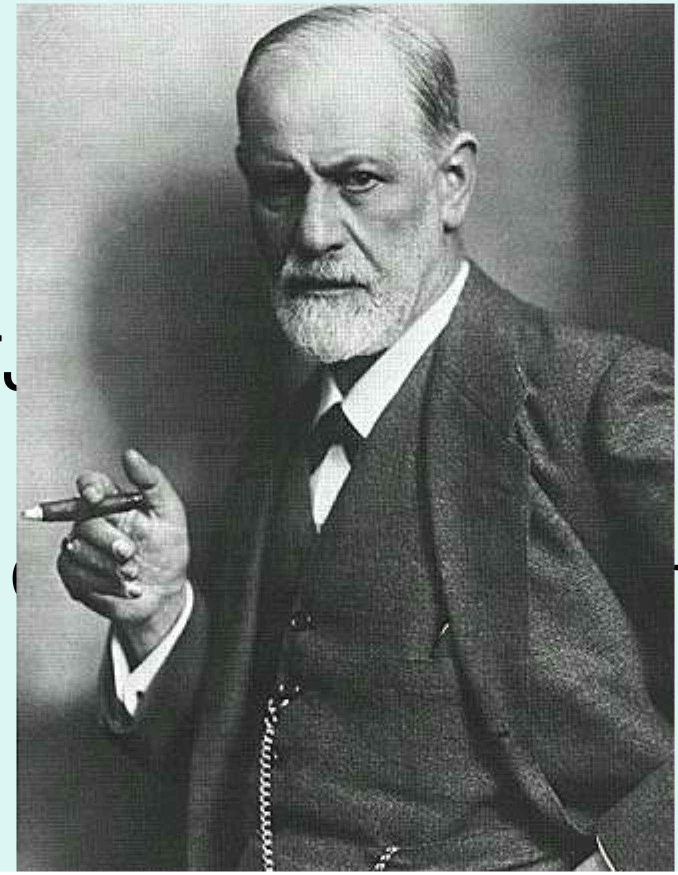
Regularization and its discontents

Graeme Smith
(IBM TJ Watson Research Center)

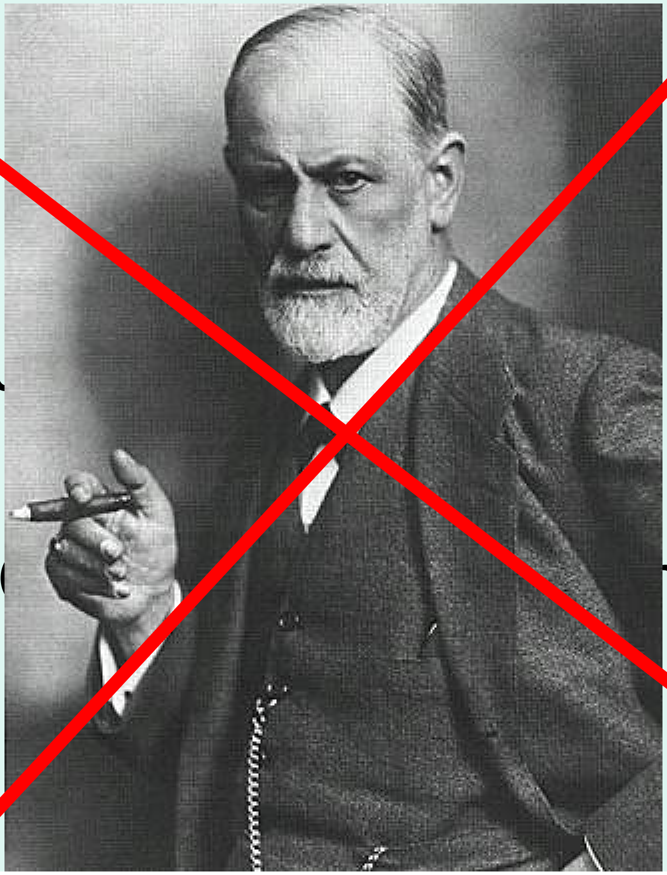
QIP Tutorial Session
January 16, 2010

Regularization and its discontents



(IBM TJ Center)

Regularization and its discontents



(IBM TJ Center)

n

Regularization and its discontents

Graeme Smith
(IBM TJ Watson Research Center)

QIP Tutorial Session
January 16, 2010

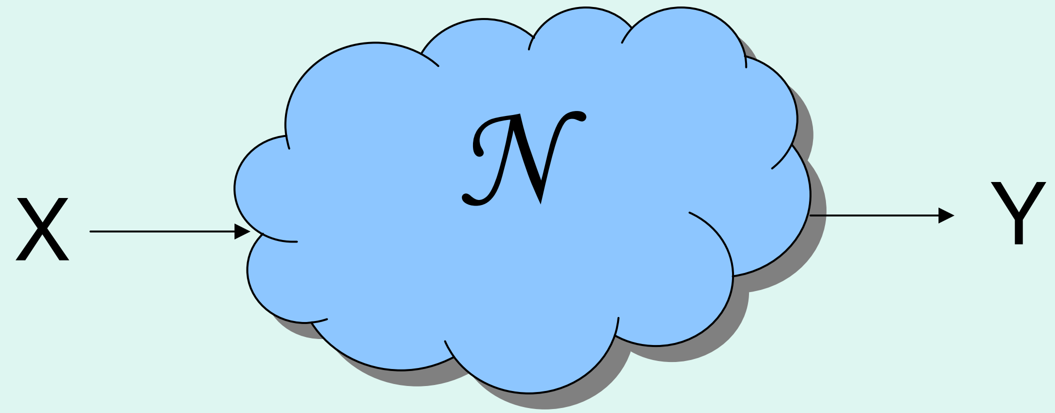
Information Theory

"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point."

-Claude Shannon 1948

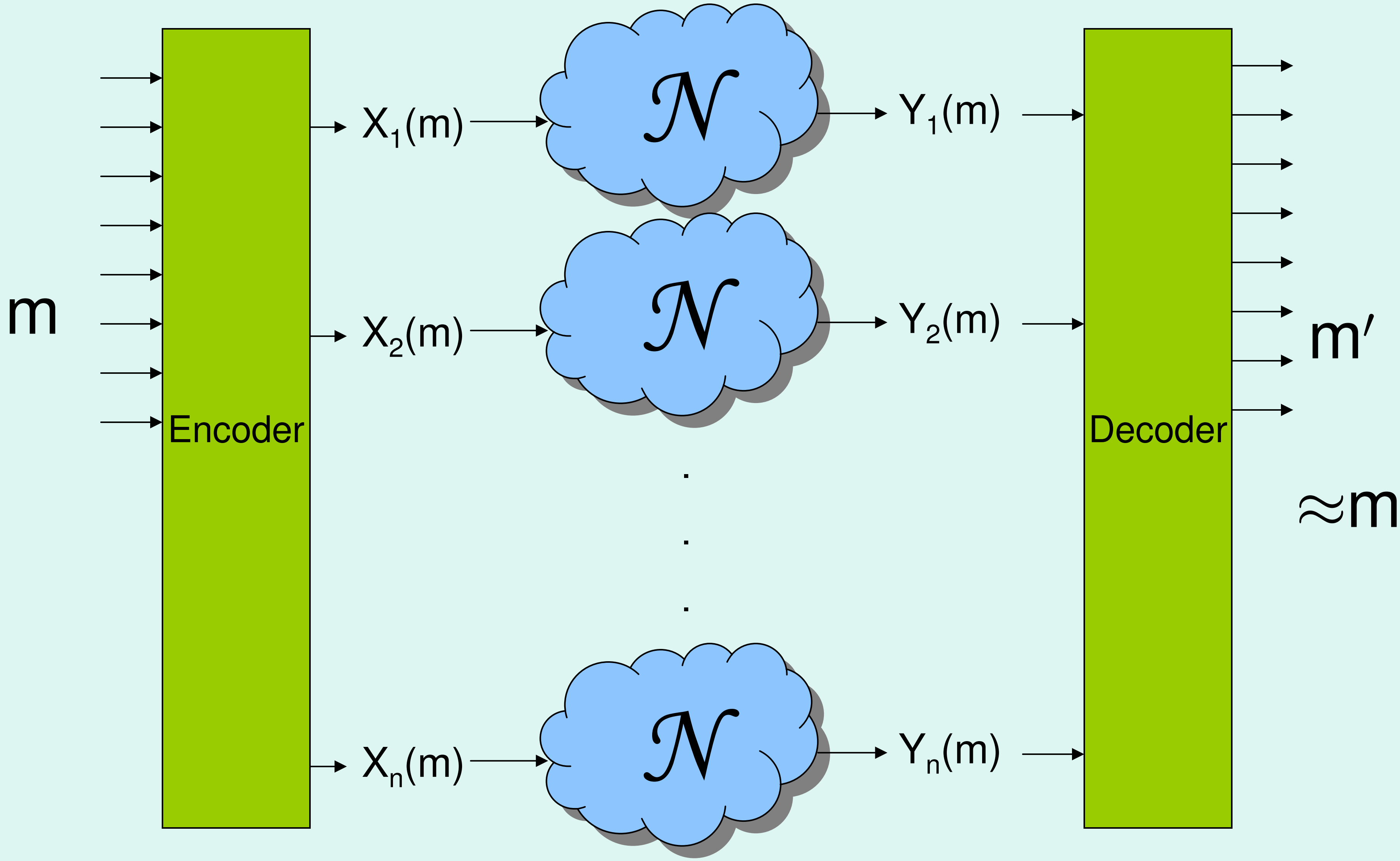
Source coding, channel coding, detection, cryptography…

Noisy Channel

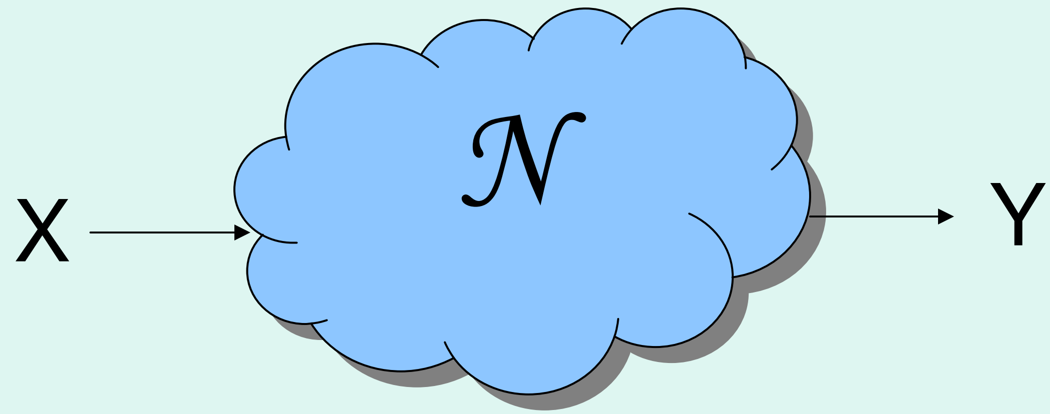


$p(y|x)$

Noisy Channel Coding



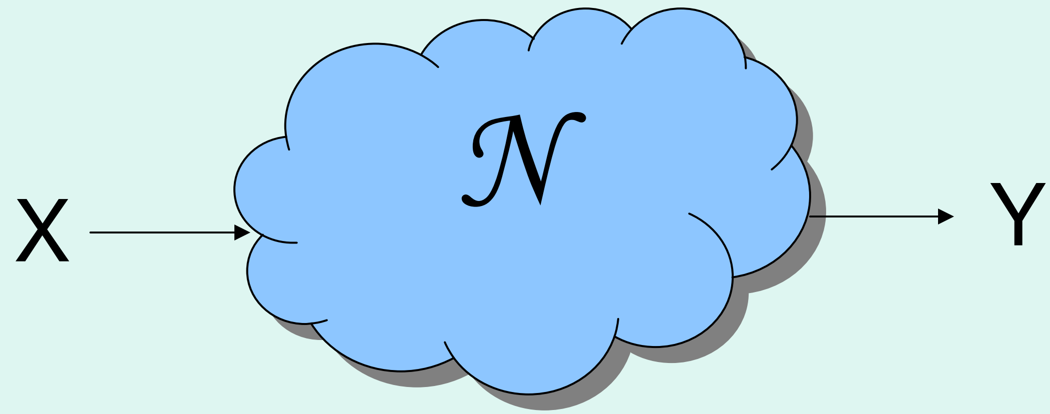
Noisy Channel Capacity



p(y|x)

Capacity: bits per channel use in the limit of many channels

Noisy Channel Capacity



$$p(y|x)$$

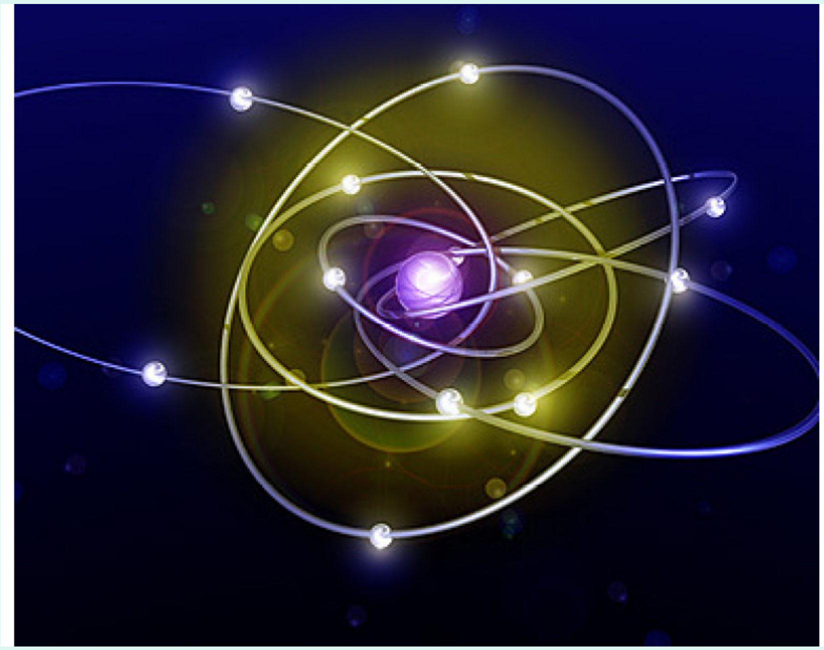
Capacity: bits per channel use in the limit of many channels

$$C = \max_X I(X;Y)$$

$I(X;Y)$ is the mutual information

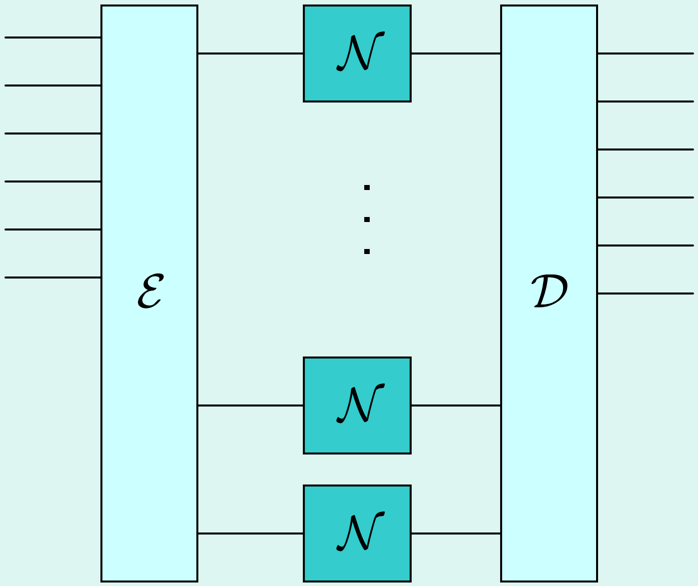
C is additive: $C(W_1 \times W_2) = C(W_1)+C(W_1)$

But the world is quantum!



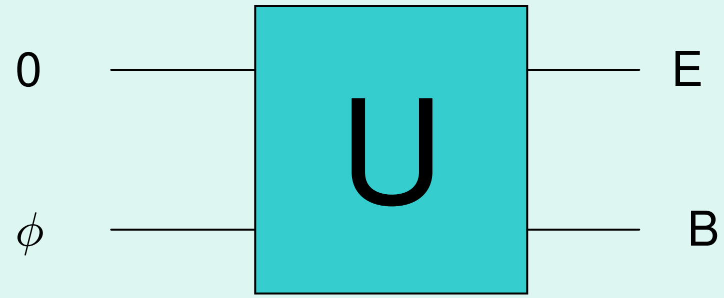
We'd like to develop a theory of information rich enough to include quantum mechanics.

Quantum Capacity



- Want to encode qubits so they can be recovered after experiencing noise.
- Quantum capacity is the maximum rate, in qubits per channel use, at which this can be done.
- We'd like a formula for Q($\mathcal{N}$) in terms of $\mathcal{N}$.

Quantum Capacity: What we know



- Coherent Information: $Q^1(\mathcal{N}) = \max S(B)-S(E)$
- $Q(\mathcal{N}) \geq Q^1(\mathcal{N})$ (Lloyd-Shor-Devetak)
- $Q(\mathcal{N}) = \lim_{n\to\infty}(1/n)\, Q^1(\mathcal{N} \otimes \ldots \otimes \mathcal{N})$
- $Q(\mathcal{N}) \neq Q^1(\mathcal{N})$ (DiVincenzo-Shor-Smolin '98)

Regularized Capacity formula vs Nonadditivity of capacity

- **Regularization**
  Do we have a simple formula for capacity?

  single-letter: $C = \max_X I(X;Y)$
  regularized: $Q(\mathcal{N}) = \lim_{n \to \infty}(1/n)\, Q^1(\mathcal{N} \otimes \ldots \otimes \mathcal{N})$

  Regularization in quantum capacity stems from nonadditivity of coherent information, Q^1, for *the same* channel: $Q^1(\mathcal{N} \otimes \mathcal{N}) > Q^1(\mathcal{N}) + Q^1(\mathcal{N})$

- **Nonadditivity of capacity**
  Can *different* channels enhance each others' capacities?
  $Q(\mathcal{N} \otimes \mathcal{M}) > Q(\mathcal{N}) + Q(\mathcal{M})$

Outline

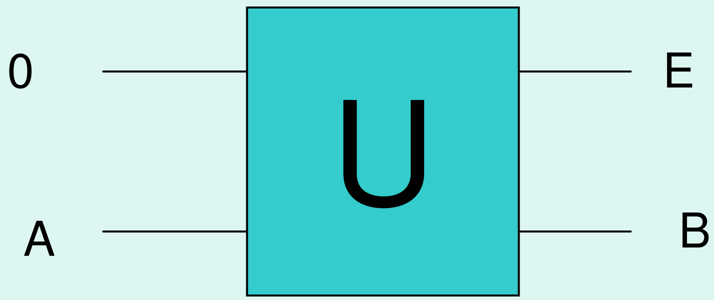
Talk One: Single letter examples---Degradable channels

- Degradable channels: Definition and Examples
- Strong Subadditivity and monotonicity of mutual information
- Additivity of coherent information for degradable channels
- Capacities of some channels

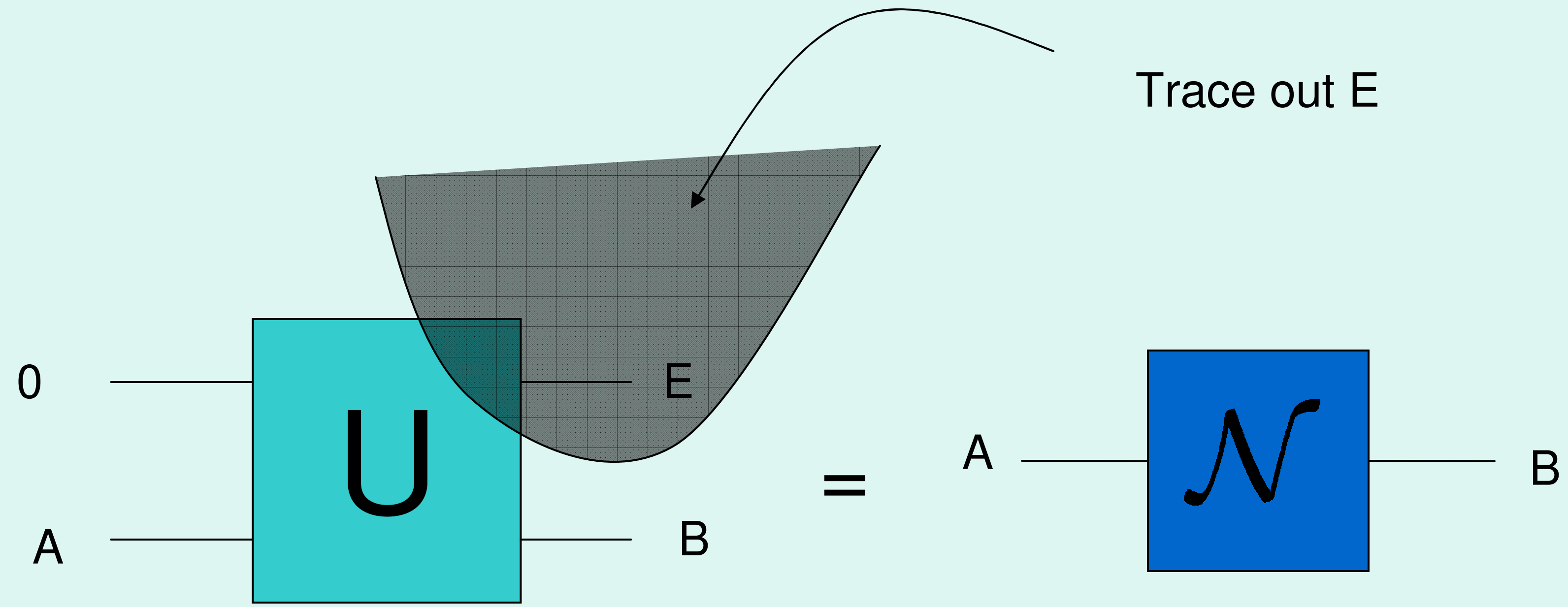
Talk Two: Where regularization really matters

- Depolarizing channel and degenerate codes
- Rocket Channels
- Switch Channels: large capacity with zero coherent information
- Back to degeneracy

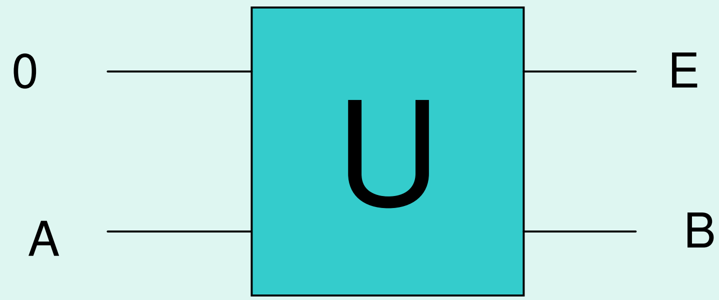
Degradable Channels



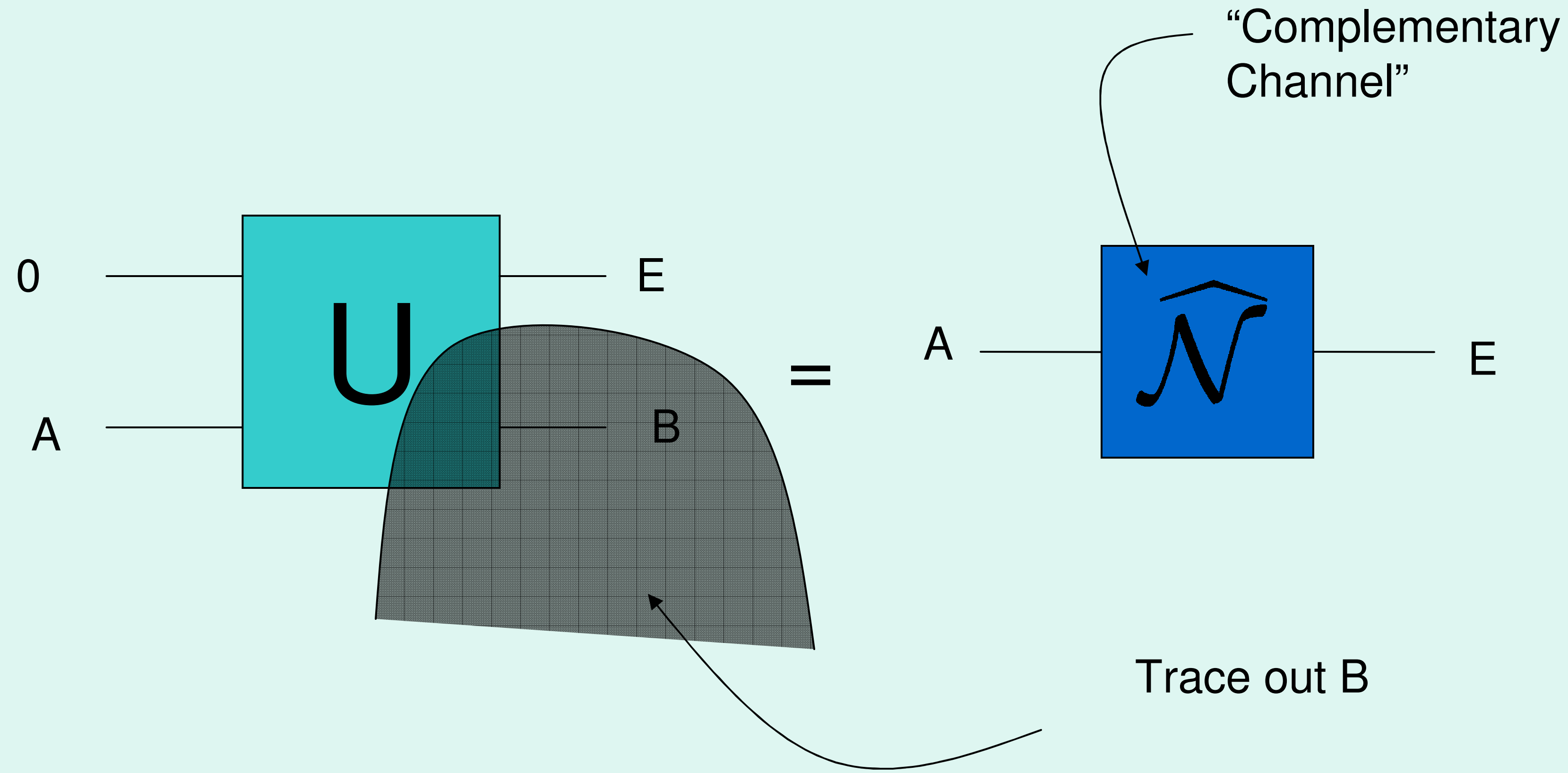
Degradable Channels



Degradable Channels

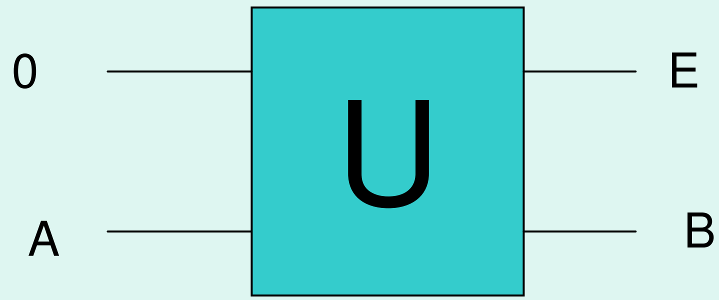


Degradable Channels



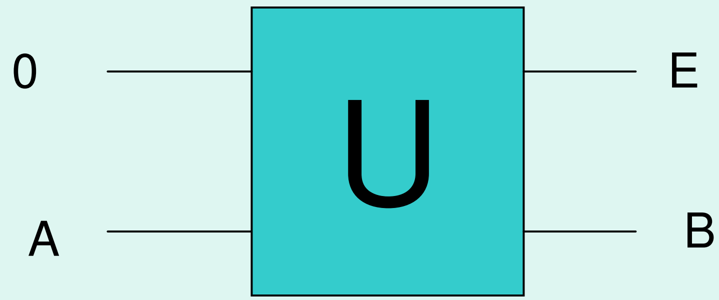
Degradable Channels

Bob can simulate Eve



Degradable Channels

Bob can simulate Eve



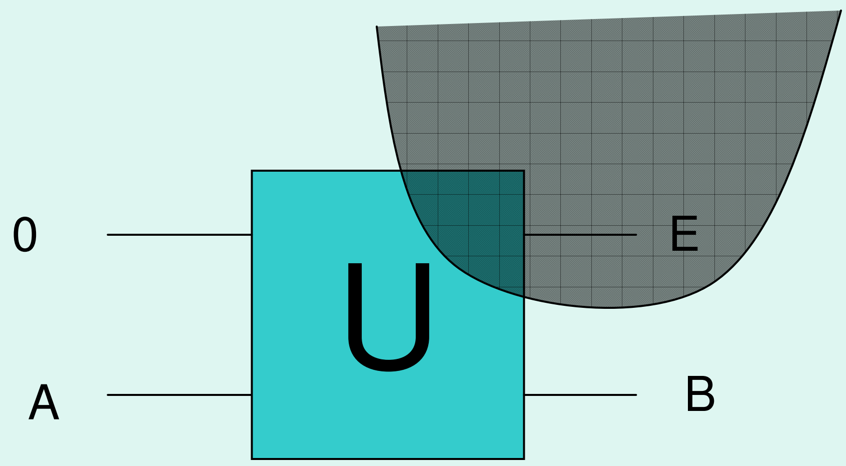
degradable if $\widehat{\mathcal{N}}$ is noisier than $\mathcal{N}$:

$$\widehat{\mathcal{N}} = \mathcal{D} \circ \mathcal{N}$$

For some "degrading channel" $\mathcal{D}$

Degradable Channels

Bob can simulate Eve



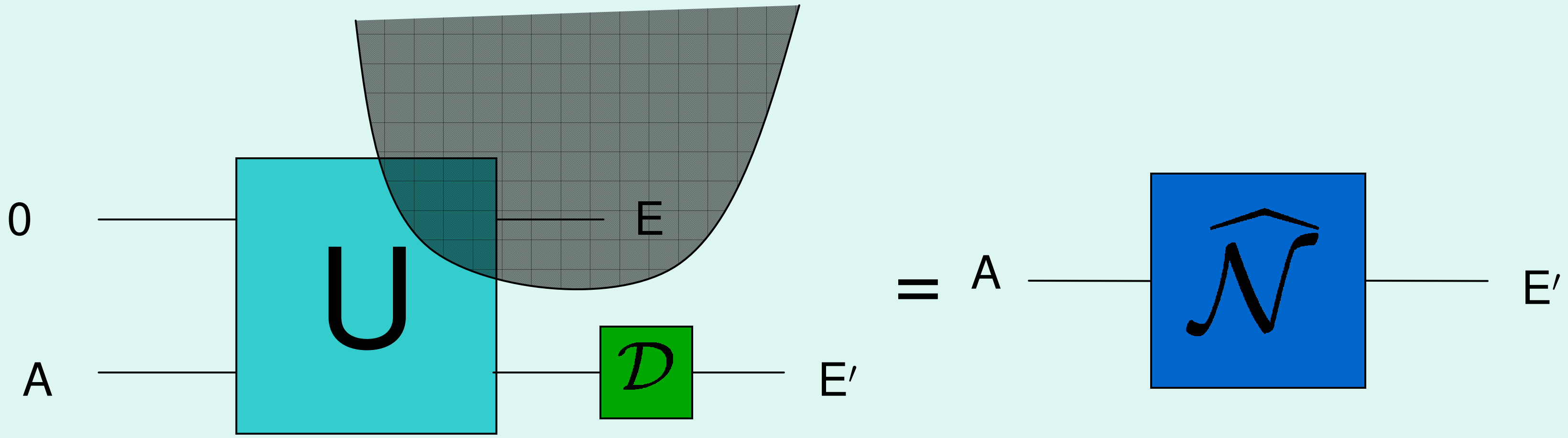
degradable if $\widehat{\mathcal{N}}$ is noisier than $\mathcal{N}$:

$$\widehat{\mathcal{N}} = \mathcal{D} \circ \mathcal{N}$$

For some "degrading channel" $\mathcal{D}$

Degradable Channels

Bob can simulate Eve



degradable if $\widehat{\mathcal{N}}$ is noisier than $\mathcal{N}$:

$$\widehat{\mathcal{N}} = \mathcal{D} \circ \mathcal{N}$$

For some "degrading channel" $\mathcal{D}$

Degradable Channels: Erasure Channel

Orthogonal erasure flag

- Erases with probability p:

$$\mathcal{E}_p(\rho) = (1-p)\rho + p|e\rangle\langle e|$$

Degradable Channels: Erasure Channel

- Erases with probability p:

Orthogonal erasure flag

$$\mathcal{E}_p(\rho) = (1-p)\rho + p|e\rangle\langle e|$$

- Envnment gets the state when its erased:

$$\widehat{\mathcal{E}}_p(\rho) = p\rho + (1-p)|e\rangle\langle e|$$

Degradable Channels: Erasure Channel

- Erases with probability p:

Orthogonal erasure flag

$$\mathcal{E}_p(\rho) = (1-p)\rho + p|e\rangle\langle e|$$

- Envnment gets the state when its erased:

$$\widehat{\mathcal{E}}_p(\rho) = p\rho + (1-p)|e\rangle\langle e|$$

- As long as p $\leq$ ½ , we can degrade:

$$\widehat{\mathcal{E}}_p = \mathcal{D}_p \circ \mathcal{E}_p$$

- $\mathcal{D}_p$ just throws away ρ with prob $\frac{1-2p}{1-p}$

Degradable Channels: Amplitude Damping

- Acts like $\mathcal{A}_\gamma(\rho) = A_0 \rho A_0^\dagger + A_1 \rho A_1^\dagger$ with $A_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}$ and $A_1 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}$. Models relaxation from excited to ground state.
- $\widehat{\mathcal{A}}_\gamma = \mathcal{A}_{1-\gamma}$ (Up to a unitary)
- So, if $\gamma \leq$ ½, can just damp more and simulate complementary channel

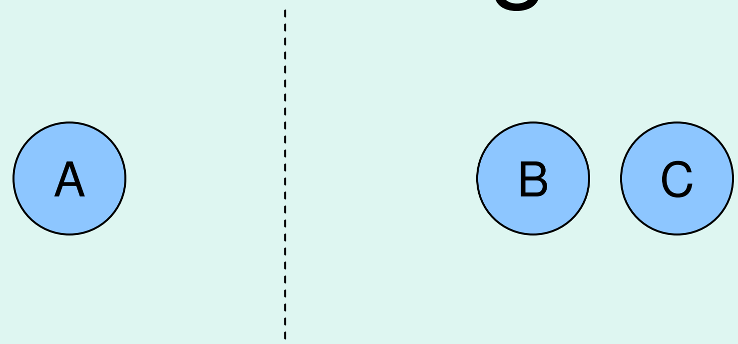
Degradable Channels: A few more

- Half of the qubit channels with two Kraus operators (the other half are reverse-degradable)
- Pure loss bosonic gaussian channel
- Channels whose complement is entanglement breaking ("Hadamard channels")
- Unruh channel??

How to check if a channel is degradable

- Let's say we have a channel $\mathcal{N}$ with complementary channel $\widehat{\mathcal{N}}$. How do we know if it's degradable?
- Well, $\mathcal{N}$ is a linear map on density matrices, so it has an inverse, $\mathcal{N}^{-1}$. Generally, this is not a CPTP map.
- If there's a degrading map, $\mathcal{D}$, it'll have to be equal to $\widehat{\mathcal{N}} \circ \mathcal{N}^{-1}$, so check if this is CPTP.

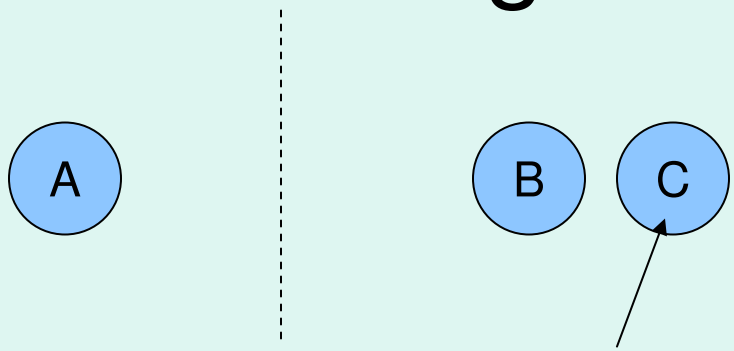
Monotonicity of Mutual Information and Strong Subadditivity



I(A;BC)

- Mutual Information:
  $I(A;B) = S(A) + S(B) - S(AB)$

Monotonicity of Mutual Information and Strong Subadditivity

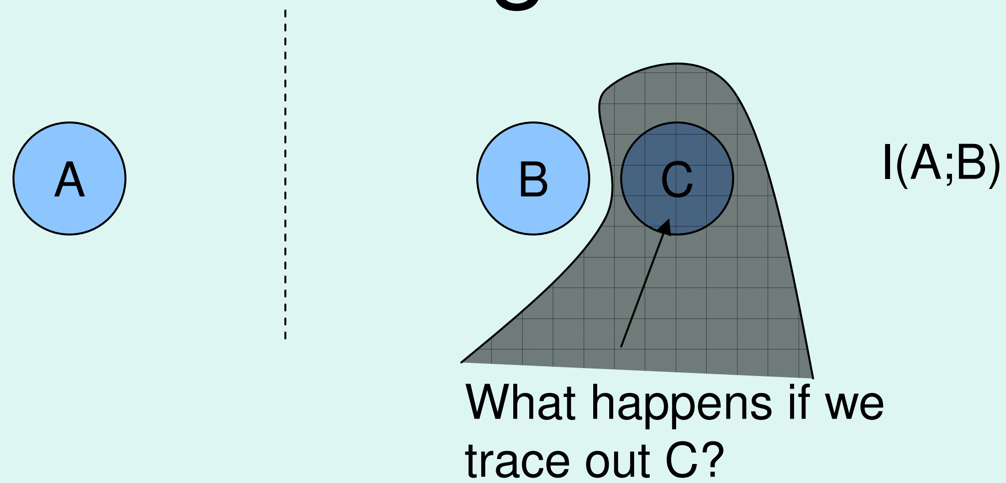


I(A;BC)

What happens if we trace out C?

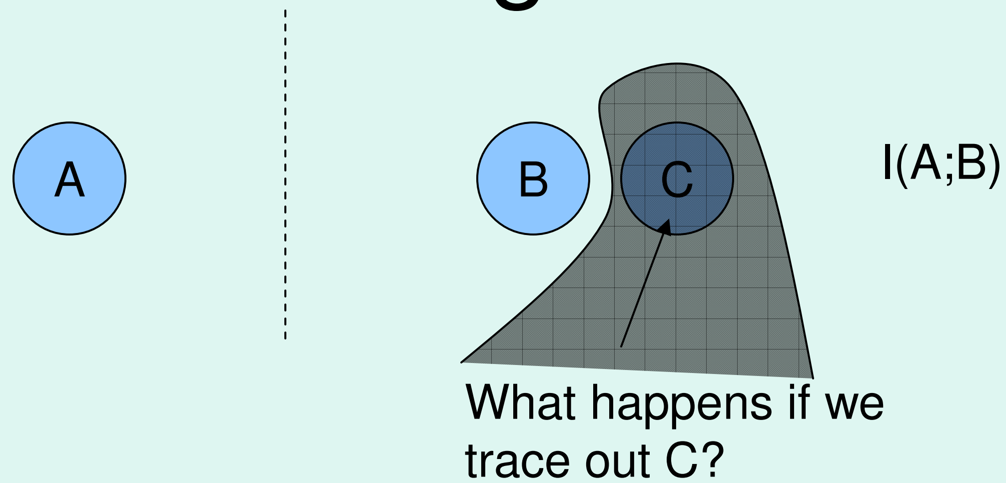
- Mutual Information:
  $I(A;B) = S(A) + S(B) - S(AB)$

Monotonicity of Mutual Information and Strong Subadditivity



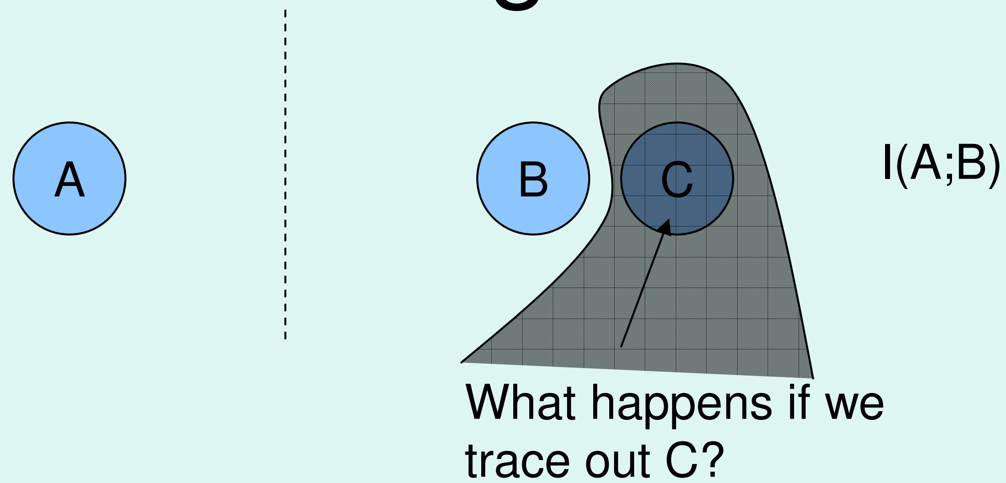
- Mutual Information:
  $I(A;B) = S(A) + S(B) - S(AB)$
- Monotonicity: $I(A;BC) \geq I(A;B)$

Monotonicity of Mutual Information and Strong Subadditivity



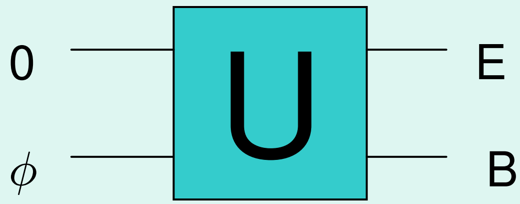
- Mutual Information:
  I(A;B) = S(A) + S(B) – S(AB)
- Monotonicity: I(A;BC) $\geq$ I(A;B)
- Equivalently: S(BC) + S(AB) $\geq$ S(ABC) + S(B)

Monotonicity of Mutual Information and Strong Subadditivity



- Mutual Information:
  I(A;B) = S(A) + S(B) – S(AB)
- Monotonicity: I(A;BC) $\geq$ I(A;B)
- Equivalently: S(BC) + S(AB) $\geq$ S(ABC) + S(B)
- This tells us mutual information can only decrease under local processing

Additivity of Coherent Information for degradable channels



Recall that $Q^1 = \max_\phi S(B) - S(E)$
ϕ is a mixed state on the input.

- The capacity is given by $Q(\mathcal{N}) = \lim_{n \to \infty} (1/n)\, Q^1(\mathcal{N} \otimes \ldots \otimes \mathcal{N})$
- If we could show $Q^1(\mathcal{N} \otimes \mathcal{N}) \leq 2Q^1(\mathcal{N})$ for any degradable channel, we'd have $Q(\mathcal{N}) = Q^1(\mathcal{N})$ for degradable channels

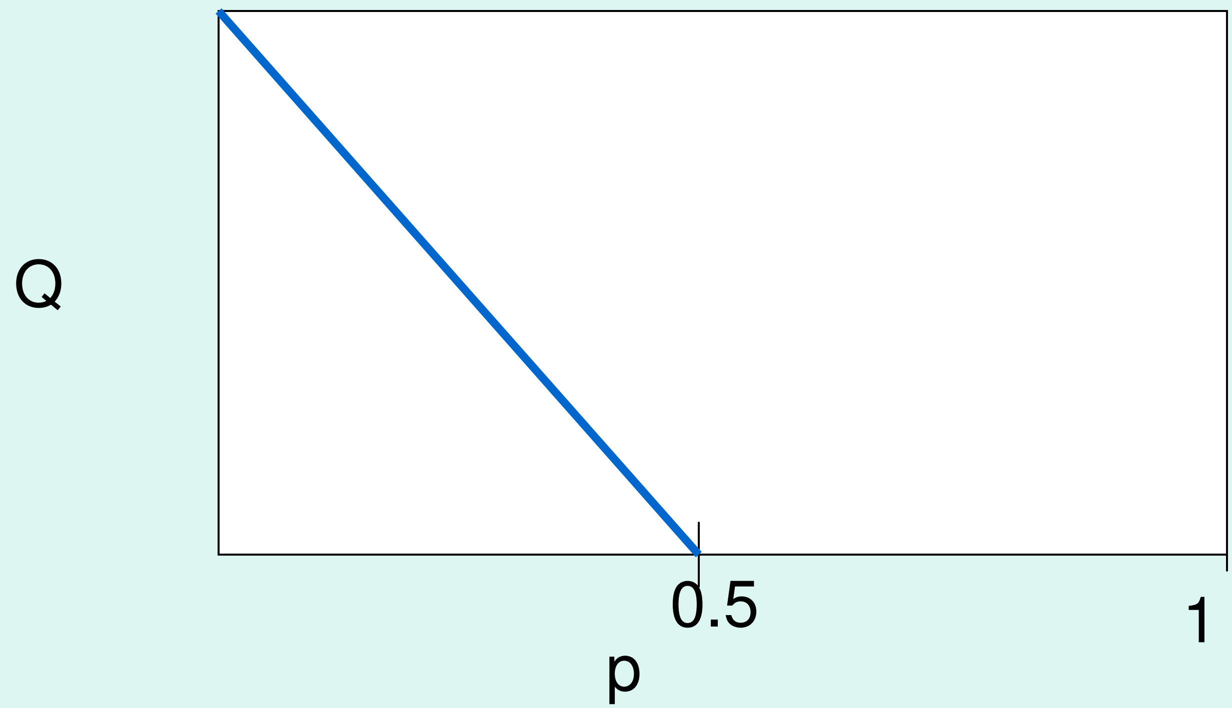
Additivity of Coherent Information for degradable channels

- Say we have ϕ_{12}, a mixed state on A_1A_2, the input of $\mathcal{N} \otimes \mathcal{N}$, with $Q^1(\mathcal{N} \otimes \mathcal{N}) = S(B_1B_2)-S(E_1E_2)$ (entropies evaluated on $\rho_{B_1B_2E_1E_2} = U \otimes U \phi_{12} U^\dagger \otimes U^\dagger$, where U is the isometric extension of $\mathcal{N}$.
- This gives us two input states to try on a single use of the channel: $\phi_1 = \mathrm{Tr}_2\, \phi_{12}$, and $\phi_2 = \mathrm{Tr}_1 \phi_{12}$. Let's evaluate how much coherent information the two of these give us.
- ϕ_1 gives us $S(B_1)-S(E_1)$, evaluated on $\rho_{B_1E_1} = U\phi_1 U^\dagger = \mathrm{Tr}_{B_2E_2}\rho_{B_1B_2E_1E_2}$, and similarly ϕ_2 gives $S(B_2)-S(E_2)$. Note that these entropies are evaluated on $\rho_{B_1B_2E_1E_2}$!
- Since $S(B_1)-S(E_1) \leq Q^1(\mathcal{N})$, and similarly for 2, now we just have to show that $S(B_1B_2) - S(E_1E_2) \leq S(B_1)-S(E_1) + S(B_2)-S(E_2)$.
- This is equivalent to showing $S(E_1)+S(E_2)-S(E_1E_2)= I(E_1; E_2) \leq I(B_1;B_2) = S(B_1)+S(B_2)-S(B_1B_2)$
- We can degrade B_1 to E_1, and B_2 to E_2, so this follows by monotonicity of mutual information!

The Capacity of Some Channels

- Erasure channel: $\mathcal{E}_p(\rho) = (1-p)\rho + p|e\rangle\langle e|$

$$Q(\mathcal{E}_p) = 1 - 2p$$

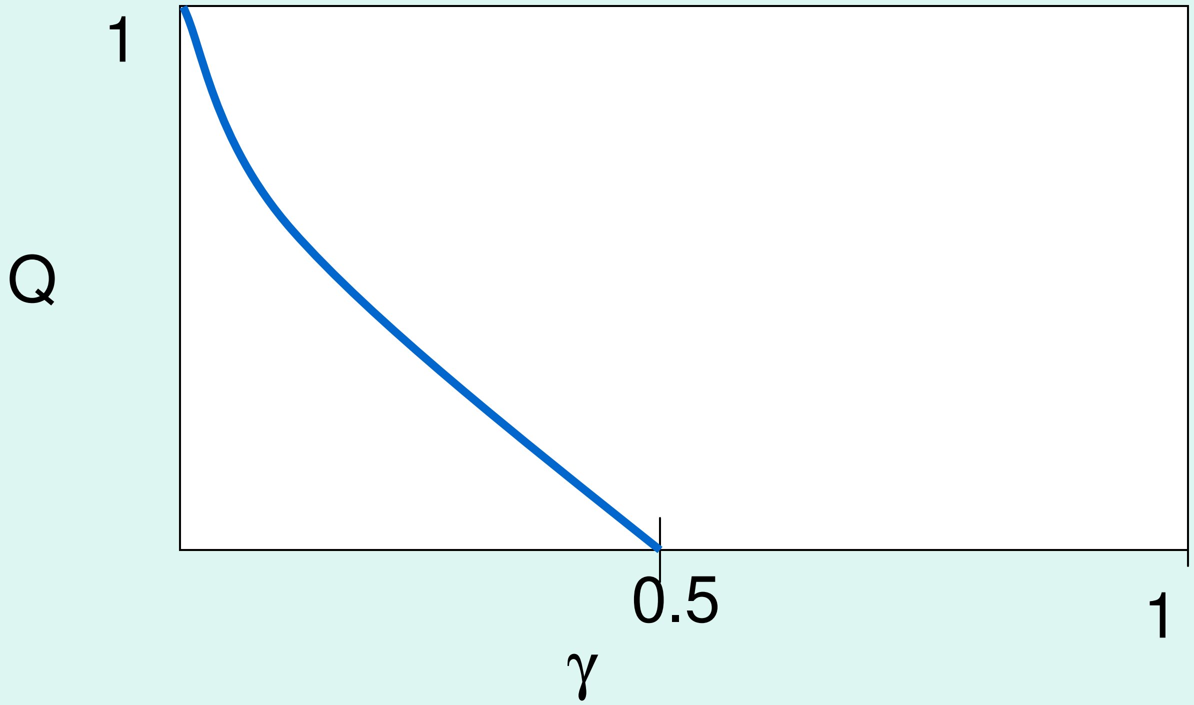


The Capacity of Some Channels

- Amplitude Damping Channel

$$\mathcal{A}_\gamma(\rho) = A_0 \rho A_0^\dagger + A_1 \rho A_1^\dagger$$

- Q = $\max_t$ ($H(t\gamma) - H(t(1-\gamma))$)



Summary of Part I: Single-letter Examples

- Capacities tell us the ultimate limits for error correction over a noisy quantum channel. The quantum capacity tells us about quantum error correcting codes.
- In general we have regularized capacity formulas, which can't be computed in general. This is annoying.
- For some special channels---e.g., degradable ones---we can actually show that the regularized formula "single-letterizes". The capacity has a simple formula.
- Some pretty interesting channels are actually degradable---erasure, amplitude damping, pure loss gaussian bosonic.
- Up next: when it's not degradable, don't expect a single-letter formula!

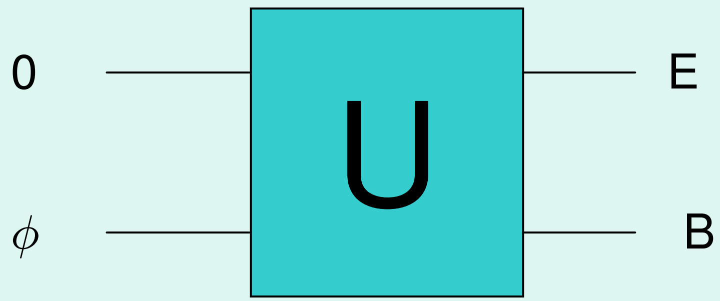
Part II: Where Regularization Really Matters

Graeme Smith
(IBM TJ Watson Research Center)

QIP Tutorial Session

January 16, 2010

Quantum Capacity: What we know



- Coherent Information: $Q^1(\mathcal{N}) = \max S(B)-S(E)$
- $Q(\mathcal{N}) \geq Q^1(\mathcal{N})$ (Lloyd-Shor-Devetak)
- $Q(\mathcal{N}) = \lim_{n \to \infty}(1/n)\, Q^1(\mathcal{N} \otimes \ldots \otimes \mathcal{N})$
- $Q(\mathcal{N}) \neq Q^1(\mathcal{N})$ (DiVincenzo-Shor-Smolin '98)

Outline

- Depolarizing channel
  and degenerate codes
- Rocket Channels
- Switch Channels: large capacity with zero coherent information
- Back to degeneracy

A blast from the past: degenerate codes and the depolarizing channel

$$\mathcal{N}_p(\rho) = (1-p)\rho + \frac{p}{3}X\rho X + \frac{p}{3}Y\rho Y + \frac{p}{3}Z\rho Z$$

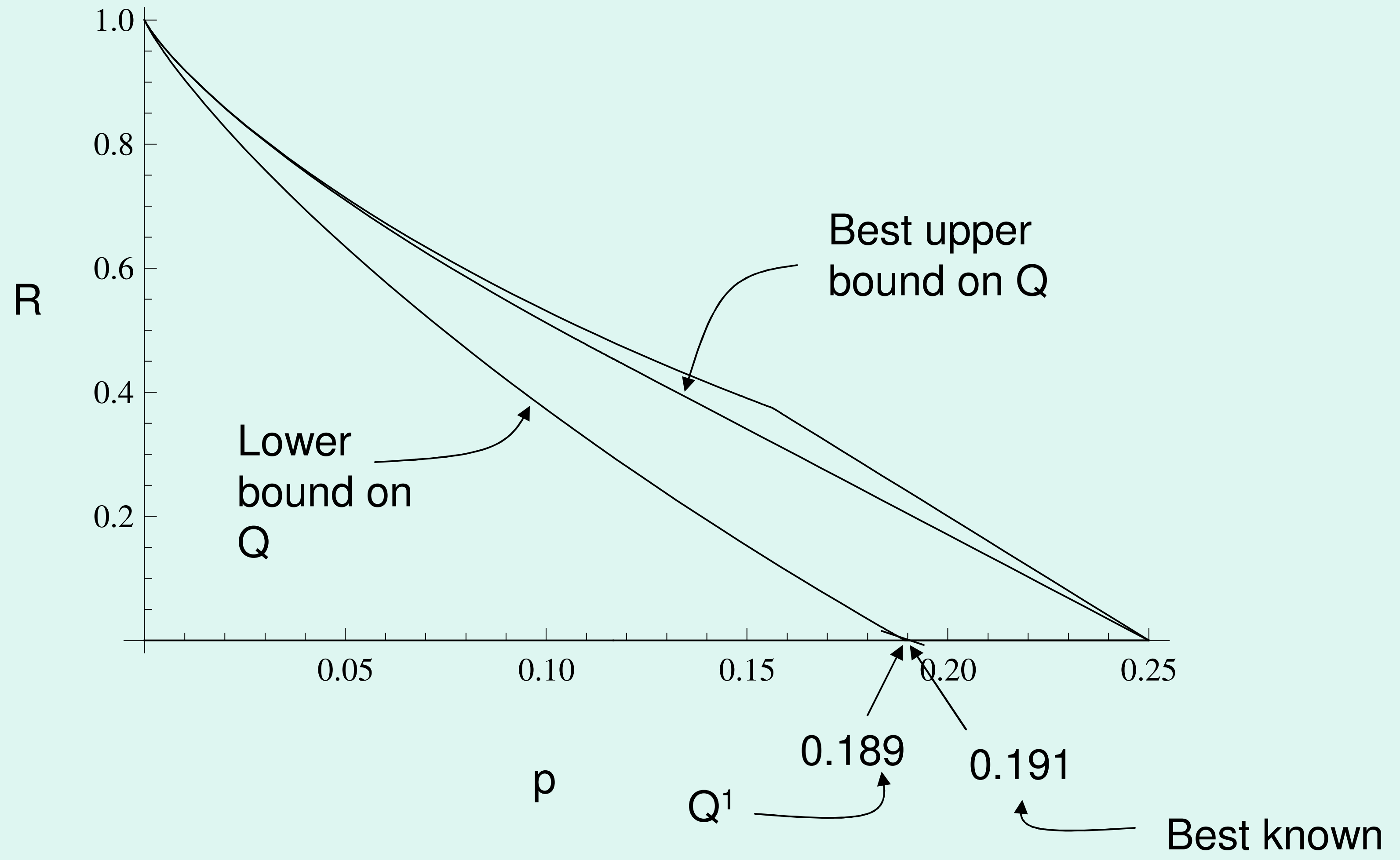
- $Q^1(\mathcal{N}_p)$ = 1- H(p) – p log 3
- This is achieved with ϕ maximally mixed.
- H(p)+p log 3 is the entropy of the errors, so if our code ensures a different syndrome for each typical error, we cant beat Q^1.

A blast from the past: degenerate codes and the depolarizing channel

A blast from the past: degenerate codes and the depolarizing channel

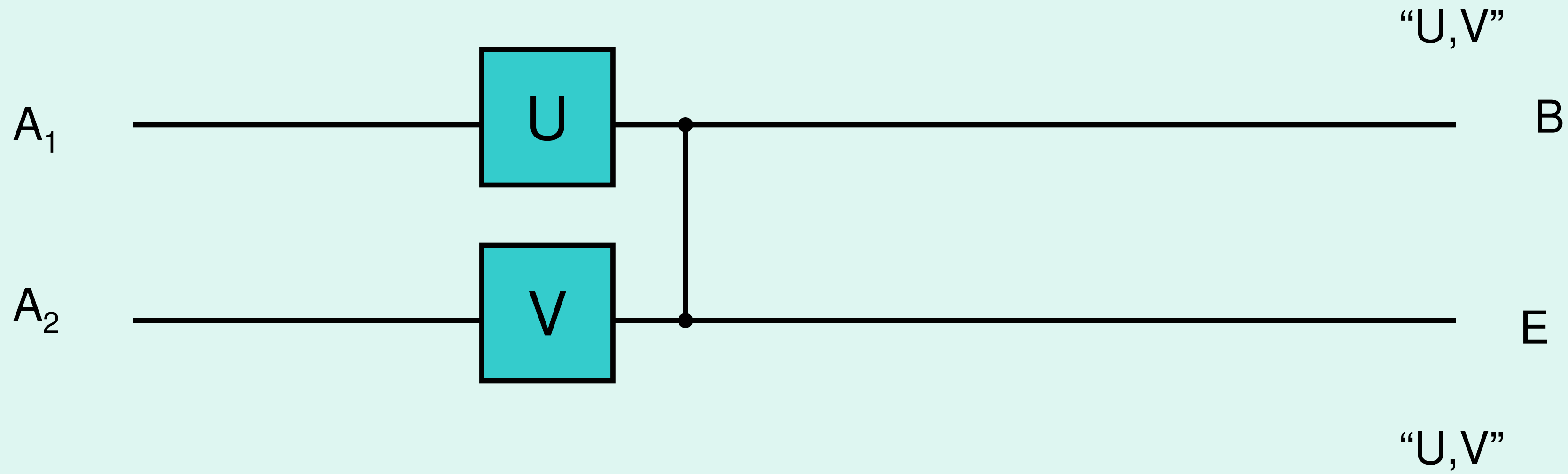
- Shor and Smolin: coding strategy that beats $Q^1(\mathcal{N}_p)$ slightly for very noisy channels.
- By choosing $\phi_5 = \frac{1}{2}\left(|0\rangle\langle 0|^{\otimes 5} + |1\rangle\langle 1|^{\otimes 5}\right)$, they showed that $Q^1(\mathcal{N}_p^{\otimes 5}) > 5\, Q^1(\mathcal{N}_p)$.
- This amounts to concatenating a 5-bit repetition code with a random stabilizer code.
- get around the idea that you have to figure out which typical error happened by manipulating the state of the environment with clever input. The environment state no longer has eigenbasis corresponding to strings of I,X,Y,Z errors.

A blast from the past: degenerate codes and the depolarizing channel



Rocket Channels

$$\mathcal{R}_d = E(\ \mathcal{R}^{U,V}{}_d \otimes |UV\rangle\langle\ UV|)$$



This channel has classical capacity ≤ 2

Proof: C'mon!

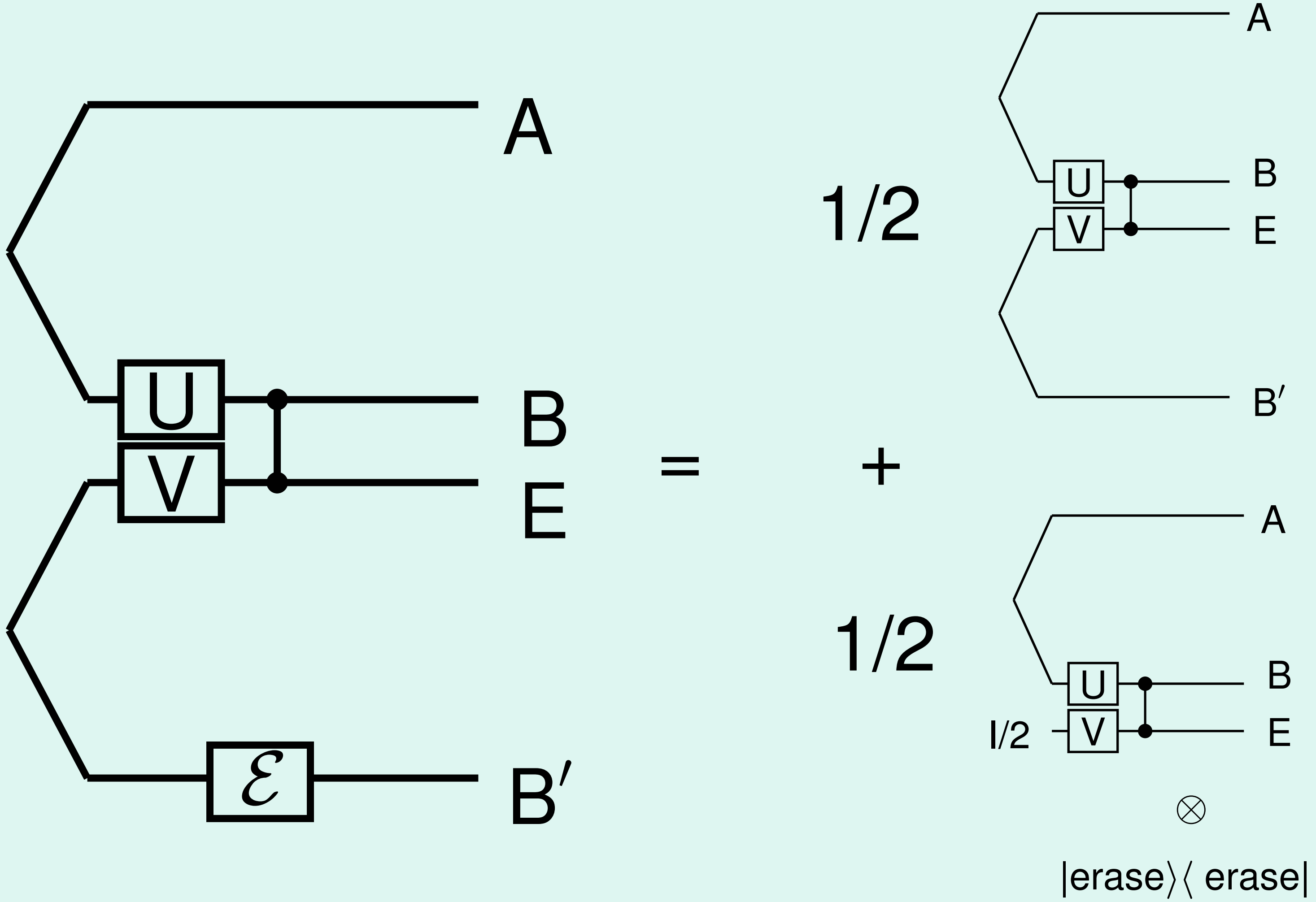
High Joint Coherent Information

- What channel should we use $\mathcal{R}_d$ with?
- We'll use a 50% erasure channel:

$$\mathcal{E}(\rho) = \tfrac{1}{2}\, \rho + \tfrac{1}{2}\, |\text{erase}\rangle\langle\text{erase}|$$

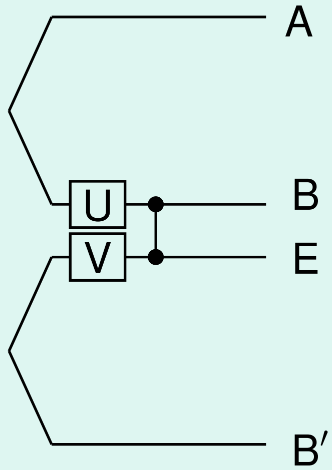
- How should we use it?

High Joint Coherent Information



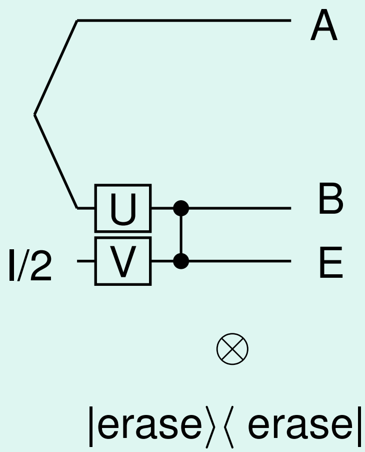
Coherent Information: average of two terms

1/2



+

1/2



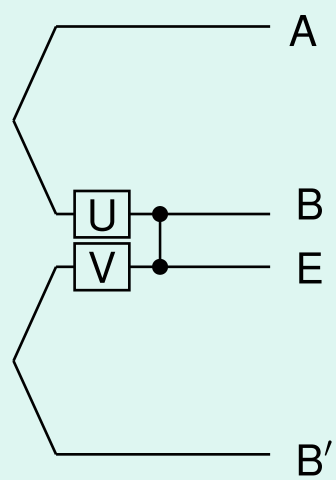
½ I^{coh}(not erased)

+

½ I^{coh}(erased)

Coherent Information: average of two terms

1/2

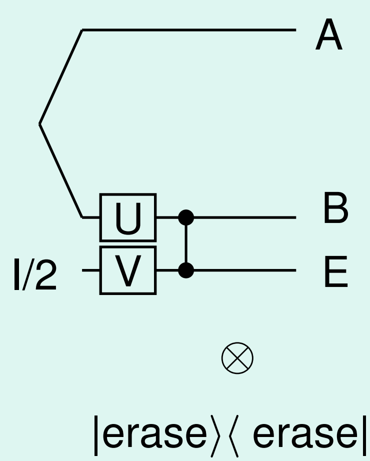


$\frac{1}{2}\, I^{coh}(\text{not erased})$

+

+

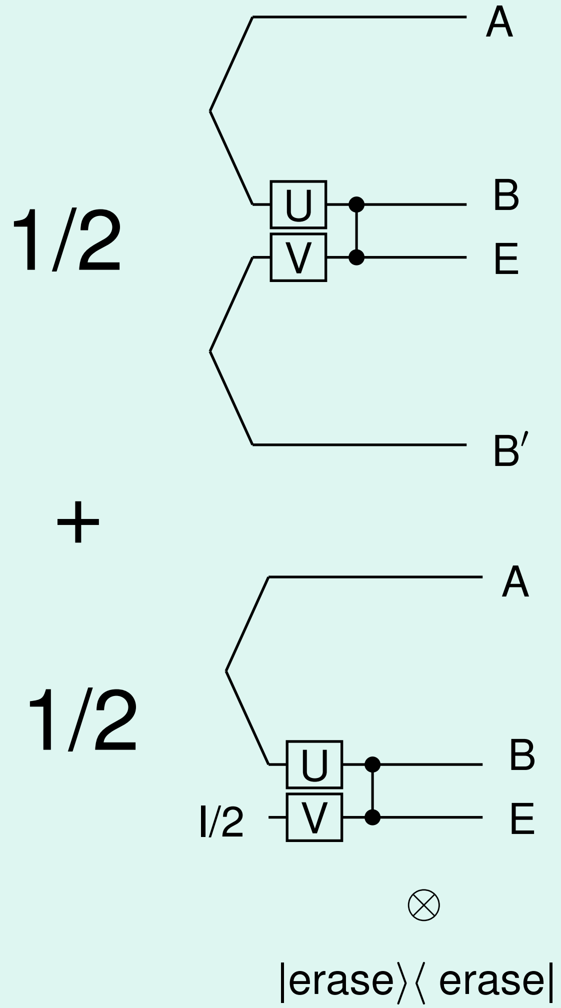
1/2



$\frac{1}{2}\, I^{coh}(\text{erased})$

This term is zero

Coherent Information: average of two terms



½ I^{coh}(not erased)

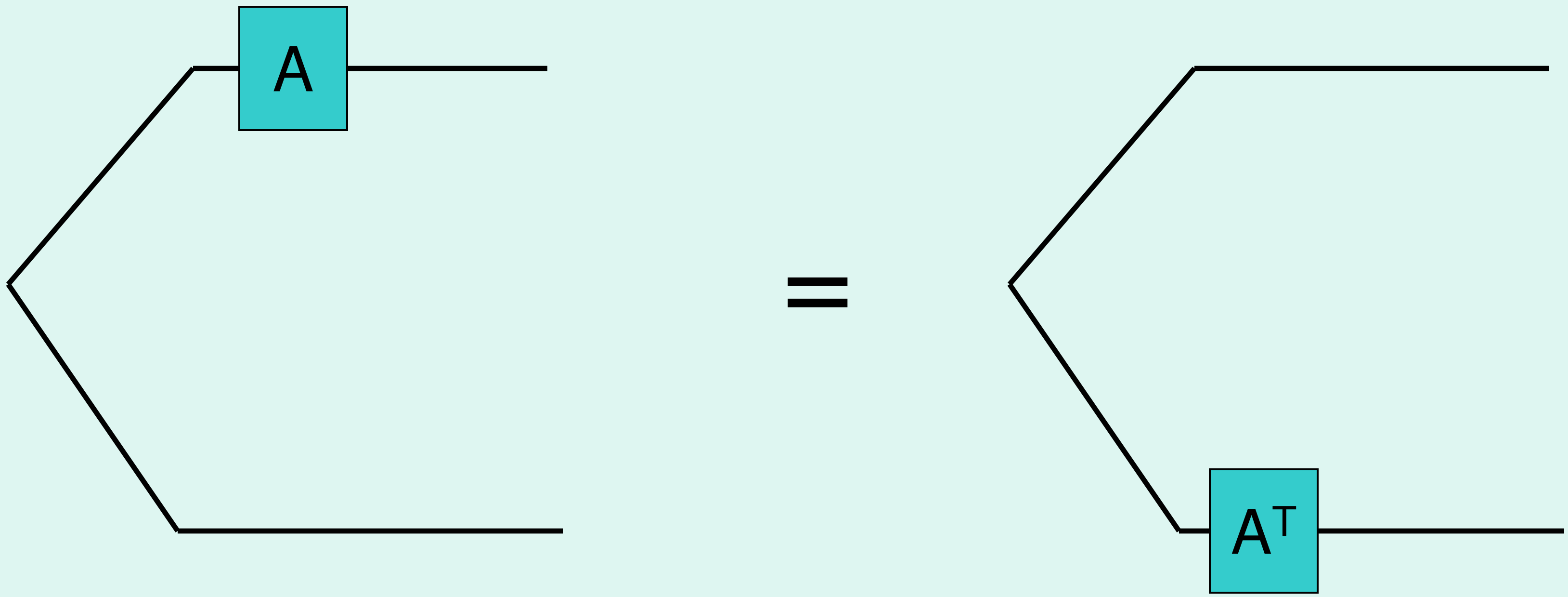
How big is this?

+

½ I^{coh}(erased)

This term is zero

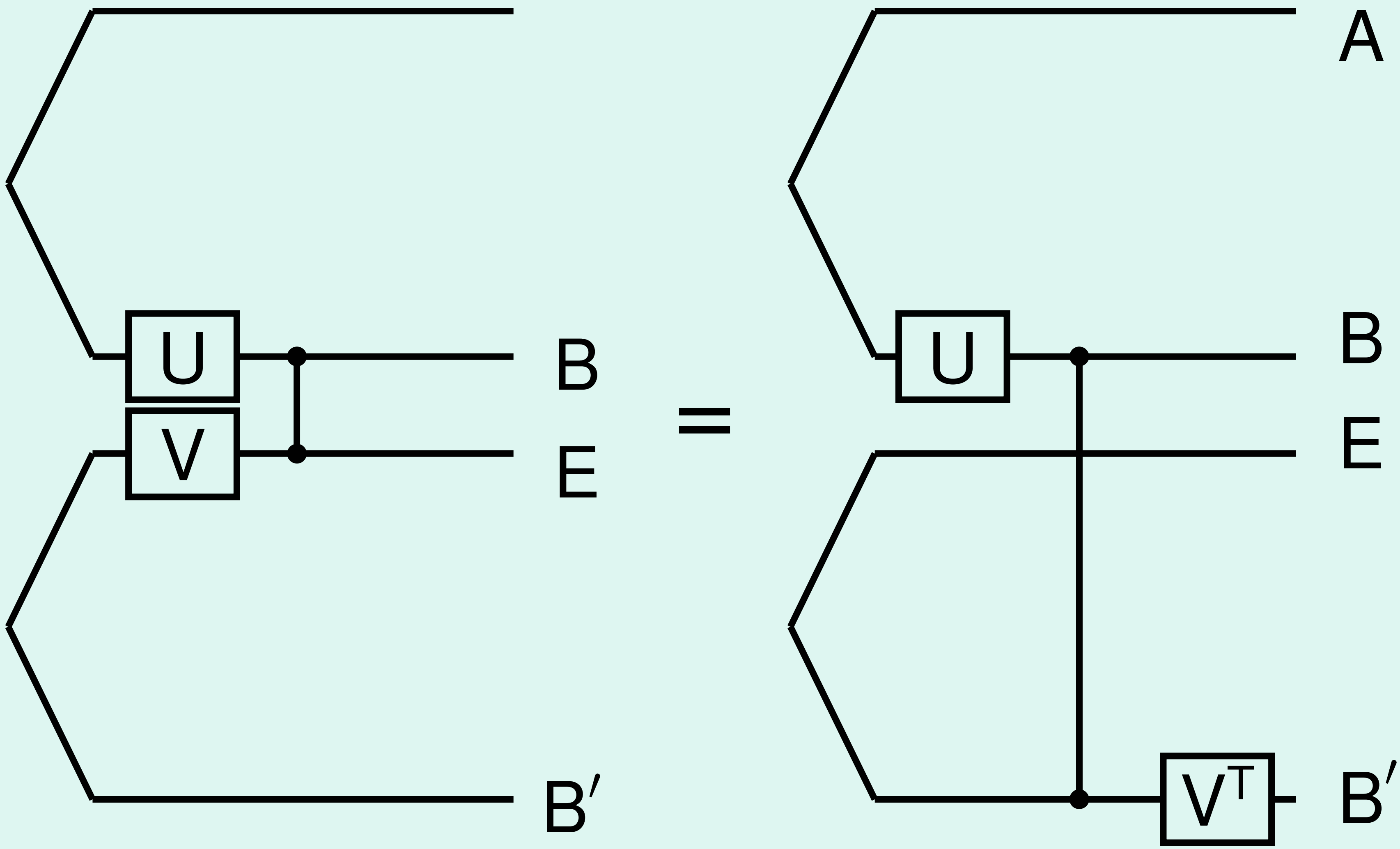
A nice identity



$$A \otimes I\,|\phi_d\rangle = \sum A_{i,j}\,|i\rangle\,|j\rangle \qquad I \otimes B\,|\phi_d\rangle = \sum B_{j,i}\,|i\rangle\,|j\rangle$$

$$|\phi_d\rangle = \sum |i\rangle\,|i\rangle$$

Bob can undo interaction



Bob can undo interaction

- Coherent information doesn't increase when Bob operates, so when there's no erasure, $I^{coh} = \log d$
- The overall coherent information, and therefore

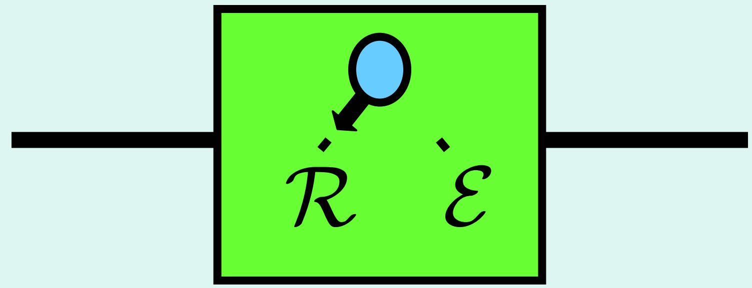
$$Q(\mathcal{R}_d \otimes \mathcal{E}) \geq Q^1(\mathcal{R}_d \otimes \mathcal{E}) \geq \tfrac{1}{2} \log d$$

Choice Channels

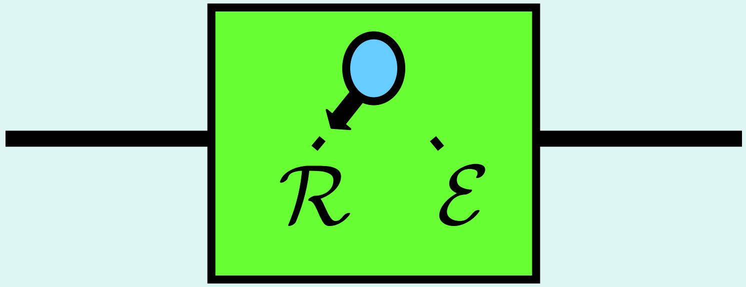
- Let's say we have two channels, R and E, with $Q^1(R)$ and $Q^1(E)$ large, but $Q^1(R \otimes E)$ large
- I want to make a single channel, T, that will show large violations of the additivity of Q^1.
- This will show that the coherent information, Q^1, is a poor approximation to the quantum capacity, Q.

Choice Channels

- Let’s say we have two channels, R and E, with $Q^1(R)$ and $Q^1(E)$ large, but $Q^1(R \otimes E)$ large
- I want to make a single channel, T, that will show large violations of the additivity of Q^1.
- This will show that the coherent information, Q^1, is a poor approximation to the quantum capacity, Q.

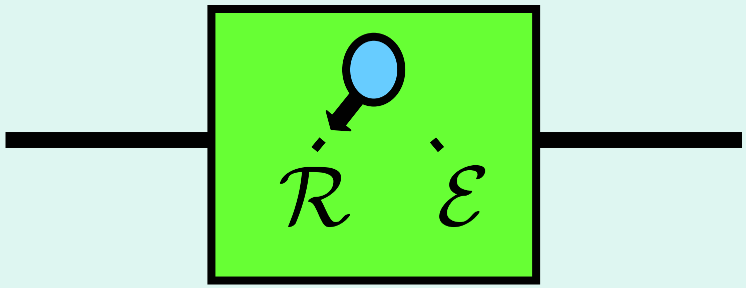


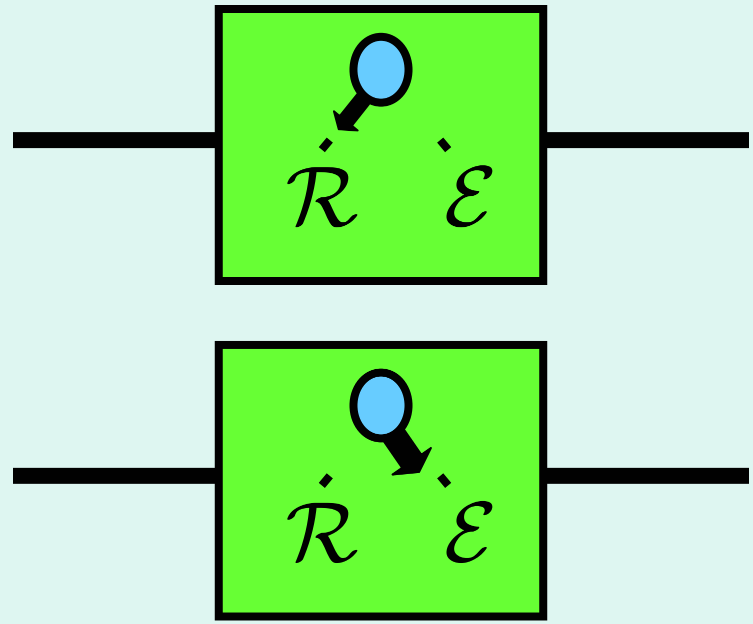
Choice Channels



$Q^1(\mathcal{T})$ is small since it's no more than max $(Q^1(\mathcal{R}), Q^1(\mathcal{E}))$

Choice Channels

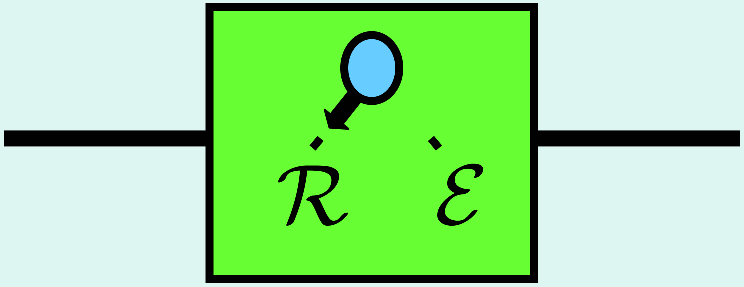


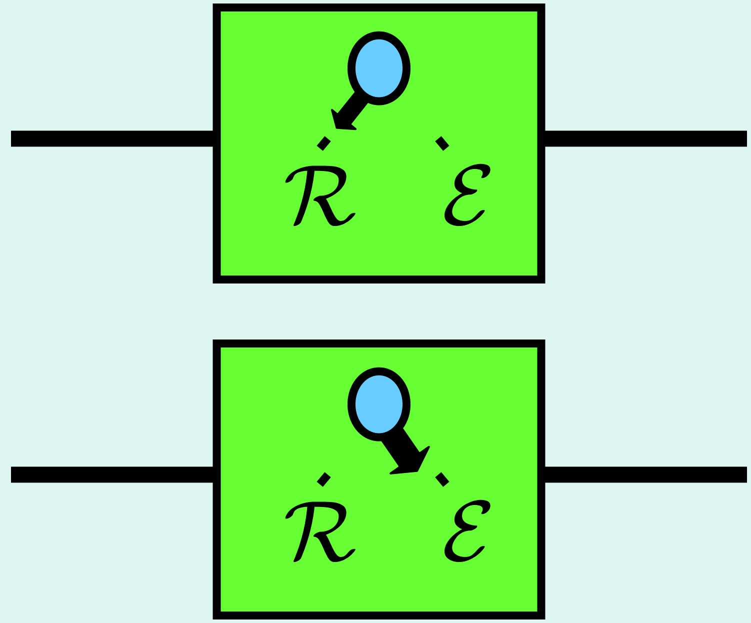


$Q^1(\mathcal{T})$ is small since it's no more than max $(Q^1(\mathcal{R}), Q^1(\mathcal{E}))$

Make one channel $\mathcal{R}$ and the other $\mathcal{E}$ to get large $Q^1(\mathcal{T} \otimes \mathcal{T})$

Choice Channels





Q^1 $(\mathcal{T})$ is small since it's no more than max $(Q^1(\mathcal{R}), Q^1(\mathcal{E}))$

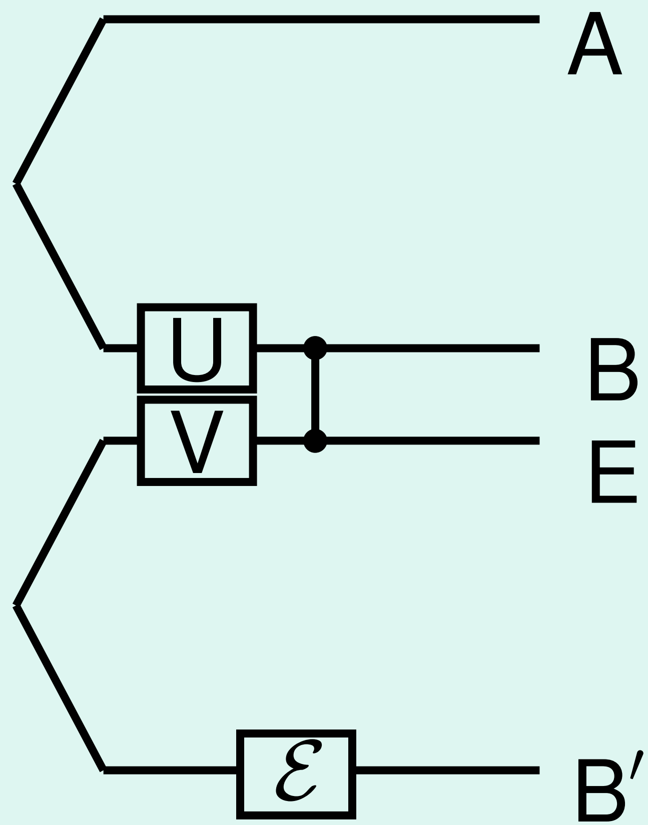
Make one channel $\mathcal{R}$ and the other $\mathcal{E}$ to get large $Q^1(\mathcal{T} \otimes \mathcal{T})$

Gives $Q^1(\mathcal{T}) \leq 2$ but

$Q(\mathcal{T}) \geq \ 1/8 \log D_{in}$

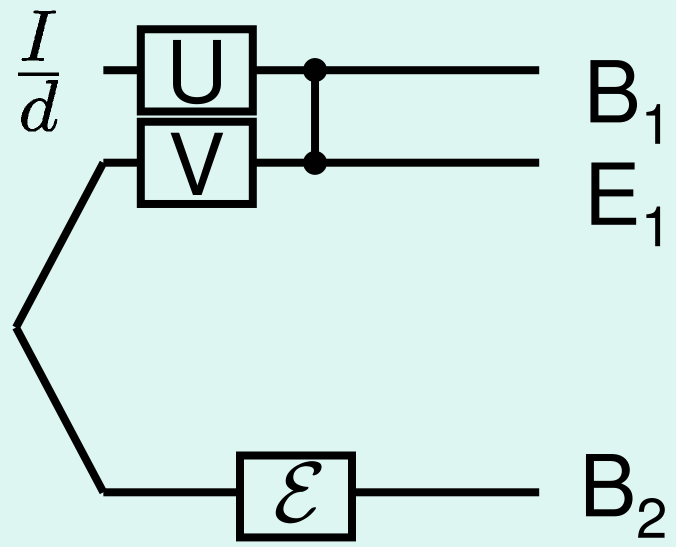
Choice Channels and Degenerate Codes

- Let's look at the states generated by using a choice channel based on the Rocket Channel. Focus on a fixed U and V.



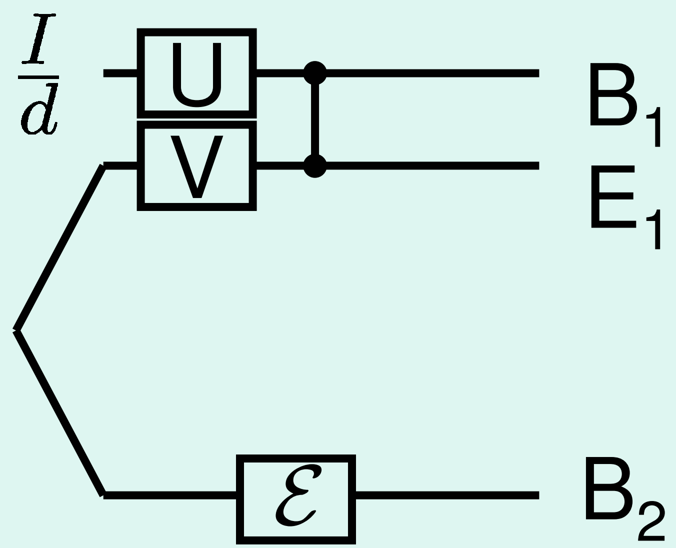
Choice Channels and Degenerate Codes

- Let’s look at the states generated by using a choice channel based on the Rocket Channel. Focus on a fixed U and V.



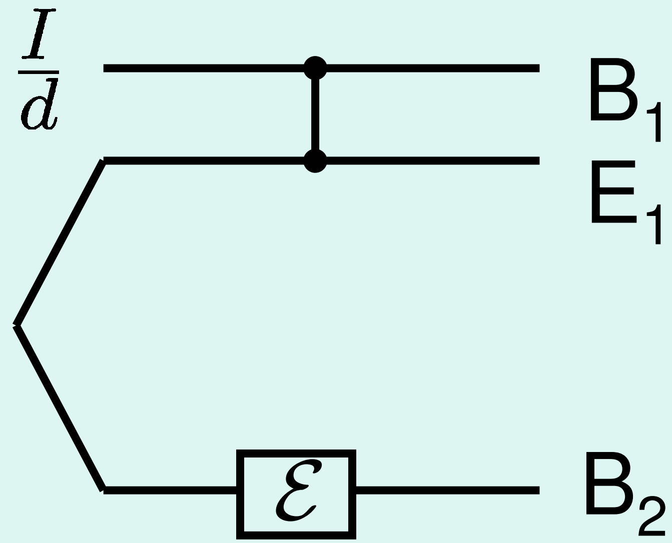
Choice Channels and Degenerate Codes

- Let's look at the states generated by using a choice channel based on the Rocket Channel. Focus on a fixed U and V.



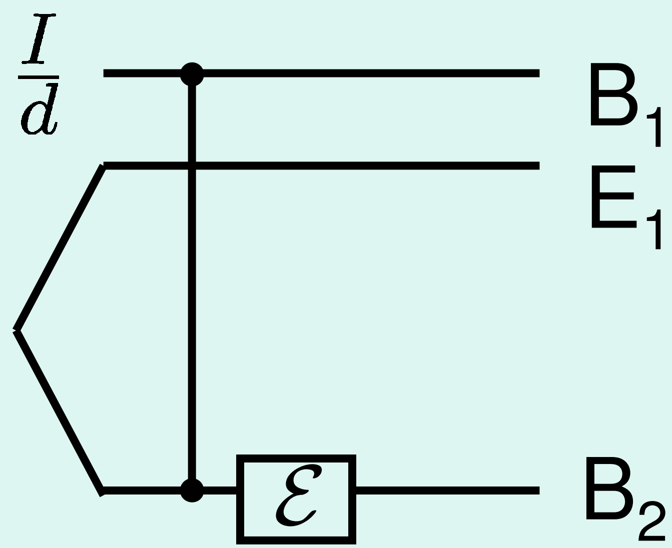
Choice Channels and Degenerate Codes

- Let's look at the states generated by using a choice channel based on the Rocket Channel. Focus on a fixed U and V.



Choice Channels and Degenerate Codes

- Let's look at the states generated by using a choice channel based on the Rocket Channel. Focus on a fixed U and V.



From this we get $\rho_{B_1B_2} = \frac{I}{d} \otimes \mathcal{E}(\frac{I}{d})$ and $\rho_{E_1E_2} = I \otimes \mathcal{E}(\sum_i \frac{1}{d}|i\rangle\langle i| \otimes |i\rangle\langle i|)$
As a result, $I(B_1;B_2) = 0$ but $I(E_1;E_2) = (1/2)\log d$ (compare: degradable case). This correlation between E_1 and E_2 means that because of our choice of input state, the effective Kraus operators of $T \otimes T$ are not just the tensor product of the individual Kraus operators---we've made them correlated even though the channels are independent to begin with. Put another way: E_1 and E_2 are correlated with each other rather than with the data (which would cause noise).

Summary

- Regularization comes from the nonadditivity of a correlation measure that we'd like to use to characterize a capacity. If we have to regularize, we can't calculate the capacity.
- Focused on the (non)additivity of coherent information, which is relevant for the capacity for quantum communication. This is related to degenerate quantum codes.
- Degradable channels are a nice class for which we get additivity of coherent information, and therefore can calculate the capacity: erasure, amplitude damping, etc.
- Similar thing happens when you're interested in classical communication: the nonadditivity of Holevo information means we have to regularize to get the capacity $\rightarrow$ No capacity formula in general. In special cases you can get lucky.
- Regularization is qualitatively different from nonadditivity of capacities from previous talk: Regularization is about whether we can compute the capacity, while nonadditivity is about whether different channels interact synergistically. (there are connections, though…)

Some things I didn't cover

- There's an intriguing relationship between quantum capacities and classical broadcast channels: in fact, degradability comes from the classical notion of a degraded broadcast channel.
- Nonadditivity of Holevo information: just like coherent information, the Holevo information is nonadditive---we understand this even less (basically one set of examples, nothing constructive)! Note, however, that we don't know whether the classical capacity is additive.
- A lot of these things get even stronger if, instead of asymptotically vanishing error, you go to the zero-error case (see Toby Cubitt's talk).
- Multiuser settings: Multiple Access Channels, Broadcast channels, etc. Very few single-letter examples. Since the single-user results are usually regularized, this isn't such a surprise.

Where to learn more

At the conference:

- Will Matthews' talk about zero-error capacity of a classical channel with entanglement/no-signaling assistance
- Toby Cubitt talks about zero-error superactivation

On the arXiv:

- Strong Subadditivity: Effros: arXiv:0802.1234
- Acheivability of coherent information: Hayden- Horodecki-Yard-Winter: quant-ph/0702005
- Degradable Channels: Devetak& Shor: quant-ph/0311131, Wolf&Perez-Garcia quant-ph/0607070, Cubitt, Ruskai, Smith: arXiv:0802.1360
- Nonadditivity of coherent information: Shor-Smolin '96, Smith-Yard '08, Smith-Smolin '08/09

THANK YOU