We will prove that quantum channel capacity is given by coherent information.

Coherent Information

$$I_c(\mathcal{N},\rho) = \cancel{\max}\ H(\mathcal{N}(\rho)) - H(I\otimes\mathcal{N}(\Phi_\rho))$$

ρ density matrix, $\mathcal{N}(\rho) = \sum_k A_k \rho A_k^\dagger$, $\sum_k A_k^\dagger A =$
is a noisy quantum channel

Quantum Channel Capacity

$$Q(\mathcal{N}) = \lim_{n\to\infty} \sup \frac{\log d}{n} \{d, n \text{ such that}$$

$\exists$ d-dimensional subspace $S \subseteq \mathcal{H}_{input}^{\otimes n}$

$$\int_{|v\rangle\in S} \langle v | \mathcal{N}(|v\rangle\langle v|) | v\rangle \, d|v\rangle > 1-\epsilon$$

This is the average fidelity criterion. Equivalent to several other fidelity criteria (see Barnum, Knill, Nielsen)

Coherent information for a channel

$$I_c(\mathcal{N}) = \max_\rho I_c(\mathcal{N}, \rho)$$
$$= \max_\rho H(\mathcal{N}(\rho)) - H(I \otimes \mathcal{N}(\Phi_\rho))$$

This is not additive. For a (qubit depolarizing) channel $\mathcal{N}$ which is nearly too noisy to carry quantum information,

$$\tfrac{1}{3} I_c(\mathcal{N}^{\otimes 3}) > I_c(\mathcal{N})$$

Easy: $I_c(\mathcal{N} \otimes \mathcal{M}) \geq I_c(\mathcal{N}) + I_c(\mathcal{M})$

Thus, to get maximum quantum channel capacity, we need to take limits

Conjecture (Schumacher, 1995)

$$\lim_{n \to \infty} \tfrac{1}{n} I_c(\mathcal{N}^{\otimes n}) = Q(\mathcal{N})$$

One direction $(\leq)$ proved by Barnum, Nielsen, Schumacher

We prove the coherent information is achievable

Proof strategy

First, do case $\rho = \mathbb{1}/d$ (maximally mixed state)

~~[illegible]~~ $\mathcal{N}(\rho) = \mathbb{1}/d_{out}$ "

Second, generalize to all ρ.

Lemma, (essentially Choi's theorem)

Let $\Phi_+ = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |e_i\rangle |e_i\rangle$

Given a channel $\mathcal{N}$, with

$$H(I \otimes \mathcal{N}(\Phi_+)) = \sum_{i=1}^{d^2} -\lambda_i \log \lambda_i$$

λ_i eigenvalues

Then, if $\rho = \frac{1}{d} \mathbb{1}$ (so Φ_+ is a purification of ρ)

There exist A_k, $\sum A_k^\dagger A_k = \mathbb{1}$

$$\mathcal{N}(\sigma) = \sum_k A_k \sigma A_k^\dagger$$

and $\operatorname{Tr} A_k \rho A_k^\dagger = \lambda_k$

$\operatorname{Tr} A_k \rho A_j^\dagger = 0, \quad j \neq k$

for $\rho = \frac{1}{d} \mathbb{1}$

Proof of lemma

Let $I \otimes \mathcal{N}(|\Phi_+\rangle\langle\Phi_+|) = \sum_k \lambda_k |v_k\rangle\langle v_k|$

so $|v_k\rangle$ are eigenvectors of $I \otimes \mathcal{N}(\cdots)$

Let $|v_k\rangle = \sum_{i,j} \alpha_{ijk} |e_i\rangle|e_j\rangle$

basis expansion of $|v_k\rangle$

Let $A_k = \sqrt{d}\sqrt{\lambda_k} \sum_{i,j} \alpha_{ijk} |e_j\rangle\langle e_i|$

Then $I \otimes A_m |\Phi_+\rangle =$

$$\frac{1}{\sqrt{d}} I \otimes A_k \sum_\ell |e_\ell\rangle|e_\ell\rangle =$$

$$\sqrt{\lambda_k} \sum_\ell |e_\ell\rangle \left(\sum_{i,j} \alpha_{ijk} |e_j\rangle\langle e_i|\right)|e_\ell\rangle$$

$$= \sqrt{\lambda_k} \sum_{j,\ell} |e_\ell\rangle \alpha_{\ell jk} |e_j\rangle = \sqrt{\lambda_k}|v_k\rangle$$

$\mathrm{Tr}\, A_k (\frac{1}{d}\mathbb{1}) A_k^\dagger$

$$= \mathrm{Tr}\, \frac{1}{d} d \lambda_k \sum_{i,j} \alpha_{ijk} |e_j\rangle\langle e_i| \sum_{i'j'k} \alpha^*_{i'j'k} |e_{i'}\rangle\langle e_{j'}|$$

$$= \lambda_k \sum_{i,j} \alpha_{ijk}\alpha^*_{ijk} = \lambda_k$$

4 and $\mathrm{Tr}\, A_k(\frac{1}{d}\mathbb{1})A_{k'}^\dagger = \sqrt{\lambda_k}\sqrt{\lambda_{k'}} \sum_{i,j} \alpha_{ijk}\alpha^*_{ijk'} = 0$

QED

We now prove, for $\rho = \frac{1}{d}\mathbb{1}$,
that we can achieve capacity

$$-\log \lambda_{max} \mathcal{N}(\rho) - I \circ \mathcal{N}(\Phi_+)$$

Here $\lambda_{max} \mathcal{N}(\rho)$ is largest eigenvalue.

This gives Theorem (coherent inf = capacity)
directly when $\rho = \frac{1}{d}\mathbb{1}$ and $\mathcal{N}(\rho) = \frac{1}{d_{out}}\mathbb{1}$

(since $-\log \lambda_{max} \frac{1}{d_{out}}\mathbb{1} = H(\frac{1}{d_{out}}\mathbb{1})$)

The coherent capacity theorem for general
ρ can be derived from this

Method of proof

1) choose random subspace
2) show there is a decoding operation
that recovers a random state
in this subspace w.h.p.

Random subspace S

Choose $|v_1\rangle, |v_2\rangle, \ldots$ randomly with $\langle v_i|v_j\rangle = \delta_{ij}$, basis vectors for random subspace S

$$\mathcal{N}^{\otimes n}(\sigma) = \sum_K A_K(\sigma) A_K^{\dagger}$$

for $A_K = A_{k_1} \otimes A_{k_2} \otimes \ldots A_{k_n}$

~~Kronsa~~ elements tensor products of individual A_{k_i}

<u>Definition</u>

A_K is <u>typical</u> if A_t appears approximately $\lambda_t n$ times in A_K, $1 \leq t \leq d^2$

Consider the set $\{A_K|v_i\rangle\}$, A_K typical, $|v_i\rangle$ basis vector of S

We show

1) There is a POVM which identifies a random element of this set w.h.p.

2) $A_K|v_i\rangle$ and $A_K|v_j\rangle$ have nearly the same length, for fixed A_K and random $|v_i\rangle, |v_j\rangle$

~~Ha~~ First, we show that we can identify $A_K|v\rangle$ w.h.p. for typical K, random i.

let $|v\rangle = \sum_i \alpha_i |v_i\rangle$ be a state in random subspace (random state)

$$A_K|v\rangle \rightarrow \sum \alpha_i A_K |v_i\rangle$$

Since we can identify A_K w.h.p., there is a measurement which ~~has high fidelity with $A_K|v\rangle$ and~~ identifies K and does not disturb $A_K|v\rangle$ much (only works if K typical, but A_K typical w.h.p.)

We thus now have a state which has high fidelity with $A_K|v\rangle$. Thus, we need only show we can restore $A_K|v\rangle$ to $|v\rangle$ with high fidelity

Might be able to use random matrix theorems for this part.
Alternatively, there is a transformation taking $A_K|v_i\rangle$ to $|v_i\rangle$ with high fidelity, since in $A_K|v_i\rangle$, i can be identified w.h.p.

Need to show can go from $A_K|v\rangle$ back to $|v\rangle$

We have $|v\rangle = \sum_i \alpha_i |v_i\rangle$, $|v_i\rangle$ are our random basis vectors of subspace.

We have square root recovery measurement $\{|m_i\rangle\langle m_i|\}$ which identifies ~~ϕ~~ i from $A_K|v_i\rangle$ with high probability

Consider recovery operator

$$R = \sum_i |v_i\rangle\langle m_i| .$$

Know $\langle m_i|A_K|v_i\rangle \geq (1-\epsilon) \big| A_K |v_i\rangle \big|$

$$R A_K |v\rangle = \sum_i \alpha_i |v_i\rangle\langle m_i|A_K|v_i\rangle + \sum_{i\neq j} \alpha_i |v_j\rangle\langle m_j|A_K|v_i\rangle$$

this term is small because α_i have random phases.

~~the~~ recovery operator thus gives state with high fidelity to

$$\sum \alpha_i |v_i\rangle\langle m_i|A_K|v_i\rangle$$
$$\approx \sum \alpha_i |v_i\rangle \big| A_K|v_i\rangle \big|$$

Need $\big|A_K|v_i\rangle\big|$ is very close for nearly all $|v_i\rangle$

At this point, we need to use Barnum, Knill, Nielsen.

They show that to correct the subspace generated by $|v_1\rangle \ldots |v_{dims}\rangle$ it is enough to show that the correction works well on two ensembles, on average (i.e., has high fidelity on these ensembles.)

The first is the ensemble $\{|v_i\rangle\}$ of basis vectors.

The second is the ensemble

$$\left\{ \frac{1}{\sqrt{dims}}\left(e^{i\theta_1}|v_1\rangle + e^{i\theta_2}|v_2\rangle + \ldots + e^{i\theta_d}|v_{dims}\rangle\right) \right\}$$

of superpositions of basis vectors with random phases.

These two conditions imply that entanglement fidelity (i.e. fidelity with maximally entangled states) is high.

From this, it follows there is a smaller subspace with high fidelity on <u>all</u> vectors.

8½

We can prove these ensembles work in our case

Gaussian random variables

We use complex Gaussian random vars.

$$g_I = g_{I_x} + i g_{I_y}$$

$$E\, g_I = 0 \,, \quad E\, g_I g_I^* = 1 \quad E\, g_I^2 g_I^{*2} = 2$$

Now, a random vector can be obtained by choosing

$$|v\rangle = \frac{\sum g_I |e_I\rangle}{\sqrt{\sum g_I g_I^*}} \approx \frac{1}{d^{n/2}} \sum g_I |e_I\rangle$$

Sum over all basis vectors $|e_I\rangle$

The rest of the proof involves manipulation of Gaussian random varibles

First, we compute $E\,|A_K|v_i\rangle|^2$

Approximating $|v\rangle \approx \frac{1}{d^{n/2}} \sum g_I |e_I\rangle$

gives a very slightly non-normalized state, so does not significantly change the result.

$$E\langle v|A_K^\dagger A_K|v\rangle =$$

$$\frac{1}{d^n} E \sum_{I,I'} \langle e_I|A_K^\dagger A_K|e_{I'}\rangle\, g_I\, g_{I'}^*$$

This expectation is 0 if $I \neq I'$, since $E\, g_I g_{I'}^* = 0$. Also $E\, g_I g_I^* = 1$.

$$= \frac{1}{d^n}\sum_{I} \langle e_I|A_K^\dagger A_K|e_I\rangle$$

$$= \frac{1}{d^n}\mathrm{Tr}\, A_K \sum_I |e_I\rangle\langle e_I| A_K^\dagger$$

$$= \mathrm{Tr}\, A_K \rho^{\otimes n} A_K^\dagger \qquad (\rho = \tfrac{1}{d}\mathbb{1}$$

$$= \prod_{i=1}^{n} \mathrm{Tr}\, A_{k_i}\rho A_{k_i}^\dagger = \prod_{i=1}^{n} \lambda_{k_i}$$

(10) $$\log \prod_{i=1}^{n} \lambda_{k_i} \approx n H(I\otimes\mathcal{N}(\Phi_+)) + \epsilon n$$

for A_K typical.

We will also need to compute

$$\mathbb{E} \langle v_i | A_K^\dagger A_L | v_i \rangle^2$$

$$\mathbb{E} \frac{1}{d^{2n}} \sum_{\substack{i_1, i_2 \\ i_3, i_4}} g_{i_1} g_{i_2}^* g_{i_3} g_{i_4}^* \langle e_{i_1} | A_K^\dagger A_L | e_{i_2} \rangle \langle e_{i_3} | A_L^\dagger A_K | e_{i_4} \rangle$$

This expectation vanishes unless
$i_1 = i_2$ and $i_3 = i_4$
or $i_1 = i_4$ and $i_2 = i_3$
(if $i_1 = i_2 = i_3 = i_4$, $\mathbb{E}\, g_{i_1} g_{i_1}^* g_{i_1} g_{i_1}^* = 2$, so we can count it in both cases)

$$= \mathbb{E} \frac{1}{d^{2n}} \sum_{i_1, i_3} g_{i_1} g_{i_1}^* g_{i_3} g_{i_3}^* \langle e_{i_1} | A_K^\dagger A_L | e_{i_1} \rangle \langle e_{i_2} | A_L^\dagger A_K | e_{i_2} \rangle$$

$$+ \mathbb{E} \frac{1}{d^{2n}} \sum_{i_1, i_2} g_{i_1} g_{i_2}^* g_{i_2} g_{i_1}^* \langle e_{i_1} | A_K^\dagger A_L | e_{i_2} \rangle \langle e_{i_2} | A_L^\dagger A_K | e_{i_1} \rangle$$

The first term is

$$\frac{1}{d^{2n}} \sum_{e_{i_1}} \langle e_{i_1} | A_K^\dagger A_L | e_{i_1} \rangle \sum_{e_{i_2}} \langle e_{i_2} | A_L^\dagger A_K | e_{i_2} \rangle$$

(1) $$= \operatorname{Tr} A_L \rho A_K^\dagger \operatorname{Tr} A_K \rho A_L^\dagger$$

$$= 0 \text{ if } L \neq K$$

[NOTE: This first term is where we really need $\rho = \frac{\mathbb{1}}{d^n}$]

The second term is

$$\sum_i \langle v_i A_K^\dagger A_L \rho^{\otimes n} A_L^\dagger A_K | v_i \rangle$$

summing over all L, we get

$$\sum_i \langle v_i | A_K^\dagger \mathcal{N}(\rho)^{\otimes n} A_K | v_i \rangle$$

summing over all typical L will give us something smaller.

If we have $|v_i\rangle \neq |v_j\rangle$, then similar computation shows

$$E\langle v_i | A_K^\dagger A_L | v_j \rangle^2 \approx \langle v_i | A_K^\dagger A_L \rho^{\otimes n} A_L^\dagger A_K | v_i \rangle$$

We have to show that choosing randomly with $\langle v_i | v_j \rangle = 0$ (i.e., random perpendicular vectors) doesn't change the value by much, as opposed to two random vectors.

Now, we can sketch the argument showing that we can identify K, i in $A_K|i\rangle$ with high probability.

Let $|u_{K,i}\rangle = \dfrac{A_K|v_i\rangle}{|A_K|v_i\rangle|}$

Use HJSWW criterion that square root measurement (pretty good measurement) works well.
[Hausladen, Jozsa, Schumacher, Westmoreland, Wootters]

HJSWW: $|u_{K,i}\rangle$ can be recovered w.h.p.

if $\displaystyle\sum_{K',i'} |\langle u_{K,i}|u_{K',i'}\rangle|^2 < \epsilon$.

$$|u_{K,i}\rangle = \frac{A_K|v_i\rangle}{|A_K|v_i\rangle|}$$

$$E\sum_{K',i'} |\langle u_{K,i}|u_{K',i'}\rangle|^2 =$$

$$E\sum_{K',i'} \frac{\langle u_{K,i}|A_{K'}|v_{i'}\rangle\langle v_{i'}|A_{K'}^{\dagger}|u_{K,i}\rangle}{\langle v_{i'}|A_{K'}^{\dagger}A_{K'}|v_{i'}\rangle}$$

$$\leq \frac{\langle u_{K,i}|\left(\sum_{K',i'} A_{K'}|v_{i'}\rangle\langle v_{i'}|A_{K'}^{\dagger}\right)|u_{K,i}\rangle}{2^{-H(I\otimes\mathcal{N}(\Phi_+))n+\epsilon n}}$$

$$\leq \dim S\ \frac{\langle u_{K,i}|\mathcal{N}(\rho)^{\otimes n}|u_{K,i}\rangle}{2^{-H(I\otimes\mathcal{N}\Phi_+)n+\epsilon n}}$$

$\dim S$ = Number of $|v_i\rangle$'s.

Want $\dim S \dfrac{\langle u_{K,i}|\mathcal{N}(\rho)^{\otimes n}|u_{K,i}\rangle}{2^{-H(I\otimes\mathcal{N}\Phi_+)n+\epsilon n}}$ small.

But $\langle u_{K,i}|\mathcal{N}(\rho)^{\otimes n}|u_{K,i}\rangle \le \lambda_{max}[\mathcal{N}(\rho)]^n$

So we need

$$\frac{\dim S \; \lambda_{max}[\mathcal{N}(\rho)]^n}{2^{-H(I\otimes\mathcal{N}\Phi_+)n+\epsilon n}} < 2^{-\epsilon n}$$

Taking logs, we have

$$\frac{1}{n}\log \dim S < -\log\lambda_{max}[\mathcal{N}(\rho)] - H(I\otimes\mathcal{N}(\Phi_+)) - \epsilon.$$

We still need to compute

$$E\langle v_i|A_K^\dagger A_K|v_i\rangle^2 - \left(E\langle v_i|A_K^\dagger A_K|v_i\rangle\right)^2$$

This will show that $|A_K|v_i\rangle|^2$ is strongly concentrated around its mean.

$$E\langle v_i | A_K^\dagger A_K | v_i \rangle^2 =$$

$$\frac{1}{d^{2n}} E \sum_{\substack{I_1 I_2 \\ I_3 I_4}} g_{I_1} g_{I_2}^* g_{I_3} g_{I_4}^* \langle e_{I_1} | A_K^\dagger A_K | e_{I_2} \rangle \langle e_{I_3} | A_K^\dagger A_K | e_{I_4} \rangle$$

For terms contributing non-zero, either $I_1 = I_2$, $I_3 = I_4$ or $I_1 = I_4$, $I_2 = I_3$

$$= \frac{1}{d^{2n}} \sum_{I_1 I_3} \langle e_{I_1} | A_K^\dagger A_K | e_{I_1} \rangle \langle e_{I_2} | A_K^\dagger A_K | e_{I_2} \rangle$$

$$+ \frac{1}{d^{2n}} \sum_{I_1 I_2} \langle e_{I_1} | A_K^\dagger A_K | e_{I_2} \rangle \langle e_{I_2} | A_K^\dagger A_K | e_{I_1} \rangle$$

The first term is $\left(E\langle v_i | A_K^\dagger A_K | v_i \rangle\right)^2$.

The second term is

$$\mathrm{Tr}\, A_K \rho^{\otimes n} A_K^\dagger A_K \rho^{\otimes n} A_K^\dagger .$$

Let $M = A_K \rho^{\otimes n} A_K^\dagger$.

$E\langle v_i | A_K^\dagger A_K | v_i \rangle = \mathrm{Tr}\, M$

(15) ~~E~~ v_i Variance " $= \mathrm{Tr}\, M^2$

But $\mathrm{Tr}\, M^2 \le (\mathrm{Tr}\, M)^2$,
equal only if M is rank 1,
which corresponds to a channel that cannot carry quantum information.

We still need to prove the theorem for general ρ

We have a capacity of
$$-\log \lambda_{max} \mathcal{N}(\rho) - H((I \otimes \mathcal{N})\phi_+)$$
for $\rho = \frac{1}{d}\mathbb{1}$

We want a capacity of
$$H(\mathcal{N}(\rho)) - H((I \otimes \mathcal{N})\phi_\rho)$$
for arbitrary ρ.

Look at typical subspace T of $\rho^{\otimes n}$.
Let π_T be density matrix projection onto T.

Consider

$$\frac{1}{n}\left[H(\mathcal{N}^{\otimes n}(\pi_T)) - H((I \otimes \mathcal{N}^{\otimes n})(\Phi_{\pi_T}))\right]$$

This has $\rho = \pi_T$, a maximally mixed input state, and goes to the desired capacity [see papers on entanglement-assisted capacity Holevo, BSST].

16 For first term, however, have
$$H(\mathcal{N}^{\otimes n}(\pi_T))$$ instead of
$$-\log \lambda_{max} \mathcal{N}^{\otimes n}(\pi_T).$$

How to get $-\log \lambda_{max}(\mathcal{N}^{\otimes n}(\Pi_T))$?

We can let do Bob perform post processing on the channel. This cannot decrease channel capacity.

If Bob projects onto the typical subspace of $\mathcal{N}(\rho)^{\otimes n}$, he gets rid of all the large eigenvalues.

We need to specify what Bob does when the projection fails, in a way that reduces $\lambda_{max}\mathcal{N}^{\otimes n}(\Pi_T)$. We can't let this increase $H(I \otimes \mathcal{N}^{\otimes n}(\Phi_{\Pi_T}))$ too much.

Answer: if projection fails, create a mixed state with eigenvalues of size $\cdot \lambda_{max}(\mathcal{N}^{\otimes n}(\Pi_T))$.