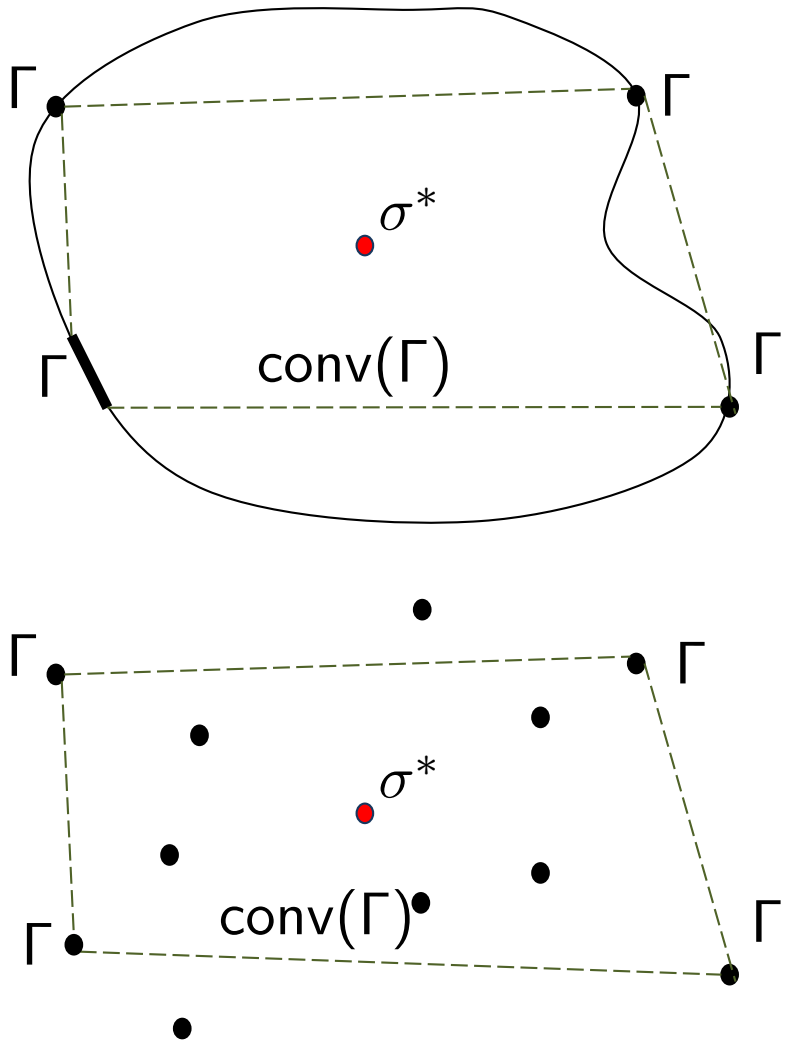
$$\sigma^*(\mathcal{W}) := \arg\min_{\sigma \in \mathcal{S}} \max_{\rho \in \mathrm{im}(\mathcal{W})} D(\rho \| \sigma)$$



- Peripheral Points: $\Gamma(\mathcal{W}) := \arg\max_{\rho \in \mathrm{im}(\mathcal{W})} D(\rho \| \sigma^*)$.

Peripheral Decompositions of the Divergence Center

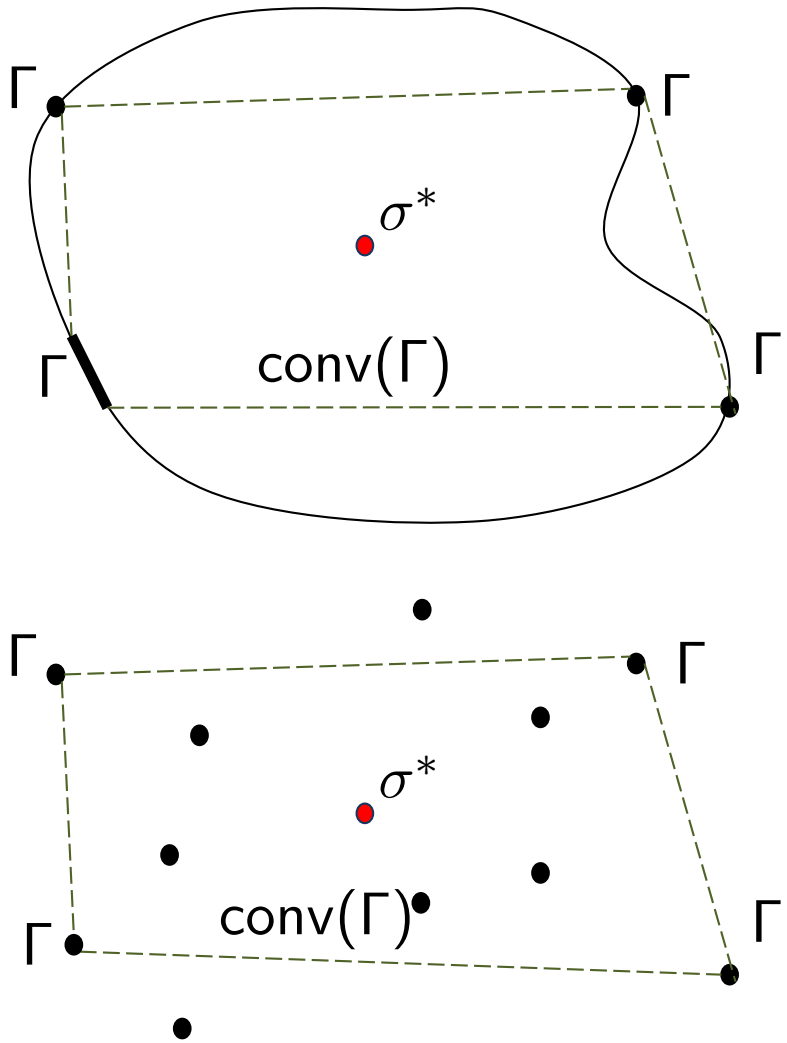
- The center $\sigma^*(\mathcal{W})$ lies in the convex hull of $\Gamma(\mathcal{W})$. (This is intuitive, but not trivial, since $D(\cdot\|\cdot)$ is not a metric.)



Peripheral Decompositions of the Divergence Center

- The center $\sigma^*(\mathcal{W})$ lies in the convex hull of $\Gamma(\mathcal{W})$. (This is intuitive, but not trivial, since $D(\cdot\|\cdot)$ is not a metric.)
- We consider the set Π of valid decompositions of σ^* using points in Γ.
- $P \in \Pi$ implies

$$\sum_{\rho\in\Gamma} P(\rho)\rho = \sigma^*.$$



From Divergence Center to Holevo's Formula

- Let P be any distribution over the set of output states $\mathcal{S}$.
- Using a minimax theorem, we find Holevo's expression:

$$\begin{aligned}\chi(\mathcal{W}) &= \max_P \left\{ \sum_\rho P(\rho)\, D\left(\rho \,\middle\|\, \sum_\rho P(\rho)\rho\right)\right\} \\ &= \max_P \left\{ H\left(\sum_\rho P(\rho)\rho\right) - \sum_\rho P(\rho) H(\rho)\right\}\end{aligned}$$

Thus, in particular

$$\chi(\mathcal{W}) = \sum_{\rho\in\Gamma} P(\rho)\, D\big(\rho \,\big\|\, \sigma^*(\mathcal{W})\big), \qquad \forall\, P \in \Pi.$$

From Divergence Center to Holevo's Formula

- Let P be any distribution over the set of output states $\mathcal{S}$.
- Using a minimax theorem, we find Holevo's expression:

$$\chi(\mathcal{W}) = \max_P \left\{ \sum_\rho P(\rho)\, D\left(\rho \,\middle\|\, \sum_\rho P(\rho)\rho\right) \right\}$$
$$= \max_P \left\{ H\left(\sum_\rho P(\rho)\rho\right) - \sum_\rho P(\rho) H(\rho) \right\}$$

Thus, in particular

$$\chi(\mathcal{W}) = \sum_{\rho \in \Gamma} P(\rho)\, D\big(\rho \,\big\|\, \sigma^*(\mathcal{W})\big), \qquad \forall\, P \in \Pi.$$

- The argument of the maximization is a mutual information.
- The decompositions $P \in \Pi$ correspond to signaling ensembles that achieve capacity.

Channel Capacity

- The HSW theorem due to Holevo [IEEE TIT **44**, 1998], and Schumacher & Westmoreland [Phys. Rev. A **56**, 1997] implies

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

Channel Capacity

- The HSW theorem due to Holevo [IEEE TIT **44**, 1998], and Schumacher & Westmoreland [Phys. Rev. A **56**, 1997] implies

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

- The strong converse due to Winter [IEEE TIT **45**, 1999] and Ogawa & Nagaoka [IEEE TIT **45**, 1999] yields

$$\lim_{n \to \infty} \frac{1}{n} \log M^*(\mathcal{W}^n; \varepsilon) = \chi(\mathcal{W}), \qquad \forall\, \varepsilon \in (0, 1).$$

- Thus, $C(\mathcal{W}) = \chi(\mathcal{W})$.

Peripheral Information Variance

- To investigate the finite block length behavior, we define

$$v_{\min}(\mathcal{W}) := \min_{P \in \Pi} \left\{ \sum_{\rho \in \mathrm{im}(\mathcal{W})} P(\rho)\, V\left(\rho \middle\| \sigma^*(\mathcal{W})\right) \right\}, \qquad \text{and}$$

$$v_{\max}(\mathcal{W}) := \max_{P \in \Pi} \left\{ \sum_{\rho \in \mathrm{im}(\mathcal{W})} P(\rho)\, V\left(\rho \middle\| \sigma^*(\mathcal{W})\right) \right\}.$$

- We find that

$$V_\varepsilon(\mathcal{W}) = \begin{cases} v_{\min}(\mathcal{W}) & \textit{if} \quad \varepsilon < \frac{1}{2} \\ v_{\max}(\mathcal{W}) & \textit{else} \end{cases} .$$

- At $\varepsilon = \frac{1}{2}$ the sign of $\Phi^{-1}(\varepsilon)$ changes, thus the preferred signaling ensemble changes with it.

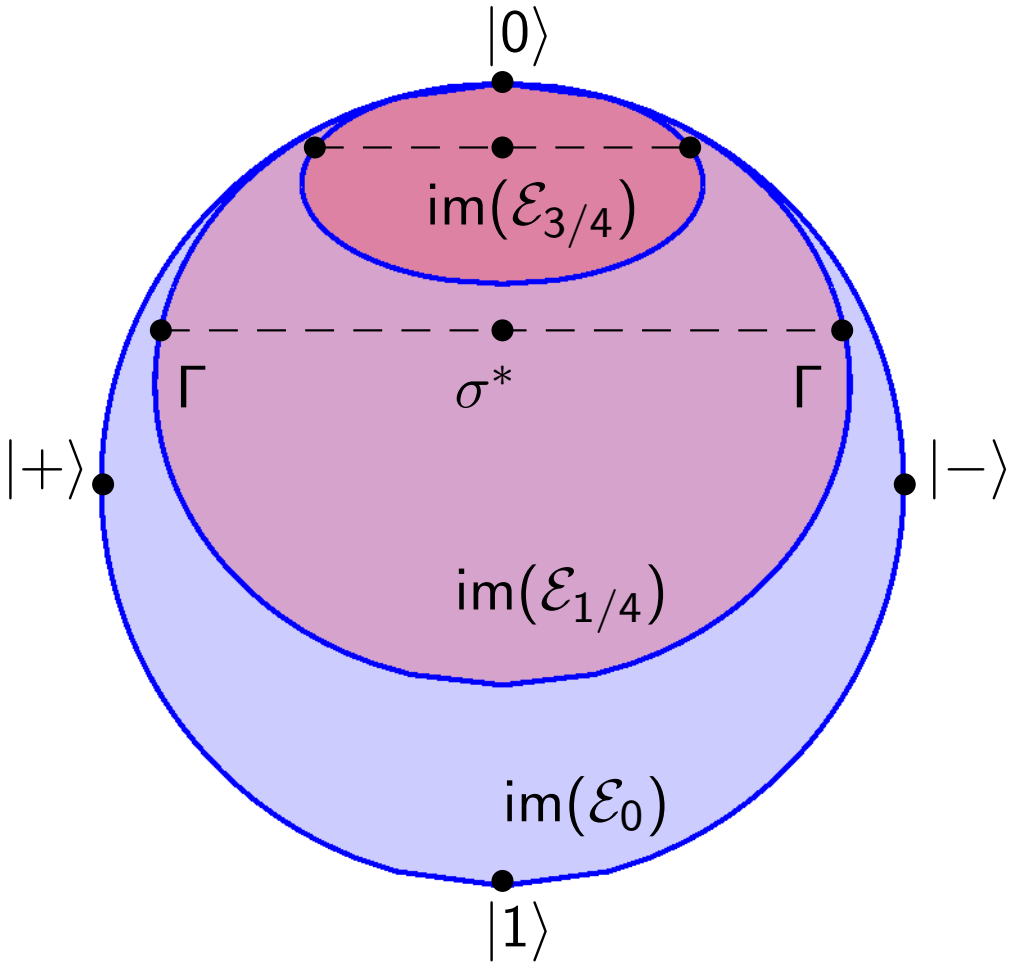
Exampe: Amplitude Damping Channel

- The amplitude damping channel for $\gamma \in [0,1]$ is given by

$$\mathcal{E}_\gamma : \rho \mapsto \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix} \rho \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix} + \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix} \rho \begin{pmatrix} 0 & 0 \\ \sqrt{\gamma} & 0 \end{pmatrix}.$$

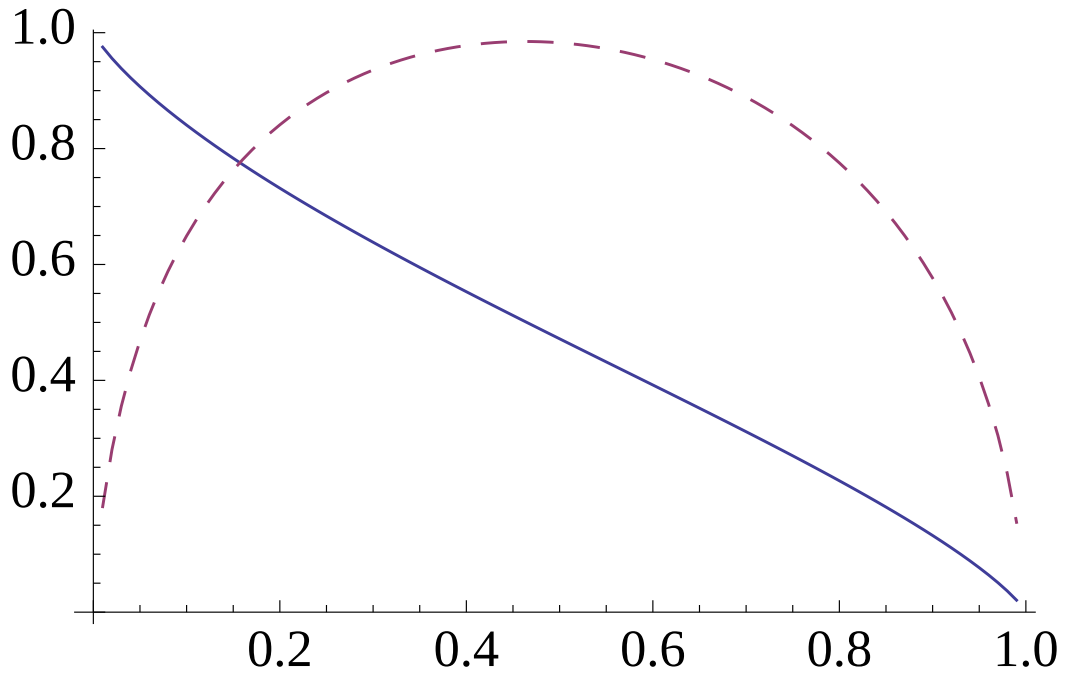
- It models the relaxation to a ground state — here $|0\rangle$ — in an open quantum system.

Example: Amplitude Damping Channel



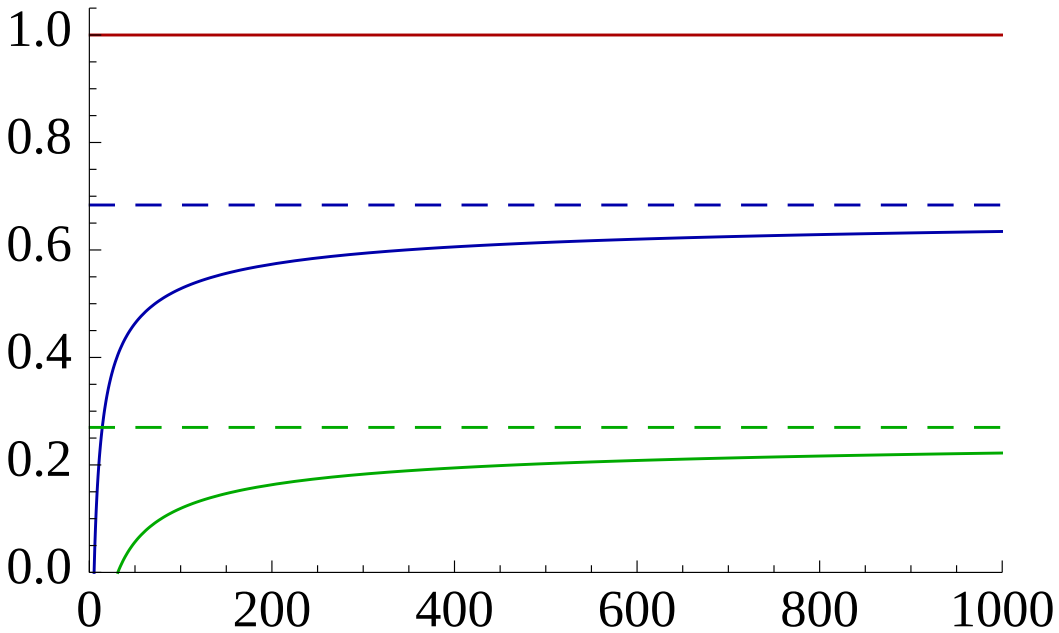
Projection of channel image onto the xz-plane of the Bloch sphere.

Example: Amplitude Damping Channel



Divergence radius, $\chi(\mathcal{E}_\gamma)$ (in bits, solid line), and peripheral information variance, $v_{\mathsf{min}}(\mathcal{E}_\gamma) = v_{\mathsf{max}}(\mathcal{E}_\gamma)$ (in bits^2, dashed line), as a function of γ.

Example: Amplitude Damping Channel



Gaussian approximation for $\frac{1}{n} \log M^* \left(\mathcal{E}_\gamma^n, \varepsilon \right)$ for $\varepsilon = 1\%$, $\gamma \in \{0, \frac{1}{4}, \frac{3}{4}\}$ (top to bottom) as a function of n. The dashed lines correspond to the asymptotic limit.

Summary

For all channels $\mathcal{W}$ and $\varepsilon \in (0,1)$, we have

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

where $C(\mathcal{W})$ and $V_\varepsilon(\mathcal{W})$ are given by simple single-letter formulas and have a geometrical interpretation.