Strong, weak and pretty strong:
Converses for quantum channel capacities

Andreas Winter (ICREA & UAB Barcelona)
1st (A)QIP: Aarhus 1998
64 participants: whom do you know? (circled: present @ QIP2014)
1. Communication

Dear Bob!

Shannon (1948): Fundamental problem is that of reproducing at one point a message selected at another point.
Dear Bob!

1. Communication (noise)
Dear Bob!

Deep throat?

(noise)
1. Channels & capacity

Noise modelled as "channel":

\[ A \xrightarrow{N} B \]
1. Channels & capacity

Channel = cptp map $N : \mathcal{L}(A) \to \mathcal{L}(B)$, where $A$, $B$ are finite-dim. Hilbert spaces.

Noise modelled as "channel":
1. Channels & capacity

Channel = cptp map \( N : \mathbb{L}(A) \rightarrow \mathbb{L}(B) \), where \( A, B \) are finite-dim. Hilbert spaces. Kraus representation, you know...
1. Channels & capacity

Channel = cptp map $\mathcal{N} : \mathcal{L}(A) \rightarrow \mathcal{L}(B)$, where $A, B$ are finite-dim. Hilbert spaces. Kraus representation, you know...

Stinespring: $\mathcal{N}(\rho) = \text{Tr}_E V \rho V^\dagger$, with an isometry $V : A \rightarrow B \otimes E$. 
1. Channels & capacity

Channel = cptp map $N: \mathcal{L}(A) \rightarrow \mathcal{L}(B)$, where $A$, $B$ are finite-dim. Hilbert spaces. Kraus representation, you know...

Stinespring: $N(\rho) = \text{Tr}_E V \rho V^\dagger$

with an isometry $V: A \rightarrow B \otimes E$.

Complementary channel:

$\hat{N}(\rho) = \text{Tr}_B V \rho V^\dagger$
1. Channels & capacity

Ex: 1) Noiseless channel = identity id$_A$.
2) Constant channel $K(\rho) = \omega_0$.
3) Depolarizing channels
4) Amplitude damping channels
5) Phase damping channels
6) Erasure channel $\xi_g(\rho) = (1-g)\rho \oplus g|\psi^\ast\rangle\langle\psi^\ast|$
1. Channels & capacity

Ex: 1) Noiseless channel = identity $id_A$.
2) Constant channel $\mathcal{K}(\rho) = \rho_0$.
3) Depolarizing channels
4) Amplitude damping channels
5) Phase damping channels
6) Erasure channel $\mathcal{E}_{\rho}(\rho) = (1-\rho)\rho \oplus \rho|0\rangle\langle0|$

(Later in this talk, we’ll look at some special classes: degradable, Hadamard, entanglement-breaking, …)
Classical capacity $C(N) := \text{maximum cbit rate } \frac{k}{n}$ for asymptotically error-free transmission over $N^\otimes n$. 
Classical capacity $C(N) := \text{maximum cbit rate } \frac{k}{n}$ for asymptotically error-free transmission over $N\otimes^n$. 

$A_1 \rightarrow N \rightarrow B_1$

$A_2 \rightarrow N \rightarrow B_2$

$\vdots$

$A_n \rightarrow N \rightarrow B_n$
Classical capacity $C(N) := \text{maximum cbit rate } \frac{k}{n}$ for asymptotically error-free transmission over $N^\otimes n$.
Classical capacity $C(N) := \text{maximum cbit rate } \frac{k}{n}$ for asymptotically error-free transmission over $N \otimes n$.

$= m$ with prob $\geq 1-\varepsilon$
...C(N) is not the only capacity:
... $C(N)$ is not the only capacity:

Private capacity $P(N) := \text{maximum cbit rate}$ as before, in addition asymptotically secret: environment almost independent.
...C(N) is not the only capacity:

Private capacity P(N) := maximum cbit rate as before, in addition asymptotically secret: environment almost independent.

Quantum capacity Q(N) := maximum qubit rate for asymptotically faithful transmission.
\( C(N) \) is not the only capacity:

Private capacity \( P(N) := \) maximum cbit rate as before, in addition asymptotically secret: environment almost independent.

Quantum capacity \( Q(N) := \) maximum qubit rate for asymptotically faithful transmission.

...and a veritable "zoo" when allowing other free resources: \( E, \leftarrow, \rightarrow, \leftrightarrow, \ldots \)
Private capacity $P(N)$ := maximum cbit rate $\frac{k}{n}$ for asymptotically error-free and secret transmission over $N$. 

$k = k(n, \varepsilon)$ bits

\[ m \xrightarrow[]{} \rho_m \xrightarrow[]{} A_1 \xrightarrow[]{} V \xrightarrow[]{} B_1 \]
\[ A_2 \xrightarrow[]{} V \xrightarrow[]{} B_2 \]
\[ \vdots \]
\[ A_n \xrightarrow[]{} V \xrightarrow[]{} B_n \]

\[ D \]

\[ m \xrightarrow[]{} \hat{m} \]

error $\leq \varepsilon$

$E_1E_2...E_n$
Private capacity $P(N) := \text{maximum cbit rate } \frac{k}{n}$ for asymptotically error-free and secret transmission over $N^\otimes n$. 

$k = k(n, \varepsilon)$ bits

Error $\leq \varepsilon$
Private capacity $P(N) := \text{maximum cbit rate } \frac{k}{n}$ for asymptotically error-free and secret transmission over $N$. 

$k = k(n, \varepsilon)$ bits

Secret: $\| N^\otimes n (\rho_m) - \omega_0 \|_1 \leq \delta$
Quantum capacity $Q(N)$ requires en- and decoding by ctp maps $E, D$.
Quantum capacity $Q(N)$ requires en- and decoding by ctp maps $E, D$:

$k = k(n, \varepsilon)$ EPR pairs
(rate still $\frac{k}{n}$)
Quantum capacity $Q(N)$ requires en- and decoding by cptp maps $E, D$: 

$k = K(n, \varepsilon)$ EPR pairs (rate still $\frac{k}{n}$)

Approximates input: $P(\Phi, \sigma) \leq \varepsilon$. 
Digression on fidelity:

\[ A(\rho, \sigma) = \| \sqrt{\rho} \sqrt{\sigma} \|_1 \]

\[ = \max |\langle \psi \mid \varphi \rangle| \text{ s.t.} \]

\[ |\psi\rangle \text{ purifies } \rho, |\varphi\rangle \text{ purifies } \sigma. \]

\[ A(\rho, \sigma) := \sqrt{1 - A(\rho, \sigma)^2} \text{ is a metric on states;} \]

\[ \ldots \text{and so is } A(\rho, \sigma) := \arcsin A(\rho, \sigma). \]

Note: Both are equivalent to the trace distance \( \| \rho - \sigma \|_1 \).

Outline

1. Quantum channels and their capacities ✓

2. Entropic capacity formulas; weak converse

3. What is a strong converse?

4. Ideal channel (warm-up); simulation argument

5. Rényi divergence paradigm: classical capacity

6. Min-entropies: “pretty strong” converse

7. End credits
2. Capacity formulas and weak converse

Thm (Holevo and Schumacher/Westmoreland, 1973 and 1996/7):

\[ C(N) = \lim_{n \to \infty} \frac{1}{n} \chi(N^\otimes n), \text{ with} \]

\[ \chi(N) = \max \ I(X:B) \ \text{wrt.} \ \{p_x, \rho_x\}, \text{ and} \]

\[ \rho_{XB} = \sum_x p_x |x><x| \otimes N(\rho_x). \]
2. Capacity formulas and weak converse

Thm (Holevo and Schumacher/Westmoreland, 1973 and 1996/7):

\[ C(N) = \lim_{n \to \infty} \frac{1}{n} \chi(N \otimes^n), \text{ with} \]

\[ \chi(N) = \max \ I(X:B) \text{ wrt. } \{p_x, \rho_x\} \text{ and} \]

\[ \rho_{XB} = \sum_x p_x |x><x| \otimes N(\rho_x). \]

Holevo information \( S(\rho_B) - \sum_x p_x S(N(\rho_x)) \)

Von Neumann entropy \( S(\rho) = -\text{Tr} \rho \log \rho \)
Unfortunately,

\[ \chi(N) = \max I(X:B) \text{ wrt. } \exists \rho_x, \rho_x \in \mathcal{H} \text{ and } \]

\[ \rho_{XB} = \sum_x \rho_x \ket{x}\bra{x} \otimes N(\rho_x) \]

is not additive in general [Hastings, Nat. Phys 2009], hence \( C(N) > \chi(N) \) possible.
Unfortunately,

\[ \chi(N) = \max I(X:B) \text{ wrt. } \xi \rho_x, \rho_x \succeq 0 \text{ and } \]

\[ \rho_{XB} = \sum_x \rho_x |x><x| \otimes N(\rho_x) \]

is not additive in general [Hastings, Nat. Phys. 2009], hence \( C(N) > \chi(N) \) possible.

However, for some classes of channels it is, and we know the classical capacity \( C(N) \) as \( \chi(N) \).
Interestingly, the upper bound ("converse") was proved first.
Interestingly, the upper bound ("converse") was proved first. In fact, Holevo [Probl. Inf. Transm. (1973), and other work in 1970's] showed that transmitting $k$ bits over $n$ uses of $N$ with error $\varepsilon$,

$$k (1 - \varepsilon) \leq 1 + \chi(N^\otimes n) \leq 1 + n C(N).$$
Interestingly, the upper bound ("converse") was proved first. In fact, Holevo [Probl. Inf. Transm. (1973), and other work in 1970's] showed that transmitting $k$ bits over $n$ uses of $N$ with error $\varepsilon$,

$$k (1 - \varepsilon) \leq 1 + \chi(N^\otimes n) \leq 1 + n C(N).$$

Uses only strong subadditivity (SSA) and continuity of von Neumann entropy (Fannes inequality): $S(\rho) = -\text{Tr} \rho \log \rho$. 
Interestingly, the upper bound ("converse") was proved first. In fact, Holevo [Probl. Inf. Transm. (1973), and other work in 1970's] showed that transmitting $k$ bits over $n$ uses of $N$ with error $\varepsilon$, 

$$k (1-\varepsilon) \leq 1 + \chi(N^\otimes n) \leq 1 + n C(N).$$

$$\frac{k}{n} \lesssim (1 + \varepsilon) C(N)$$
Interestingly, the upper bound ("converse") was proved first. In fact, Holevo [Probl. Inf. Transm. (1973), and other work in 1970's] showed that transmitting $k$ bits over $n$ uses of $N$ with error $\varepsilon$, 

$$k(1-\varepsilon) \leq 1 + \chi(N^\otimes n) \leq 1 + n C(N).$$

\[ \frac{k}{n} \lesssim (1 + \varepsilon) C(N) \]

"weak converse"
Interestingly, the upper bound ("converse") was proved first. In fact, Holevo [Probl. Inf. Transm. (1973), and other work in 1970’s] showed that transmitting \( k \) bits over \( n \) uses of \( N \) with error \( \varepsilon \),

\[
k (1-\varepsilon) \leq 1 + \chi(N^\otimes n) \leq 1 + n C(N).
\]

\[
\frac{k}{n} \lesssim (1 + \varepsilon) C(N)
\]

"weak converse" ...is the implied tradeoff real?
Analogous formulas for $P(N)$ and $Q(N)$:

**Thm (Devetak and Cai/Yeung/AW, 2003):**

$$P(N) = \lim_{n \to \infty} \frac{1}{n} P^{(1)}(N \otimes^n),$$

with

$$P^{(1)}(N) = \max I(X:B) - I(X:E) \text{ wrt. } \mathbb{E}p_x, \rho_x.$$
Analogous formulas for $P(N)$ and $Q(N)$:

Thm (Devetak and Cai/Yeung/AW, 2003):

$$P(N) = \lim_{n \to \infty} \frac{1}{n} P^{(1)}(N \otimes^n), \text{ with}$$

$$P^{(1)}(N) = \max I(X:B) - I(X:E) \text{ wrt. } \mathcal{E}_p_x, \rho_x$$

Thm (Schumacher and Lloyd-Shor-Devetak, 1996-2003):

$$Q(N) = \lim_{n \to \infty} \frac{1}{n} Q^{(1)}(N \otimes^n), \text{ with}$$

$$Q^{(1)}(N) = \max I(A:B) = \max S(N(\rho)) - S(\hat{N}(\rho)) \text{ wrt. } \rho$$
Thm (Devetak and Cai/Yeung/AW, 2003):
\[ P(N) = \lim_{n \to \infty} \frac{1}{n} P^{(1)}(N^\otimes n) \]

Thm (Schumacher and Lloyd-Shor-Devetak, 1996-2003):
\[ Q(N) = \lim_{n \to \infty} \frac{1}{n} Q^{(1)}(N^\otimes n) \]
Thm (Devetak and Cai/Yeung/AW, 2003):
\[ P(N) = \lim_{n \to \infty} \frac{1}{n} P^{(1)}(N \otimes^n) \]

Thm (Schumacher and Lloyd-Shor-Devetak, 1996–2003):
\[ Q(N) = \lim_{n \to \infty} \frac{1}{n} Q^{(1)}(N \otimes^n) \]

Have analogous weak converses for \( P(N) \) and \( Q(N) \), and for much every other capacity we know how to characterize.
Thm (Devetak and Cai/Yeung/AW, 2003):
\[ P(N) = \lim_{n \to \infty} \frac{1}{n} P^{(1)}(N \otimes^n) \]

Thm (Schumacher and Lloyd-Shor-Devetak, 1996-2003):
\[ Q(N) = \lim_{n \to \infty} \frac{1}{n} Q^{(1)}(N \otimes^n) \]

Have analogous weak converses for \( P(N) \) and \( Q(N) \), and for much every other capacity we know how to characterize.

(Btw: also additivity issue with these!)
Thm (Devetak and Cai/Yeung/AW, 2003):
\[ P(N) = \lim_{n \to \infty} \frac{1}{n} P^{(1)}(N \otimes^n) \]

Thm (Schumacher and Lloyd-Shor-Devetak, 1996-2003):
\[ Q(N) = \lim_{n \to \infty} \frac{1}{n} Q^{(1)}(N \otimes^n) \]

Important to know: For all these capacities, at rates below, the error goes to zero exponentially, always!
Thm (Devetak and Cai/Yeung/AW, 2003):
\[ P(N) = \lim_{n \to \infty} \frac{1}{n} P^{(1)}(N^n) \]

Thm (Schumacher and Lloyd-Shor-Devetak, 1996-2003):
\[ Q(N) = \lim_{n \to \infty} \frac{1}{n} Q^{(1)}(N^n) \]

Important to know: For all these capacities, at rates below, the error goes to zero exponentially, always!

...so what about rates above capacity?
3. Strong converse?

The strong converse - in the sense of Wolfowitz [Ill. J. Math. 1:591 (1957)] -, is the statement that there is no rate-error trade-off. Viz., for rates $R$ above the capacity, the error converges to 1.
3. Strong converse?

The strong converse - in the sense of Wolfowitz [Ill. J. Math. 1:591 (1957)] -, is the statement that there is no rate-error trade-off. Viz., for rates $R$ above the capacity, the error converges to 1.

By contrapositive: If error < 1, then asymptotically the rate $\frac{k}{n}$ is bounded by the capacity.
Strong converse: If error < 1, then asymptotically the rate $\frac{k}{n}$ is bounded by the capacity.
Strong converse: If error < 1, then asymptotically the rate $\frac{k}{n}$ is bounded by the capacity.

Progress over the years:

- Classical channels [Shannon-Wolfowitz]
Strong converse: If error < 1, then asymptotically the rate $\frac{k}{n}$ is bounded by the capacity.

Progress over the years:

- Classical channels [Shannon-Wolfowitz]
- Classical capacity with product state inputs [Ogawa/Nagaoka; AW, IEEE-IT 45(7), 1999] - i.e., cq-channels
Strong converse: If error $< 1$, then asymptotically the rate $\frac{k}{n}$ is bounded by the capacity.

Progress over the years:

- Classical channels [Shannon-Wolfowitz]
- Classical capacity with product state inputs [Ogawa/Nagaoka; AW, IEEE-IT 45(7), 1999] - i.e., cq-channels
- Classical capacity of covariant channels [Koenig/Wehner, PRL 103:070504 (2009)]
Rate vs asymptotic error:

Definition/coding theorem (HSW)
Rate vs asymptotic error:

- **Definition/coding theorem (HSW)**
- **Weak converse: error bound**

Diagram with axes labeled:
- \( R: \text{cbit rate} \)
- \( C(N) \)
- \( \Pr\{\text{Err}\} \)
Rate vs asymptotic error:

- Definition/coding theorem (HSW)
- Weak converse: error bound
- Strong converse

\[ \Pr[\text{Err}] \]

\[ C(N) \]

\[ R: \text{ cbit rate} \]
4. Ideal channel

As a warm-up, prove strong converse for the noiseless qubit channel \( \text{id}_2 \). Note:

quantum code \( \Rightarrow \) private code \( \Rightarrow \) classical code
4. Ideal channel

As a warm-up, prove strong converse for the noiseless qubit channel $id_2$. Note:

quantum code $\Rightarrow$ private code $\Rightarrow$ classical code

Hence $Q(N) \leq P(N) \leq C(N)$ in general.

Since $Q(id_2) = P(id_2) = C(id_2) = 1$, enough to show it for the classical capacity.
Warm-up: strong converse for the noiseless qubit channel $\text{id}_2$.

Encode $M$ message into $\text{id}_2$ via states $\rho_m$ and POVM elements $D_m$ to decode:

[Nayak, Proc. 40th FOCS, pp. 369-376 (1999)]
Warm-up: strong converse for the noiseless qubit channel \( \text{id}_2 \).

Encode \( M \) message into \( \text{id}_2 \) via states \( \rho_m \) and POVM elements \( D_m \) to decode:

\[
1 - \varepsilon \leq \frac{1}{M} \sum_{m=1}^{M} \text{Tr}(\rho_m D_m)
\]

[Nayak, Proc. 40th FOCS, pp. 369-376 (1999)]
Warm-up: strong converse for the noiseless qubit channel \( \text{id}_2 \).

Encode \( M \) message into \( \text{id}_L \) via states \( \rho_m \) and POVM elements \( D_m \) to decode:

\[
1 - \varepsilon \leq \frac{1}{M} \sum_{m=1}^{M} \text{Tr}(\rho_m D_m) \leq \frac{1}{M} \sum_{m=1}^{M} \text{Tr} D_m
\]

[Nayak, Proc. 40th FOCS, pp. 369-376 (1999)]
Warm-up: strong converse for the noiseless qubit channel \( \text{id}_2 \).

Encode \( M \) message into \( \text{id}_L \) via states \( \rho_m \) and POVM elements \( D_m \) to decode:

\[
1 - \varepsilon \leq \frac{1}{M} \sum_{m=1}^{M} \text{Tr}(\rho_m D_m) \leq \frac{1}{M} \sum_{m=1}^{M} \text{Tr} D_m = \frac{L}{M}.
\]

[Nayak, Proc. 40th FOCS, pp. 369-376 (1999)]
Warm-up: strong converse for the noiseless qubit channel $\text{id}_2$.

Encode $M$ message into $\text{id}_L$ via states $\rho_m$ and POVM elements $D_m$ to decode:

$$1 - \varepsilon \leq \frac{1}{M} \sum_{m=1}^{M} \text{Tr}(\rho_m D_m) \leq \frac{1}{M} \sum_{m=1}^{M} \text{Tr} D_m = \frac{L}{M}.$$ 

For $n$ uses of the channel and rate $R > 1$:

$L = 2^n$ and $M = 2^{nR}$, so

$$\varepsilon \geq 1 - 2^{-n(R-1)}.$$  \textit{QED}

[Nayak, Proc. 40th FOCS, pp. 369-376 (1999)]
The simulation argument: If you can simulate a channel $N$ by $id_2$ at rate $K$, then $C(N) \leq K$ and for rates $R > K$, the error $\varepsilon \geq 1 - 2^{-n(R-K)}$.

In particular: If $K = C(N)$, strong converse holds.
The simulation argument: If you can simulate a channel $N$ by $id_2$ at rate $K$, then $C(N) \leq K$ and for rates $R > K$, the error $\epsilon \geq 1 - 2^{-n(R-K)}$.

In particular: If $K = C(N)$, strong converse holds. Almost only trivial cases, except:

Thm (Wilde/AW, 1308.6732): For pure loss optical channel w/ transmissivity $\eta$ and maximum mean photon number $P$, $C = g(\eta P)$, and the strong converse holds.
The simulation argument: If you can simulate a channel $N$ by id$_2$ at rate $K$, then $C(N) \leq K$ and for rates $R > K$, the error $\varepsilon \geq 1 - 2^{-n(R-K)}$.

More interesting with free resources, eg.

$C_e(N) =$ ent.-assisted classical capacity

= minimal simulation cost assisted by ent. (“Qu. Reverse Shannon Thm.”)

I.e. strong converse holds for $C_e$.

[Bennett et al., IEEE-IT 48:2637 (2002); Bennett et al. 0912.5537]
[Cf. Berta et al., IEEE-IT 59:6770 (2013) - $P(N)$ bound]
5. Rényi divergences for $C$

What can we do for $C(N)$? Nothing general it seems...
5. Rényi divergences for $C$

What can we do for $C(N)$? Nothing general it seems... However, unifying and extending the earlier results of Ogawa/Nagaoka, AW and König/Wehner:

**Thm (Wilde/AW/Yang, 1306.1586):** If $N$ is entanglement-breaking (EB) or Hadamard (H), then for any code $w$ rate $R > C(N)$, $Pr[err]^{3}$ converges to 1, exponentially fast in the number $n$ of channel uses.
There exists $t \geq \Omega((R-C(N))^2)$ s.t.

$$1-P_{\text{err}} \leq \exp(-tn).$$
There exists $t \geq \Omega \left( (R - C(N))^2 \right)$ s.t.

$$1 - P_{\text{err}} \leq \exp(-tn).$$

In other words, these channels satisfy the strong converse.
Hold on! I haven't even told you what these “EB” and “H” things are...
Hold on! I haven’t even told you what these “EB” and “H” things are...

Entanglement-breaking (EB) channels:

Complementary to these:

Hadamard channels (H)
Entanglement-breaking (EB) channels:

\[ N(A'B) = \text{separable} \]

Fact: \( N \) entanglement-breaking iff

\[
N(\rho) = \sum_i \text{Tr}(\rho M_i) \sigma_i \quad \text{s.t.} \quad \sum_i M_i = 1
\]

\[
= \sum_j \beta_j \langle \alpha_j | \rho | \alpha_j \rangle \beta_j^* \]
Entanglement-breaking (EB) channels:

\[ \phi_{A'A} \]

\[ \sigma_{AB} \text{ separable} \]
$\mathcal{N}(\rho) = \sum_i T_K(\rho M_i) \sigma_i \quad \text{s.t.} \quad \sum_i M_i = 1$

$= \sum_j |\beta_j \rangle \langle \alpha_j | \rho | \alpha_j \rangle \langle \beta_j |$

Stinespring: $V: |\phi \rangle_A \rightarrow \sum_j <\alpha_j | \phi > |\beta_j \rangle_B \langle j |_E$
\[ N(\rho) = \sum_i \mathcal{T}_i(\rho M_i) \sigma_i \quad \text{s.t.} \quad \sum_i M_i = 1 \]

\[ = \sum_j |\beta_j\rangle \langle \alpha_j| \rho \langle \alpha_j| \langle \beta_j| \]

\text{Stinespring: } \forall |\phi\rangle \in A \rightarrow \sum_j \langle \alpha_j | \phi \rangle |\beta_j\rangle \langle \beta_j| \epsilon

\[ \hat{N}(\rho) = \sum_{jk} |j\rangle \langle k| \langle \alpha_j | \rho \langle \alpha_k| \langle \beta_k| \beta_j| \]

\[ \mathcal{N}(\rho) = \sum_i \mathcal{T}_i(\rho M_i) \sigma_i \quad \text{s.t.} \quad \sum_i M_i = 1 \]

\[ = \sum_j |\beta_j\rangle \langle \alpha_j| \rho |\alpha_j\rangle \langle \beta_j| \]

**Stinespring:** \( V : |\phi\rangle \rightarrow \sum_j \langle \alpha_j| \phi \rangle |\beta_j\rangle \}_{B} |j\rangle \rangle_{\mathcal{E}} \)

\[ \hat{\mathcal{N}}(\rho) = \sum_{jk} \langle j| \langle k| \langle \alpha_j| \rho |\alpha_k\rangle < \beta_k | \beta_j | \]

\[ = \rho \mathcal{U} \mathcal{U}^\dagger \mathcal{S} \]
\[ N(\rho) = \sum_i Tr(\rho M_i) \sigma_i \quad \text{s.t.} \quad \sum_i M_i = I \]

\[ = \sum_j |\beta_j><\alpha_j| \rho |\alpha_j><\beta_j| \]

Stinespring: \( V |\phi\rangle \rightarrow \sum_j <\alpha_j| \phi \rangle |\beta_j\rangle \)

\[ \hat{N}(\rho) = \sum_{jk} |j><k| \sum_i <\alpha_i| \rho |\alpha_k\rangle <\beta_k| |\beta_j\rangle \]

\[ = U \rho U^\dagger \circ \mathcal{L} \]

isometry \( U = \sum |j><\alpha_j| : A \rightarrow \mathcal{E} \)
$\mathcal{N}(\rho) = \sum_i T_\mathcal{X}(\rho M_i) \sigma_i \quad \text{s.t.} \quad \sum_i M_i = 1$

$= \sum_j \beta_j \langle \alpha_j| \rho | \alpha_j \rangle \langle \beta_j |$

Stinespring: $V: |\phi \rangle_A \rightarrow \sum_j \langle \alpha_j| \phi \rangle | \beta_j \rangle_B | j \rangle_\mathcal{E}$

$\hat{\mathcal{N}}(\rho) = \sum_{jk} | j \rangle \langle k | \langle \alpha_j| \rho | \alpha_k \rangle \langle \beta_k | \beta_j |$

$= U \rho U^\dagger \circ S$

isometry $U = \sum_i | i \rangle \langle \alpha_i | : A \rightarrow \mathcal{E}$

Schur product
\[ \mathcal{N}(\rho) = \sum_i T_K(\rho M_i) \sigma_i \quad \text{s.t.} \quad \sum_i M_i = 1 \]
\[ = \sum_j |\beta_j><\alpha_j| \rho |\alpha_j><\beta_j| \]

Stinespring: \( V:|\phi\rangle \rightarrow \sum_j <\alpha_j|\phi > |\beta_j\rangle \}

\[ \mathcal{N}(\rho) = \sum_{j,k} |j><k| \langle \alpha_j|\rho |\alpha_k > <\beta_k|\beta_j > \]
\[ = \sum_j |j><\alpha_j| \rho |\alpha_j><\beta_j| \]

isometry \( U = \sum_j |j><\alpha_j| : A \rightarrow E \)

\[ \text{Channels of this form: Hadamard channels} \]
Examples - Entanglement-breaking channels:

1) cq-channels, i.e. classical input determines state preparation at output.
Examples - Entanglement-breaking channels:

1) cq-channels, i.e. classical input
determines state preparation at output

2) qc-channels, i.e. measurement with classical output
Examples - Entanglement-breaking channels:

1) cq-channels, i.e. classical input determines state preparation at output
2) qc-channels, i.e. measurement with classical output

Hadamard channels:

3) Phase damping channels, more generally Schur multipliers
Examples - Entanglement-breaking channels:

1) cq-channels, i.e. classical input determines state preparation at output
2) gc-channels, i.e. measurement with classical output

Hadamard channels:

3) Phase damping channels, more generally Schur multipliers
The proof is beautiful but a bit long...
The proof is beautiful but a bit long...

Departure point minimax characterization of $\chi(N)$: [Schumacher/Westmoreland, PRA 2000]

$$\chi(N) = \min_{\sigma} \max_{\rho} D(N(\rho) \parallel \sigma)$$
The proof is beautiful but a bit long...

Departure point minimax characterization of $\chi(N)$: [Schumacher/Westmoreland, PRA 2000]

$$\chi(N) = \min_{\sigma} \max_{\rho} D(N(\rho) \parallel \sigma)$$

Relative entropy:

$$D(\rho \parallel \sigma) = \text{Tr} \rho (\log \rho - \log \sigma)$$
The proof is beautiful but a bit long...

Departure point minimax characterization of $\chi(N)$: [Schumacher/Westmoreland, PRA 2000]

$$\chi(N) = \min_{\sigma} \max_{\rho} D(\rho \| \sigma)$$

Relative entropy:

$$D(\rho \| \sigma) = \text{Tr} \, \rho \left( \log \rho - \log \sigma \right)$$

Note: For EB and H channels $N$ this is additive, and so $C(N) = \chi(N)$.

[Shor, JMP 2002 (EB); King et al., quant-ph/0509126 (H)]
Relative entropy

$$\mathcal{D}(\rho \parallel \sigma) = \text{Tr} \rho (\log \rho - \log \sigma)$$

is a special case of a whole family of "generalized divergences"...

[Cf. Petz, 0909.3647; Müller-Lennert et al., 1306.3142]

(⇒ talk by Marco Tomamichel, Fri 10:20)
Relative entropy

\[ D(\rho \| \sigma) = \text{Tr} \, \rho (\log \rho - \log \sigma) \]

is a special case of a whole family of "generalized divergences"...

[Cf. Petz, 0909.3647; Müller-Lennert et al., 1306.3142]

(> talk by Marco Tomamichel, Fri 10:20)

Fundamental property is monotonicity:
For any ctp map \( N \),

\[ D(\rho \| \sigma) \geq D(N(\rho) \| N(\sigma)) \geq 0. \]
Compare code for $N^\otimes n$ with trivial channel:

$\Pr\{m = \hat{m}^3\} \geq 1 - \varepsilon$

$\Pr\{m = \hat{m}^3\} = \frac{1}{M}$

Compare code for $N^\otimes n$ with trivial channel:

\[
\begin{align*}
M & \xrightarrow{\rho_m} N^\otimes n \xrightarrow{D} \hat{m} \\
\text{msg's} & \xrightarrow{\Pr\exists m=\hat{m}^3} \geq 1-\varepsilon
\end{align*}
\]

vs

\[
\begin{align*}
M & \xrightarrow{\rho_m} \sigma \xrightarrow{D} \hat{m} \\
\text{msg's} & \xrightarrow{\Pr\exists m=\hat{m}^3} = \frac{1}{M}
\end{align*}
\]

\[
\tilde{\chi}(1-\varepsilon \| 1/M) \leq \min_{\sigma} \max_{\rho} \tilde{\chi}(N^\otimes n(\rho) \| \sigma)
\]

Compare code for $N^\otimes n$ with trivial channel:

$$\tilde{H}(1-\varepsilon \| M) \leq \min_{\sigma} \max_{\rho} \tilde{H}(N^\otimes n(\rho) \| \sigma)$$

$$= \chi_D(N^\otimes n)$$
Compare code for $N^\otimes n$ with trivial channel:

$$\tilde{\chi}(1-\varepsilon \| 1/M) \leq \min_{\sigma} \max_{\rho} \tilde{\chi}(N^\otimes n(\rho) \| 1 \sigma)$$

$$=:\chi_\tilde{\chi}(N^\otimes n)$$

(For usual relative entropy, we recover the previous weak converse.)
Compare code for $N^\otimes n$ with trivial channel:

$$\tilde{\chi}(1-\varepsilon \| M) \leq \min_{\sigma} \max_{\rho} \tilde{\chi}(N^\otimes n(\rho)\|\sigma)$$

$$=: \chi_{\tilde{\mathcal{D}}}(N^\otimes n)$$

Sandwiched $\alpha$-Rényi relative entropy ($\alpha > 1$):

$$\tilde{\mathcal{D}}_{\alpha}(\rho\|\sigma) := \frac{1}{\alpha-1} \log \text{Tr} \left( \sigma^{\frac{1-\alpha}{\alpha}} \rho \sigma^{\frac{1-\alpha}{\alpha}} \right)^\alpha$$

[Cf. Müller-Lennert et al., 1306.3142; Beigi 1306.5920; Frank/Lieb 1306.5358]
Compare code for $N^\otimes n$ with trivial channel:

$$\tilde{\chi}(1-\varepsilon \| I_1/M) \leq \min_{\sigma} \max_{\rho} \tilde{\chi}(N^\otimes n(\rho) \| I_1 \sigma)$$

$$=: \chi_{\tilde{\chi}}(N^\otimes n)$$

**Sandwiched $\alpha$-Rényi relative entropy ($\alpha > 1$):**

$$\tilde{\chi}_{\alpha}(\rho \| I_1 \sigma) := \frac{1}{\alpha - 1} \log \text{Tr} \left( \sigma^{1-\alpha} \rho \sigma^{1-\alpha} \right)^{\alpha}$$

[Cf. Müller-Lennert et al., 1306.3142;
Beigi 1306.5920; Frank/Lieb 1306.5358]

**Crucially additive:** $\tilde{\chi}_{\alpha}(N^\otimes n) = n \tilde{\chi}_{\alpha}(N)$.

(> talk by Marco Tomamichel, Fri 10:20)
\[\tilde{\mathcal{D}}(1-\varepsilon \| M) \leq \min_\sigma \max_\rho \tilde{\mathcal{D}}(N^{\otimes n} \rho \| \sigma)\]
\[= : \chi_{\tilde{\mathcal{D}}}(N^{\otimes n})\]

Sandwiched $\alpha$-Rényi relative entropy ($\alpha > 1$):

\[\tilde{\mathcal{D}}_\alpha(\rho \| \sigma) := \frac{1}{\alpha - 1} \log \text{Tr} \left( \sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha\]

[Cf. Müller-Lennert et al., 1306.3142; Beigi 1306.5920; Frank/Lieb 1306.5358]

Crucially additive:

\[\tilde{\mathcal{X}}_\alpha(N^{\otimes n}) = n \tilde{\mathcal{X}}_\alpha(N)\]

(x EB & H channels!)
\[ \tilde{\mathcal{D}}(1-\varepsilon 111/M) \leq \min_{\sigma} \max_{\rho} \tilde{\mathcal{D}}(\mathcal{N}^{\otimes n}(\rho) \parallel \sigma) \]
\[ =: \chi_{\tilde{D}}(\mathcal{N}^{\otimes n}) \]

Sandwiched $\alpha$-Rényi relative entropy ($\alpha > 1$):

\[ \tilde{\mathcal{D}}_\alpha(\rho \parallel \sigma) := \frac{1}{\alpha - 1} \log \text{Tr} \left( \sigma^{\frac{1-\alpha}{\alpha}} \rho \sigma^{\frac{1-\alpha}{\alpha}} \right)^\alpha \]

[Cf. Müller-Lennert et al., 1306.3142; Beigi 1306.5920; Frank/Lieb 1306.5358]

Crucially additive: $\tilde{\chi}_\alpha(\mathcal{N}^{\otimes n}) = n \tilde{\chi}_\alpha(\mathcal{N})$.

...and converges to $\chi(\mathcal{N})$ as $\alpha \to 1$!

(x EB & H channels!)
\[ \tilde{\chi}(1-\epsilon \| M) \leq \min_{\sigma} \max_{\rho} \tilde{\chi}(N^\otimes n(\rho) \| \sigma) \]

\[ =: \chi_{\tilde{D}}(N^\otimes n) \]

Sandwiched \( \alpha \)-Rényi relative entropy \((\alpha > 1)\):

\[ \tilde{\chi}_\alpha(\rho \| \sigma) := \frac{1}{\alpha - 1} \log \text{Tr} \left( \sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \]

[Cf. Müller-Lennert et al., 1306.3142; Beigi 1306.5920; Frank/Lieb 1306.5358]

Crucially additive: \( \tilde{\chi}_\alpha(N^\otimes n) = n \tilde{\chi}_\alpha(N) \).

...and converges to \( \chi(N) \) as \( \alpha \to 1 \)!

L.h.s.: \( \frac{\alpha}{\alpha - 1} \log(1-\epsilon) + \log M \).
For $\log M = nR$: 

$$1 - \varepsilon \leq \exp \left[ -n \frac{\alpha - 1}{\alpha} (R - \bar{\chi}_{\alpha}(N)) \right],$$

which is exponentially small for $R > \chi(N)$ and $\alpha > 1$ small enough. $\text{QED}$
For log \( M = nR \):

\[
1 - \varepsilon \leq \exp \left( -n \frac{\alpha - 1}{\alpha} (R - \bar{x}_\alpha(N)) \right),
\]

which is exponentially small for \( R > \chi(N) \) and \( \alpha > 1 \) small enough.

Combining with simulation argument and recent additivity of minimum output (Rényi) entropies [Giovannetti/García-Patrón/Cerf/Holevo, 1312.6225]: Strong converse for covariant Gaussian channels.

[Bardhan/García-Patrón/Wilde/AW, 1401.4161]
6. Min-entropies: “pretty strong” converse for $\mathcal{Q}$

**Stinespring:** $\mathcal{N}(\rho) = \text{Tr}_E V \rho V^\dagger$
with an isometry $V : A \rightarrow B \otimes E$.

**Complementary channel:**
$\hat{\mathcal{N}}(\rho) = \text{Tr}_B V \rho V^\dagger$
6. Min-entropies: “pretty strong” converse for Q

Stinespring: $N(\rho) = \text{Tr}_E V \rho V^\dagger$
with an isometry $V: A \rightarrow B \otimes E$.

Complementary channel:
$\hat{N}(\rho) = \text{Tr}_B V \rho V^\dagger$

$N$ is degradable if there exists a cptp map $D$
s.t. $\hat{N} = D \circ N$. Vice-versa: anti-degradable.
Degradability in the Church of the Larger Hilbert Space:
Degradability in the Church of the Larger Hilbert Space:

Apply degrading map (Stinespring form)
Degradability in the Church of the Larger Hilbert Space:
Degradability in the Church of the Larger Hilbert Space:

\[ \phi_{A'A'} \rightarrow A' \rightarrow \Box \rightarrow \psi_{ABE} \rightarrow E \rightarrow E' \rightarrow \Box \rightarrow \psi_{AEE'E'} \rightarrow A \rightarrow E \leftrightarrow E' \] symmetric
Examples:

1) Phase damping channel, more generally Schur multipliers and Hadamard channels
Examples:

1) Phase damping channel, more generally Schur multipliers and Hadamard channels
2) Amplitude damping channel
Examples:

1) Phase damping channel, more generally Schur multipliers and Hadamard channels

2) Amplitude damping channel

3) Symmetric channels, i.e. trivial $F$, for instance 50% erasure channel
Why we are interested in degr. channels:

$Q^{(1)}(N)$ is additive and so $Q(N) = Q^{(1)}(N)$, and the latter is a convex optimisation.

[Devetak/Shor, CMP 256:287 (2005)]
Why we are interested in degr. channels:
$Q^{(1)}(N)$ is additive and so $Q(N) = Q^{(1)}(N)$, and the latter is a convex optimisation.

[Devetak/Shor, CMP 256:287 (2005)]

(For anti-degradable $N$: $Q(N) = 0$.)
A previous result [via E. Rains, IEEE-IT 47(7):2921-2933 (2001)]: If $N$ is PPT entanglement-binding, then of course $Q(N)=0$, and strong converse holds (with error converging exponentially to 1).
A previous result [via E. Rains, IEEE-IT 47(7):2921-2933 (2001)]: If $N$ is PPT entanglement-binding, then of course $Q(N)=0$, and strong converse holds (with error converging exponentially to 1).

Note: Already for symmetric (degradable & anti-degradable) channels - for which also $Q(N)=0$ - not clear at all.
Exercise: Strong converse for noiseless qubit $id_2$, even assisted by classical communication.
Exercise: Strong converse for noiseless qubit id₂, even assisted by classical communication.

Implies: Error goes to one for rates above \( E_c(N) \), the entanglement cost of simulating the channel with free classical communication [Berta et al., IEEE-IT 59(10):6779-6795 (2013)].
Exercise: Strong converse for noiseless qubit id₂, even assisted by classical communication.

Implies: Error goes to one for rates above \( E_c(N) \), the entanglement cost of simulating the channel with free classical communication [Berta et al., IEEE-IT 59(10):6779-6795 (2013)].

Still doesn’t take care of 50% erasure channel, dephasing channels, etc!
Thm (Morgan/AW, 1301.4927): For any degradable channel $N$, all codes with rate $R > Q(N)$ have error at least $0.707$, asymptotically.
Thm (Morgan/AW, 1301.4927): For any degradable channel $N$, all codes with rate $R > \chi(N)$ have error at least 0.707, asymptotically. I.e., at $\chi(N)$, the error has a finite “jump”:

![Diagram showing the rate vs. error trade-off for degradable channels. The graph illustrates a sharp increase in error rate at $\chi(N)$.]
Thm (Morgan/AW, 1301.4927): For any degradable channel $N$, all codes with rate $R > Q(N)$ have error at least 0.707, asymptotically. I.e., at $Q(N)$, the error has a finite "jump":
Thm: For any degradable channel \( N \), codes with rate \( R > Q(N) \) have error at least 0.707, asymptotically.

Error/fidelity achieved by a single 50% erasure channel - without encoding.
Thm: For any degradable channel $N$, codes with rate $R > Q(N)$ have error at least 0.707, asymptotically.

- Error/fidelity achieved by a single 50% erasure channel - without encoding.

On the other hand: For larger error, any i.i.d. symmetric channel allows coding of $k = c\sqrt{n}$ qubits, by random codes. More?
Thm: For any degradable channel $N$, codes with rate $R > Q(N)$ have error at least 0.707, asymptotically.

Similar result for private capacity:

Thm (1301.4927): For degradable channel $N$, if decoding error and distance from perfect privacy are both below some universal threshold, then the rate is asymptotically bounded by $R(N) = Q(N)$. 
Thm: For any degradable channel $N$, codes with rate $R > Q(N)$ have error at least $0.707$, asymptotically.

Significance of symmetric channels:

Thm (1301.4927): If symmetric channels (whose quantum capacity is 0) obey a strong converse, then so do all degradable channels $N$: for error below 1, rate would be asymptotically bounded by $Q(N)$. 
Proof uses tight finite block length characterization of P and Q via (smooth) min-entropies & some tricks: symmetrization, de Finetti theorem, asymptotic equipartition property...

Proof uses tight finite block length characterization of $P$ and $Q$ via 
(sMOOTH) MIN-ENTROPIES & some tricks: symmetrization, de Finetti theorem, 
asymptotic equipartition property...


Can be viewed as a complicated version of the proof of additivity: $P(N) = Q(N) = Q^{(1)}(N)$ 
for degradable $N$... :-/

[Devetak/Shor, CMP 256:287 (2005)]
Lesson: To get more precise understanding of code performance have to abandon von Neumann entropy and embrace non-standard entropies (Rényi entropy, min-entropy, ...)

Price to pay: Each channel and each capacity requires its own approach. Many open - e.g. multi-user channels...
7. Conclusion (sort of...)

The trick with the sandwiched channel reduces the additivity of $\chi(N)$ to that of the minimum output Rényi entropy of an associated family of cp (trace non-preserving) maps. Can it be applied to other channels? Other divergences?

[Wilde/AW/Yang, 1306.1586]

Can we also get “2nd order” behaviour?

[Cf. Tomamichel/Tan, 1308.6503 for cq-channels]
7. Conclusion (sort of...)

Big open problem: “pretty strong” is pretty ugly - how to get full strong converse for Q of degradable channels!? Bottleneck are the symmetric channels, e.g. 50% erasure channel...

How to prove strong converses without additivity? Note that neither P, Q nor P(1), Q(1), X are generally additive!

(Not known for C.)
A. Proof ideas for \( C \)

A goody first: minimax characterisation of \( \chi(N) \): [Schumacher/Westmoreland, PRA 2000]

\[
\chi(N) = \min_{\sigma} \max_{\rho} \mathcal{D}(N(\rho) || \sigma)
\]

**Note**: For EB and H channels \( N \) this is additive, and so \( C(N) = \chi(N) \).

[Shor, JMP 2002 (EB); King et al., quant-ph/0509126 (H)]
A. Proof ideas for C

A goody first: minimax characterisation of $\chi(N)$: [Schumacher/Westmoreland, PRA 2000]

$$\chi(N) = \min_{\sigma} \max_{\rho} D(N(\rho) \| \sigma)$$

Relative entropy:

$$D(\rho \| \sigma) = Tr \rho (\log \rho - \log \sigma)$$
Relative entropy

$$\mathcal{D}(\rho \parallel \sigma) = Tr \rho (\log \rho - \log \sigma)$$

is a special case of a whole family of “generalised divergences”.

[Cf. Petz, 0909.3647; Müller-Lennert et al., 1306.3142]
Relative entropy

\[ D(\rho \parallel \sigma) = \text{Tr} \, \rho (\log \rho - \log \sigma) \]

is a special case of a whole family of "generalised divergences".

[Cf. Petz, 0909.3647; Müller-Lennert et al., 1306.3142]

Fundamental property is monotonicity: for any cptp map \( N \),

\[ D(\rho \parallel \sigma) \geq D(\rho \parallel N(\sigma)) \geq 0. \quad (\star) \]
Relative entropy
\[ D(\rho \parallel \sigma) = \text{Tr} \, \rho (\log \rho - \log \sigma) \]
is a special case of a whole family of “generalised divergences”.
[Cf. Petz, 0909.3647; Müller-Lennert et al., 1306.3142]

Fundamental property is monotonicity: for any cptp map \( N \),
\[ \tilde{D}(\rho \parallel \sigma) \geq \tilde{D}(N(\rho) \parallel N(\sigma)) \geq 0. \quad (\ast) \]

Notation: for binary distributions \( P=(p,1-p) \) and \( Q=(q,1-q) \), write \( \tilde{D}(P \parallel Q) = \tilde{D}(p \parallel q) \).
Assume furthermore that

\[ \tilde{\mathcal{X}}(\bigoplus_{x} p_{x} \rho_{x} \parallel \bigoplus_{x} p_{x} \sigma_{x}) = \sum_{x} p_{x} \tilde{\mathcal{X}}(\rho_{x} \parallel \sigma_{x}). \]
Assume furthermore that
\[ \mathcal{F}(\bigoplus_{i} p_{i} \rho_{i} \| \bigoplus_{i} p_{i} \sigma_{i}) = \sum_{i} p_{i} \mathcal{F}(\rho_{i} \| \sigma_{i}). \quad (+) \]

Then, for a code with \( M \) msg's, error \( \leq \varepsilon \), and
\[ \rho_{XB} = \frac{1}{M} \sum_{m} |m\rangle \langle m| \otimes N(\rho_{m}): \]
Assume furthermore that

\[ \mathcal{H}(\bigoplus_x p_x \rho_x \parallel \bigoplus_x p_x \sigma_x) = \sum_x p_x \mathcal{H}(\rho_x \parallel \sigma_x). \]  

Then, for a code with \( M \) msg's, error \( \leq \varepsilon \), and \( \rho_{XB} = \frac{1}{M} \sum_m |m><m| \otimes \mathcal{N}(\rho_m) \):

\[ \mathcal{H}(1-\varepsilon \parallel /M) \overset{(*)}{\leq} \mathcal{H}(\rho_{XB} \parallel \rho_X \otimes \sigma) \]

\[ \overset{(+)}{\leq} \frac{1}{M} \sum_m \mathcal{H}(\mathcal{N}(\rho_m) \parallel \sigma) \]

\[ \leq \max_{\rho} \mathcal{H}(\mathcal{N}(\rho) \parallel \sigma) =: \chi_{\mathcal{H},\sigma}(\mathcal{N}) \]
Assume furthermore that
\[ \sum p_x \tilde{X}(\rho_x \parallel \sigma_x) = \sum p_x \tilde{X}(\rho_x \parallel \sigma_x). \] (+) 

Then, for a code with \( M \) msg's, error \( \leq \varepsilon \), and \( \rho_{XB} = \frac{1}{M} \sum m \rho_{m} \otimes N(\rho_m) \):

\[ \tilde{X}(1-\varepsilon \parallel 1/M) \leq \tilde{X}(\rho_{XB} \parallel \rho \otimes \sigma) \]

\[ \leq \frac{1}{M} \sum m \tilde{X}(N(\rho_m) \parallel \sigma) \leq \max \tilde{X}(N(\rho) \parallel \sigma) =: \chi_{\tilde{X},\sigma}(N). \]

[Cf. Nagaoka (≈2000); Polyanskiy/Verdú (2010); Sharma/Warsi, 1205.1712.]
\( \tilde{D}(1 - \varepsilon \| M) \leq \max_{\rho} \tilde{D}(\mathcal{N}(\rho) \| \sigma) =: \chi_{\tilde{D}, \sigma} (\mathcal{N}) \)
\[ \tilde{D}(1 - \varepsilon \ l\ l\ l / M) \leq \max_{\rho} \tilde{D}(N(\rho) \parallel \sigma) =: \chi_{\tilde{D},\sigma}(N) \]

Everything depends on right choice of \( \tilde{D} \):
Everything depends on right choice of $\tilde{D}$:

Sandwiched $\alpha$-Rényi relative entropy $(\alpha > 1)$

$$\tilde{D}_\alpha (\rho \parallel \sigma) := \frac{1}{\alpha - 1} \log \text{Tr} \left( \sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha$$
\[ \tilde{D}(1-\varepsilon \| M) \leq \max_{\rho} \tilde{D}(N(\rho) \| \sigma) =: \chi_{\tilde{D},\sigma}(N) \]

Everything depends on right choice of \( \tilde{D} \):

Sandwiched \( \alpha \)-Rényi relative entropy (\( \alpha > 1 \))

\[ \tilde{D}_\alpha(\rho \| \sigma) := \frac{1}{\alpha - 1} \log \mathrm{Tr} \left( \sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \]

[Cf. Müller-Lennert et al., 1306.3142; Beigi 1306.5920; Frank/Lieb 1306.5358]

It's monotonic, has property (+) and is

\[ \leq D_\alpha(\rho \| \sigma) := \frac{1}{\alpha - 1} \log \mathrm{Tr} \rho^\alpha \sigma^{1-\alpha}, \text{ with} \]

which it coincides when states commute.
\[ \tilde{D}_\alpha(\rho \parallel \sigma) := \frac{1}{\alpha - 1} \log \text{Tr} \left( \sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right)^{\alpha} \]

\[ \tilde{D}_\alpha(1-\varepsilon \parallel M) \leq \max_{\rho} \tilde{D}_\alpha(\mathcal{N}(\rho) \parallel \sigma) =: \chi_{\alpha, \sigma}(\mathcal{N}) \]
\[ \tilde{D}_\alpha(\rho \parallel \sigma) := \frac{1}{\alpha - 1} \log \operatorname{Tr} \left( \sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right) \alpha \]

\[ \tilde{D}_\alpha(1-\varepsilon \parallel \rho) \leq \max_\rho \tilde{D}_\alpha(\mathcal{N}(\rho) \parallel \sigma) =: \chi_{\alpha, \sigma}(\mathcal{N}) \]

**Lhs:** \[ \tilde{D}_\alpha(1-\varepsilon \parallel \rho) \geq \log M + \frac{\alpha}{\alpha - 1} \log(1-\varepsilon) \]
\[ D_\alpha(\rho \parallel \sigma) := \frac{1}{\alpha - 1} \log \text{Tr} \left( \rho^{1-\alpha} \sigma^{1-\alpha} \right)^\alpha \]

\[ \tilde{D}_\alpha(1-\varepsilon \parallel 1/M) \leq \max_\rho \tilde{D}_\alpha(\mathcal{N}(\rho) \parallel \sigma) =: \chi_{\alpha,\sigma}(\mathcal{N}) \]

\[ \text{Lhs: } \tilde{D}_\alpha(1-\varepsilon \parallel 1/M) \geq \log M + \frac{\alpha}{\alpha - 1} \log(1-\varepsilon) \]

\textbf{Crucial: } \chi_{\alpha,\sigma}(\mathcal{N}) \text{ is the minimum } \alpha \text{-Rényi output entropy of a perturbed cp map } \mathcal{N}', \]

\[ \mathcal{N}'(\rho) = \sigma^{2\alpha} \mathcal{N}(\rho) \sigma^{2\alpha}. \]
Have:

\[ \log(1 - \varepsilon) \leq \left(1 - \frac{1}{\alpha}\right) \left( x_{\alpha, \sigma}(N) - \log M \right) \]
Have:

\[
\log(1-\varepsilon) \leq (1- \frac{1}{\alpha}) \left( x_{\alpha, \sigma}(N) - \log M \right)
\]

Now apply this to \( N^\otimes n \), \( 0^\otimes n \), and \( M=2^{nR} \).
Have:
\[
\log(1 - \varepsilon) \leq (1 - \frac{1}{\alpha}) \left( \chi_{\alpha, 0}(N) - \log M \right)
\]

Now apply this to \( N^n, 0^n, \) and \( M = 2^{nR} \).

Key observation: Sandwiched channel is \( (N')^n \), and \( N' \) is EB if \( N \) is.
\[
\log(1 - \varepsilon) \leq \left(1 - \frac{1}{\alpha}\right) \left( \chi_{\alpha, \sigma}(N) - \log M \right)
\]

Now apply this to \(N^n\), \(\sigma^n\), and \(M = 2^{nR}\).

Key observation: Sandwiched channel is \((N^n) \otimes n\), and \(N^n\) is EB if \(N\) is.

\[\Rightarrow \text{Additivity, } \chi_{\alpha, \sigma}(N^n) = n \chi_{\alpha, \sigma}(N).\]

(Because of identity with min output entropy of \(N^n\))

Get, for \( n \) uses of \( N \) at rate \( R \):

\[
\log(1 - \varepsilon) \leq n \left( 1 - \frac{1}{\alpha} \right) (\chi_{\alpha, \sigma}(N) - R).
\]
Get, for $n$ uses of $N$ at rate $R$:

$$\log(1-\varepsilon) \leq n \left(1- \frac{1}{\alpha}\right) \left(\chi_{\alpha,\sigma}(N) - R\right). \quad (\&)$$

To complete the proof, need only to observe convergence of $\chi_{\alpha,\sigma}(N)$ to $\chi(N)$;
Get, for $n$ uses of $N$ at rate $R$:
\[
\log(1 - \varepsilon) \leq n (1 - \frac{1}{\alpha}) (\chi_{\alpha, 0}(N) - R).
\]  \(\&\)

To complete the proof, need only to observe convergence of $\chi_{\alpha, 0}(N)$ to $\chi(N)$; hence can make r.h.s. of \(\&\) $\leq -nt$, $t > 0$, by choosing $\alpha > 1$ close enough to 1.
Get, for $n$ uses of $N$ at rate $R$:

$$\log(1-\varepsilon) \leq n \left(1-\frac{1}{\alpha}\right) \left(\chi_{\alpha,0}(N) - R\right). \quad (\&)$$

To complete the proof, need only to observe convergence of $\chi_{\alpha,0}(N)$ to $\chi(N)$; hence can make r.h.s. of $(\&)$ $\leq -nt$, $t > 0$, by choosing $\alpha > 1$ close enough to 1.

Takes care of EB channels; $H$ similar but requires another small trick (…)
Get, for \( n \) uses of \( N \) at rate \( R \):
\[
\log(1 - \varepsilon) \leq n \left(1 - \frac{1}{\alpha}\right) \left( \chi_{\alpha,\sigma}(N) - R \right). \tag{\&}
\]

To complete the proof, need only to observe convergence of \( \chi_{\alpha,\sigma}(N) \) to \( \chi(N) \); hence can make r.h.s. of (\&) \( \leq -nt, \ t > 0 \), by choosing \( \alpha > 1 \) close enough to 1.

Takes care of EB channels; \( H \) similar but requires another small trick (\ldots) \quad \text{QED}
B. Proof ideas for $Q \& P$

(smooth) min-entropies, symmetrisation, de Finetti theorem, AEP
Ideas: (smooth) min-entropies, symmetrisation, de Finetti theorem, AEP

1) Use code - for simplicity subspace - with maximally entangled state $\Phi$ of $k$ qubits:
Maximally entangled state $\Phi$ of $k$ qubits:
Maximally entangled state $\Phi$ of $k$ qubits:

$$k \leq H_{\min}^\epsilon (A|E)$$
Maximally entangled state $\Phi$ of $k$ qubits:

$$k \leq H^\epsilon_{\min}(A|E) = -H^\epsilon_{\max}(A|E'|F)$$
Maximally entangled state $\Phi$ of $k$ qubits:

$$k \leq H_{\min}^\epsilon(A|E) = -H_{\max}^\epsilon(A|E|F)$$

\[ k \leq H_{\min}^{\varepsilon}(AIE) = -H_{\max}^{\varepsilon}(AIE'F) \]
\[ k \leq H^\varepsilon_{\min}(AIE) = -H^\varepsilon_{\max}(AIEF) \]

[Cf. also Buscemi/Datta, IEEE-IT 56(3), 2010; Datta/Hsieh, 1103.1135]
$k \leq H_{\min}^\epsilon(AIE)$

$= -H_{\max}^\epsilon(AIE^c F)$

Note: If we knew that for \( n \) channel uses, the maximum min-entropy is attained on a tensor product input, we'd be done by AEP (= asymptotic equipartition property)...
\[ k \leq H_{\min}^{\epsilon}(AIE) \]
\[ = -H_{\max}^{\epsilon}(AIE^{'F}) \]
\[ \leq H_{\max}^{\lambda}(FIE^{'}) - H_{\max}^{\delta}(AFI E^{'}) + O(1) \]
\[
-k \leq H_{\min}^\epsilon(AIE) \\
= -H_{\max}^\epsilon(AIE'F) \\
\leq H_{\max}^\lambda(FIE') - H_{\max}^\delta(AFIE') + O(1)
\]

*Chain rule, \( \delta = \epsilon + 3 \lambda \).*
\[ k \leq H_{\text{min}}^{\epsilon}(AIE) \]
\[ = -H_{\text{max}}^{\epsilon}(AIE'F) \]
\[ \leq H_{\text{max}}^{\lambda}(FIE') - H_{\text{max}}^{\delta}(AFIE') + O(1) \]

Chain rule, \( \delta = \epsilon + 3\lambda \).
\[ \leq H_{\text{max}}^{\lambda}(FIE') + O(1) \]
\[ k \leq H_{\min}^\epsilon(A1E) \]
\[ = -H_{\max}^\epsilon(A1E'F) \]
\[ \leq H_{\max}^\lambda(FIE') - H_{\max}^\delta(AFIE') + O(1) \]

**Chain rule,** \( \delta = \epsilon + 3 \lambda. \)

\[ \leq H_{\max}^\lambda(FIE') + O(1) \]

...if \( \delta < 0.707, \) by inequality \( H_{\min} \) vs. \( H_{\max}, \) and using symmetry between \( E \) and \( E'. \)
2) For $n$ channel uses, have restricted concavity of $H^\lambda_{\text{max}}$:

$$k \leq H^\lambda_{\text{max}}(F^nIE^n) + O(1)$$
2) For $n$ channel uses, have restricted concavity of $H^\lambda_{\max}$:

For $n$ channel uses, have restricted concavity of $H^\lambda_{\max}$:

\[
k \leq H^\lambda_{\max}(F^n | E^n) + O(1)
\]

\[
\leq H_{\max}^\lambda(F^n | E^n)_{\rho_{(n)}^{(n)}} + O(1)
\]
2) For $n$ channel uses, have restricted concavity of $H_{\text{max}}^\lambda$:

$$k \leq H_{\text{max}}^\lambda(F^n | E^n) + O(1)$$

$$\leq H_{\text{max}}^{\lambda'}(F^n | E^n)_{\rho_A^{(n)}} + O(1)$$

W.r.t. a permutation symmetric input state and $\lambda' = \lambda / \sqrt{2}$
2) For $n$ channel uses, have restricted concavity of $H^\lambda_{\text{max}}$:

\[
    k \leq H^\lambda_{\text{max}}(F^n | E^n) + O(1)
\]

\[
    \leq H^\lambda_{\text{max}}(F^n | E^n)_{\rho_A^{(n)}} + O(1)
\]

3) By de Finetti theorem


\[
    k \leq \max_{\rho_A} H^\lambda_{\text{max}}(F^n | E^n)_{\rho_A^\otimes n} + o(n)
\]
4) By AEP (asymptotic equipartition property) [M. Tomamichel, arXiv:1203.2142]:

\[ k \leq \max_{\rho_A} \max_{\lambda'} H^{\lambda'} (F^n | E^n)_{\rho \otimes n} + o(n) \]
4) By AEP (asymptotic equipartition property) [M. Tomamichel, arXiv:1203.2142]:

\[ k \leq \max_{\rho_A} \chi^\prime \left( F^n | E^n \right) \rho \otimes n + o(n) \]

\[ = \max_{\rho_A} n S(F | E) \rho + o(n) \]
4) By AEP (asymptotic equipartition property) [M. Tomamichel, arXiv:1203.2142]:

\[
k \leq \max_{\rho_A} \max_{\lambda''} H^{\lambda''}_{\rho_A} (F^n E^n) \rho_A \otimes n + o(n)
= \max_{\rho_A} \max_n S(F^n E^n) \rho_A + o(n)
= n Q^{(1)}(N) + o(n)
\]
4) By AEP (asymptotic equipartition property) [M. Tomamichel, arXiv:1203.2142]:

\[ k \leq \max \max_{\rho_A} \lambda''(F^{n}IE^{n})_{\rho} \otimes n + o(n) \]

\[ = \max n \rho_A S(FIE)_{\rho} + o(n) \]

\[ = n Q^{(1)}(N) + o(n) \]

(by the degradability argument)
4) By AEP (asymptotic equipartition property) [M. Tomamichel, arXiv:1203.2142]:

\[ k \leq \max_{\rho_A} \max_{\lambda'} H^\lambda_{\max} (F^n I E^n)_{\rho_A \otimes n} + o(n) \]

\[ = \max_{\rho_A} n S(F I E^n)_{\rho_A} + o(n) \]

\[ = n Q^{(1)}(N) + o(n) \]  

(by the degradability argument) 

QED