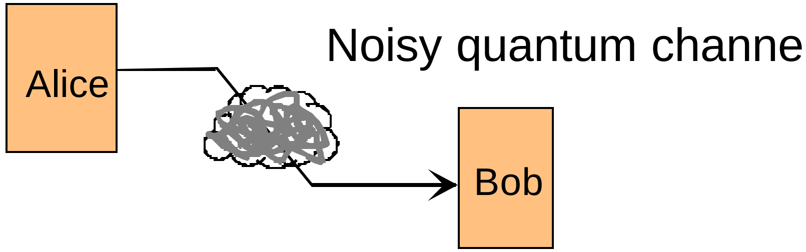
Recent progress on the

Quantum Reverse Shannon Theorem,

by Andreas Winter and Igor Devetak ,
urged on by C.H. Bennett (*IBM Research Yorktown*)
building on previous work of Peter Shor and Aram Harrow

SUNY Stony Brook 28 May 03

Multiple Capacities of Quantum Channels



Q plain quantum capacity = qubits faithfully trasmitted per channel use, via quantum error correcting codes

C plain classical capacity = bits faithfully trasmitted per channel use

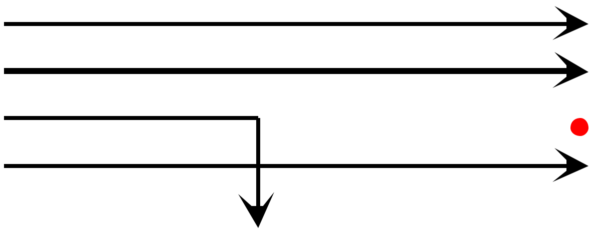
$\boldsymbol{Q_2}$ classically assisted quantum capacity, i.e. qubit capacity in the presence of unlimited 2-way classical communication, (e.g. using entanglement distillation and teleportation)

$\boldsymbol{C_E}$ entanglement assisted classical capacity i.e. bit capacity in the presence of unlimited prior entanglement between sender and receiver.

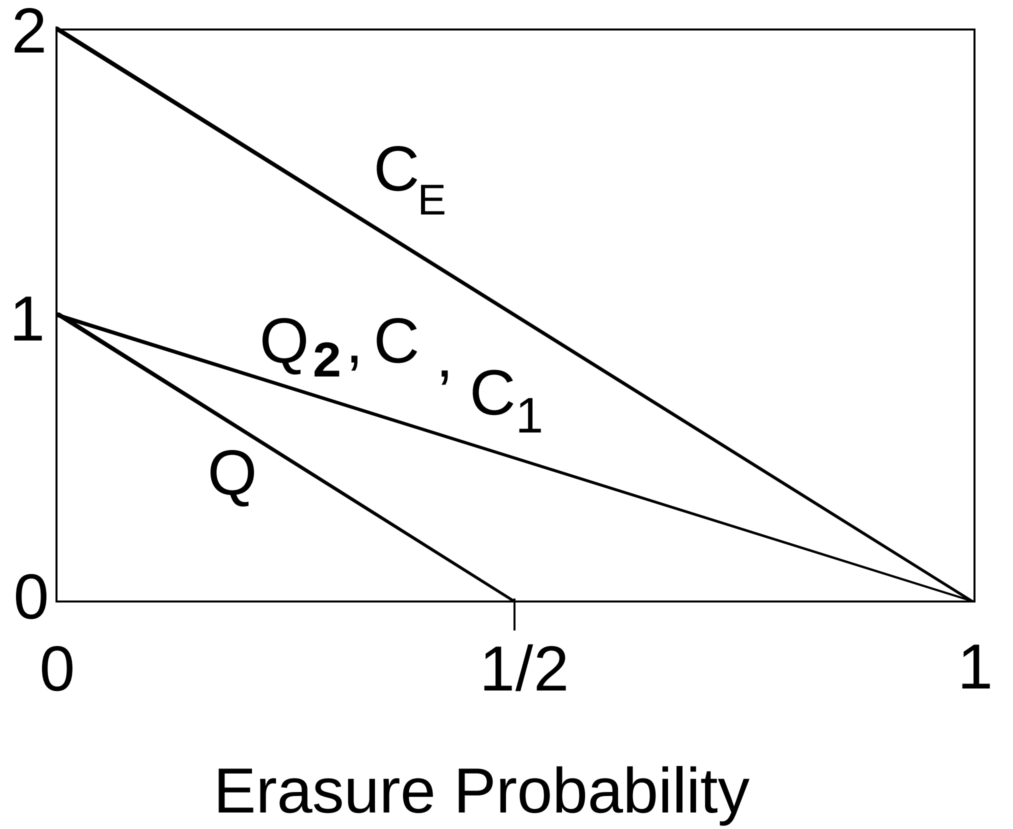
(+ special capacities, eg with restricted encoding/decoding C_{11}, C_{1A})

Quantum Erasure Channel

input qubit sometimes lost



Capacities of Quantum Erasure Channel



Inequalities among major capacities

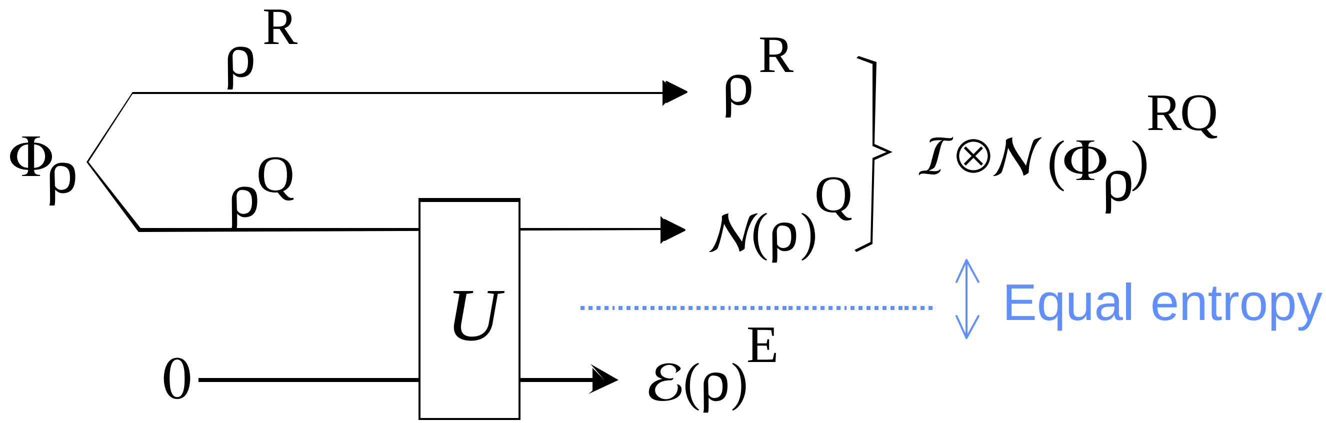
All inequalities may be = or < depending on channel

$$Q \leq Q_2 \leq C \leq C_E = 2Q_E$$

conjectured

by definition

by teleportation
and superdense
coding



= if max Entropy output is additive Shor '03

Entropic quantities related to channel capacities.

Shor '02

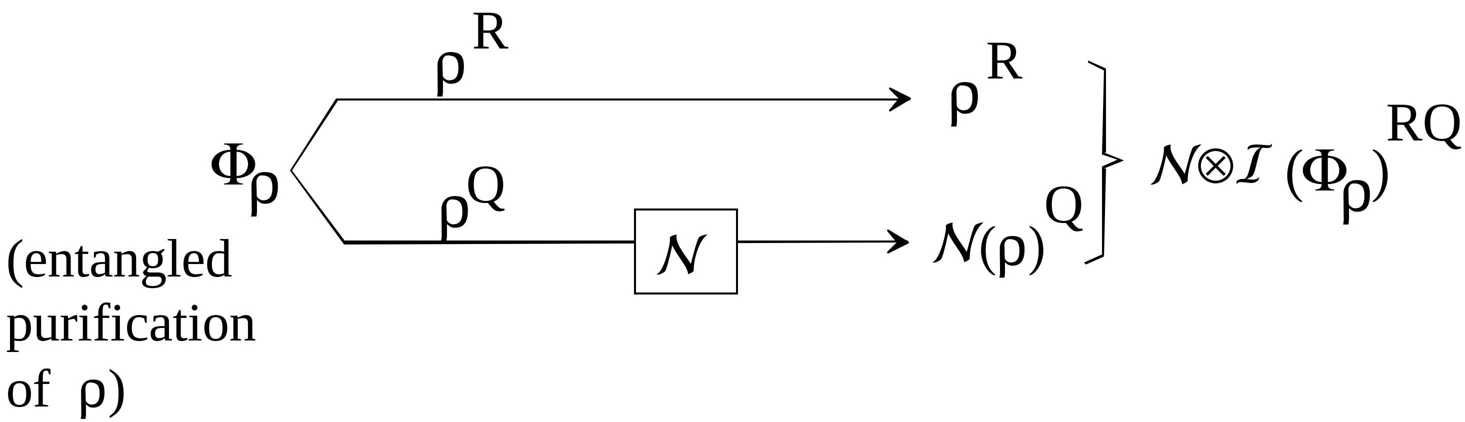
C =? Holevo capacity = $\max\limits_{\{p_i,\rho_i\}} \; S(\mathcal{N}(\rho)) - \Sigma p_i\, S(\mathcal{N}(\rho_i))$

Q = Coherent Information = $\lim\limits_{n\to\infty} \max\limits_{\rho} \; S(\mathcal{N}(\rho)) - S(\mathcal{E}(\rho))$

C_E = Quantum Mutual Info. = $\max\limits_{\rho} \; S(\rho) + S(\mathcal{N}(\rho)) - S(\mathcal{E}(\rho))$

$Q_2 \approx$ Distillable entanglement = $\max\limits_{\rho} D(\mathcal{I}\otimes\mathcal{N}(\Phi\rho))$ = ?

(LOCC-distillable entanglement *D* has no simple expression, may be nonconvex)



$$C_E(\mathcal{N}) \;=\; \max_\rho \quad S(\rho) \;+ S(\mathcal{N}(\rho)) \;-\; S(\mathcal{N}\otimes\mathcal{I}(\Phi_\rho))$$

Entanglement-Assisted capacity C_E of a quantum channel $\mathcal{N}$ is equal to the maximum, over channel inputs ρ, of the input (von Neumann) entropy plus the output entropy minus their "joint" entropy (more precisely the joint entropy of the output and a reference system entangled with the late input) (BSST 0106052, Holevo 0106075).

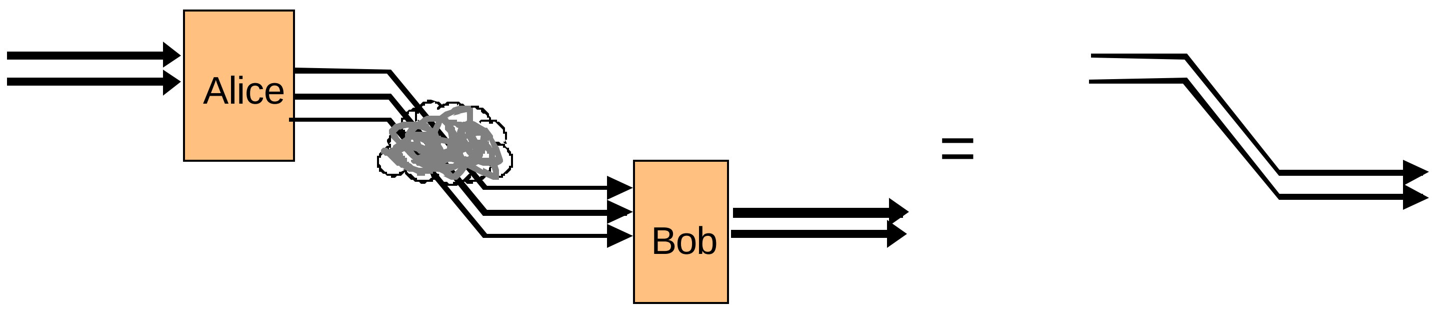
Thus, in retrospect, entanglement-assisted capacity is the natural quantum generalization of the classical capacity of a classical channel.

Simplification: $C_E = 2Q_E$ for all channels, by teleportation & superdense coding.

Classical Reverse Shannon Theorem (0106052)

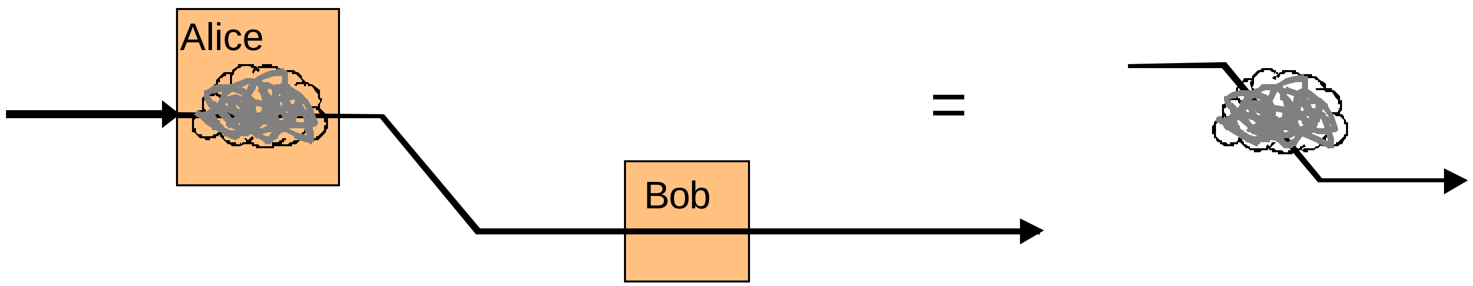
Classical Shannon Theorem:

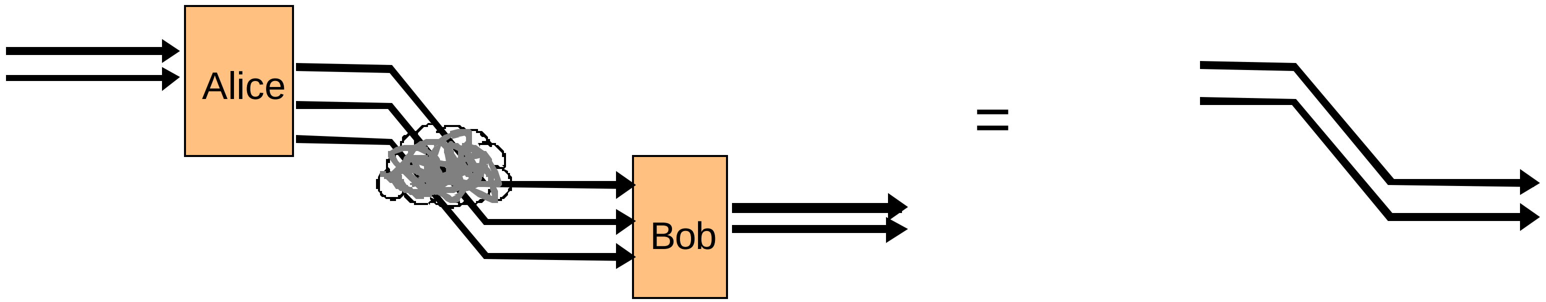
A noisy channel can simulate a noiseless channel



Homer Simpson's Reverse Shanon's Theorem:

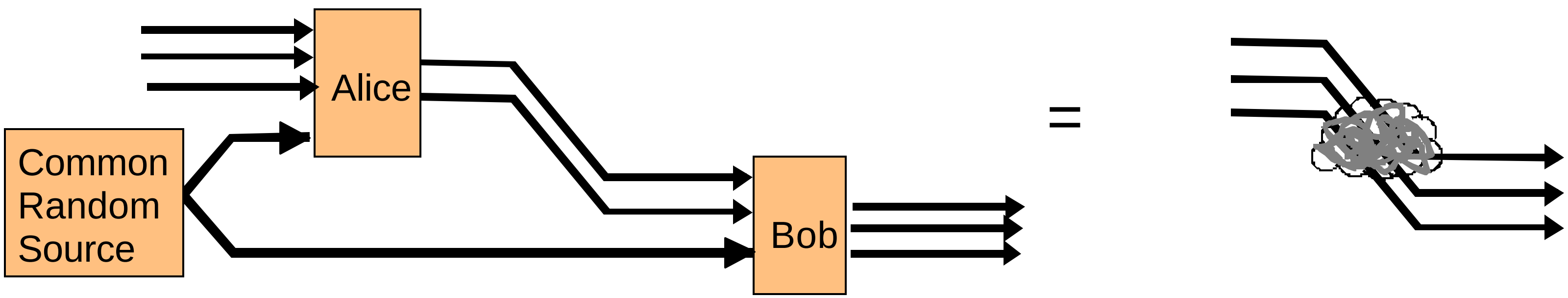
A noiseless channel can simulate a noisy channel.



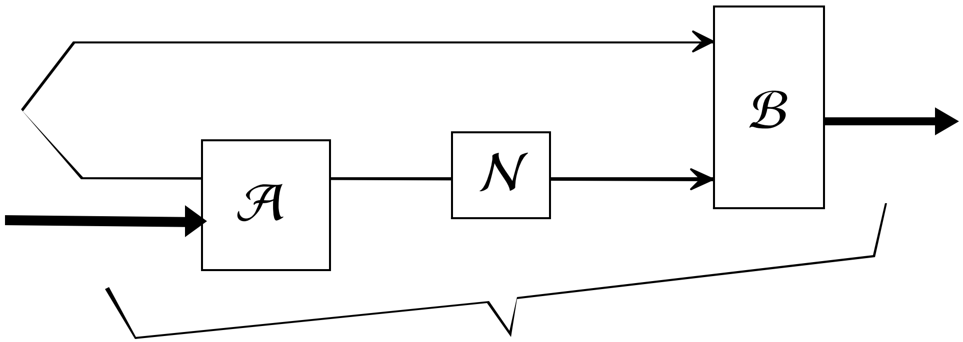


A Better Reverse Shannon Theorem *(quant-ph/0106052)*

In the presence of shared random information between sender and receiver,
a noiseless channel can asymptotically simulate a noisy one ***of equal capacity*****.**

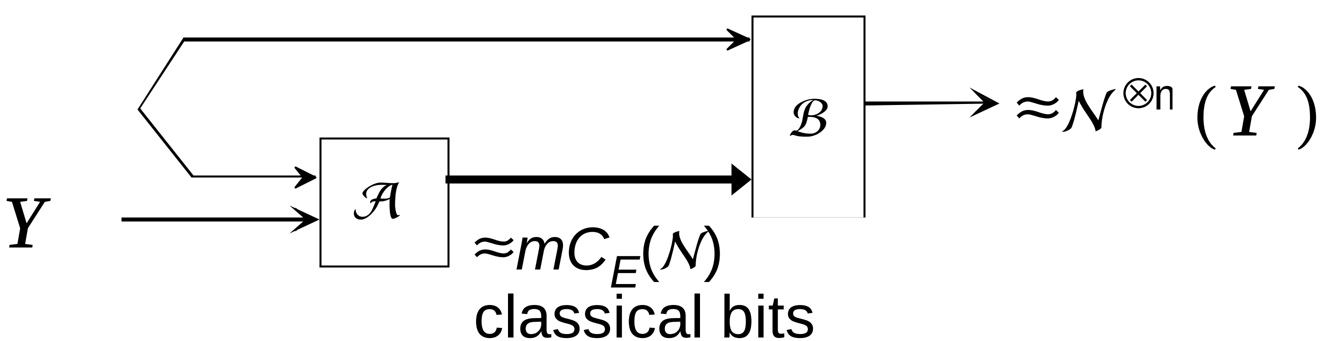


Therefore, in the presence of shared random information,
all classical noisy channels are asymptotically equivalent.

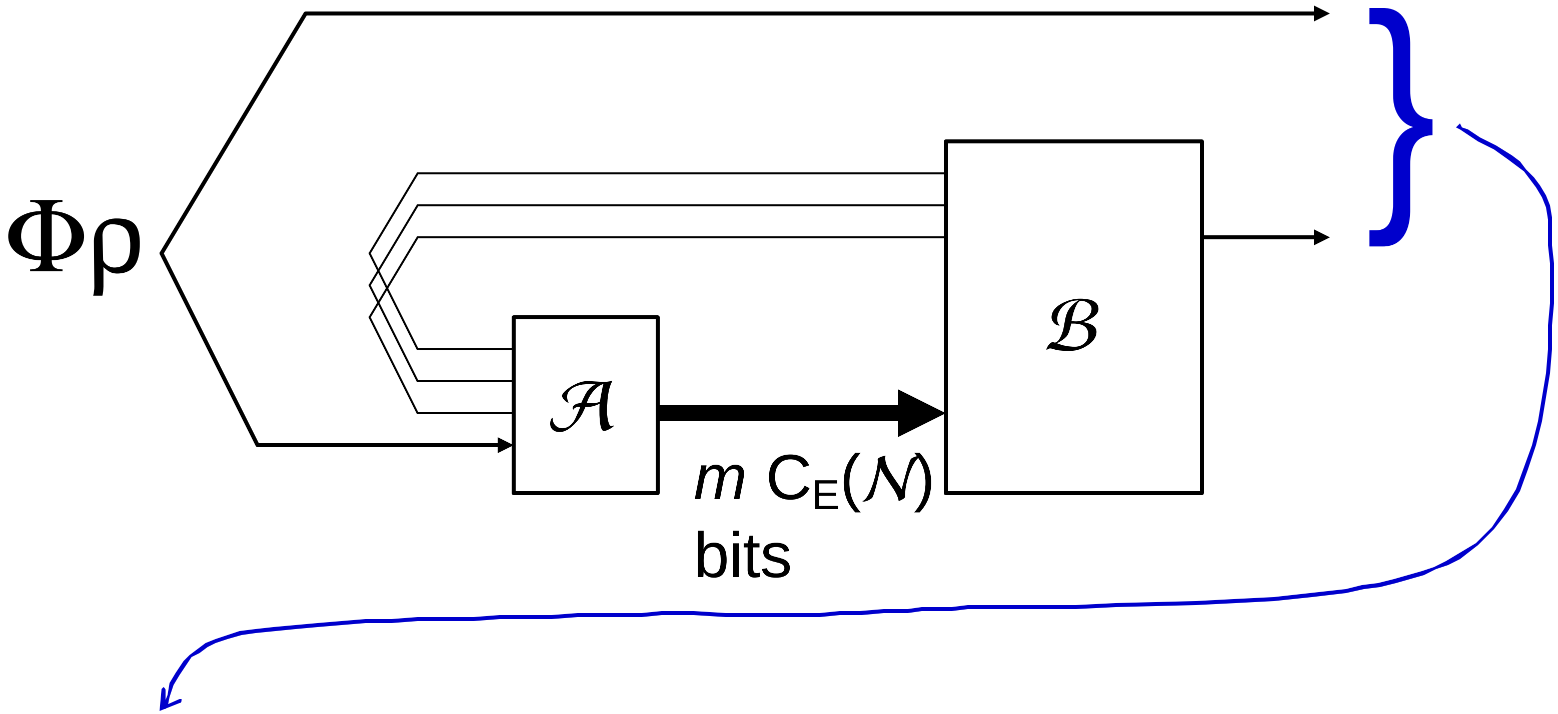


$C_E(\mathcal{N})$

The complicated theory of quantum channel capacity would be greatly simplified if the Quantum Reverse Shannon Theorem (QRST) were true: any quantum channel can be asymptotically simulated by prior entanglement and an amount of classical communication equal to its entanglement assisted capacity. Then, in a world full of entanglement, all **quantum** channels would be qualitatively equivalent, and quantitatively could be characterized by a single parameter.



More generally, we should demand high fidelity on entangled purifications of a mixed state input ρ



Output of simulation, including reference system, should have high fidelity with respect to $(\mathcal{N}\text{Ä}\mathcal{I})^{\otimes m}(\mathrm{F}_{\mathrm{r}})$, the output on the same input of m copies of the channel being simulated.

Last year Shor showed that the QRST holds for all (quantum discrete memoryless) channels when their inputs are drawn from a known fixed distribution ρ. This is the quantum analog of a classical IID source. Recent work of Devetak and Winter has generalized this to non-IID sources, known or unknown, provided the source is not entangled between channel inputs.

For many channels, the QRST holds for arbitrary sources even if the inputs are allowed to be entangled across multiple instances of the quantum channel being simulated.

The ability to properly simulate a completely general source is important because, for a channel simulation to be considered faithful, it ought to accurately simulate what the channel would do even on atypical inputs which a malicious adversary might send to expose the weaknesses of the simulation.

Kinds of sources:

Tensor Power (analogous to classical IID): $\mathrm{r} = \mathrm{r}^{\otimes m}$

Tensor Product: $\mathrm{r} = \mathrm{r}_1 \otimes \mathrm{r}_2 \otimes \mathrm{r}_3 \otimes \ldots$
with each factor in H_{in}

(Arbitrary pure: y = a general pure state in $H_{\mathrm{in}}^{\text{Ä}m}$)

Most general: any pure state Y in $H_{\mathrm{in}}^{\text{Ä}m} \text{Ä} H_{\mathrm{in}}^{\text{Ä}m}$
(the worst an adversary could send)

m channel inputs

Purifying reference system

Winter's Measurement Compression

Given a density matrix ρ and a POVM $\boldsymbol{a} = \{a_j\}$, define the one-shot output probabilities $\lambda_j = \mathrm{Tr}\, \rho a_j$, and the square root ensemble $\rho_j = (\sqrt{\rho})\, a_j\, (\sqrt{\rho}) / \lambda_j$ realizing ρ. Then for any tolerance $\varepsilon > 0$, there exists a block size l and a POVM $\boldsymbol{B}$, which is a good approximation to $\boldsymbol{A} = \boldsymbol{a}^{\otimes l}$, and where $\boldsymbol{B}$ can be expressed as a convex combination $\boldsymbol{B} = \Sigma_\nu x_\nu \boldsymbol{B}^\nu$ of constituent POVMs $\boldsymbol{B}^\nu$ each having at most M outcomes, where $\log M \approx l\,(S(\rho) - \Sigma_j \lambda_j S(\rho_j))$ is the Holevo information of the square root ensemble. The approximation is good in the sense that for any entangled purification Φ of $\rho^{\otimes l}$,

$F((\boldsymbol{A} \otimes I)\, \Phi, (\boldsymbol{B} \otimes I)\, \Phi) > 1 - \varepsilon$.

On any tensor power source ρ, the POVM $\boldsymbol{a}$, regarded as QC channel, can be asymptotically simulated by shared randomness and an amount of forward classical communication approaching the quantum mutual information of $\boldsymbol{a}$ on ρ.

= QRST for QC channels on known IID sources

$$\mathrm{QMI}\,(\boldsymbol{a}, \rho) \equiv S(\rho) + S(\boldsymbol{a}(\rho)) - S\,(\boldsymbol{a} \otimes \mathcal{I}(\Phi_\rho)).$$

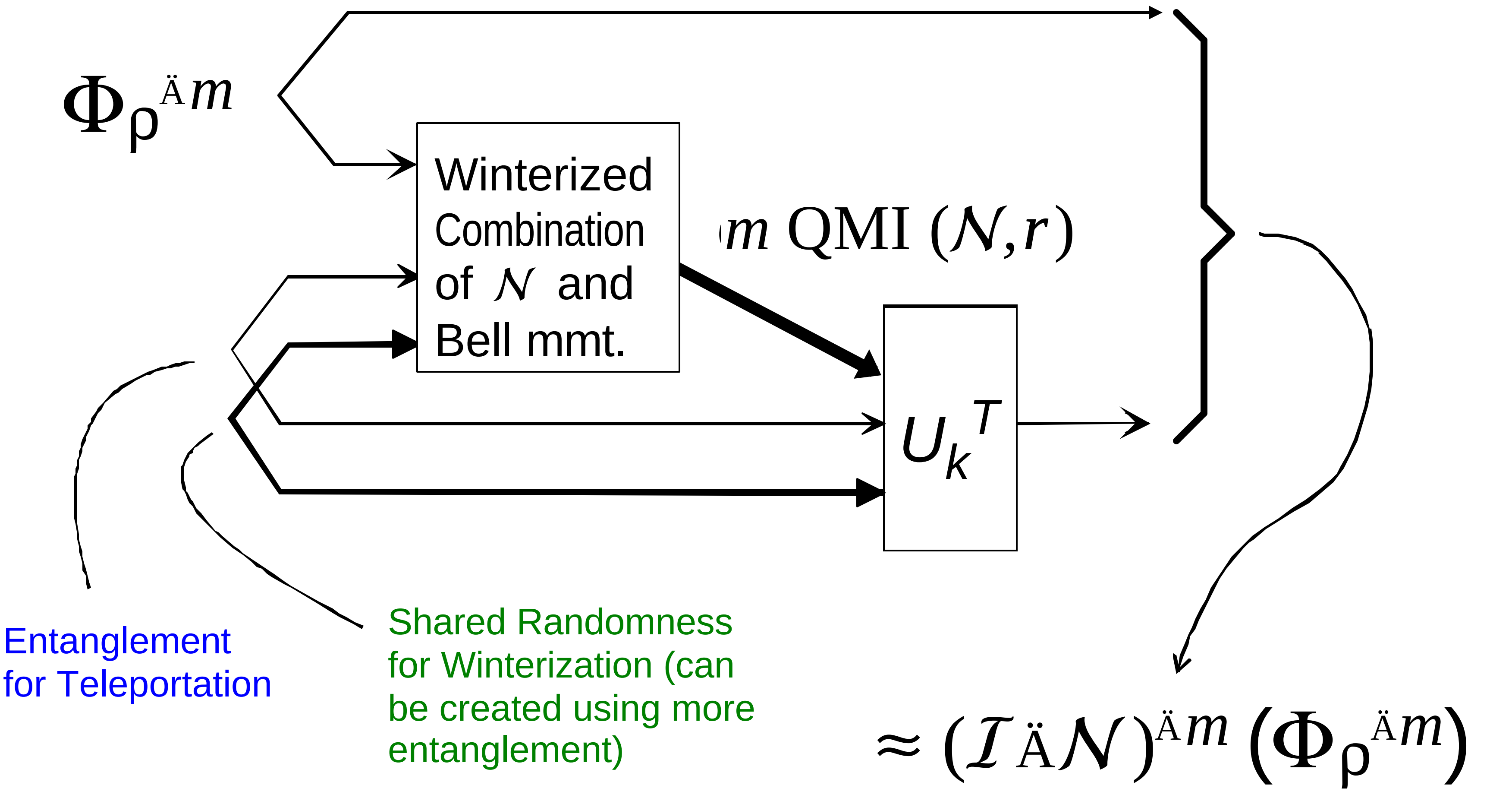
$$= S(\boldsymbol{a}(\rho)) + \Sigma_j \lambda_j S(\rho_j)$$

Sketch of Shor's proof of QRST for tensor power sources, using Winter's compression theorem. Alice's wants to simulate a general noisy channel $\mathcal{N}$, using shared entanglement and as little classical communication to Bob as possible. Let $\mathcal{N}$ be defined by the Kraus operators $\{N_k : k=1\ldots\delta\}$ so on input state ρ the channel output is $\Sigma_k N_k \rho N_k^\dagger$. Let Φ_{in} and Φ_{out} denote projectors onto maximally entangled states sized to the input and output dimensions of $\mathcal{N}$. Let U_j be d_{out} dimensional generalized Pauli matrices.

Generalized Teleportation: Alice performs a POVM with elements $(I\otimes U^*_j N^*_k)\,\Phi_{\text{in}}\,(I\otimes N^T_k U^T_j)$ on the input and her half of a specimen of Φ_{out}, after which she tells Bob only j, the index of which Pauli she performed. He undoes the Pauli, and is left with $\mathcal{N}(\rho)$. This uses 2 2log d_{out} bits of classical communication.

Measurement compression: For large block size m, Alice and Bob approximate this POVM by another with an intrinsic cost of $m\,(\text{QMI}\,(\mathcal{N},\rho)) + o(m)$

Overall picture



Lower bounds. For any channel and any tensor power source the entanglement assisted cost of simulating the channel on that source must be at least the QMI of the channel-source combination. Otherwise causality would be violated. In particular, the cost of simulating a channel on an unconstrained source must be at least C_E

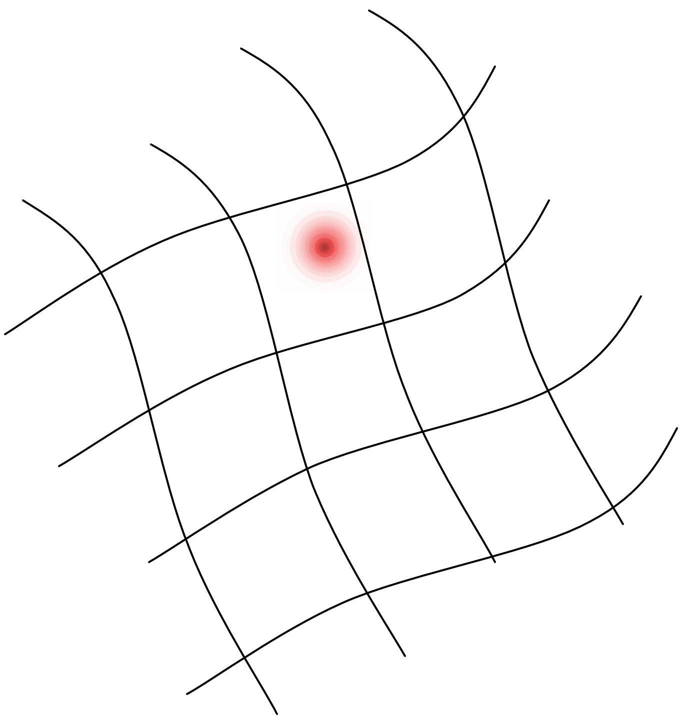
This establishes QRST for a general channel on known IID source.

For a general channel on an unknown IID source, we use gentle tomography on a large block of inputs to estimate the source without disturbing it significantly.

For a CQ channel on an arbitrary source, Alice performs the initial C part of the channel on a large block of m inputs and makes a copy of the results. These results will be unentangled between channel instances, but may not be IID. Using $o(m)$ bits, Alice tells Bob the frequency distribution (type class) of the measured results and they then simulate the full CQ channel on this type class. (Alternatively, this may be viewed as remote state preparation of mixed states which can be done at the cost of the Holevo information of the ensemble, which equals the QMI.)

For a Bell-diagonal channels on arbitrary sources, the noisy quantum channel is directly equivalent to teleportation through a noisy classical channel, which can be simulated using the classical reverse Shannon theorem.

To extend QRST to an unknown tensor power source we use gentle tomography to estimate ρ from a large number *m* of copies of ρ without much disturbing the global state. (Hayashi & Matsumoto 0202001, Presnell & Jozsa EQIS02, A. Harrow in prep).



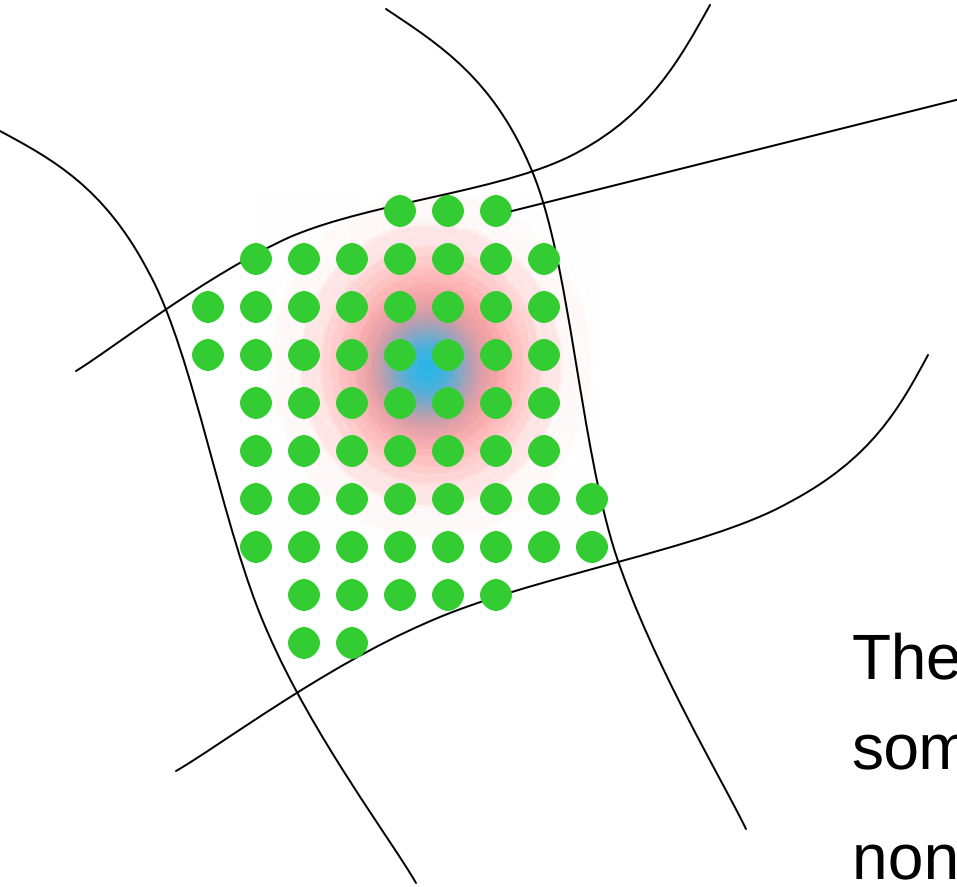
This may be viewed roughly as choosing a random mesh on the parameter space of ρ coarse enough $(\propto 1/\sqrt{m})$ so that for any ρ, a measurement on ρ^m, of *which cell* the average falls into will almost always yield the same result. This measurement, when conducted coherently, will therefore scarcely disturb the global state.

With gentle tomography, we get an estimate ρ_{est} of the unknown density matrix ρ and its quantum mutual information.
But unfortunately the typical subspace of $\rho_{est}^{\text{Ä}m}$ has little overlap with the typical subspace of the true $\rho^{\text{Ä}m}$.

So (crudely speaking): we do a compressed teleportation using a version of Winter's theorem designed not for the estimated

source $\rho_{est}^{\text{Ä}m}$, but rather for a (non tensor power) source ρ_{cell} corresponding to the average over a finer mesh $\rho_1^{\text{Ä}m}\ldots\ \rho_N^{\text{Ä}m}$ of density matrices in the same tomographic cell as ρ_{est} .

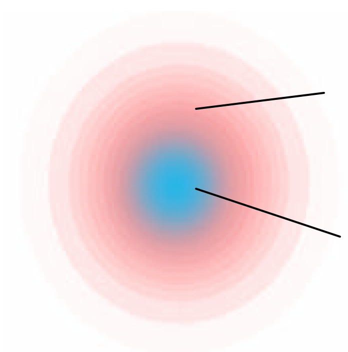
This finer mesh has only polynomially many (in *m*) points but is still fine enough so that the true density matrix source $\rho^{\text{Ä}m}$ has good fidelity with at least one of the $\rho_k^{\text{Ä}m}$.



Fine mesh of $N = poly(m)$ density matrices ρ_k covering the original tomographic cell

$$\rho_{\mathrm{cell}} = (1/N) \sum \rho_k{}^{\text{Ä}m}$$

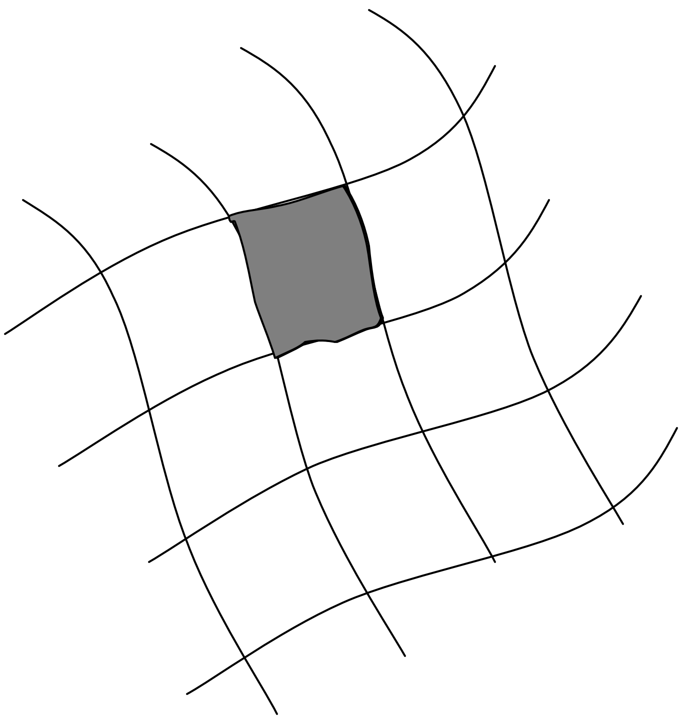
The true $\mathrm{r}^{\text{Ä}m}$ has high fidelity to some $\mathrm{r}_k{}^{\text{Ä}m}$, which in turn has nonnegligible participation in ρ_{cell}.

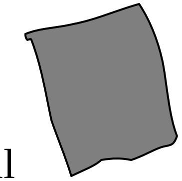
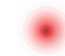


region of tomographic uncertainty of ρ

region of high fidelity between $\rho^{\text{Ä}m}$ and $\rho^{\text{¢}\,\text{Ä}m}$

Finally, we use the fact that the fidelity of measurement compression approaches 1 exponentially with increasing block size, for any forward communication rate R exceeding the QMI.



Even though ρ_{cell}  consists mostly of states atypical of $\rho^{\text{Ä}m}$  , nevertheless, for any forward classical communication rate R exceeding the QMI on ρ_{est} , the

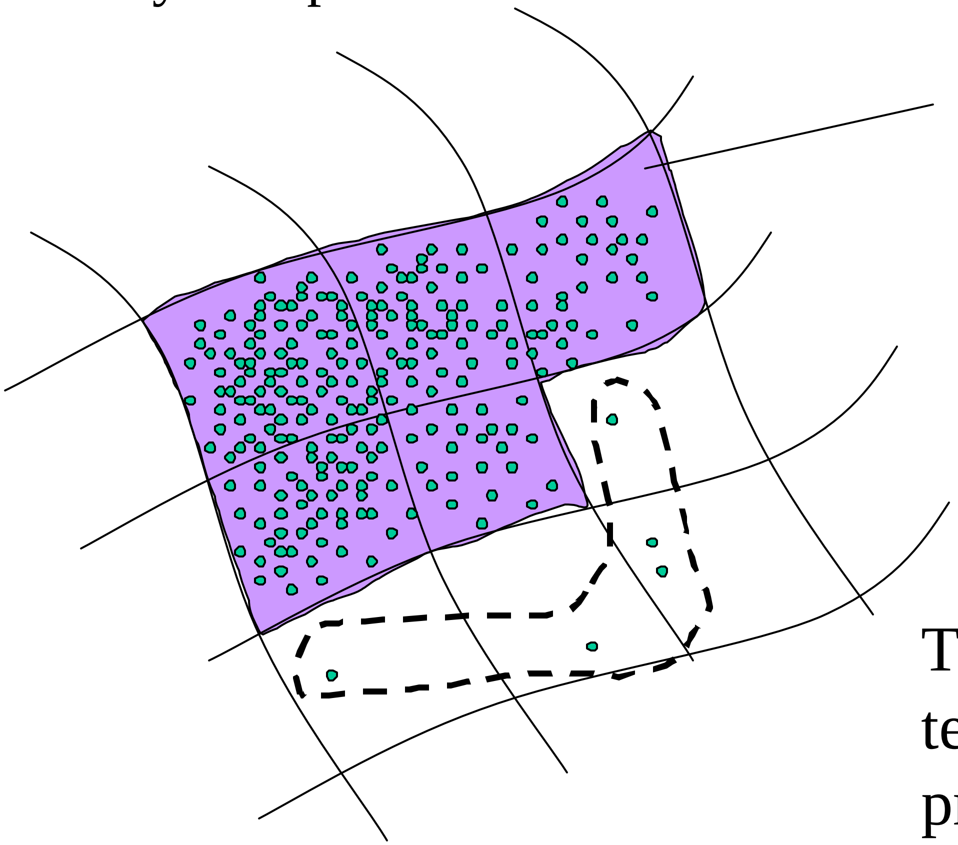
fidelity (1–ε) of simulation on ρ_{cell} is so good that it must be pretty good on $\rho^{\text{Ä}m}$ also, approaching unity in the limit of large m.

$$\rho_{\mathrm{cell}} = \delta(m)\ \rho_k{}^{\text{Ä}m} + (1 - \delta(m))\ \rho_{\mathrm{other}}$$

with $\delta(m) \to 0$ subexponentially. But $\varepsilon(m) \to 0$ exponentially, so $\varepsilon(m) / \delta(m) \to 0$.

Extension to a known tensor product source: $\mathbf{r} = r_1 \otimes r_2 \otimes r_3 \otimes \ldots$

Divide the parameter space into coarse cells and note in which cell each known tensor factor r_i falls. For large m, most r_i will fall into heavily occupied cells.



Each heavily occupied cell is coded by applying measurement compression to the known tensor product of density matrices it contains, at a cost of the QMI for that cell.

The few remaining points are then teleported exactly, without compressing, at negligible extra cost.

$$\text{Total cost} = \Sigma_i \, \text{QMI}(r_i) = \text{QMI}(\mathbf{r})$$

Extension to an ***Unknown*** tensor product source

$$r = r_1 \otimes r_2 \otimes r_3 \otimes \ldots\ r_m$$

(Assume for now that the factors are drawn from a finite set that does not increase with increasing block size *m*. This assumption will be removed later).

Define ρ_{permuted} as an equal mixture of randomly permutated versions of ρ, and r_{ave} as the average of the single-symbol density matrices in ρ. Unfortunately ρ_{permuted} 1 $r_{\text{ave}}{}^{\text{Ä}m}$.

But fortunately the fractional participation of ρ_{permuted} in $r_{\text{ave}}{}^{\text{Ä}m}$ decreases only polynomially with *m*. Thus a good simulation on $r_{\text{ave}}{}^{\text{Ä}m}$ guarantees a good simulation on ρ_{permuted} and therefore (since the simulation is symmetric) on ρ.

If the unknown tensor factors do not come from a finite set,

it can be shown there is a state with high fidelity to ρ_{permuted} whose participation ration in $r_{\mathrm{ave}}^{\text{Ä}m}$ decreases more slowly (albeit still exponentially) than the rate of convergence of measurement compression.

Classical cost of entanglement assisted channel simulation

<table>
<tr><td>channel
source</td><td>Bell diagonal</td><td>Classical or CQ</td><td>General Channel</td></tr>
<tr><td>Known tensor power</td><td colspan="3" rowspan="3">Cost = quantum mutual information (QMI)
of source-channel combination
$\mathrm{QMI}\,(\mathcal{N},\rho) = S\,(\mathrm{r}) + S\,(\mathcal{N}(\mathrm{r})) - S\,(\mathcal{N}\text{Ä}\mathcal{I}(\mathrm{F}_{\mathrm{r}}))$</td></tr>
<tr><td>Unknown tensor power</td></tr>
<tr><td>Known tensor Product</td></tr>
<tr><td>Unknown tensor product (general separable source)</td><td rowspan="2">$\leq\ C_E$, sometimes equal</td><td rowspan="2">≤ QMI of collapsed source (after initial von Neumann mmt.) sometimes equal</td><td>≤ QMI of randomly permuted source, equal to m times the average QMI of the single-symbol sources</td></tr>
<tr><td>Unconstrained (inseparable) source</td><td>$\leq 2 \log \min\{d_{\mathrm{in}}, d_{\mathrm{out}}\}$
(crude teleportation bound)</td></tr>
</table>

Open questions:

Capacity relations, e.g. Q_2 £ C, Q_2 £ C_E ,

Additivity question for unassisted classical capacity C, entanglement of formation, or maximal output entropy.

Prove QRST for the most general source model, with inter-symbol entanglement, or else find a counterexample, i.e. a channel that, for some (entangled) source, requires more classical communication to simulate than the C_E of the channel.

The question of whether the QRST is violated with inter-source entanglement is reminiscent of, but will probably not help solve, the question of whether inter-symbol entanglement increases classical capacity (the famous additivity question).