

## Lecture 17

①

### Q Method of types

Suppose  $P = \sum_x p_{\mathbf{x}}(x) |x\rangle\langle x|$  is the orthogonal decomposition, and

$$P^{\otimes n} = \sum_{\mathbf{x}^n} p_{\mathbf{x}^n}(x^n) |x^n\rangle\langle x^n| \quad \left\{ \begin{array}{l} x^n = x_1 \dots x_n \\ p_{\mathbf{x}^n}(x^n) = \prod_{j=1}^n p_{\mathbf{x}}(x_j) \\ |x^n\rangle = |x_1\rangle \dots |x_n\rangle \end{array} \right.$$

Then all  $\{|x^n\rangle\}$  with the same type  $t$  have the same probability:

$$\overline{T}_t^{X^n} = \text{span} \{ |x^n\rangle : \cancel{\text{other } x^n} \quad x^n \in T_t^{X^n} \}$$

$\Pi_t^{X^n}$  = projector onto  $\overline{T}_t$ .

So we can write

$$P^{\otimes n} = \sum_t p_t \Pi_t^{X^n}, \quad p_t = p_{\mathbf{x}^n}(x^n) \text{ for any } x^n \in \overline{T}_t.$$

From the classical results about the size of type classes, we can see that for a typical type  $t$ ,

$$\text{Tr} \{ \Pi_t^{X^n} \} \geq 2^{n[H(\mathbf{x}) - \eta(d) - d \frac{1}{n} \log(n+1)]}$$

where  $\eta(q) \rightarrow 0$  as  $q \rightarrow 0$ ,  $d = \dim \mathcal{H}$ .

(2)

## Entanglement Concentration

Let  $|\Psi\rangle_{AB}$  be a bipartite pure state. In Schmidt form

$$|\Psi\rangle_{AB} = \sum_{j=1}^d \sqrt{p_j} |j\rangle_A |j\rangle_B.$$

Suppose Alice & Bob have  $n$  copies:  $|\Psi\rangle_{AB}^{\otimes n}$ . They wish to transform these into maximally entangled states.

1. Alice and Bob each do a ~~type~~ measurement:  $\{\Pi_f^{X^n}\}$  for all types  $f$ . After the result  $t$ , the state is

$$|\Psi_t\rangle_{AB} = \frac{1}{\sqrt{D_t}} \sum_{j^n \in T_f^{X^n}} |j^n\rangle_A |j^n\rangle_B$$

↑ evenly weighted!  $D_t = \text{Tr} \{\Pi_f^{X^n}\}$

2. Based on the outcome, A & B do an isometry to

$$\sum_{l=1}^{D_t} \frac{1}{\sqrt{D_t}} |l\rangle_A |l\rangle_B \leftarrow \text{a maximally entangled state of dim } D_t.$$

3. To map this state onto ebits we need the dimension to be a power of 2. Subdivide the subspace into subspaces whose dimensions are powers of 2, then do a subspace measurement.

4. If the dimension is  $2^k$ , this <sup>state</sup> can be mapped onto  $k$  ebits:  $(\Phi_+)^{\otimes k}_{AB}$ . ③

With probability  $> 1 - \epsilon$ , the type measurement will result in a typical type. Since  $D_f = \text{Tr} \{ \Pi_{+}^{\otimes n} \} \geq 2^{n[H(X) - \eta(d\delta) - d\frac{1}{n} \log(n+1)]}$ , with high probability the protocol will produce  $\approx n[H(X) - \delta']$  ebits, with  $\delta' \rightarrow 0$  as  $n$  becomes large.

### The Packing Lemma

Consider an ensemble  $\mathcal{E} = \{(p_x(x), \sigma_x)\}_{x \in \mathcal{X}}$ . Suppose  $\exists$  a codeword subspace projector  $\Pi$ , and codeword subspace projectors  $\{\Pi_x\}$ , s.t.

$$\text{Tr} \{ \Pi \sigma_x \} \geq 1 - \epsilon \quad \forall x,$$

$$\text{Tr} \{ \Pi_x \sigma_x \} \geq 1 - \epsilon \quad \forall x,$$

$$\text{Tr} \{ \Pi_x \} \leq d \quad \forall x,$$

$$\Pi \sigma \Pi \leq \frac{1}{D} \Pi, \quad \sigma = \sum_x p_x(x) \sigma_x,$$

where  $0 < d < D$ .

Choose a random code  $C = \{C_m\}_{m \in M}$  for a set of  $|M|$  messages  $\{m\}$ , where the  $C_m$  are drawn at random from the ensemble  $\mathcal{E}$ . ④

Then there exists a POVM  $\{\Lambda_m\}$  that reliably distinguishes between the states  $\{\sigma_{C_m}\}$ :

$$\mathbb{E}_C \left[ \frac{1}{|M|} \sum_{m \in M} \text{Tr} \{ \Lambda_m \sigma_{C_m} \} \right] \geq 1 - 2(\epsilon + 2\sqrt{\epsilon}) - 4 \left( \frac{D}{d|M|} \right)^{-1}$$

when  $D/d \gg 1$ ,  $|M| \ll D/d$ , and  $\epsilon$  is arbitrarily small.

The proof is given in Chapter 15. The desired POVM elements are given by a "square root" or "pretty good" measurement:

$$\Gamma_x \equiv (\Pi)(\Pi_x)(\Pi)$$

$$\Lambda_m \equiv \left( \sum_{m'=1}^{|M|} \gamma_{C_{m'}} \right)^{-1/2} \gamma_{C_m} \left( \sum_{m'=1}^{|M|} \gamma_{C_{m'}} \right)^{-1/2}.$$

There is a derandomized version of the lemma, which states that there exists some code  $C_0 = \{C_m\}$  s.t.  $\exists$  a POVM  $\{\Lambda_m\}$  where

$$\forall m \in M \quad \text{Tr} \{ \Lambda_m \sigma_{C_m} \} \geq 1 - 4(\epsilon + 2\sqrt{\epsilon}) - 8 \left( \frac{D}{d|M|} \right)^{-1}$$

## Classical Comm by Noisy Quantum Channels

(5)

We now have the pieces we need for the classical capacity of a quantum channel.

The main tools will be typical subspaces and the packing lemma, and the capacity will be defined in terms of the Holevo information.

We will start with a naive approach, that works, but does not approach capacity in general; then a more sophisticated approach that uses collective measurements; and finally, the most general approach, that also uses entangled inputs.

We will see that we can find a concrete formula for the capacity, but unlike the classical case, this is not a "single letter" formula, which makes it difficult to evaluate in general. This phenomenon shows up for many different QIT protocols. The question of which Q channels are "additive" is an open and active area of research.

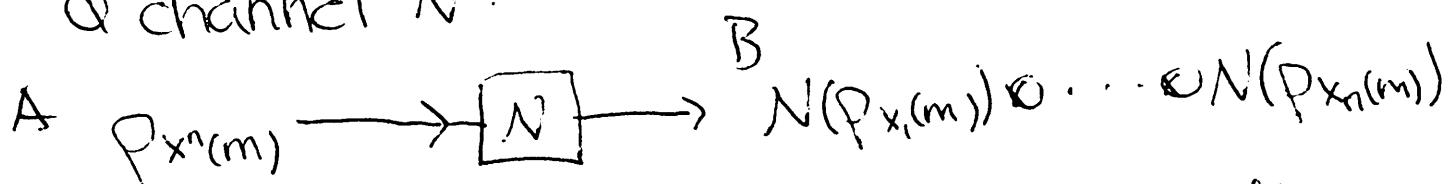
(6)

## Naive encoding and decoding

Suppose there are  $|M|$  possible messages  $m \in M$  that are equally likely. Alice and Bob agree on a codebook of strings  $\{x^n(m)\}$ , and Alice uses an alphabet of  $Q$  states  $\{p_x\}$ :

$$x^n(m) \rightarrow P_{x^n(m)} = p_{x_1(m)} \otimes p_{x_2(m)} \otimes \dots \otimes p_{x_n(m)}.$$

Alice sends the  $n$  systems through a noisy  $Q$  channel  $N$ :



Bob measures these systems individually by some POVM  $\{\Lambda_y\}$ . So in this scheme there is, effectively, a noisy classical channel

$$x \rightarrow y \quad P_{Y|X}(y|x) = \text{Tr} \{ \Lambda_y \otimes \otimes N(p_x) \}.$$

$$\Rightarrow P_{Y^n|X^n}(y^n|x^n) = \prod_{j=1}^n P_{Y|X}(y_j|x_j).$$

By the classical Shannon theorem, then, the capacity is

$$C = \max_{\{P_X(x), \{\Lambda_y\}, \{p_x\}\}} I(X; Y) \equiv I_{\text{acc}}(N)$$

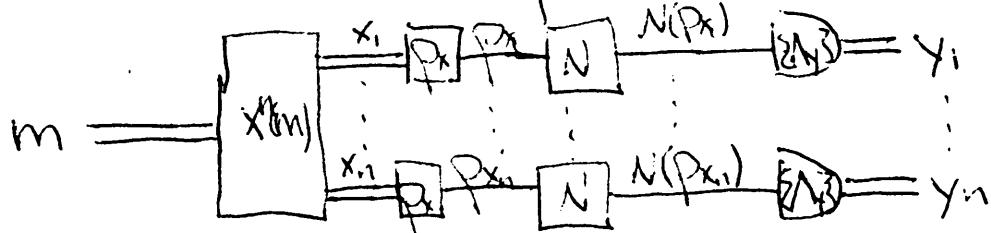
Here we have defined the accessible information of a channel  $N$  to be the maximum over all ensembles  $\{(P_X(x), N(p_x))\}$  of  $I_{\text{acc}}(S)$ .

As we showed earlier, the accessible information  $\mathcal{I}_{\text{acc}}(\mathcal{E})$  is upper-bounded by the Holevo information:

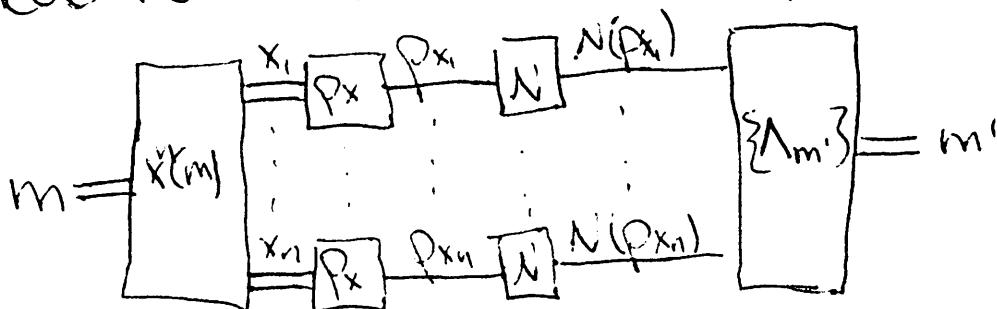
$$\mathcal{I}_{\text{acc}}(\mathcal{E}) \leq X(\mathcal{E}).$$

This means that  $C \leq \max_{\{\mathcal{P}_X(x), P_x\}} X(\mathcal{E}) = X(N)$   
 $\mathcal{E} = \{\mathcal{P}_X(x), N(P_x)\}$

From a circuit point of view the naive scheme is



It turns out that we can achieve the rate  $X(N)$  if we allow Bob to make collective measurements on the output  $N(P_{x_1}) \otimes \dots \otimes N(P_{x_n})$ .



Thm (Holevo-Schumacher-Westmoreland) — simple version  
 The supremum over all achievable rates for a Q channel  $N$  using ~~connected~~ unentangled inputs and collective measurements is  $X(N)$  in the limit of many channel uses.

## Direct coding theorem

Alice chooses an ensemble  $\{P_{\bar{x}}(x), P_x\} \quad x \in \mathcal{X}$ ,  
 and generates a random code for the messages  
 $m \in \mathbb{N} : m \rightarrow x^n(m) \rightarrow p_{x_1} \otimes \dots \otimes p_{x_n}$ , where  
 $P_{\bar{x}}(x^n) = \prod_{j=1}^n P_{\bar{x}}(x_j).$   $\sigma_{x^n}^{B^n} = N(p_{x_1}) \otimes \dots \otimes N(p_{x_n})$

If, by chance, any of the  $\{x^n(m)\}$  generated  
 is not <sup>strongly</sup> typical sequence, Alice discards that  
 string and chooses again.

We now will use the Packing Lemma. Define  
 the following projectors:

$\Pi \equiv \Pi_b^{B^n} \leftarrow$  codaspace projector = projector  
 on typical subspace

$\Pi_m, \Pi_{\text{cond}} = \Pi_s^{B^n|x^n(m)} \leftarrow$  codeword projectors = projectors  
 on conditionally typical  
 subspaces

These projectors satisfy the requirements  
 of the lemma:

$$\text{Tr}\{\sigma_{x^n}^{B^n} \Pi_s^{B^n}\} \geq 1 - \epsilon$$

$$\text{Tr}\{\sigma_{x^n}^{B^n} \Pi_s^{B^n|x^n}\} \geq 1 - \epsilon$$

$$\text{Tr}\{\Pi_s^{B^n|x^n}\} \leq 2^{n(H(B|\mathcal{X}) + CS)} = d$$

$$\Pi_s^{B^n} \sigma_{x^n}^{B^n} \Pi_s^{B^n} \leq \underbrace{(1-\epsilon)^{-n(H(B)-CS)}}_{= \frac{1}{d}} \Pi_s^{B^n}$$

Then by the derandomized version of the  
Packing Lemma, there exists a POVM  $\{\Lambda_m\}$   
such that the maximum probability of error is

$$p_e^* \equiv \max_m \text{Tr} \left\{ (I - \Lambda_m) \sigma_{X(m)}^{B^n} \right\}$$

$$\leq 4(\epsilon + 2\sqrt{\epsilon}) + \delta(1-\epsilon)^n 2^{-n(I(X;B)-2c\delta)}$$

So we choose the rate of communication to be

$$\frac{1}{n} \log_2 |M| = I(X;B) - 3c\delta$$

and  $p_e^* \leq 4(\epsilon + 2\sqrt{\epsilon}) + \delta(1-\epsilon)^n 2^{-nc\delta}$

which  $\rightarrow 0$  as  $n \rightarrow \infty$ .

This achievable rate is

$$\max_{\substack{\epsilon \\ \{p_{X(x)}, p_X\}}} I(X;B) = \max_{\{p_{X(x)}, p_X\}} X(\epsilon) = X(N).$$

□

Next time we will consider the effect  
of entangled inputs across multiple channel  
uses; regularization; and additivity of  $X(N)$ .