Entropies & Information Theory

LECTURE III

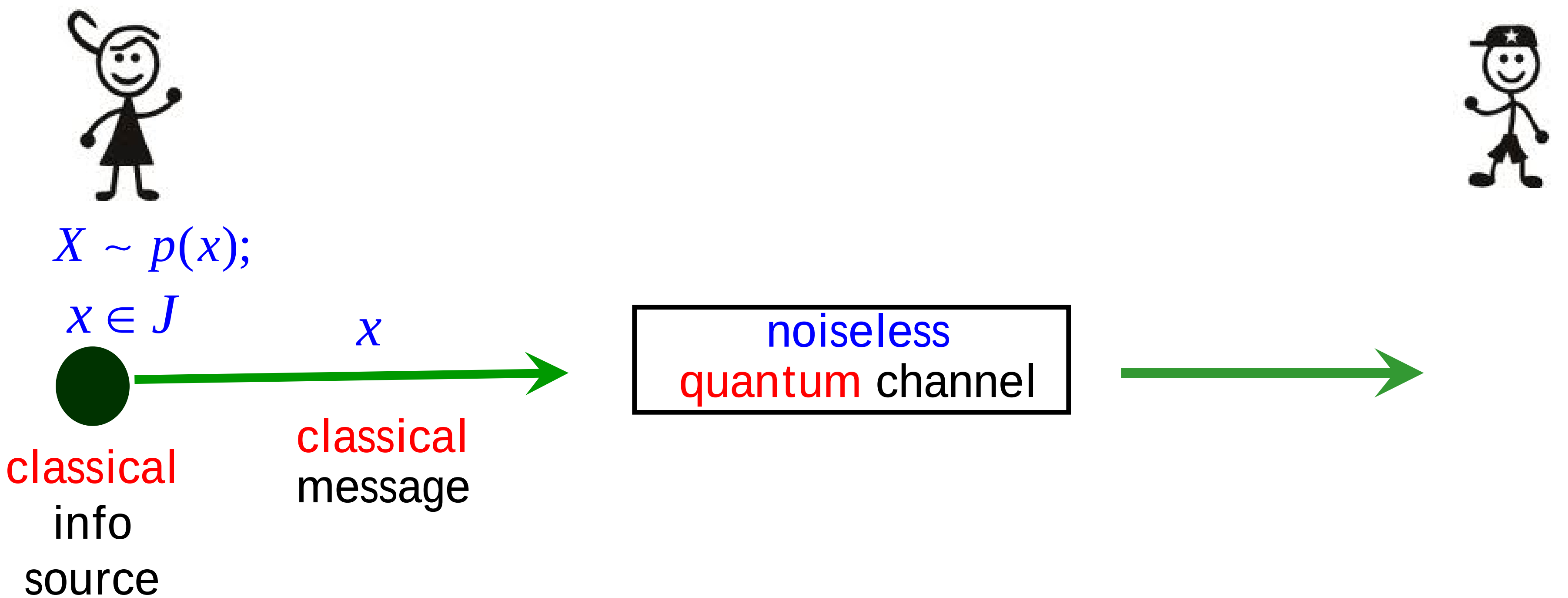
Nilanjana Datta

University of Cambridge,U.K.

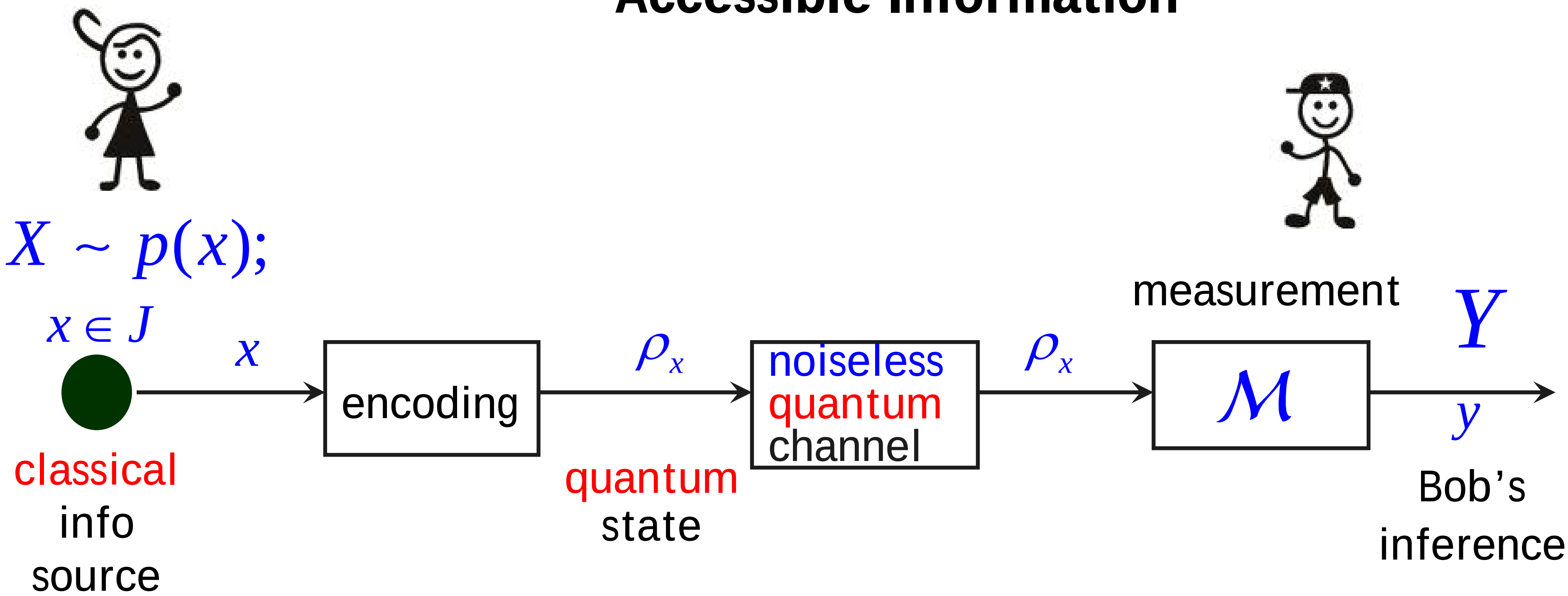
UNIVERSITY OF CAMBRIDGE

Transmission of information

Transmission of classical info through a noiseless quantum channel



Accessible Information



- Bob receives the ensemble: $\mathcal{E} = \{ p(x), \rho_x \}$

- The maximum amount of info Bob can extract

Accessible Information: $$I_{acc}(\mathcal{E}) = \max_{\mathcal{M}} I(X:Y)$$

(classical) mutual info

UNIVERSITY OF CAMBRIDGE

Holevo Bound

$$I_{acc}(\mathcal{E}) \leq \chi\left(\{p(x), \rho_x\}\right)$$

The maximum amount of info Alice can send to Bob

using the ensemble $\mathcal{E} = \{p(x), \rho_x\}$

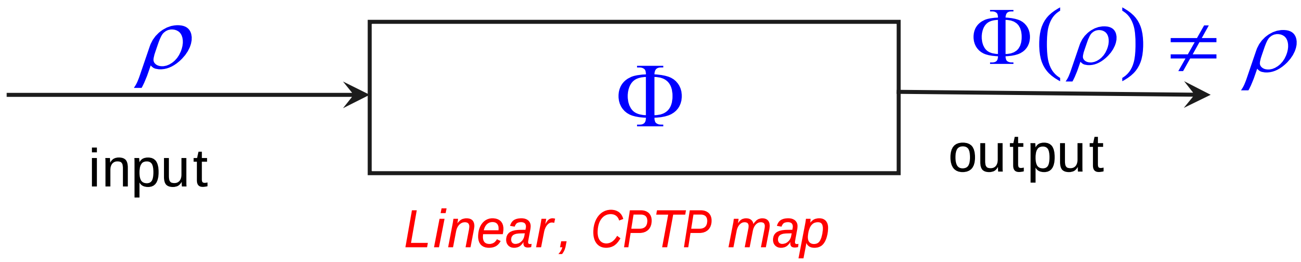
- Holevo $\chi-$quantity of the ensemble of states $\{p(x), \rho_x\}$

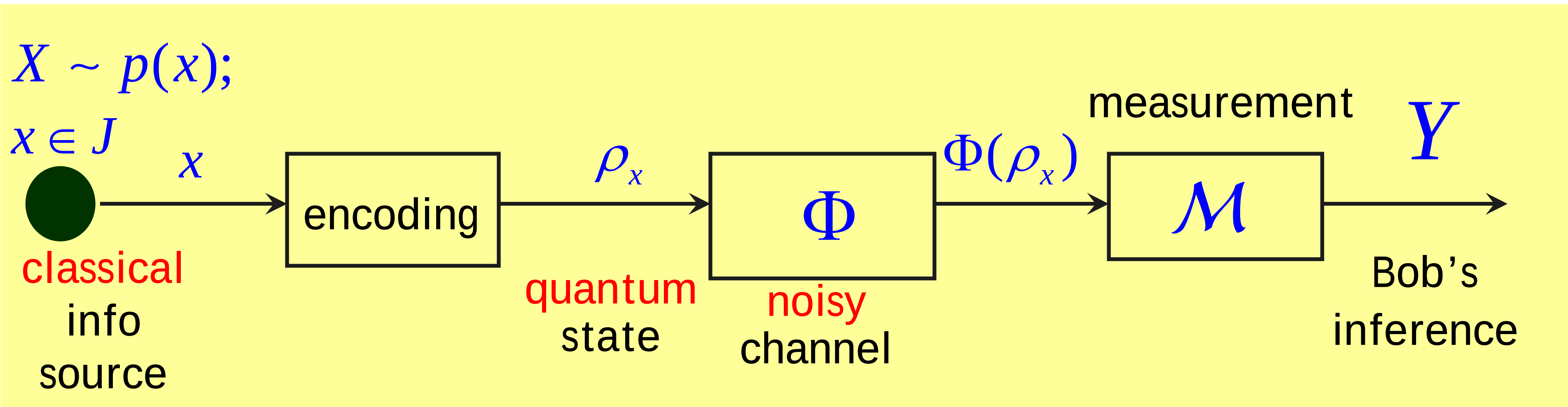
$$\chi\left(\{p(x), \rho_x\}\right) := S\left(\sum_x p(x)\rho_x\right) - \sum_x p(x) S\left(\rho_x\right)$$

If the ρ_x are pure :

$\chi\left(\{p(x), \rho_x\}\right) = S(\rho)$; where $\rho := \sum_x p(x)\rho_x$

Noisy Quantum Channels





- Bob receives the ensemble: $\mathcal{E} = \{ p(x), \Phi(\rho_x) \}$

$$I_{acc}(\mathcal{E}) \leq \chi\left(\{ p(x), \Phi(\rho_x) \} \right)$$

Capacities of a Noisy Quantum Channel

$$\xrightarrow[\text{input}]{\rho} \boxed{\;\Phi\;} \xrightarrow[\text{output}]{\Phi(\rho)}$$

- A classical channel has a unique capacity

BUT

a quantum channel has various different capacities

-- This is due to the greater flexibility in the use of a quantum channel

Memoryless quantum channel

n successive uses : $$\boxed{\Phi^{(n)} = \Phi^{\otimes n}}$$

- The different capacities depend on:
  - the nature of the transmitted information

    (classical or quantum)
  - the nature of the input states

    (entangled or product states)
  - the nature of the measurements done on the outputs

    (collective or individual)
  - the presence or absence of any additional resource

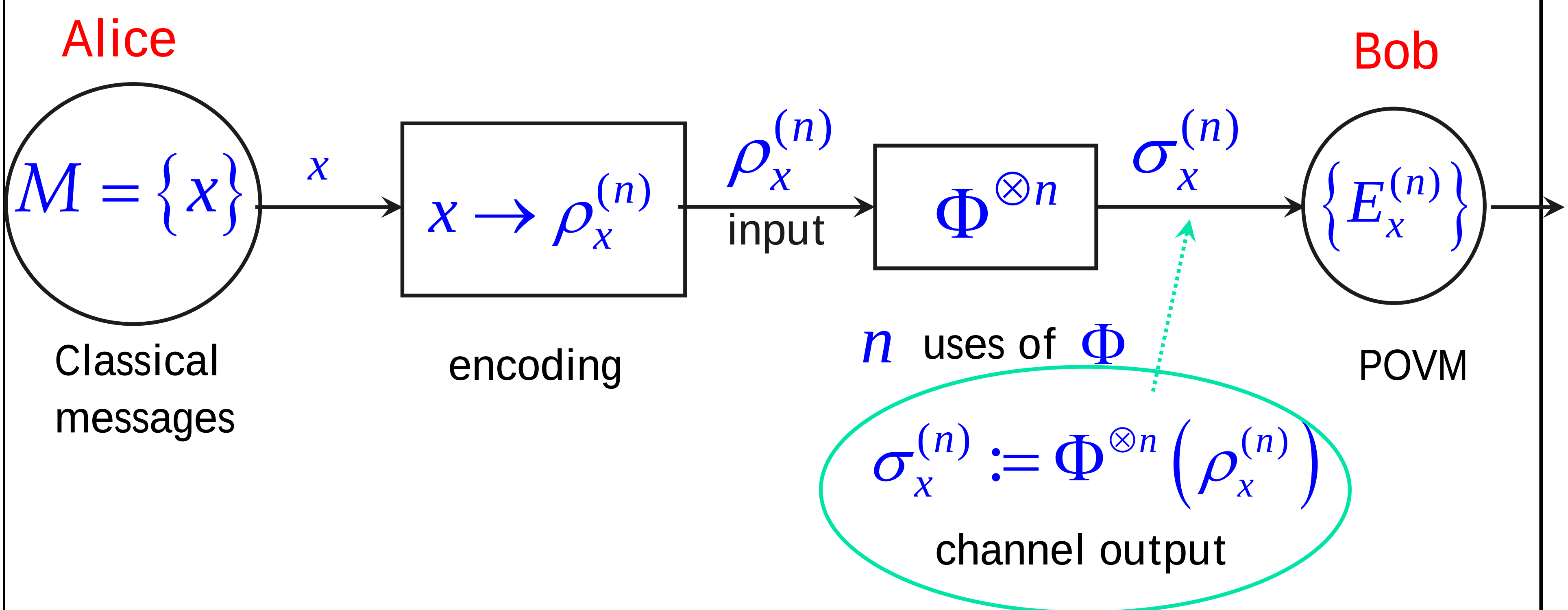
    (e.g. prior shared entanglement between Alice & Bob)
  - Etc.

- *Capacities evaluated in the "asymptotic memoryless setting"*

$$\Phi^{(n)} = \Phi^{\otimes n}; \quad n \to \infty$$

UNIVERSITY OF CAMBRIDGE

Transmission of Classical Info through a quantum channel



- Probability (Bob infers x correctly)= $\mathrm{Tr}\left(E_x^{(n)}\sigma_x^{(n)}\right)$
- Average probability of error:

$$\boxed{p_{av}^{(n)} = \frac{1}{|M|}\sum_{x\in M}\left[1-\mathrm{Tr}\left(E_x^{(n)}\sigma_x^{(n)}\right)\right]}$$

UNIVERSITY OF CAMBRIDGE

If $p_{av}^{(n)} \to 0$ *as* $n \to \infty$ ……(1) : *information transmission is reliable*

Classical capacity of the memoryless quantum channel

$C(\Phi) :=$ *maximum number of bits of classical message sent per use of the quantum channel*

- If Alice restricts her codewords to product states, i.e., if

$$x \to \rho_x^{(n)} = \rho_{x_1} \otimes \rho_{x_2} \otimes \ldots\ldots \otimes \rho_{x_n}$$

- And Bob does a collective measurement (POVM) on

$$\sigma_x^{(n)} := \Phi^{\otimes n}\left(\rho_x^{(n)}\right)$$: the output of n uses of the channel

$$= \Phi(\rho_{x_1}) \otimes \Phi(\rho_{x_2}) \otimes \ldots\ldots \otimes \Phi(\rho_{x_n})$$

Capacity **:** **product state capacity** $C_p(\Phi)$

- **Holevo-Schumacher-Westmoreland (HSW) Theorem**

$$\boxed{C_p(\Phi) = \max_{\{p_x,\rho_x\}} \chi\left(\{p_x, \Phi(\rho_x)\}\right) = \chi^*(\Phi)}$$

Holevo Capacity

UNIVERSITY OF CAMBRIDGE

HSW Theorem

$$\boxed{C_p(\Phi) = \max_{\{p_x,\rho_x\}} \chi\left(\{p_x, \Phi(\rho_x)\}\right) = \chi^*(\Phi)}$$

Holevo Capacity

$p_{av}^{(n)} \to 0$ as $n \to \infty$

$\chi^*(\Phi)$

$p_{av}^{(n)} \not\to 0$ as $n \to \infty$

R (rate)

UNIVERSITY OF CAMBRIDGE

- **Classical capacity of a memoryless channel Φ :**

(without the restriction of inputs being product states):

$$\boxed{C(\Phi) = \lim_{n\to\infty} \frac{1}{n} \chi^*\left(\Phi^{\otimes n}\right)}$$

regularised Holevo capacity

$\chi^*\left(\Phi^{\otimes n}\right)$ *Holevo Capacity* of the block $\Phi^{\otimes n}$ of n channels

(This generalization is obtained by considering inputs which are product states over blocks of n channels but which may be entangled within each block)

UNIVERSITY OF CAMBRIDGE

$$\boxed{C(\Phi) = \lim_{n\to\infty} \frac{1}{n} \chi^*\left(\Phi^{\otimes n}\right)}$$

(Q) Can the classical capacity of a memoryless quantum channel be increased by using entangled states as inputs ?

$$\boxed{\chi^*(\Phi_1 \otimes \Phi_2) \geq \chi^*(\Phi_1) + \chi^*(\Phi_2)}$$

Holevo capacity is superadditive

$$\Rightarrow \chi^*(\Phi^{\otimes n}) \geq n\chi^*(\Phi)$$

$$\Rightarrow C(\Phi) = \lim_{n\to\infty} \frac{1}{n} \chi^*\left(\Phi^{\otimes n}\right) \geq \lim_{n\to\infty} \frac{1}{\cancel{n}} \cancel{n} \chi^*(\Phi) \geq \chi^*(\Phi) = C_p(\Phi)$$

$$C(\Phi) \geq C_p(\Phi) \Rightarrow$$ *entangled inputs could help!*

UNIVERSITY OF CAMBRIDGE

(Q) Do entangled inputs really help?

$$C(\Phi) \overset{?}{>} C_p(\Phi)$$

- *This is related to :*

The (global) additivity conjecture of the Holevo capacity :

$$\forall \Phi_1, \Phi_2 \qquad \chi^*(\Phi_1 \otimes \Phi_2) = \chi^*(\Phi_1) + \chi^*(\Phi_2)$$

$$\Rightarrow \chi^*(\Phi^{\otimes n}) = n\chi^*(\Phi)$$

$$\Rightarrow \boxed{C(\Phi) = \lim_{n\to\infty} \frac{1}{n} \chi^*\left(\Phi^{\otimes n}\right) = \lim_{n\to\infty} \frac{1}{\cancel{n}} \cancel{n} \chi^*(\Phi) = \chi^*(\Phi) = C_p(\Phi)}$$

- *IF the Holevo capacity is additive then using entangled inputs would not increase its classical capacity!*

UNIVERSITY OF CAMBRIDGE

- **Additivity conjecture disproved by Matt Hastings 2008**

There exist channels in which using entangled inputs help in transmitting classical information through a quantum channel!!

UNIVERSITY OF CAMBRIDGE

- **Asymptotics to One-shot Information Theory**

UNIVERSITY OF CAMBRIDGE

In Quantum information theory, initially one evaluated:

- optimal rates of info-processing tasks, e.g.,
  - data compression,
  - transmission of information through a channel, etc.

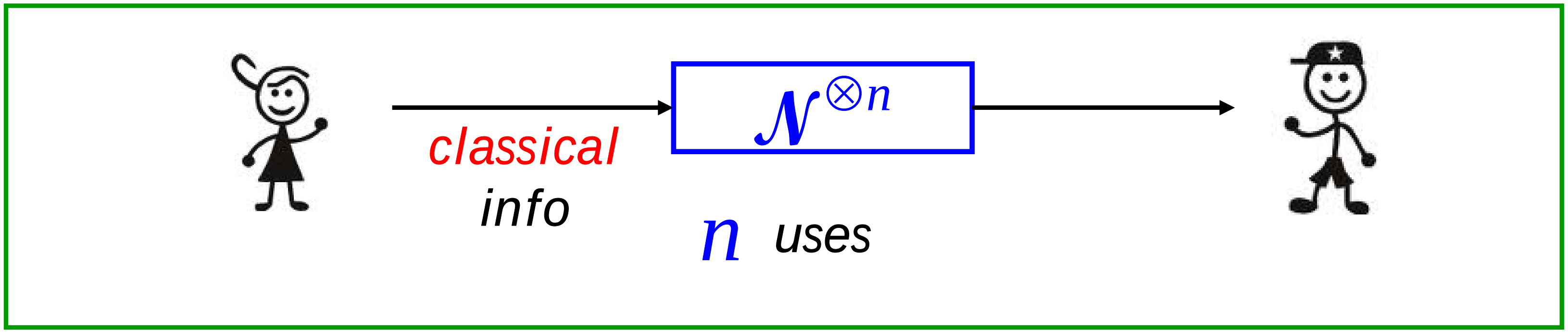
under the assumption of an *"asymptotic, memoryless setting"*

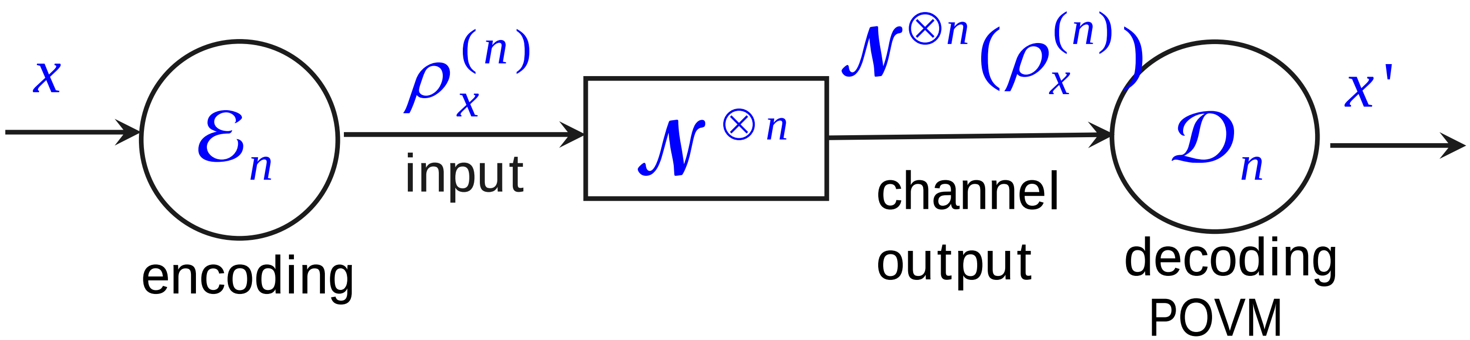
- information sources & channels were memoryless
- they were used an infinite number of times (asymptotic limit) $n \to \infty$

UNIVERSITY OF CAMBRIDGE

"asymptotic, memoryless setting"

- *To evaluate* $C(\mathcal{N})$: *classical capacity*





- One requires : **prob. of error** $p_e^{(n)} \to 0$ as $n \to \infty$

Entropic Quantities

Optimal rates of information-processing tasks in the

"asymptotic, memoryless setting"

- *Compression of Information*:

Memoryless quantum info. source $\{\rho, \mathcal{H}\}$

- Data compression limit: $S(\rho)$

- *Info Transmission thro' a memoryless quantum channel* $\mathcal{N}$

- Classical capacity $C(\mathcal{N})$

  --given in terms of the Holevo capacity ;

- Quantum capacity $Q(\mathcal{N})$

  --given in terms of the coherent information ;

These entropic quantities are all obtainable from a single parent quantity;

Quantum relative entropy: For $\rho, \sigma \geq 0; \quad \mathrm{Tr}\rho = 1$

$$\boxed{D(\rho \| \sigma) := \mathrm{Tr}\left(\rho \log \rho\right) - \mathrm{Tr}\left(\rho \log \sigma\right)}$$

e.g. Data compression limit:

$$S(\rho) := -\mathrm{Tr}\left(\rho \log \rho\right) = -D(\rho \| I) \qquad (\sigma = I)$$

$D(\rho \| \sigma)$: *acts as a parent quantity for optimal rates in the "asymptotic, memoryless setting"*

UNIVERSITY OF CAMBRIDGE

In real-world applications

"asymptotic memoryless setting" not necessarily valid

- **In practice:** information sources & channels are used a finite number of times;
- there are unavoidable correlations between successive uses *(memory effects)*

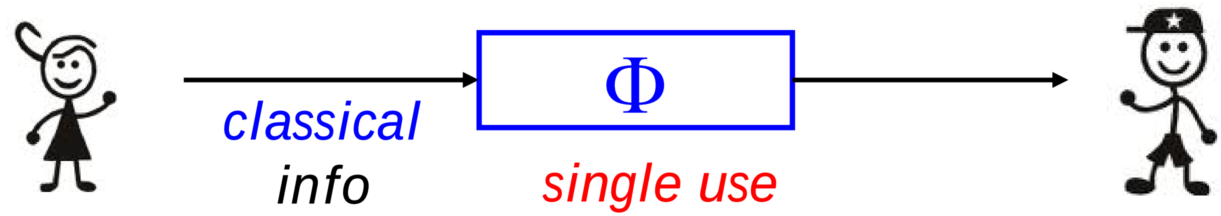
Hence it is important to evaluate optimal rates for *finite number of uses (or even a single use)* *of an arbitrary source or channel*

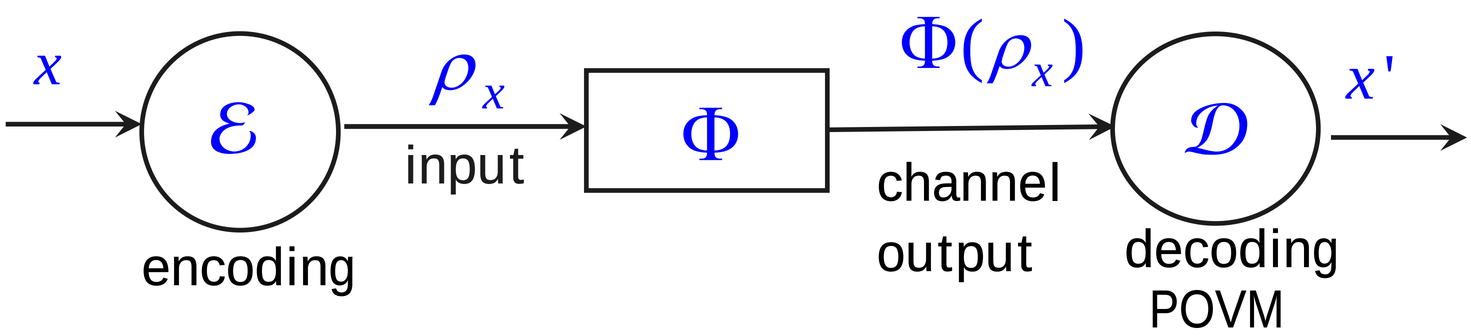
- Evaluation of corresponding optimal rates:

One-shot information theory

UNIVERSITY OF CAMBRIDGE

- **An example**: **One-shot information theory**





One-shot classical capacity := *max. number of bits that can be transmitted on a single use*

$$C_{\varepsilon}^{(1)}(\Phi)$$

Prob. of error: $p_e \leq \varepsilon$ for some $\varepsilon > 0,$

UNIVERSITY OF CAMBRIDGE

Introduce 2 generalized relative entropies

Min- & Max relative entropies: $D_{\min}(\rho \| \sigma)$, $D_{\max}(\rho \| \sigma)$

act as parent quantities for one-shot rates of protocols

just as

Quantum relative entropy: $D(\rho \| \sigma)$

acts as a parent quantity for asymptotic rates of protocols

UNIVERSITY OF CAMBRIDGE

- *Definition 1:* The max- relative entropy of a state ρ & a positive operator σ is

$$D_{\max}(\rho \| \sigma) := \inf \left\{ \gamma : \rho \leq 2^{\gamma} \sigma \right\}$$

$$\text{supp}\, \rho \subseteq \text{supp}\, \sigma$$

$$(2^{\gamma} \sigma - \rho) \geq 0$$

$$D_{\max}(\rho \| \sigma) = \log\left(\lambda_{\max}(\sigma^{-1/2} \rho \sigma^{-1/2}) \right)$$

pseudoinverse

- *Definition 2:* The min- relative entropy of a state ρ & a positive operator σ is

$$D_{\min}(\rho \| \sigma) := -\log \mathrm{Tr}\,(\pi_\rho \sigma)$$

where π_ρ denotes the projector onto the support of ρ $(\mathrm{supp}\;\rho)$

- *Remark:* The min- relative entropy

$$\boxed{D_{\min}(\rho \| \sigma) := -\log\left(\mathrm{Tr}\,(\pi_\rho \sigma)\right)}$$

is expressible in terms of: *quantum relative Renyi entropy*

$$\boxed{D_\alpha(\rho \| \sigma) := \frac{1}{\alpha - 1} \log\left(\mathrm{Tr}\,(\rho^\alpha \sigma^{1-\alpha})\right)} \qquad \alpha \neq 1$$

$$\boxed{D_{\min}(\rho \| \sigma) = \lim_{\alpha \to 0^+} D_\alpha(\rho \| \sigma) \; = D_0(\rho \| \sigma)}$$

relative Renyi entropy of order 0

UNIVERSITY OF CAMBRIDGE

$$\boxed{D_{\max}(\rho\|\sigma) \geq D_{\min}(\rho\|\sigma)}$$

- *Proof:*

$$\boxed{D_{\max}(\rho\|\sigma) := \inf\{\gamma : \rho \leq 2^{\gamma}\sigma\} \quad = \gamma_0}$$

$$\rho \leq 2^{\gamma_0}\sigma, \qquad (2^{\gamma_0}\sigma - \rho) \geq 0,$$ *Also* $\pi_\rho \geq 0$

$$\mathrm{Tr}\,[\pi_\rho(2^{\gamma_0}\sigma - \rho)] \geq 0 \qquad \because A, B \geq 0 \Rightarrow \mathrm{Tr}\,(AB) \geq 0$$

$$\boxed{2^{\gamma_0}\,\mathrm{Tr}\,[\pi_\rho\sigma] \geq \mathrm{Tr}\,[\pi_\rho\rho] = 1}$$

$$\gamma_0 + \log\,[\mathrm{Tr}(\pi_\rho\sigma)] \geq 0$$

$$\gamma_0 \geq -\log\,[\mathrm{Tr}(\pi_\rho\sigma)]$$

$$\boxed{D_{\max}(\rho\|\sigma) \geq D_{\min}(\rho\|\sigma)}$$

Why are $D_{\min}(\rho\|\sigma)$ & $D_{\max}(\rho\|\sigma)$ relative entropies?

- Like $D(\rho\|\sigma)$ we have for $* = \max,\ \min$

  $\boxed{D_*(\rho\|\sigma) \geq 0}$ for ρ,σ states

  $\boxed{D_*(\Lambda(\rho)\|\Lambda(\sigma)) \leq D_*(\rho\|\sigma)}$ for any CPTP map Λ

- Also $\boxed{D_*(\rho\|\sigma) = D_*(U\rho U^\dagger\|U\sigma U^\dagger)}$ for any unitary operator U

- Most interestingly

$$\boxed{D_{\min}(\rho\|\sigma) \leq D(\rho\|\sigma) \leq D_{\max}(\rho\|\sigma)}$$

- Also act as parent quantities for other entropies..........

$$H_{\min}(\rho) := -D_{\max}(\rho \| I)$$
$$= -\log \| \rho \|_{\infty}$$

$$H_{\max}(\rho) := -D_{\min}(\rho \| I)$$
$$= \log\ \operatorname{rank}(\rho)$$

[Renner]

Just as:

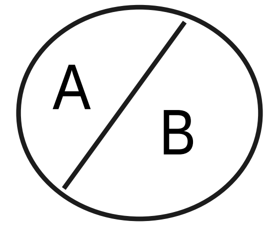
von Neumann entropy

$$S(\rho) = -D(\rho \| I)$$

$$H_{\max}(\rho) \geq H_{\min}(\rho)$$

UNIVERSITY OF CAMBRIDGE

For a bipartite state ρ_{AB} :



- *Conditional min-entropy* *[Renner]*

$$\boxed{H_{\min}(A|B)_\rho := \max_{\sigma_B}\{-D_{\max}(\rho_{AB} \,||\, I_A \otimes \sigma_B)\}}$$

just as: *Quantum conditional entropy*

$$\boxed{S(A|B) = -D(\rho_{AB} \,||\, I_A \otimes \rho_B) = \max_{\sigma_B}\{-D(\rho_{AB} \,||\, I_A \otimes \sigma_B)\}}$$

- *Max-information* *[Berta, Christandl, Renner]*

$$\boxed{I_{\max}(A:B)_\rho := \min_{\sigma_B} D_{\max}(\rho_{AB} \,||\, \rho_A \otimes \sigma_B)}$$

just as: *Quantum mutual information* *[Buscemi & ND]*

$$\boxed{I(A:B) = D(\rho_{AB} \,||\, \rho_A \otimes \rho_B) = \min_{\sigma_B} D(\rho_{AB} \,||\, \rho_A \otimes \sigma_B)}$$

Operational significance of $D_0(\rho\|\sigma)$

- *State Discrimination:* Bob receives a state

$$\rho \quad \text{or} \quad \sigma$$

- He does a measurement to infer which state it is

POVM $\Pi\;[\rho]$ & $(I-\Pi)\;[\sigma]$

- *Possible errors*

Possible errors	*inference*	*actual state*
Type I	σ	ρ
Type II	ρ	σ

hypothesis testing

- *Error probabilities*

$$\alpha = \mathrm{Tr}((I-\Pi)\rho) \qquad \textit{Type I}$$

$$\beta = \mathrm{Tr}(\Pi\sigma) \qquad \textit{Type II}$$

- *Suppose* $\Pi = \pi_\rho$ (POVM element)

Prob(Type I error)

$$\alpha = \mathrm{Tr}((I - \Pi)\rho)$$
$$= 0$$

Prob(Type II error)

$$\beta = \mathrm{Tr}(\Pi\sigma)$$
$$= \mathrm{Tr}(\pi_\rho \sigma)$$

Bob never infers the state

to be σ *when it is* ρ

BUT

$$\boxed{D_{\min}(\rho \| \sigma) := -\log \mathrm{Tr}\, \pi_\rho \sigma}$$

Hence $\beta = 2^{-D_{\min}(\rho\|\sigma)}$ *when* $\alpha = 0$

= Prob(Type II error | Type I error = zero)

- Compare with the operational significance of $D(\rho \| \sigma)$

  arises in asymptotic hypothesis testing

- Suppose Bob is given many (n) identical copies of the state

  - He receives $\rho^{\otimes n}$ or $\sigma^{\otimes n}$

- For any $\delta > 0,$ for n large enough,

  - $Prob(Type\ II\ error\ |\ Type\ I\ error < \delta)$

$$\beta_\delta^{(n)} \approx 2^{-n\, D(\rho\|\sigma)}$$

[Quantum Stein's Lemma]

- Hence, $$D_{\min}(\rho \| \sigma) \ \& \ D(\rho \| \sigma)$$

have similar interpretations in terms of *Prob(Type II error)*

$$D_{\min}(\rho \| \sigma):$$

- a single copy of the state
- *Prob(Type I error)* $= 0$

$$D(\rho \| \sigma):$$

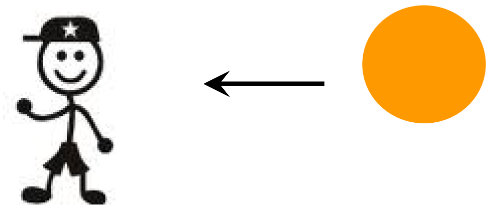
- n copies of the state
- *Prob(Type I error)*

$$\underset{n \to \infty}{\longrightarrow} 0$$

UNIVERSITY OF CAMBRIDGE

Operational interpretations of the max-relative entropy (i)

- *Multiple state discrimination problem:*



Bob

a quantum system *& told*

its state	with prob.
ρ_1	$1/M$
ρ_2	$\vdots$
$\vdots$	
ρ_M	$1/M$

- He does measurements to infer the state: **POVM**

$$\{E_1, .., E_M\}: \ 0 \le E_i \le I; \ \sum_{i=1}^{M} E_i = I$$

- *His optimal average success probability:*

$$p^*_{succ} := \max_{\{E_1,..,E_k\}} \frac{1}{M} \sum_{i=1}^{M} \mathrm{Tr}(E_i \rho_i)$$

UNIVERSITY OF CAMBRIDGE

- *Theorem 3 [M.Mosonyi & ND]:*

The optimal average success probability in this multiple state discrimination problem is given by:

$$p^*_{succ} = \frac{1}{M} \min_{\sigma} \max_{1 \le i \le M} 2^{D_{\max}(\rho_i \| \sigma)}$$

UNIVERSITY OF CAMBRIDGE

Operational interpretations of the max-relative entropy (ii)

- *Separability of a bipartite state*

[Lewenstein, Sanpera] :The state $\sigma = \sigma_{AB}$ of any bipartite system can always be written as a weighted average of a separable state ρ_s and another (possibly entangled) state ω,

$$\boxed{\sigma = \lambda\rho_s + (1-\lambda)\omega}$$

such that the weight λ is maximal.

ρ_s : Best separable approximation (BSA) of the state σ

λ : separability of the state σ *[Wellen & Kus]*

UNIVERSITY OF CAMBRIDGE

$$\sigma = \lambda\rho_s + (1-\lambda)\omega$$

- *Theorem 2 [ND,T.Rudolph]:*

  The separability of the state σ of a bipartite system is given by:

$$\lambda = \max_{\rho \in \mathcal{S}(\mathcal{H})} 2^{-D_{\max}(\rho\|\sigma)}$$

set of separable states

(I) Product-state classical capacity $C_p(\Phi)$

- Encoding restricted to product states, i.e.,

$$\mathcal{E}_n : \qquad x \rightarrow \rho_x^{(n)} = \rho_{x_1} \otimes \rho_{x_2} \otimes \ldots\ldots \otimes \rho_{x_n}$$

HSW Theorem

$$C_p(\Phi) = \chi^*(\Phi) \quad \textit{Holevo Capacity}$$

$$= \max_{\{p_x, \rho_x\}} \min_{\sigma_B} D\left(\rho_{XB} \,\|\, \rho_X \otimes \sigma_B\right)$$

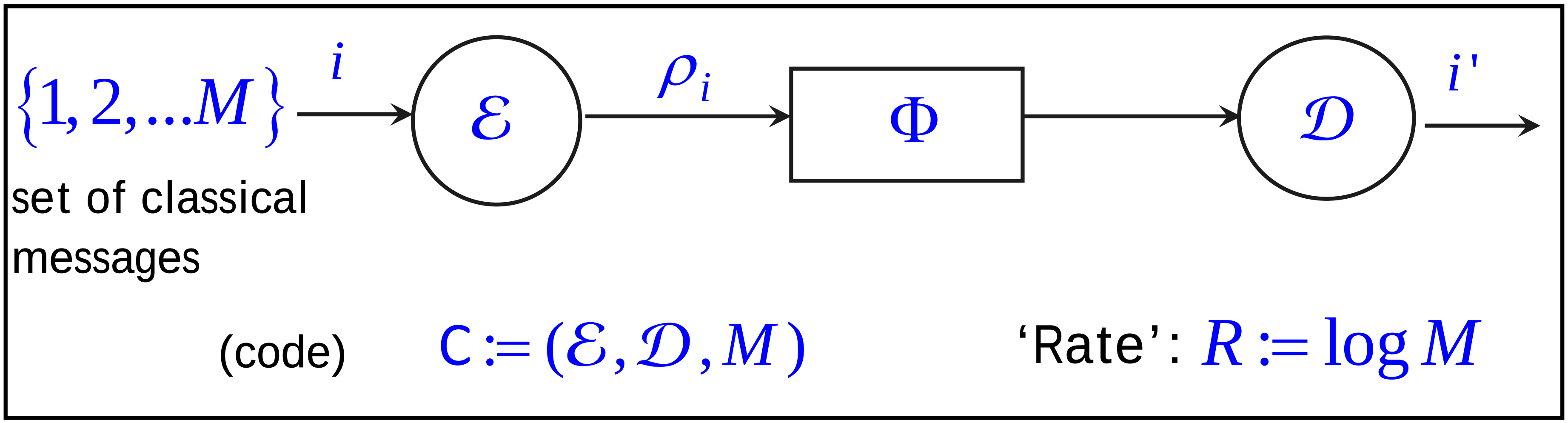
where

$$\boxed{\rho_{XB} = \sum_x p_x |x\rangle\langle x| \otimes \Phi(\rho_x);}$$

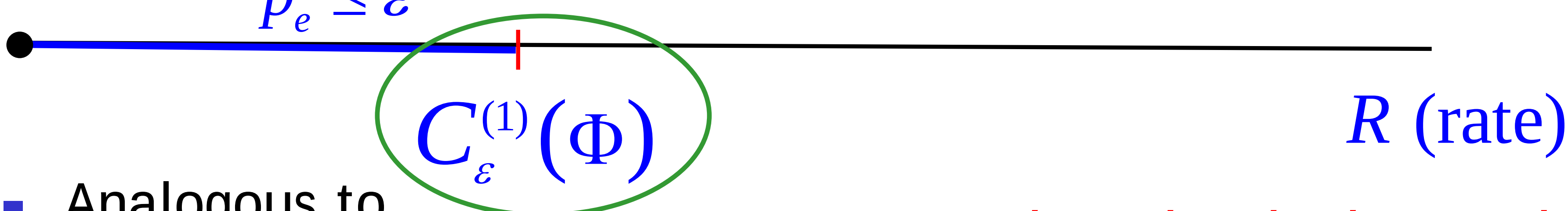
$$\rho_X = \mathrm{Tr}_B \rho_{XB};$$

UNIVERSITY OF CAMBRIDGE

One-shot classical capacity

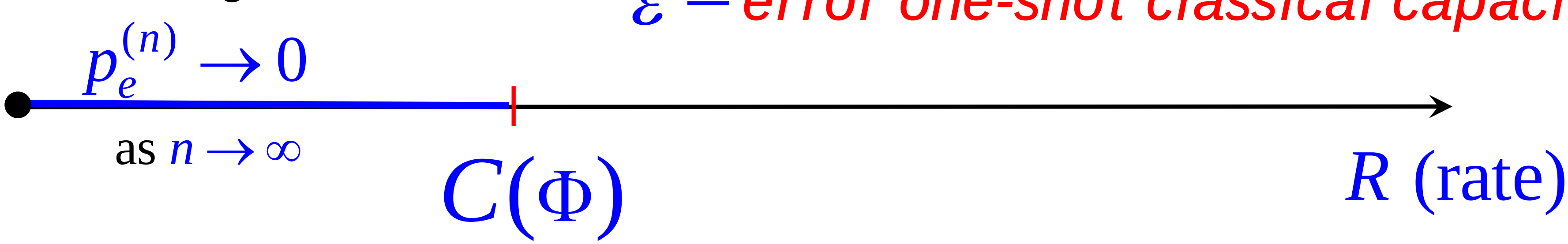


$0 < \varepsilon < 1$



- Analogous to

$\varepsilon -$ *error one-shot classical capacity*



UNIVERSITY OF CAMBRIDGE

[HSW Theorem]

$$C_p(\Phi) = \chi^*(\Phi) = \max_{\{p_x,\rho_x\}} \min_{\sigma_B} D\left(\rho_{XB} \,\|\, \rho_X \otimes \sigma_B\right)$$

Holevo-capacity

$$\rho_{XB} = \sum_x p_x |x\rangle\langle x| \otimes \Phi(\rho_x);$$

$\forall\, 0 < \varepsilon < 1$

[ND, Mosonyi, Hsieh, Brandao]

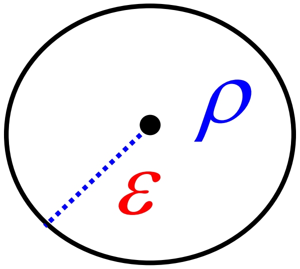
$$C_\varepsilon^{(1)}(\Phi) \approx \chi^*_{\max,\varepsilon}(\Phi) = \max_{\{p_x,\rho_x\}} \min_{\sigma_B} D^\varepsilon_{\max}\left(\rho_{XB} \,\|\, \rho_X \otimes \sigma_B\right)$$

smooth max-Holevo capacity

[See also Wang & Renner]

UNIVERSITY OF CAMBRIDGE

Smooth max-relative entropy



$$\boxed{D_{\max}^{\varepsilon}(\rho \| \sigma) := \min_{\bar{\rho} \in B^{\varepsilon}(\rho)} D_{\max}(\bar{\rho} \| \sigma)}$$

$$B^{\varepsilon}(\rho) := \left\{ \bar{\rho} \geq 0, \mathrm{Tr}\,\bar{\rho} = 1, \rho \overset{\varepsilon}{\simeq} \bar{\rho} \right\}$$

UNIVERSITY OF CAMBRIDGE

From one-shot to the asymptotic i.i.d. setting

$$\forall \varepsilon > 0, \quad \limsup_{n\to\infty} \frac{1}{n} D_{\max}^{\varepsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) \equiv D(\rho \| \sigma)$$

One-shot bounds $\longrightarrow$ *asymptotic, i.i.d. result*

(Relative entropy version of the

Quantum Asymptotic Equipartition Property

[Colbeck, Renner, Tomamichel]; [ND, Mosonyi, Hsieh, Brandao]

Why are one-shot results important?

- One-shot results yield the known results of the asymptotic case, on taking:

$$\boxed{n \to \infty \quad \text{and then} \quad \varepsilon \to 0}$$

- Hence the one-shot analysis is more general !

- One-shot results also take into account effects of correlation (or memory) in sources, channels etc.

- In fact, one-shot results can be looked upon as the *fundamental building blocks of Quantum Info. Theory*

UNIVERSITY OF CAMBRIDGE

Other occurrences of smooth max-relative entropy

- One-shot quantum state splitting *[M.Berta et al]*
- Single-shot thermodynamics *[J.Oppenheim, M.Horodecki]*

Min- and Max- relative entropies : parent quantities for

- One-shot state merging *[M.Berta et al]*
- One-shot hypothesis testing *[Wang & Renner]*
- One-shot quantum capacity *[ND, F.Buscemi; ND, M-H. Hsieh]*
- One-shot entanglement cost under LOCC *[ND, F.Buscemi]*
- One-shot entanglement-assisted classical & quantum capacities *[ND, M-H. Hsieh]* *etc.*

UNIVERSITY OF CAMBRIDGE

- <u>**Unifying the different relative entropies**</u>