Quantum Reverse Shannon Theorem

Resources Required
to simulate
a general quantum channel
on arbitrary inputs

Enlarged June 08
version of QIP 2007
talk in Brisbane,
Australia 2 Feb 07

Igor Devetak *(USC),* A. Harrow *(Bristol), P.* Shor *(MIT),*
A. Winter *(Bristol), &* CHB *(IBM)*

research supported by NSA, IBM, Bristol, USC, MIT

Outline

Classical Shannon Reverse Shannon Theorem: Given free shared randomness between sender and receiver, classical channels can efficiently and reversibly simulate one another

Quantum Channels and their Multiple Capacities

Entanglement-Assisted Capacity as the natural generalization of classical capacity

Quantum Reverse Shannon Theorem: simulating a general quantum channel using shared entanglement and noiseless classical or quantum communication

- on IID (tensor power) sources

- on general sources, sometimes requiring a stronger entanglement resource, such as entanglement embezzling states.

Simulating a Classical or Quantum Channel using limited or no shared randomness/entanglement: Relation to Wyner Common Information and Entanglement of Purification

$$C = \max_X [\, H(X) + \mathrm{H}(\mathcal{N}(X)) - \mathrm{H}(X,\mathcal{N}(X)) \,]$$

Shannon showed that a classical channel's capacity, operationally defined, is given by the maximum, over input distributions X, of the input:output *mutual information*.

The analogous quantum quantity was studied long ago by Cerf & Adami, but its operational significance remained unclear. (CA97 , Phys.Rev.Lett. **79**, 5194)

Shannon also showed that neither shared randomness nor back communication increases the capacity of a classical channel.

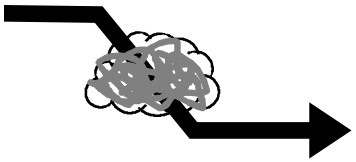
$$C_R = C_B = C$$

As Shannon also showed, shared randomness is useful for encryption. Another more recently discovered use for it is in the *Classical Reverse Shannon Theorem* (quant-h/0106052, 0208131) according to which, in the presence of shared randomness, the asymptotic capacity of a channel ***M*** to simulate another channel ***N*** is simply the ratio of their plain capacities.

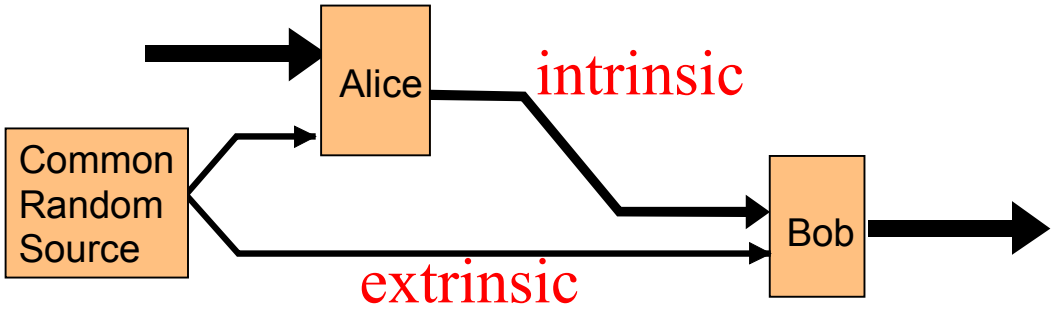
$$C_R(\boldsymbol{M},\boldsymbol{N}) \;=\; C(\boldsymbol{M}) \,/\, C(\boldsymbol{N})$$

(Without the shared randomness, the simulation could still be done, but it would not be efficient: there would be a net loss in using channel ***M*** to simulate ***N*** to simulate ***M*** again.)

In the large n limit, each noisy channel use



can be simulated by sending about $I(X:Y)$ noiseless "intrinsic" bits, which Alice chooses with the help of the input,



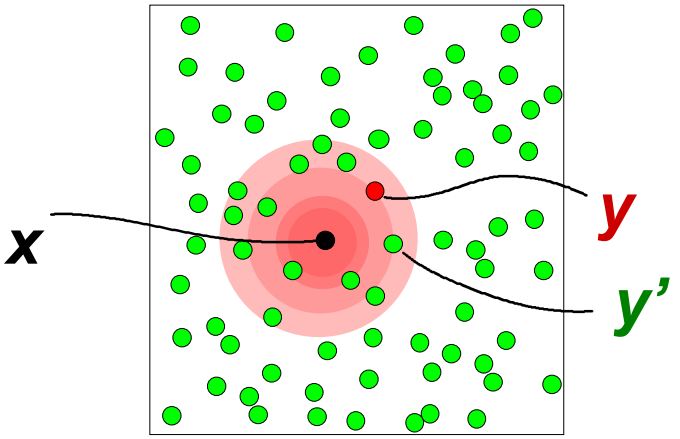
and about $H(Y|X)$ "extrinsic" random bits, which have nothing to do with the protocol input, and so can be preagreed before Alice receives the input.

Simulating a classical channel of capacity C

101010 input
110011 output

Input Type: 3 0's, 3 1's
Joint Type: 1 (0,0), 2 (1,1), 2 (0,1), 1 (1,0)

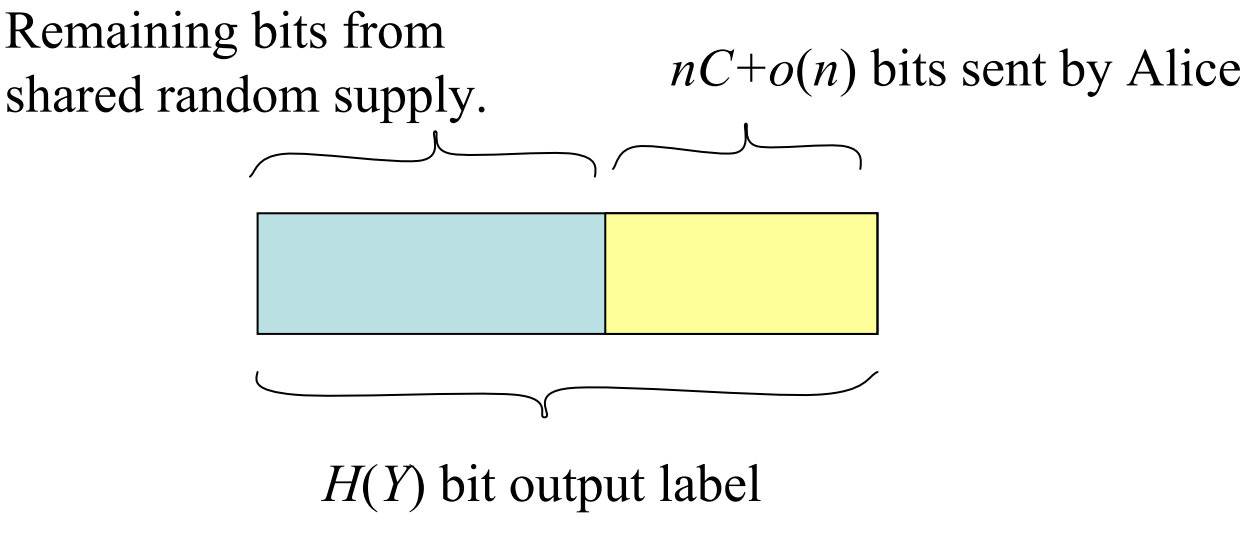
Receiving an n-character input **x**, Alice computes the input's *type* (letter frequency distribution), and simulates the channel locally to determine an output ***y*** and also the output type and *input:output joint type* to which (***x,y***) belongs. She communicates the output type to Bob using $O(\log n)$ bits.



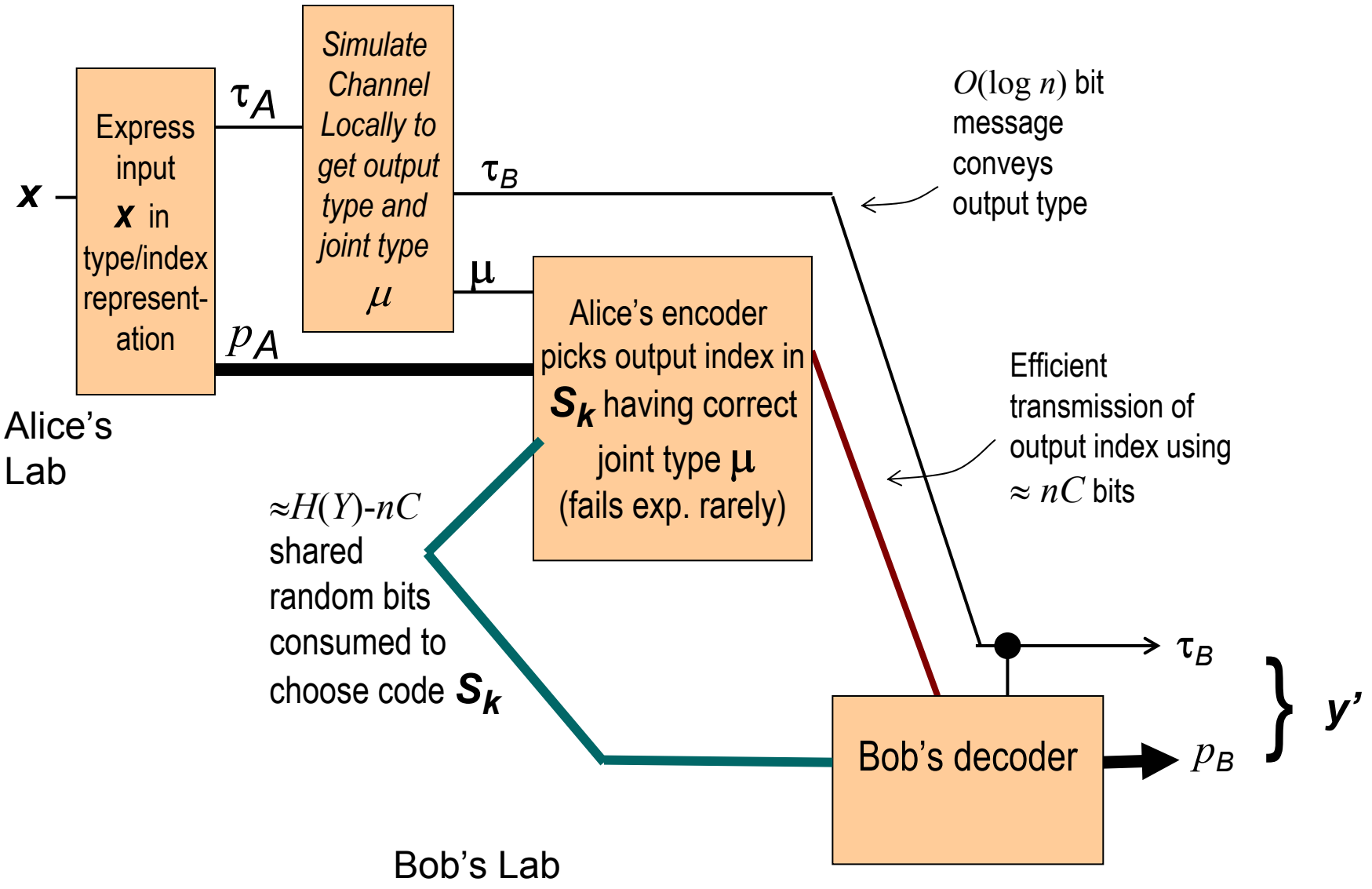
Alice and Bob have pre-agreed on a partition of the output space into random codes $\{S_1, S_2 \ldots\}$ each comprising $2^{nC+o(n)}$ strings. Using their shared randomness, they pick one of the codes S_k (green dots in the figure). Alice then uses $nC+o(n)$ bits of forward communication to specify a random word **y'** in S_k such that (**x**,**y'**) belongs to same joint type as (**x**,**y**) did. In the large n limit this **y'** will be correctly distributed with respect to ***x***, giving an asymptotically faithful simulation.

If no such **y'** exits, she tells Bob **y** directly, using $O(n)$ bits, but that happens exponentially rarely in the limit of large n.

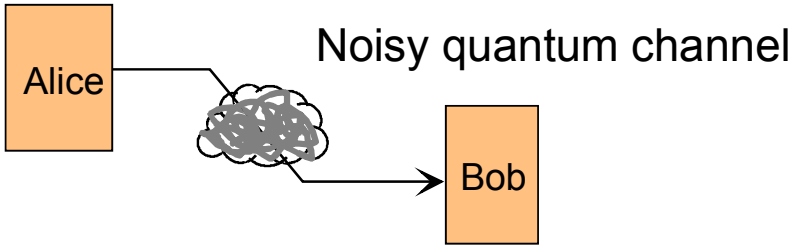
Bob's decoder can be very simple. Let the strings in the output space be labeled in a preagreed fixed random order by bit strings of length $H(Y)$. To get the label of the desired output ***y'*** , Bob simply tops up the $nC+o(n)$ bits he received from Alice with $H(Y)-nC-o(n)$ bits from their shared random supply.



Circuit for typewise simulation underlying the classical reverse Shannon theorem



Multiple capacities of Quantum Channels



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

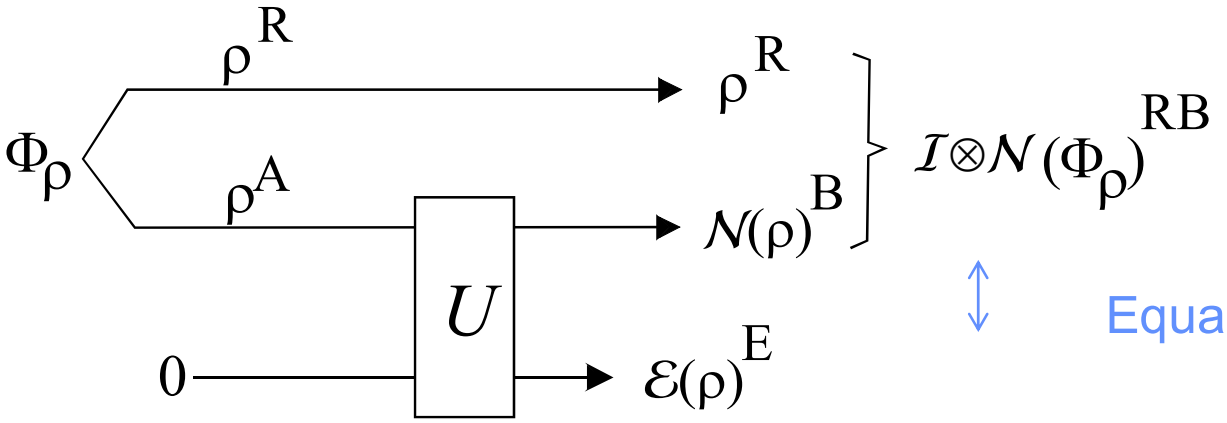
C plain classical capacity = bits faithfully trasmitted per channel use

Q_B quantum capacity assisted by classical back communication

Q_2 quantum capacity assisted by classical two-way communication (also C_B and C_2 -- ask me after class)

C_E entanglement assisted classical capacity i.e. bit capacity in the presence of unlimited prior entanglement between sender and receiver.

For quantum channels, these assisted capacities can be greater than the corresponding unassisted capacities.



Equal entropy

Entropic quantities related to channel capacities.

Famous Additivity problem

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

LSD

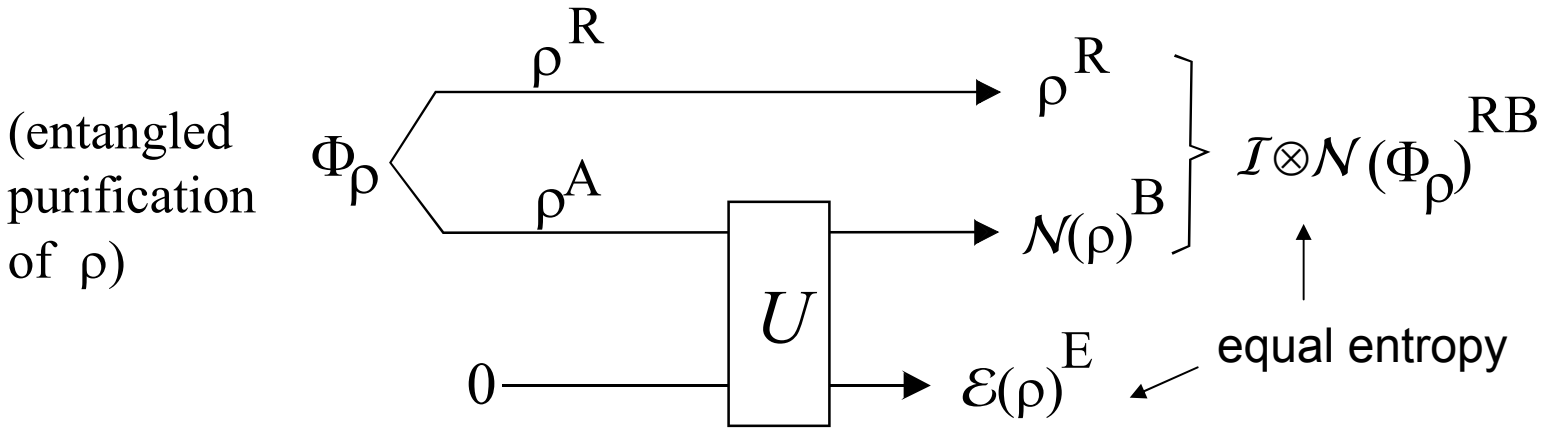
Q = Coherent Info. = $\lim_{n\to\infty} \max_{\rho} S(\mathcal{N}^{\otimes n}(\rho)) - S(\mathcal{E}^{\otimes n}(\rho))/n$

C_E = Quantum Mutual Info. = $\max_{\rho} S(\rho) + S(\mathcal{N}(\rho)) - S(\mathcal{E}(\rho))$ BSST '01

$Q_2 \approx$ Distillable entanglement = $\lim_{n\to\infty} \max_{\rho} D_2(\mathcal{I}\otimes\mathcal{N}^{\otimes n}(\Phi_\rho))/n$ = ??

(Unfortunately Q_2 has no simple expression, may be nonconvex)

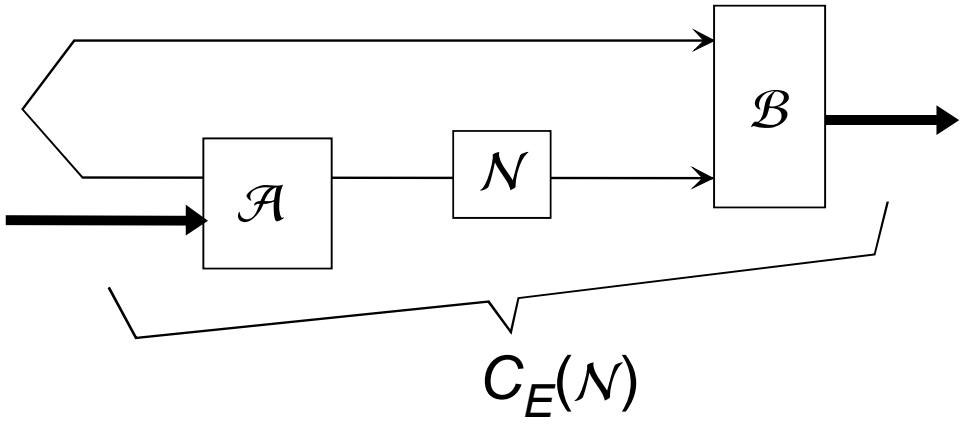
Even Q ~~may be~~ is nonconvex, ~~if~~ two channels with Q=0 can activate one another. (SY08)



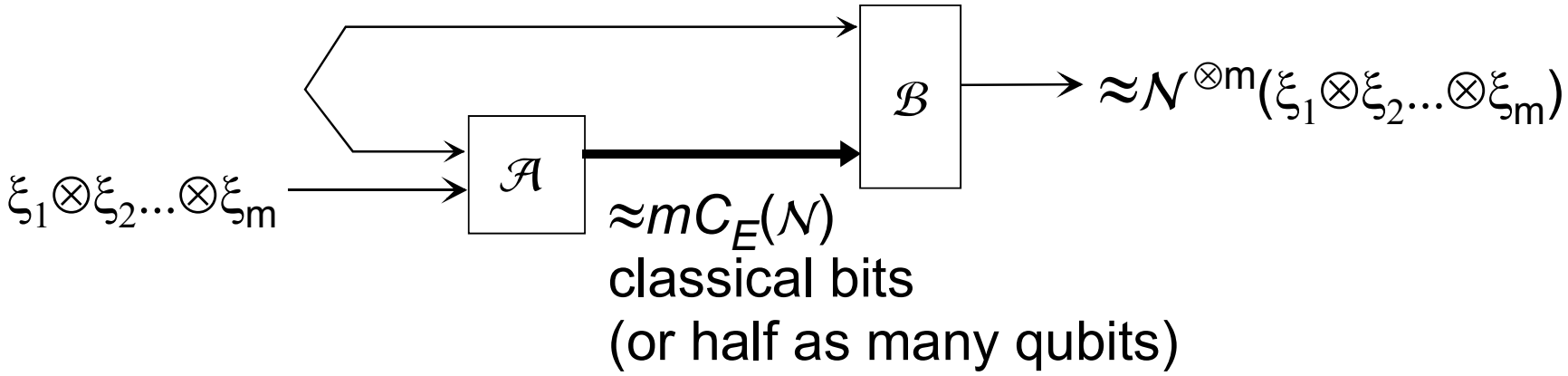
$$C_E(\mathcal{N}) = \max_\rho \left(S(\rho) + S(\mathcal{N}(\rho)) - S(\mathcal{E}(\rho)) \right)$$

In retrospect, ***entanglement-assisted capacity***, not plain classical capacity, is the natural quantum generalization of the classical capacity of a classical channel. Thus Shannon's great discovery was a simple expression for entanglement assisted capacity. He failed to appreciate this because for the classical channels he was thinking about, C and C_E happen to coincide.

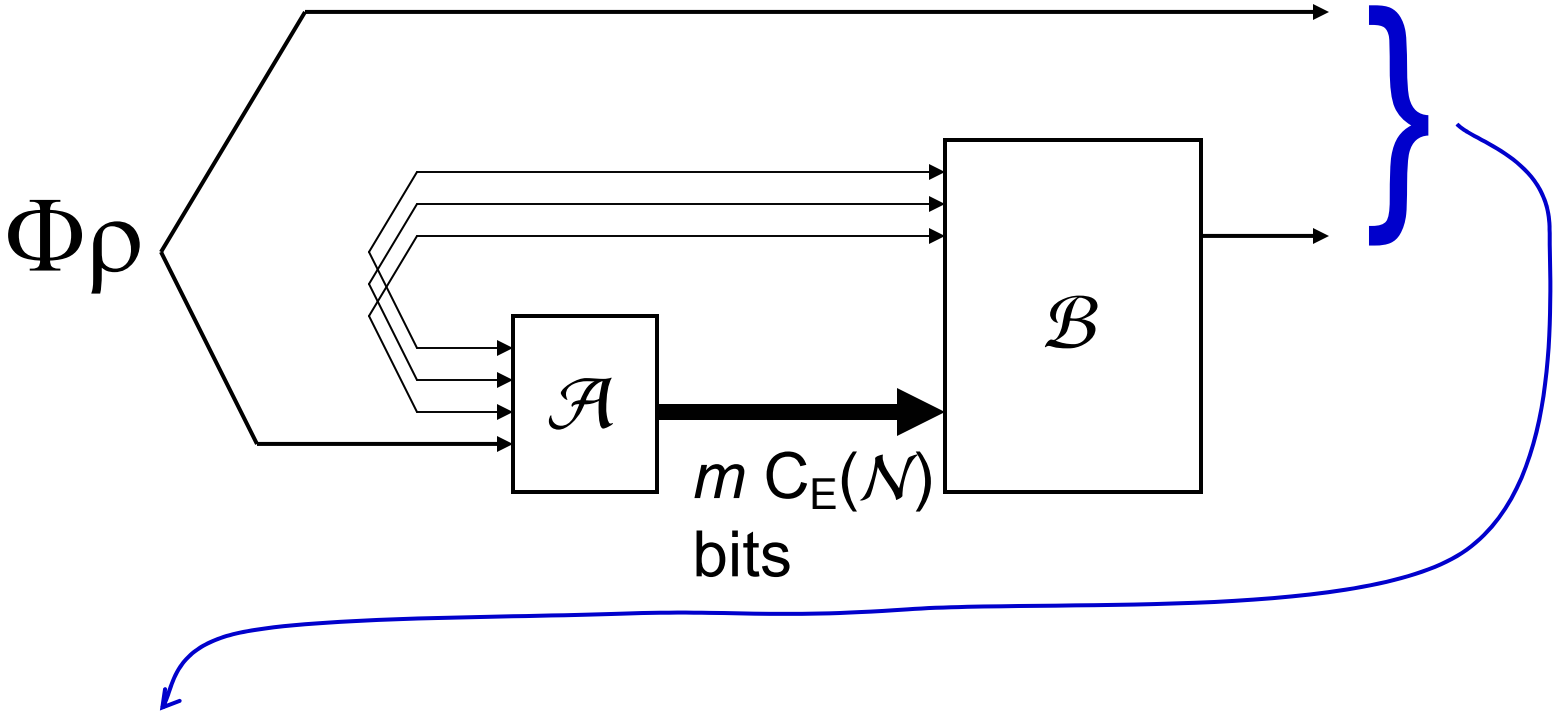
$Q_E = C_E / 2$ for all channels, by teleportation & superdense coding.



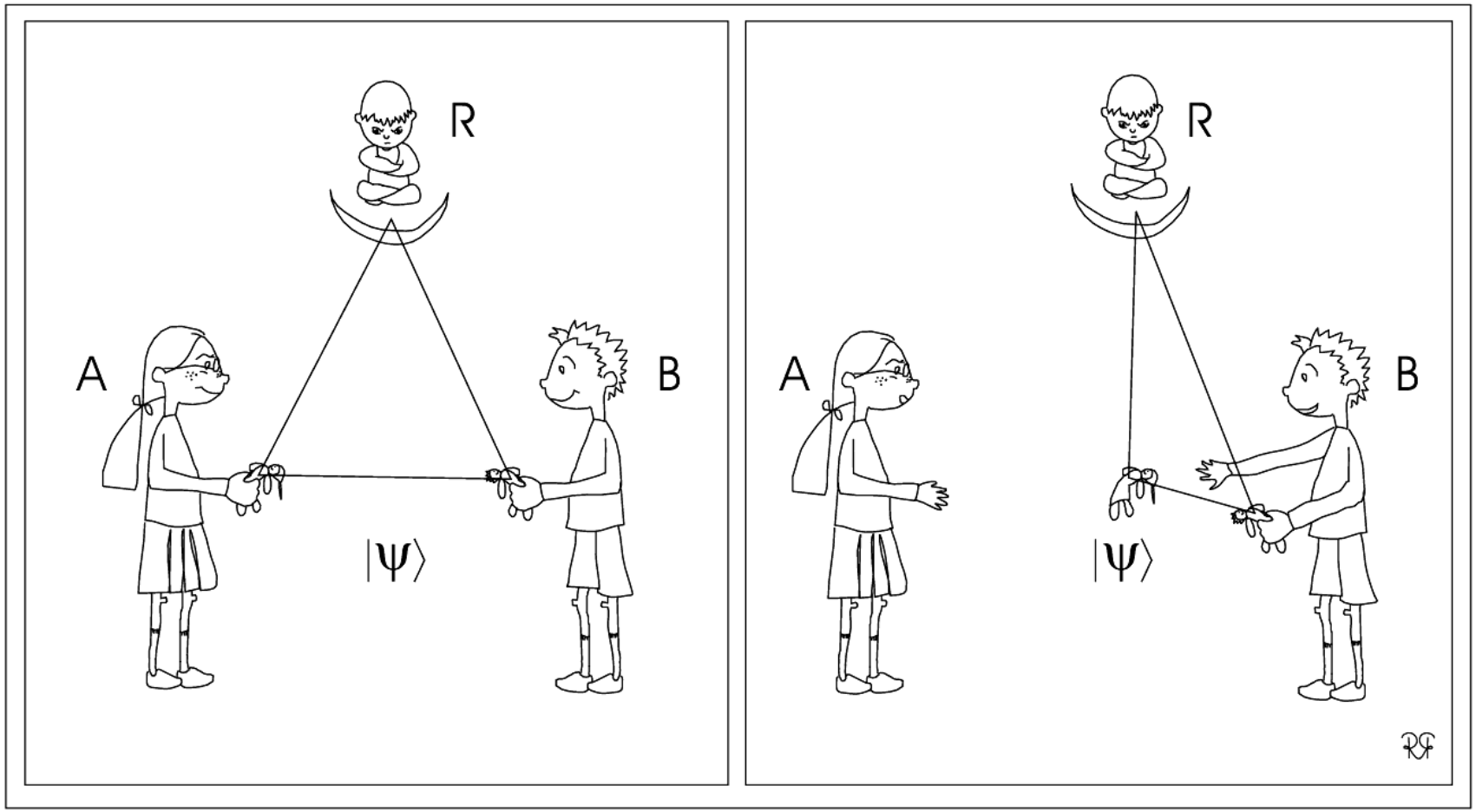
Quantum Reverse Shannon Theorem (QRST): Any quantum channel can be simulated by shared entanglement and an amount of classical communication equal to the channel's entanglement-assisted capacity. Therefore, in a world full of entanglement, all quantum channels are qualitatively equivalent, and quantitatively can be characterized by a single parameter (as classical channels can in world full of shared randomness, by the CRST).



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

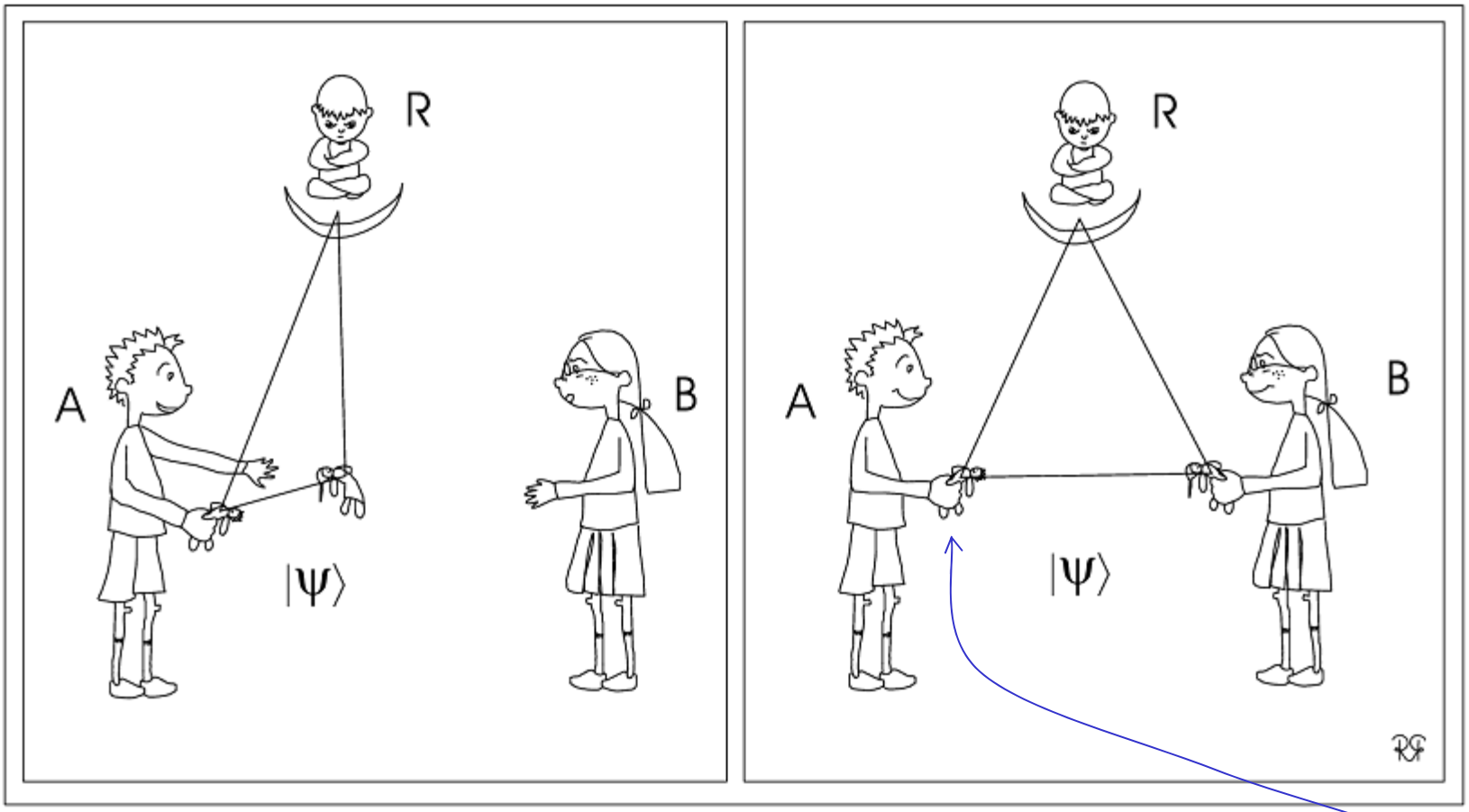


Output of simulation, including reference system, should have high fidelity with respect to $(\mathcal{N}\otimes\mathcal{I})^{\otimes m}(\Phi_\rho)$, the output on the same input of m copies of the channel being simulated.



Andreas told you about State Merging a.k.a. Fully Quantum Slepian Wolf, and its reverse, state splitting. State splitting is in fact the quantum reverse Shannon theorem for known tensor power sources. Interesting complications emerge when attempting to generalize it, as we would like to do, to unknown and non-tensor power sources.

State Splitting (asymptotically faithful for many copies of same state)

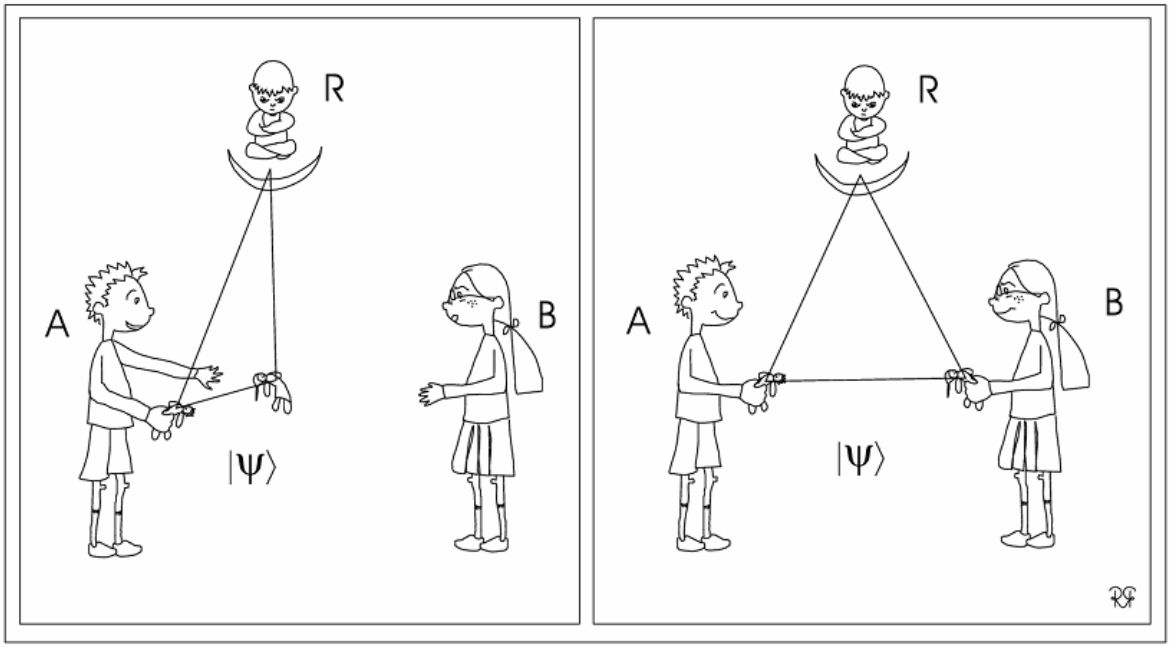


Apologies to Roberta Rodriquez

By State Splitting, Albert can transfer his share of ψ to Betty using $I(B:R) = S(B) + S(R) - S(B)$ bits of A=>B communication and $S(B)$ ebits of entanglement.

⇔QRST ⇔ asymptotically faithful and efficient simulation of channel from A to B on known tensor power source, with A retaining state of channel environment

Alternate communication resources for State Splitting



I(B:R) classical bits A=>B
S(B) ebits (A:B)

Advantage: Resources strictly incomparable
Disadvantage: Protocol not reversible

Make classical communication coherent, using relation 2 cobits = 1 qubit + 1 ebit

½ I(B:R) = qubits A=>B
½ I(A:B) = ebits (A:B)

Advantage: Protocol is reversible
Disadvantage: Resources partly comparable (ebits can be made from qubits).

Generalizing Classical Reverse Shannon Theorem to quantum case encounters several difficulties:

1. A classical encoder can measure the input type without disturbing it; a quantum encoder cannot.

Solution: Measure the quantum type (Schur basis representation) coherently, and unmeasure it before the protocol is done. Protocols will then still work on a superposition of input types.

2. On non-IID sources, different branches of a classical simulation can use very different amounts of shared randomness. This is OK classically, but if different branches of a quantum protocol use different amounts of entanglement, and the environment finds out, the branches will decohere and the simulation will fail.

Solution: On non-IID sources supplement ordinary entanglement by entanglement embezzling states (van Dam & Hayden 0201041) or classical back (or forward) communication to obfuscate how much entanglement each branch uses, thereby preventing decoherence.

Classical cost of entanglement assisted channel simulation

<table>
<tr><th>channel / source</th><th>Bell diagonal or rank-1 QC</th><th>Classical or CQ</th><th>General Channel</th></tr>
<tr><td>Known tensor power</td><td colspan="3" rowspan="3">Cost = quantum mutual information (QMI) of source-channel combination
QMI $(\mathcal{N},\rho)$ = $S(\rho) + S(\mathcal{N}(\rho)) - S(\mathcal{N}\otimes\mathcal{I}(\Phi_\rho))$</td></tr>
<tr><td>Unknown tensor power source</td></tr>
<tr><td>Known tensor product source</td></tr>
<tr><td>Unknown tensor product source</td><td colspan="3">$\leq$ QMI of channel on average source $\leq C_E$</td></tr>
<tr><td>Unconstrained source</td><td>$\leq C_E$</td><td>$\leq$ QMI on deco-hered source resulting from channel's initial measurement</td><td>$\leq C_E$ (in general requires use of back communication or entanglement embezzling states)</td></tr>
</table>

symmetries of $(\mathbb{C}^d)^{\otimes n}$

$$U \in \mathcal{U}_d \rightarrow U \otimes U \otimes U \otimes U$$

$$(\mathbb{C}^d)^{\otimes 4} = \mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^d$$

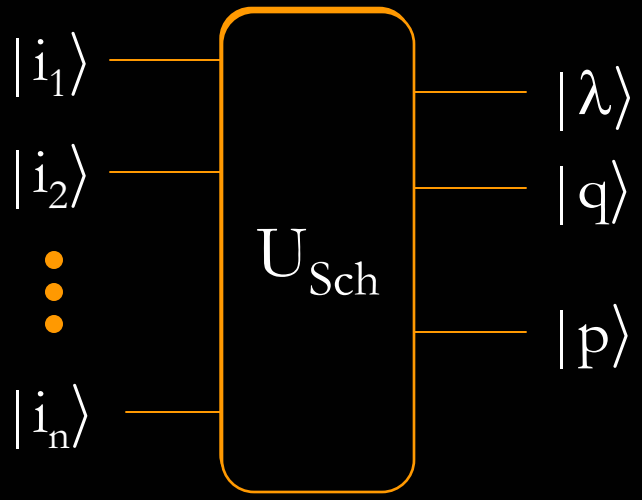
$$(1324) \in \mathcal{S}_4 \rightarrow$$

Schur duality

$$(\mathbb{C}^d)^{\otimes n} \cong \bigoplus_\lambda \mathcal{Q}_\lambda \otimes \mathcal{P}_\lambda$$

A. Harrow's slide

the Schur transform

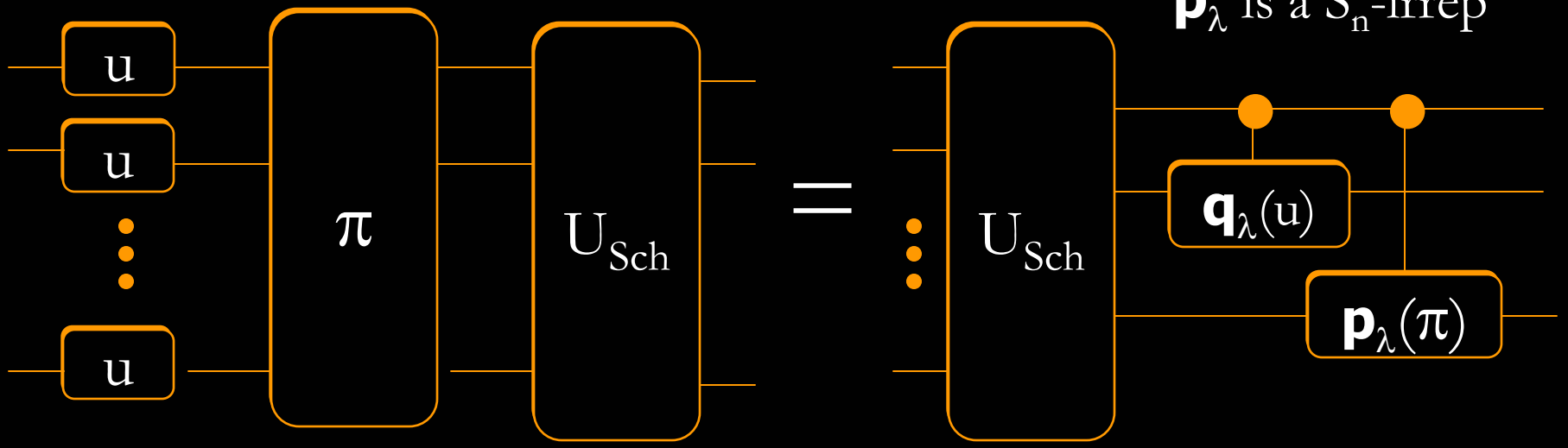


$u \in U(d)$

$\pi \in S_n$

$\mathbf{q}_\lambda$ is a U(d)-irrep

$\mathbf{p}_\lambda$ is a S_n-irrep



A. Harrow's slide

Schur basis ≅ classical types

Types for classical strings:

$x \in [d]^n$ can be written as (τ,p), where $\tau=(n_1,\ldots,n_d)$ is the type (letter frequencies) and p ranges from 1 to $n!/n_1!\ldots n_d!$.

We can further split T into (λ,q) where $\lambda_1 \geqslant \ldots \geqslant \lambda_d \geqslant 0$ are the frequencies and $q \in [d!/\lambda'_1!\ldots\lambda'_d!]$ map these frequencies to the alphabet [d].

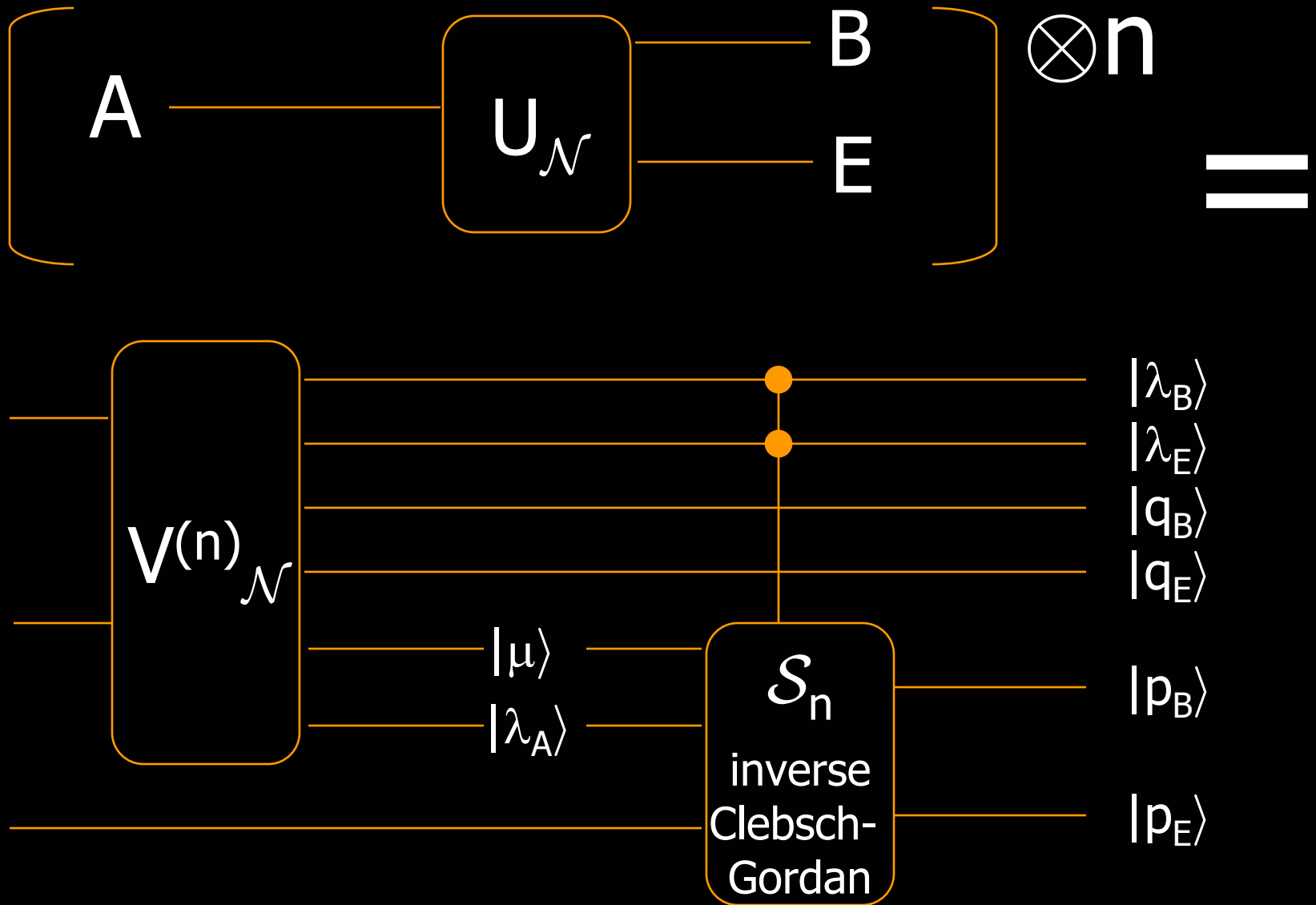
Key features:

Permuting x affects only p. Relabeling the alphabet changes only q. λ determines the range of p and q.

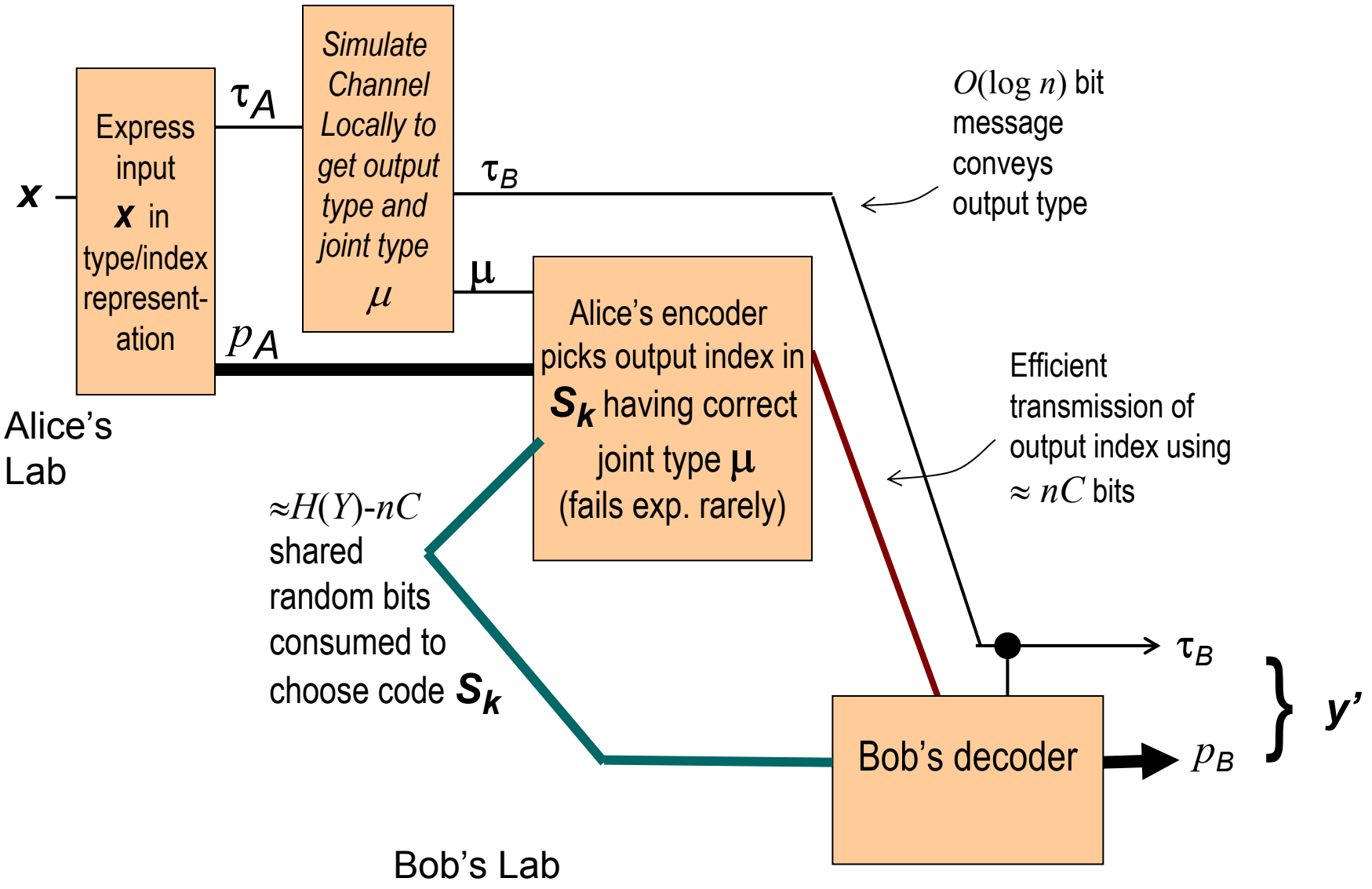
Quantum analogue:

λ has range $n^{O(d)}$. dim $\mathcal{Q}_\lambda = n^{O(d^2)}$. log dim$\mathcal{P}_\lambda \approx nH(\lambda/n)$

normal form of i.i.d. channels

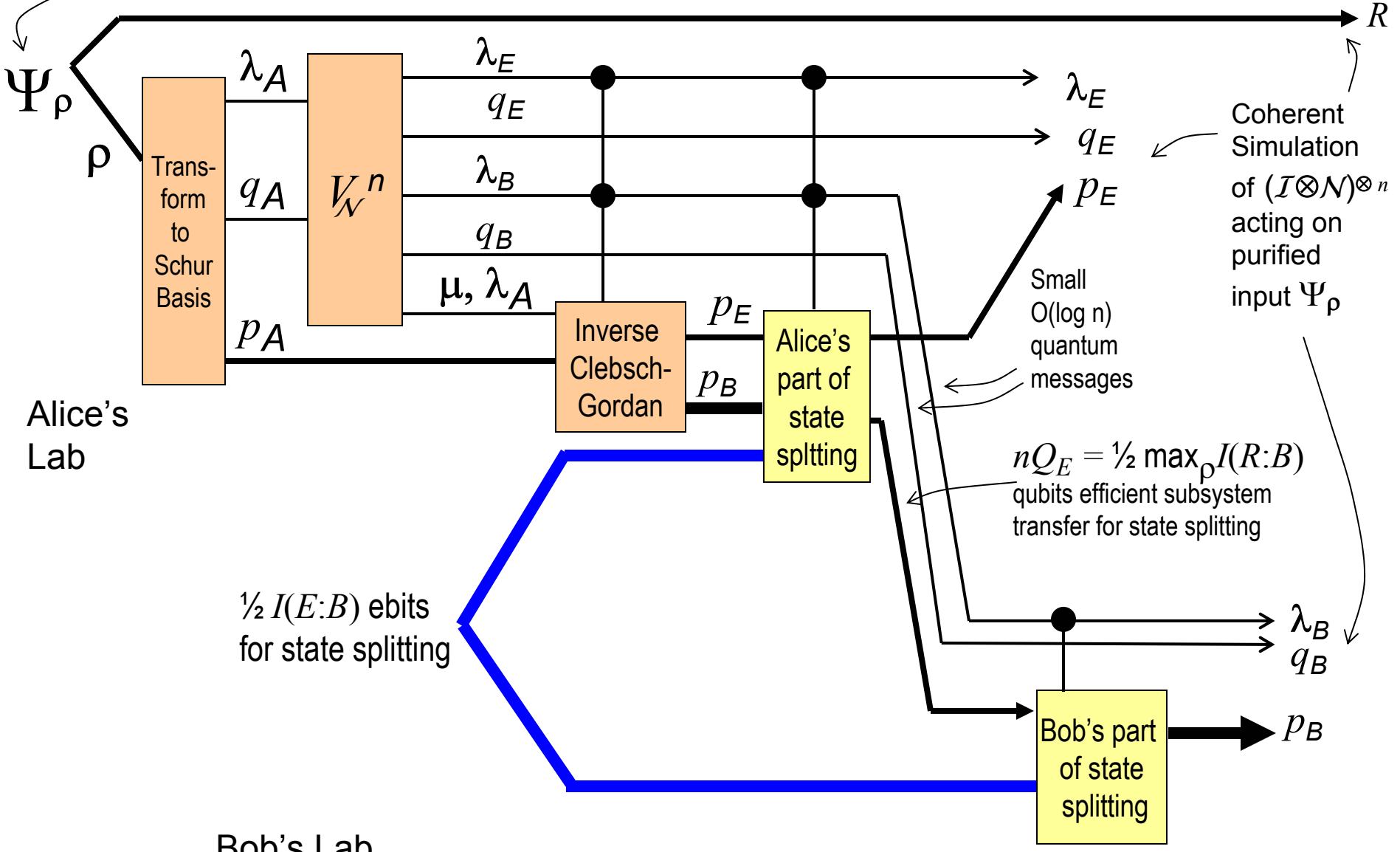


Recall typewise simulation underlying classical reverse Shannon theorem



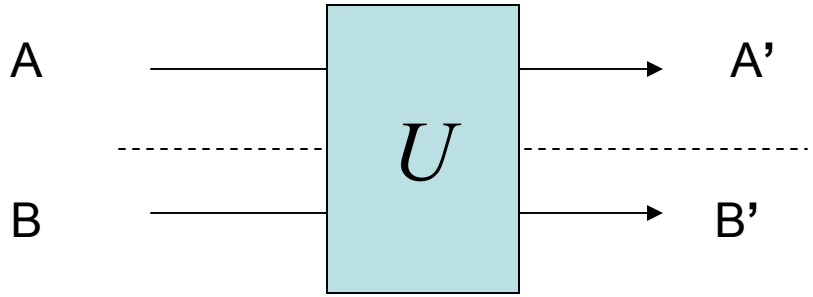
Ananlgous Quantum RST in Schur Basis, for unknown tensor power source

Purification of a tensor power state ρ input to n instances of general channel $\mathcal{N}$

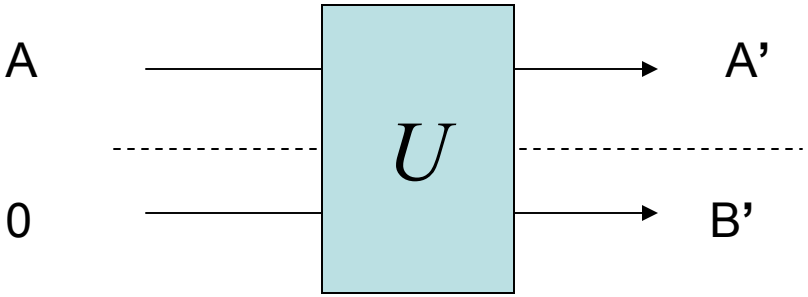


Gates, Isometries and Channels

Gate

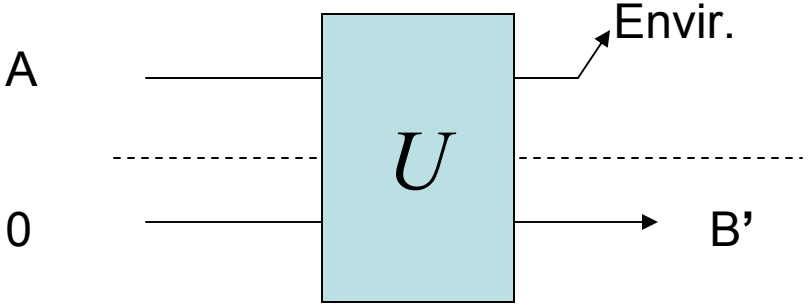


Isometry,
a.k.a.
channel
with quantum
feedback

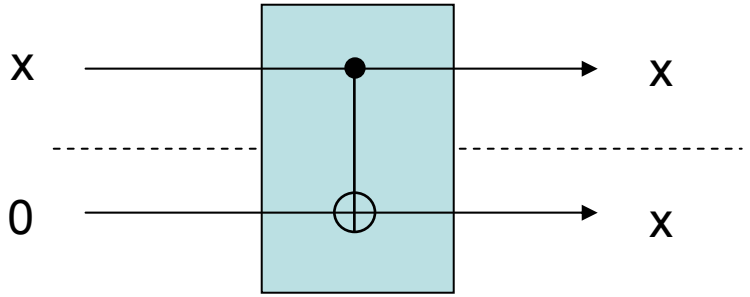


Cf classical feedback, where wlog Alice gets a *copy* of output. Here the input is *divided* between Alice and Bob.

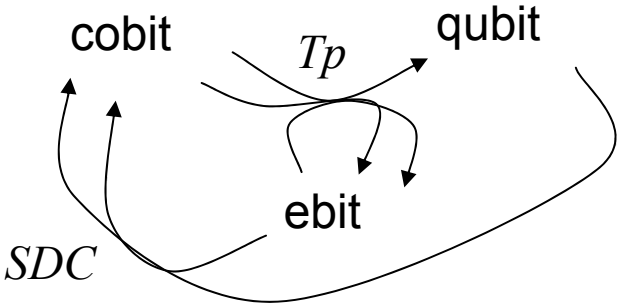
Channel,
a.k.a.
CPTP map



A particularly useful isometry is the coherent bit, or “cobit” (Harrow 0307091), where U=CNOT.



For input x = 0 or 1, Alice sends the bit to Bob and keeps a copy for herself (instead of leaking it to the environment as she would have done in an ordinary classical channel). Performing classical communication coherently in this fashion allows the classical bits in teleportation to be recycled as entanglement, so that teleportation and superdense coding become asymptotically inverse operations.



Asymptotically,

1 qubit + 1 ebit = 2 cobits

Embezzlement to the Rescue!

"Embezzling states" (van Dam & Hayden quant-ph/0201041)

$$\mu_n = \sum_{j=1}^{n} |j\,j\rangle_{AB} / \sqrt{j}$$

have a very broad Schmidt spectrum.

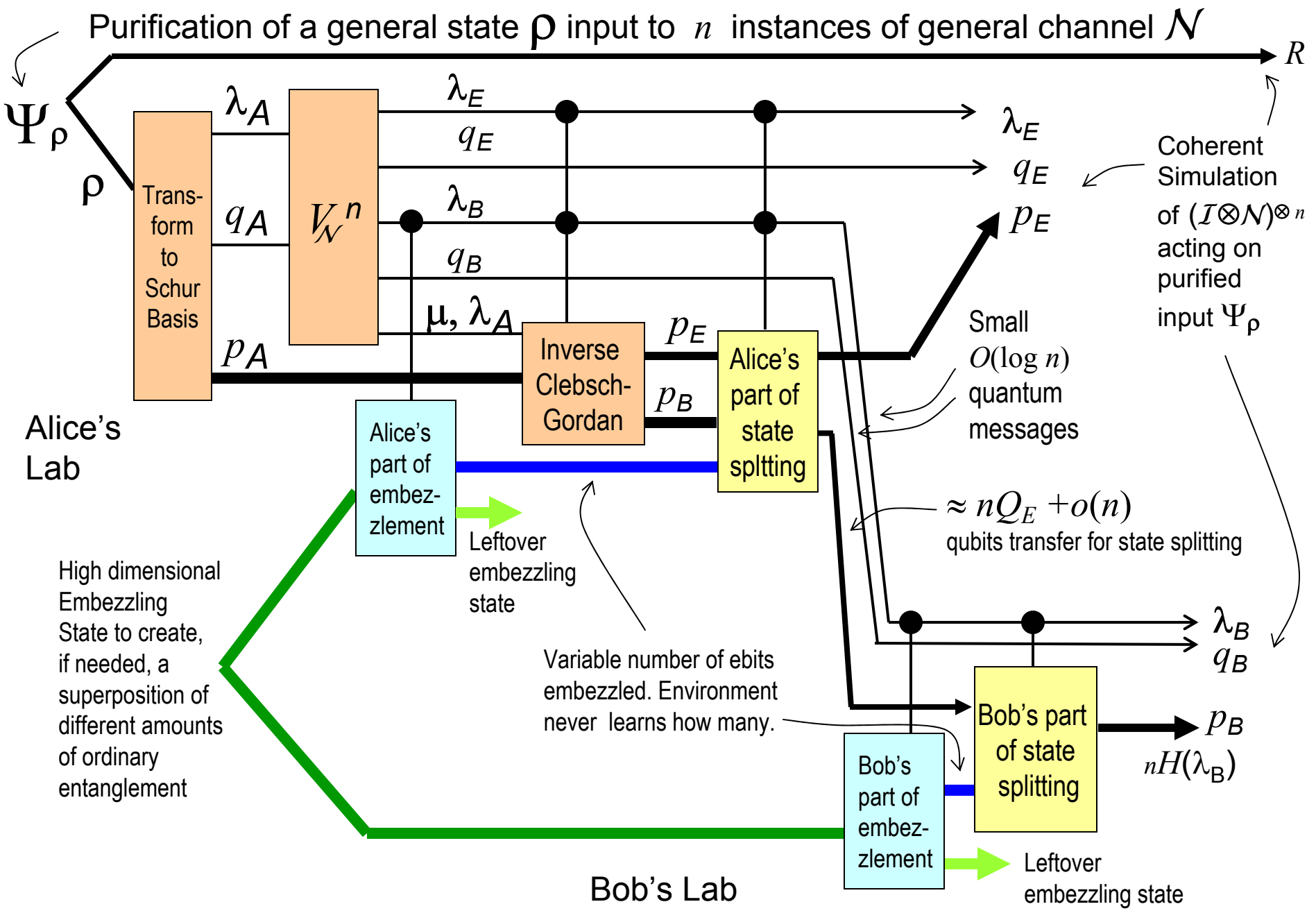
Any bipartite pure state φ_{AB} on a d x d Hilbert space can be created, without communication, from an embezzling state, leaving the embezzling state almost unchanged.

$\mu_n \Rightarrow \mu_n \varphi$ with fidelity $>1-\varepsilon$ in the limit of large n.

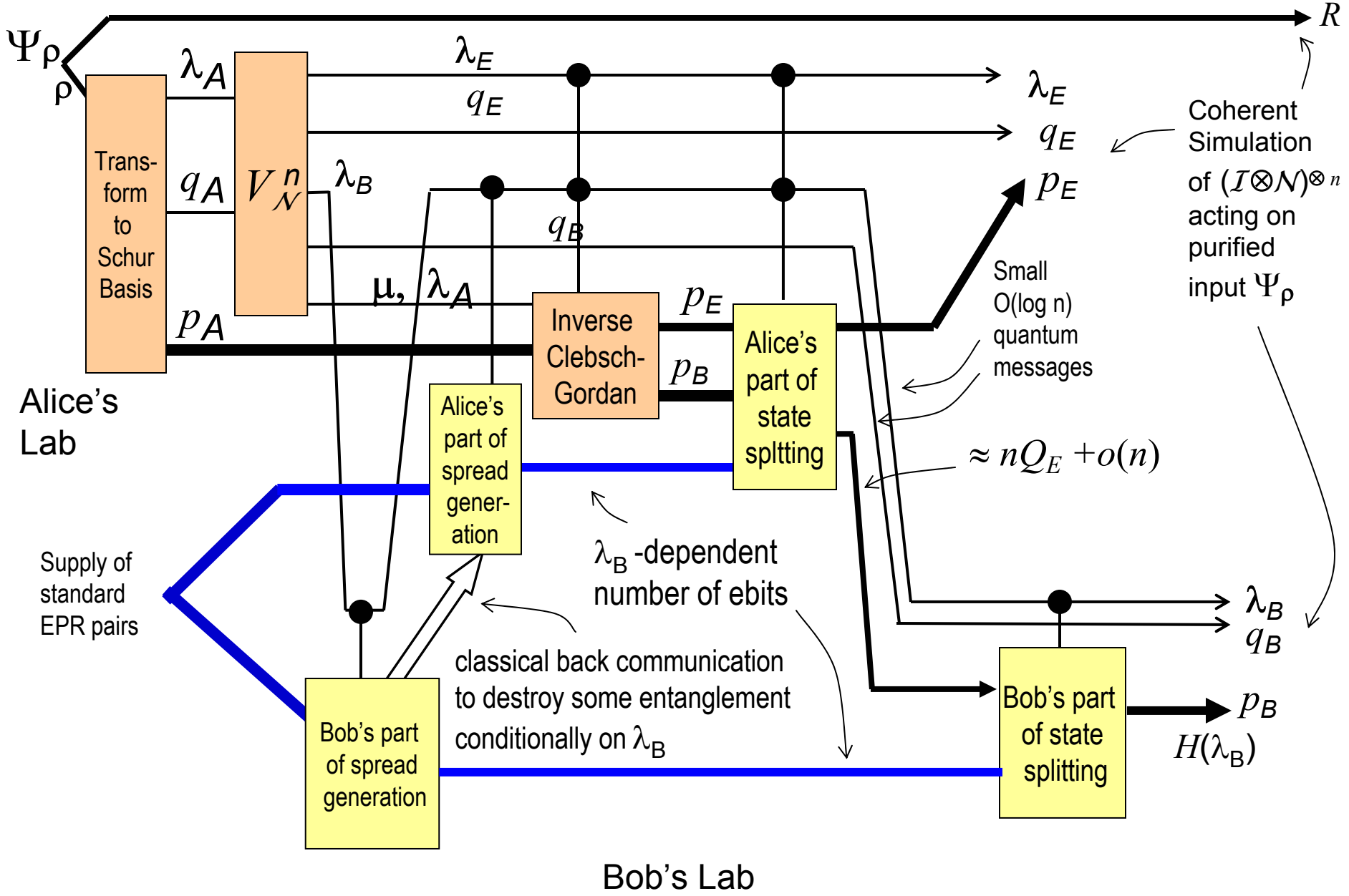
How big an n is needed?

Approximately $d^{1/\varepsilon}$, so $\log n \approx (1/\varepsilon) \log d$

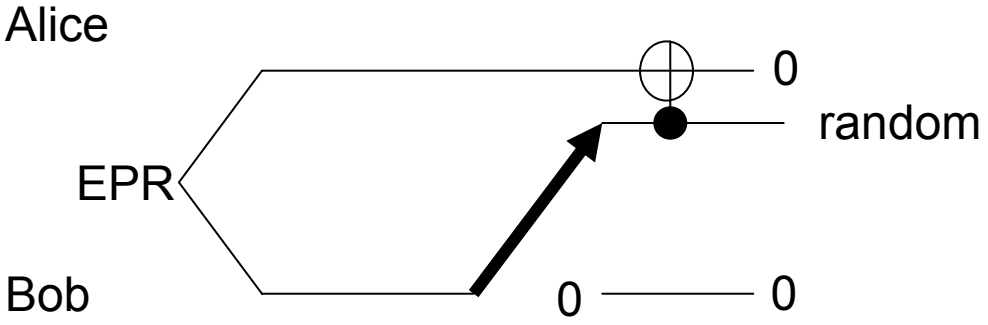
Embezzlement-assisted QRST in Schur Basis, for general source and channel



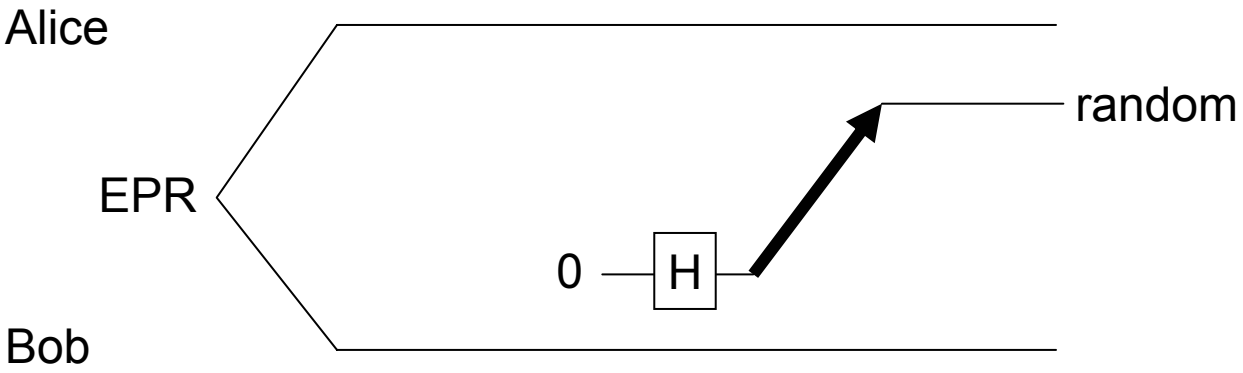
Back-communication assisted QRST in Schur Basis, for general source and channel



How can classical communication destroy entanglement without the environment finding out?



On a branch where Bob and Alice wish to destroy an ebit, Bob sends his half of the ebit through a classical channel to Alice, who XORs it with her half. Meanwhile Bob makes a fresh 0.



On a branch where Bob and Alice want to keep the ebit, Bob prepares a bit in the 0+1 state and sends it to Alice through a classical channel. Environment can't distinguish these 2 situations.

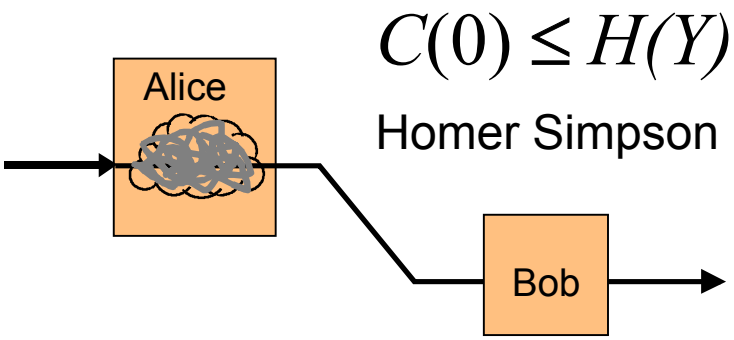
The CRST and QRST show that classical (resp. quantum) channels can simulate one another efficiently in the presence of unlimited shared randomness (resp. entanglement).

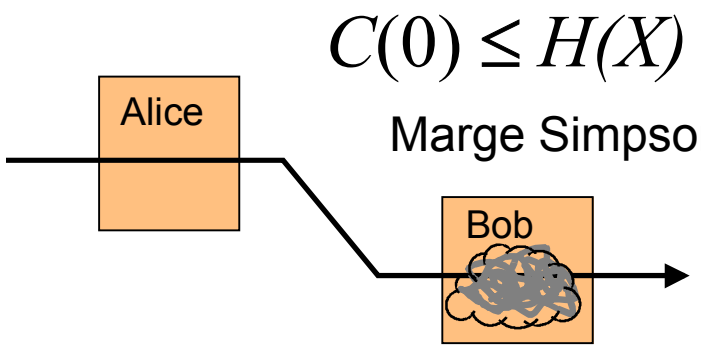
What is the tradeoff between shared randomness r and noiseless forward communication $C(r)$ necessary and sufficient to simulate a general classical noisy channel?

Similarly, for quantum channels, what is the tradeoff between shared entanglement e and noiseless forward quantum communication $Q(e)$ required to simulate the quantum channel.

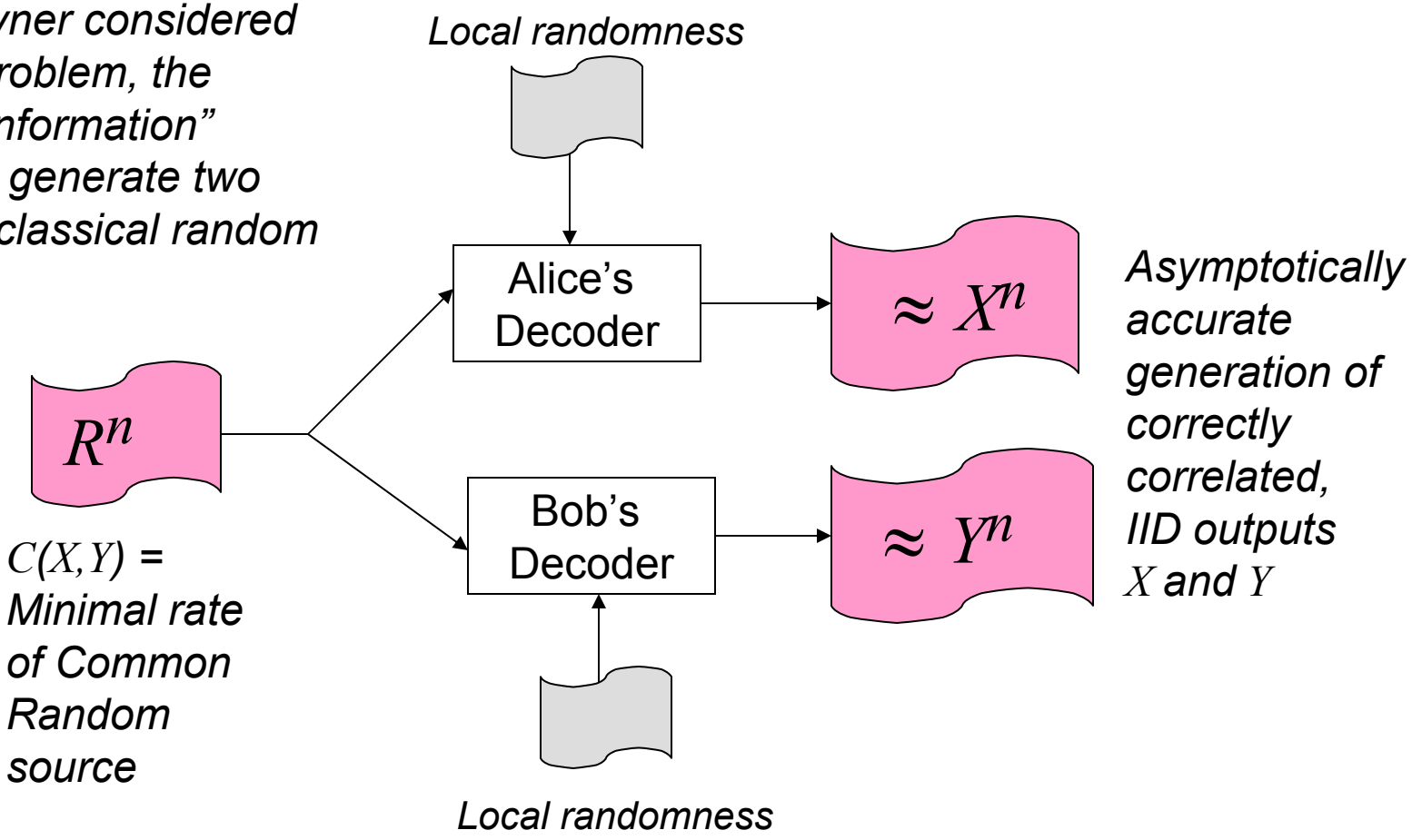
In particular, what are the simulation costs $C(0)$ and $Q(0)$ without any shared randomness or entanglement?

Simple upper bounds on C(0) from the Simpson family





In 1975 Wyner considered a related problem, the "common information" required to generate two correlated classical random variables.



Asymptotically accurate generation of correctly correlated, IID outputs X *and* Y

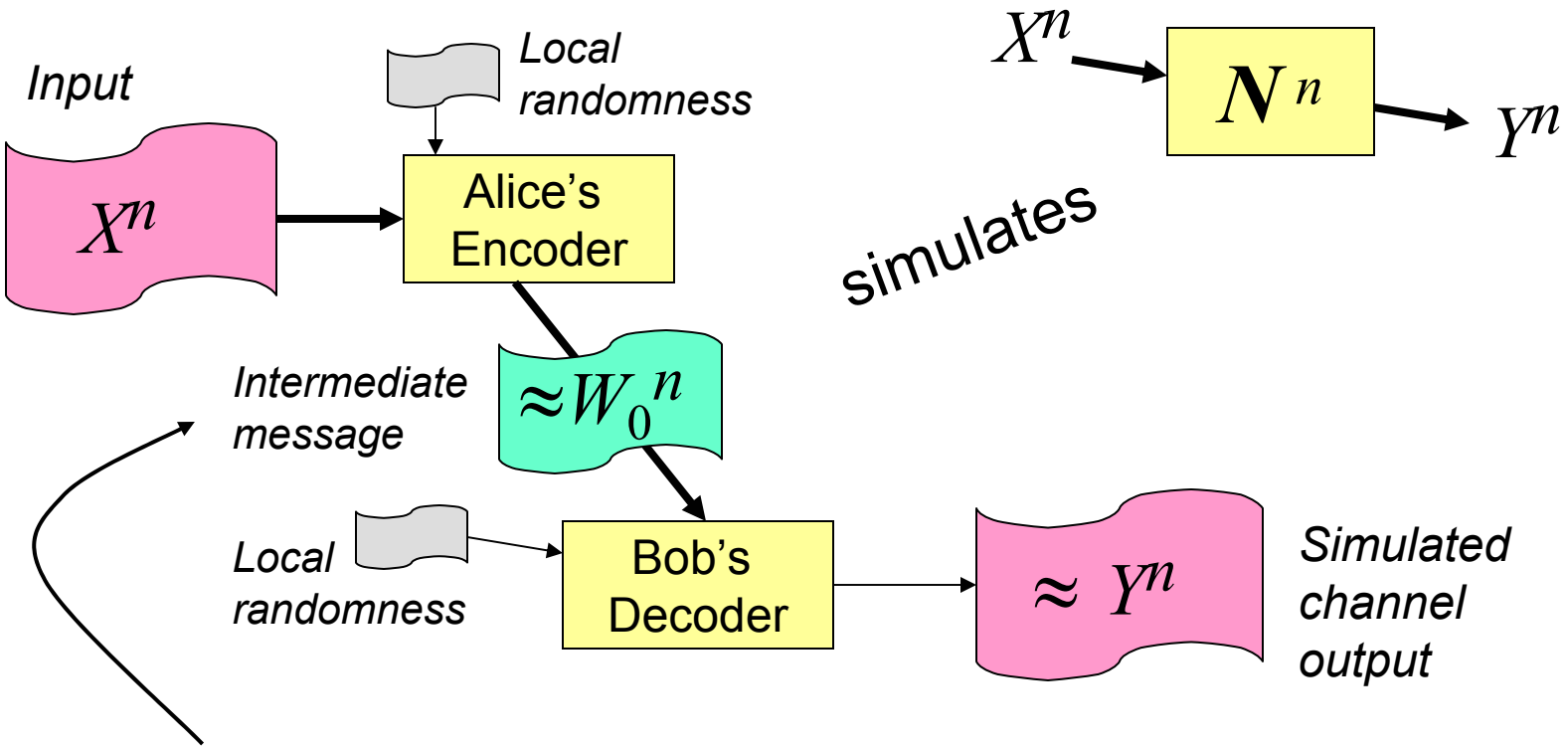
$C(X,Y)$ = *Minimal rate of Common Random source*

He found a single-letter formula for it: $C(X,Y) = \min_{W \text{ such that } XWY \text{ forms a Markov chain: } I(X:Y|W)=0} I(XY:W)$

Let $W_0(X,Y)$ *denote an arbitrary random variable jointly distributed with* X *and* Y *so as to satisfy this constraint and achieve this minimum.*

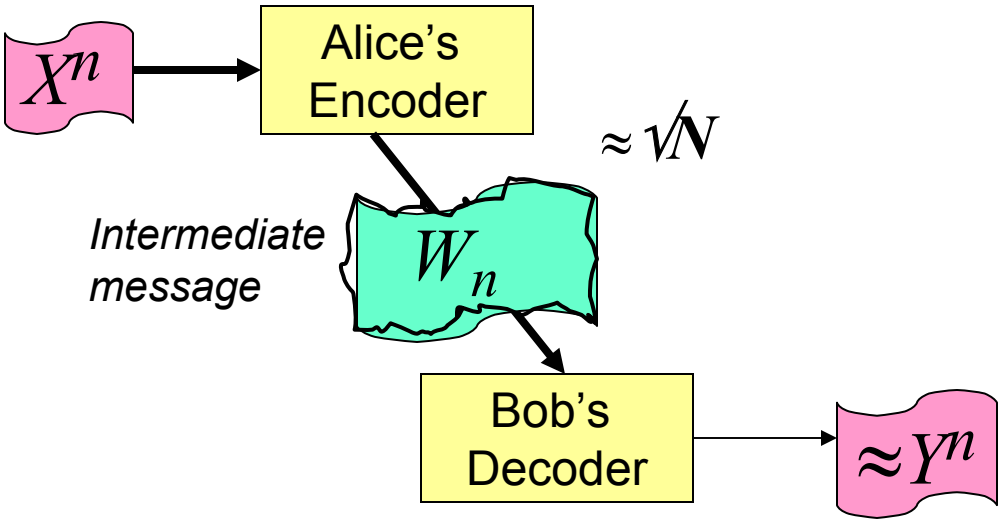
A.D. Wyner *IEEE Trans.* **IT-21**, 163-179 (1975)

Reversing the direction of Alice's operations, Wyner's formula also applies to the problem of simulating a noisy channel $\mathcal{N}$ by noiseless forward communication and *local* randomness. Here Alice block-encodes an IID input X^n to get an intermediate message which she sends noiselessly to Bob, who then decodes it to get an approximation to Y^n.

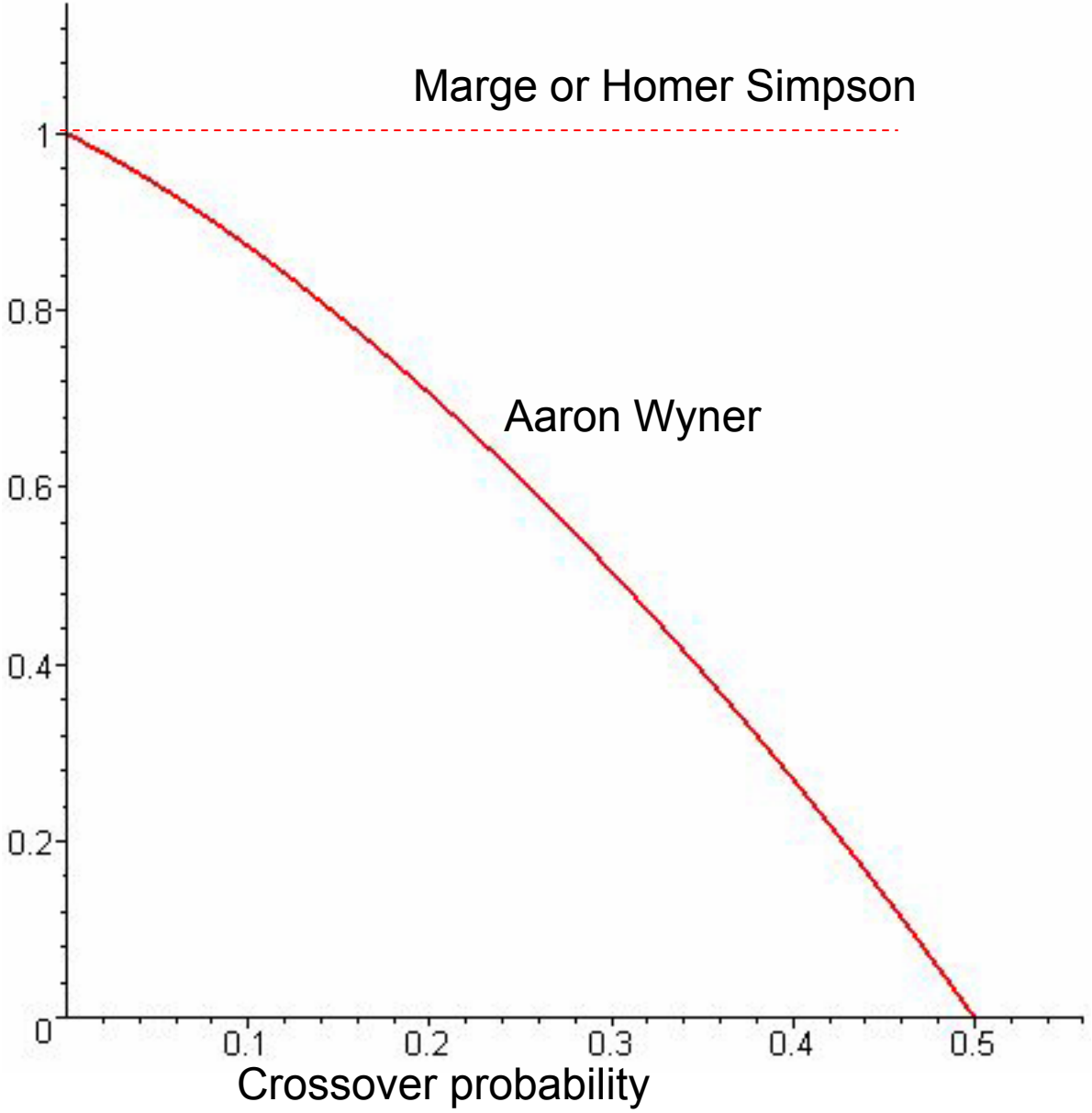


Minimal forward communication for this simulation is $I(XY:W_0)$, where W_0 is a random variable jointly distributed with X and Y so as to minimize $I(XY:W)$, while satisfying $I(X:Y|W)=0$.

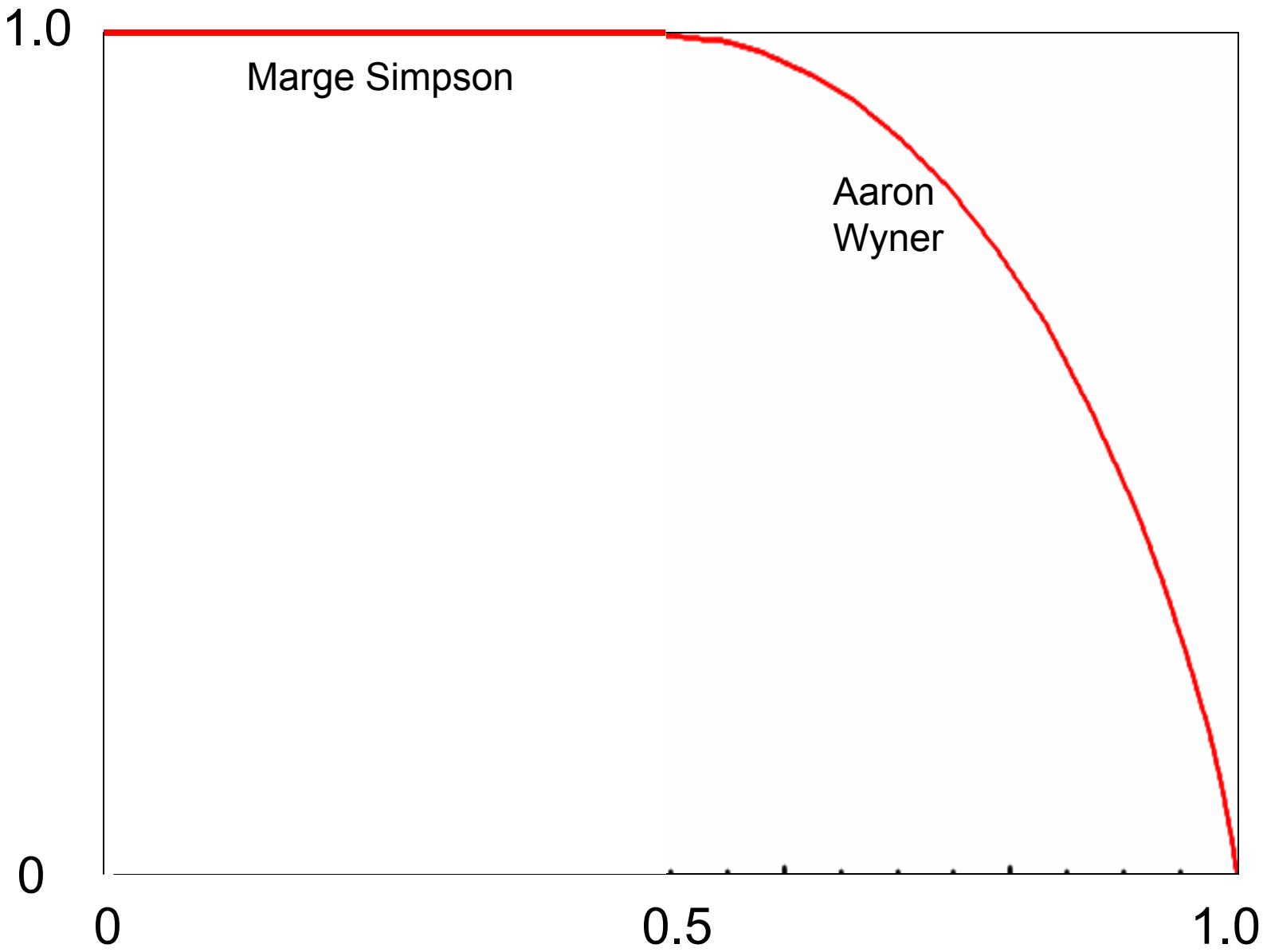
For a BSC channel N, the intermediate variable W_0 is just what one would get by having Alice apply a less noisy BSC $\sqrt{N}$ to the input X. After receiving the intermediate message $W_n \approx {W_0}^n$, Bob applies $\sqrt{N}$ again to get Y. At first sight, this seems like a step backward, since the less noisy first stage $\sqrt{N}$ should require ***more*** forward communication to simulate than the original channel N we are trying to simulate. In fact, the first stage mapping, from X^n to $W_n \approx {W_0}^n$, can be done on the cheap, because its output is only destined for further degradation, and so can be of shoddy quality, without harming the overall mapping from X^n to Y^n. If a bakery knows its pies are destined only to be thrown in people's faces, it can use cheaper ingredients and no one will notice. This may be viewed as a case of partial derandomization.



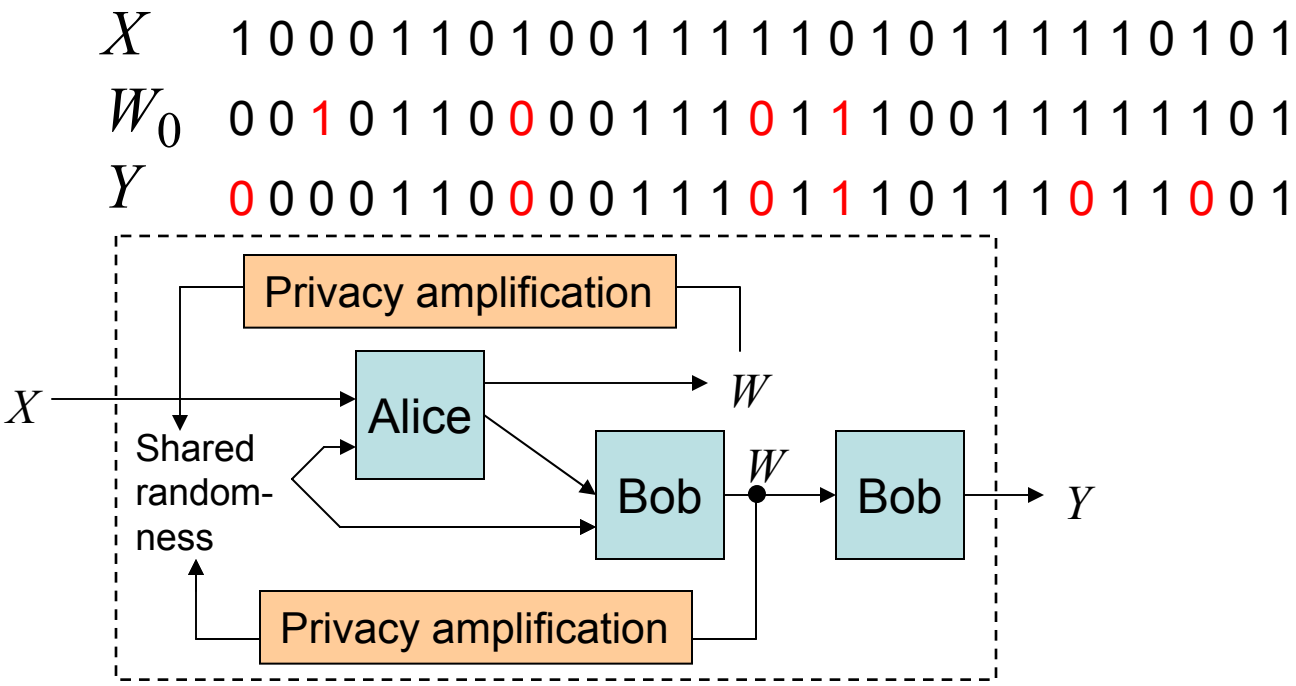
Forward communication cost of simulating a binary symmetric channel without shared randomness.



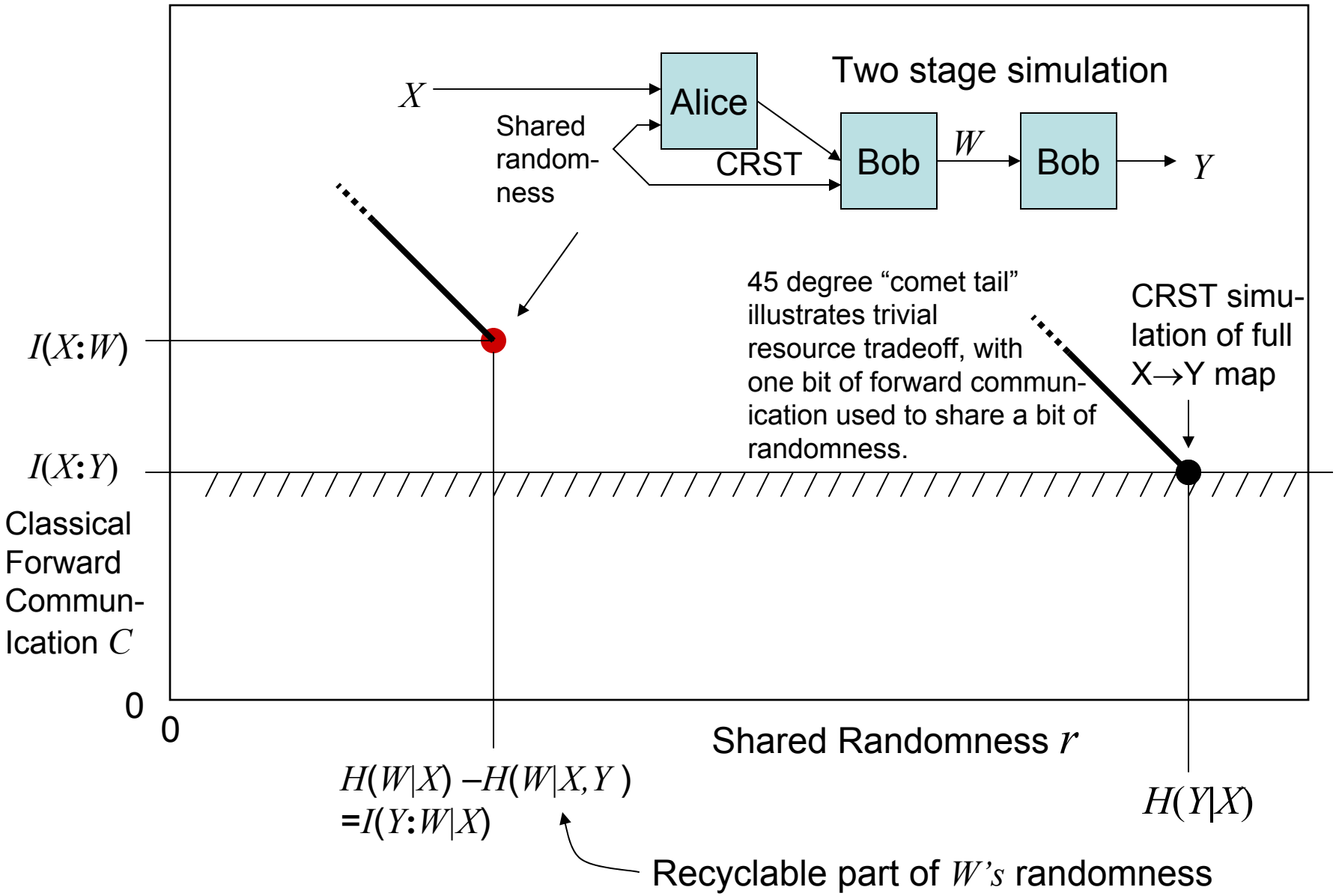
Forward communication cost of simulating binary erasure channel without shared randomness.



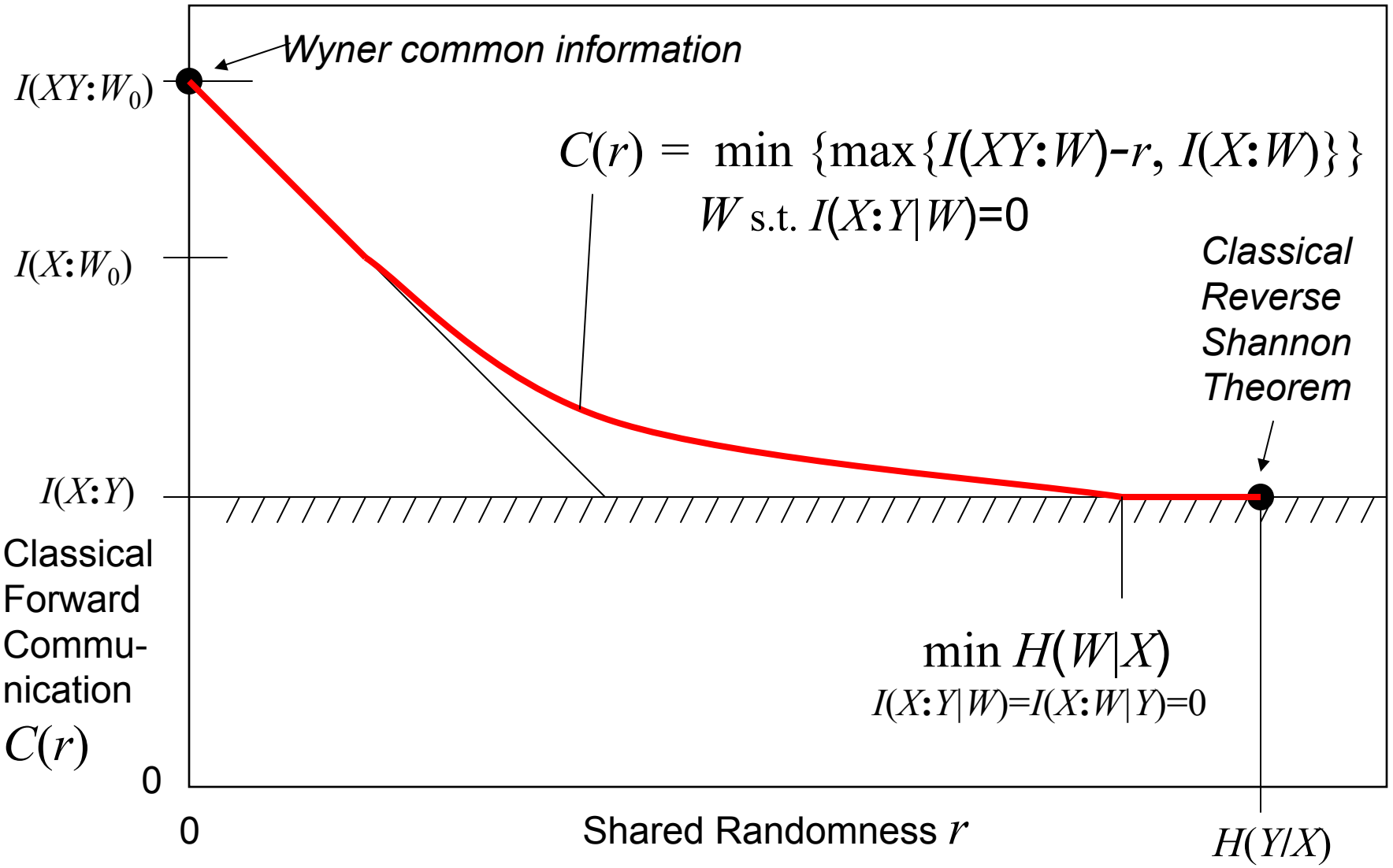
A more quantitative way to view the saving on the first stage mapping is via a simulation involving *catalytic* shared randomness and privacy amplification. Let the first stage $\sqrt{N}$ be simulated exactly, using secret shared randomness that Alice and Bob borrow from the bank, and generating W_0^n as the intermediate message. After the protocol is over, X^n and Y^n are known publicly, but only partial information has leaked out about W_0^n. How much? Clearly $n(1\text{-}H(W_0|XY))$. Alice and Bob use their noiseless channel to share this amount of fresh random information, privacy-amplify what they borrowed from the bank to compensate for what has been leaked, and return it to the bank, none the worse for wear.

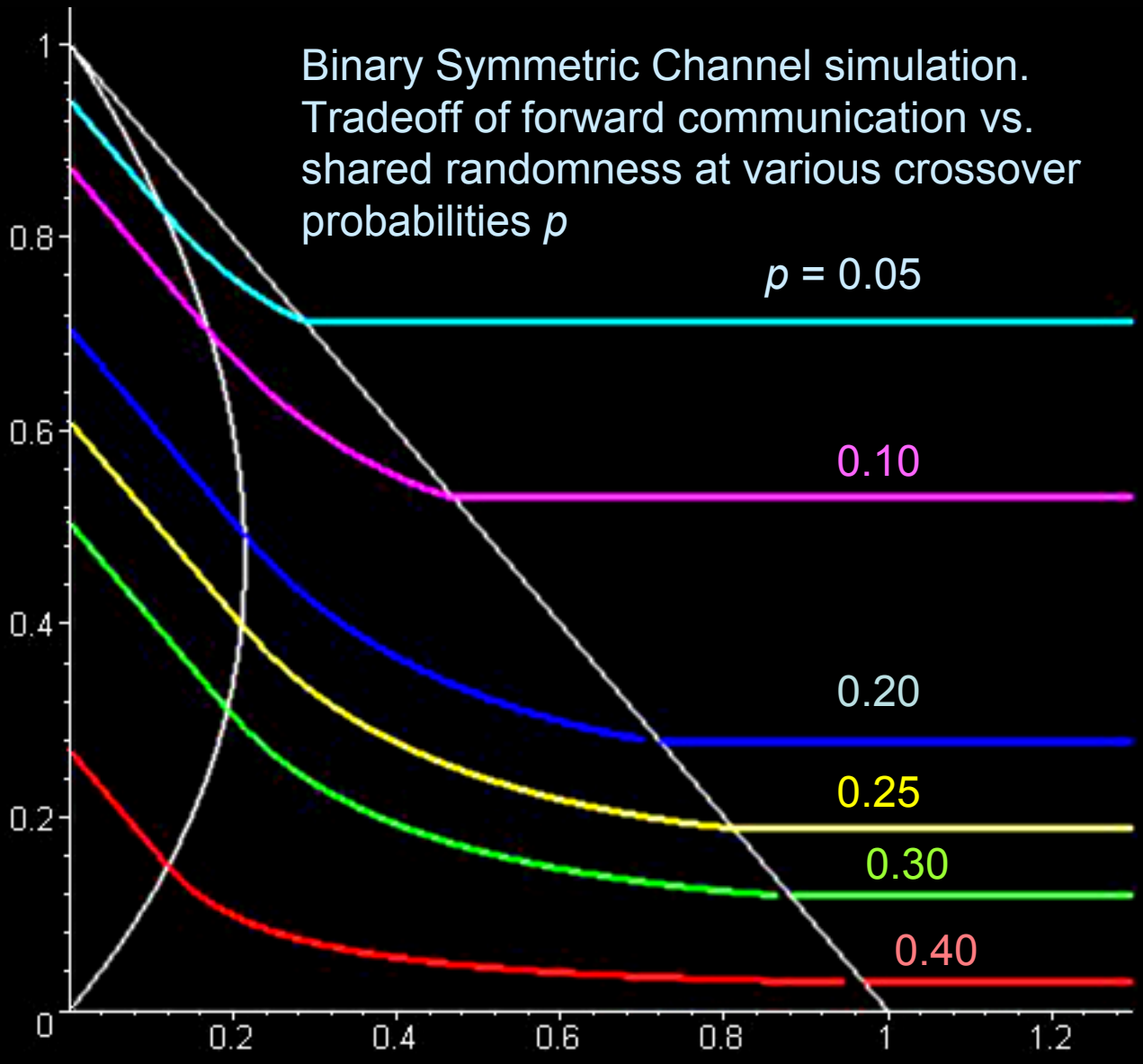


Two-stage simulation of a classical noisy channel. Alice and Bob use CRST to simulate a first stage, mapping X to some intermediate random variable W. Then Bob finishes the job by locally mapping W to Y. This can yield a nontrivial resource tradeoff.

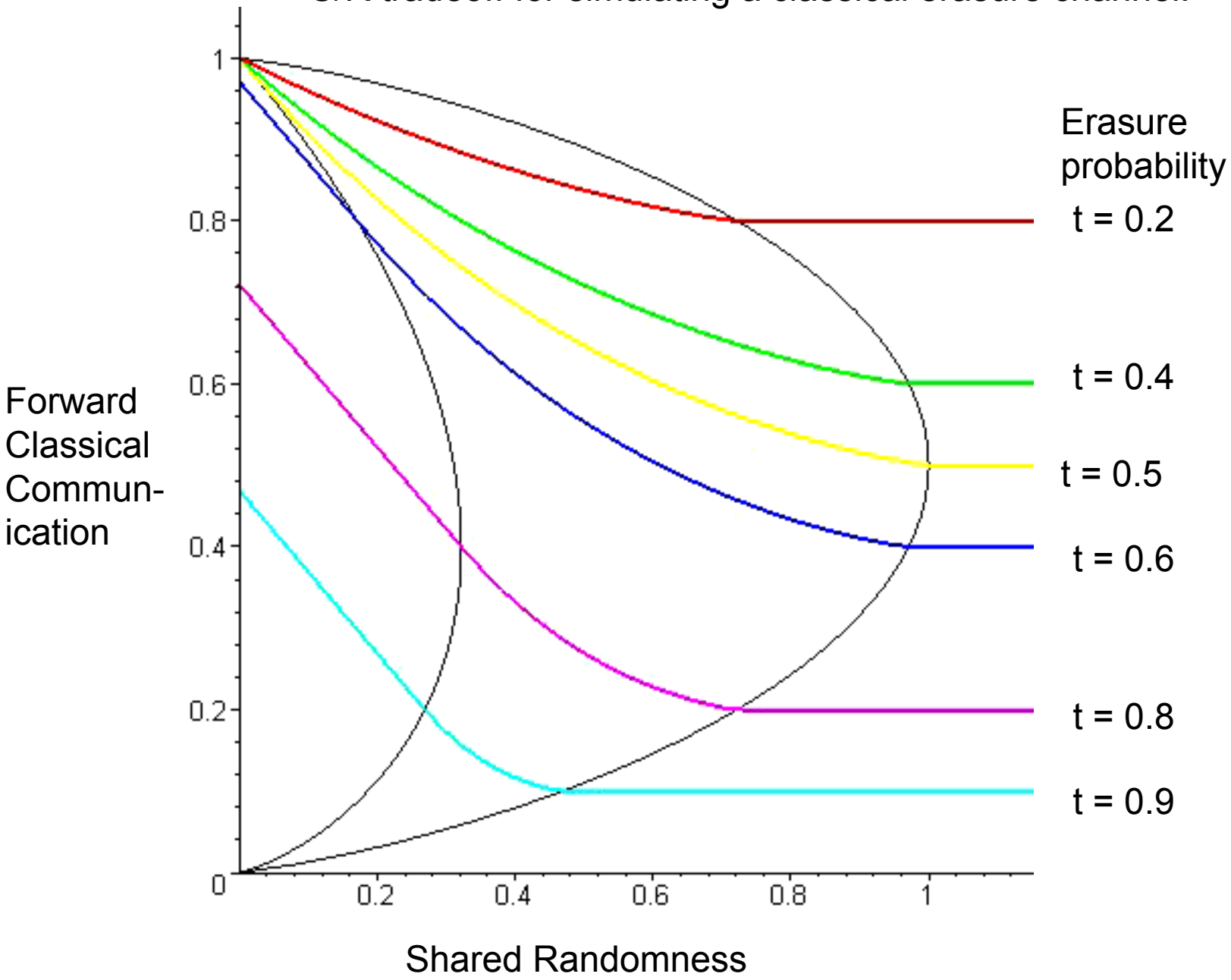


Tradeoff between Forward Communication & Shared Randomness necessary & sufficient to simulate a noisy classical channel mapping input X to output Y.

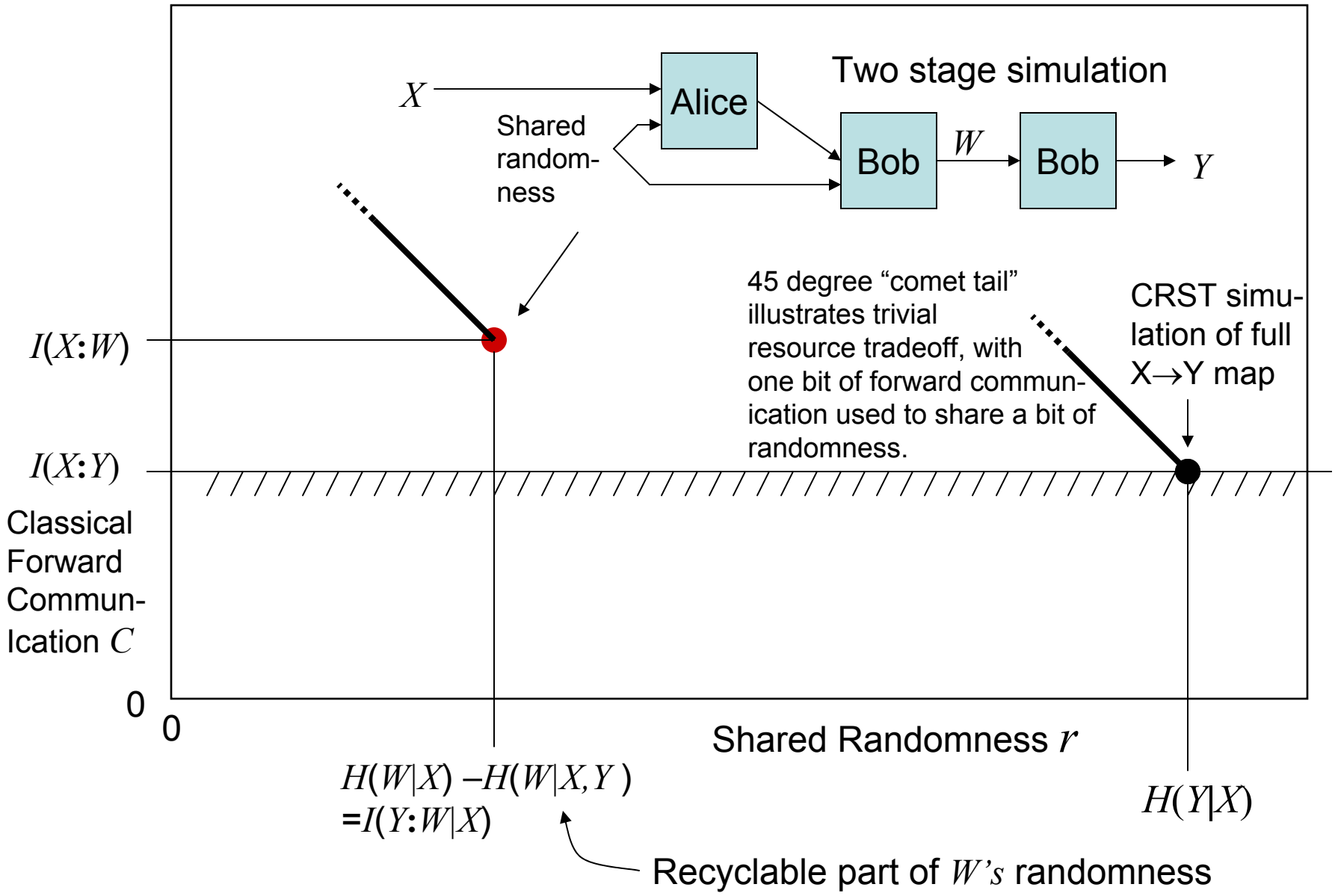



Binary Symmetric Channel simulation.
Tradeoff of forward communication vs.
shared randomness at various crossover
probabilities p
p = 0.05
0.10
0.20
0.25
0.30
0.40
1
0.8
0.6
0.4
0.2
0
0.2
0.4
0.6
0.8
1
1.2

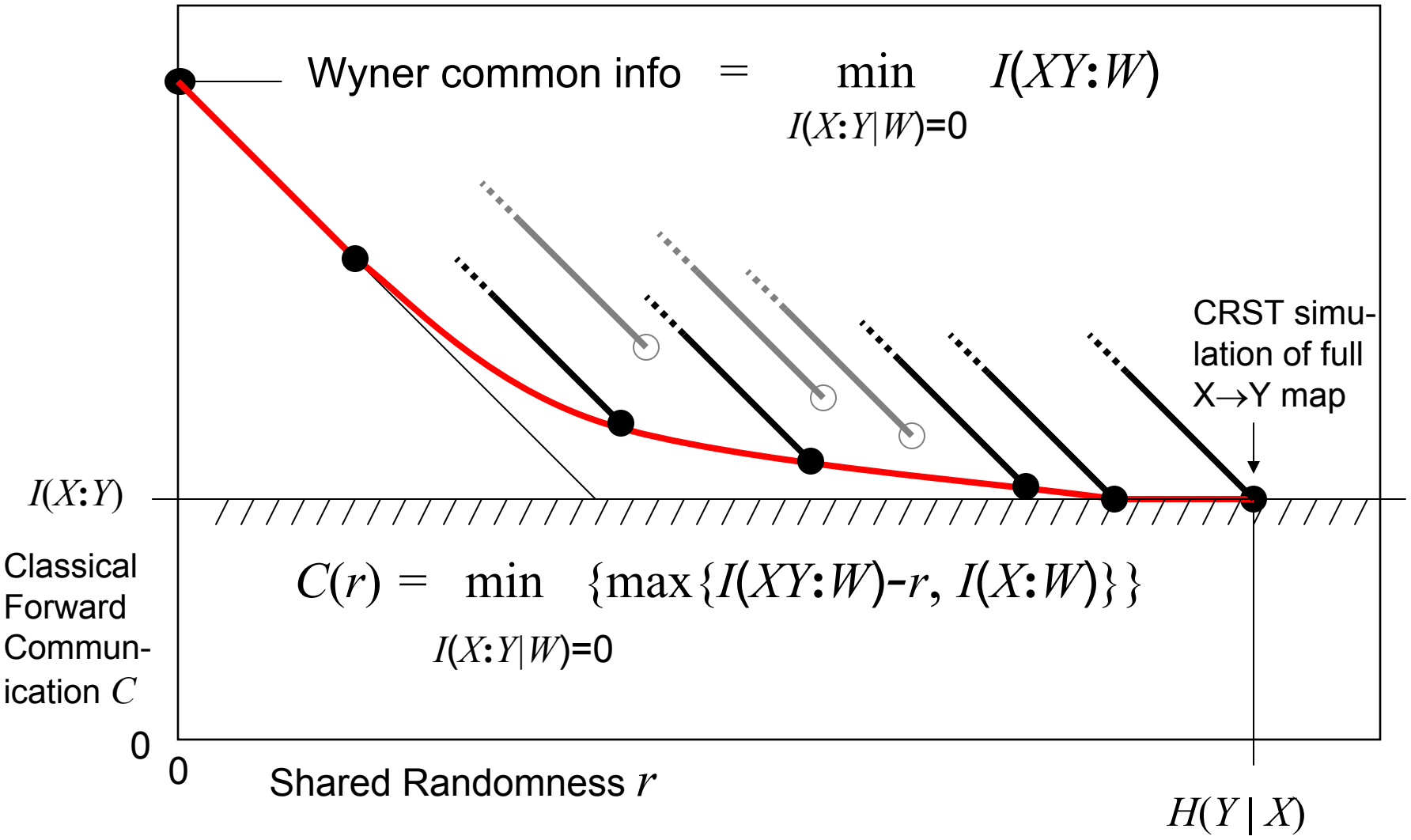
C/R tradeoff for simulating a classical erasure channel.



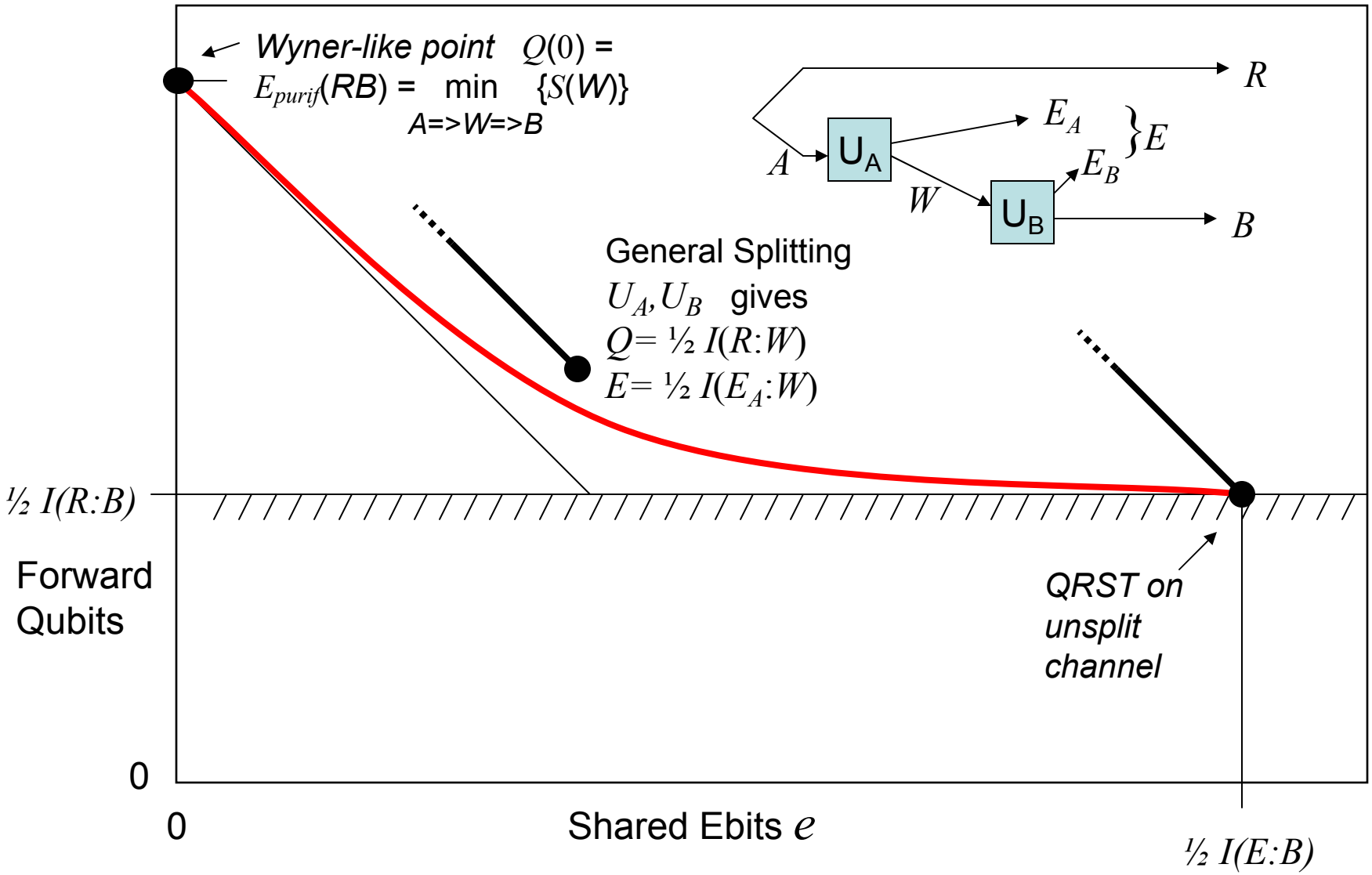
Two-stage simulation of a classical noisy channel. Alice and Bob use CRST to simulate a first stage, mapping X to some intermediate random variable W. Then Bob finishes the job by locally mapping W to Y. This can yield a nontrivial resource tradeoff.



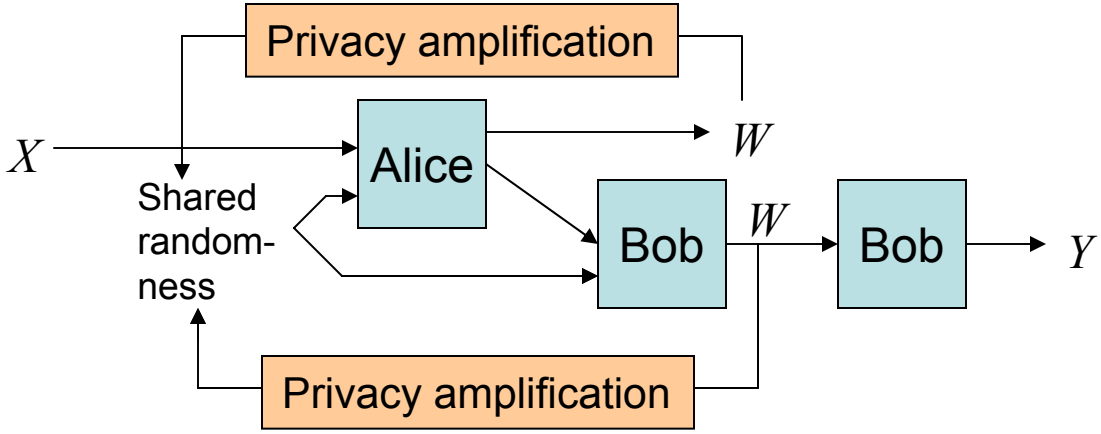
General form of randomness vs. communication tradeoff for simulating a classical channel



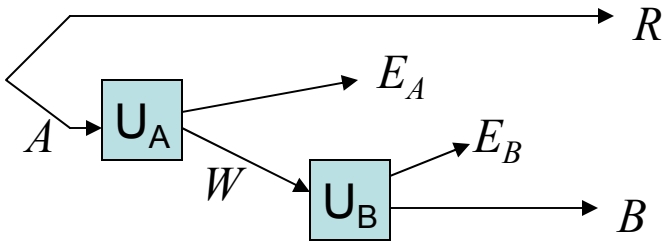
Splitting environment between Alice and Bob can give nontrivial tradeoff between Forward Qubits and Shared Ebits to simulate a quantum channel



Classical: Nontrivial tradeoff due to randomness recycling.

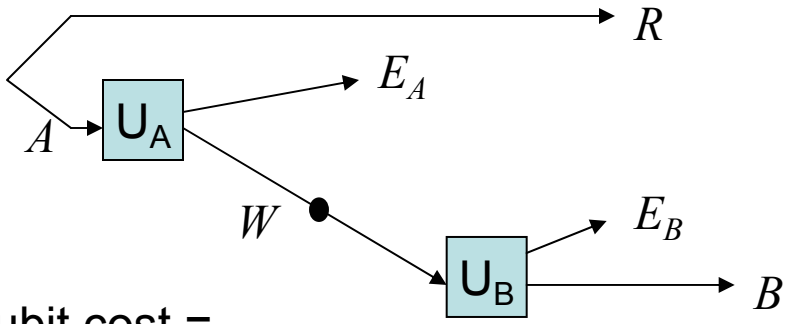


Quantum: Nontrivial tradeoff due to $W=(B,E_B)$ sometimes being cheaper to send (less entropy) than one of its parts, namely B alone.



$Q(0)$ = qubit cost without entanglement =

$E_p(\rho_{R,B})$ the entanglement of purification of RB

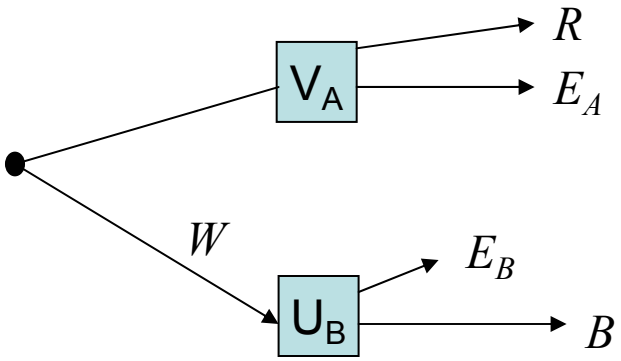


First stage qubit cost =

½ min $\{I(R{:}W)|\ I(E_A{:}W)=0\}$

by splitting formula

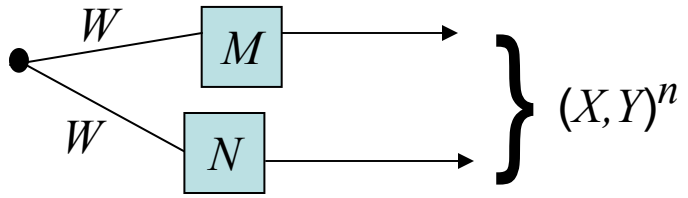
(second stage free—done by Bob)



Classical Wyner Common Info:

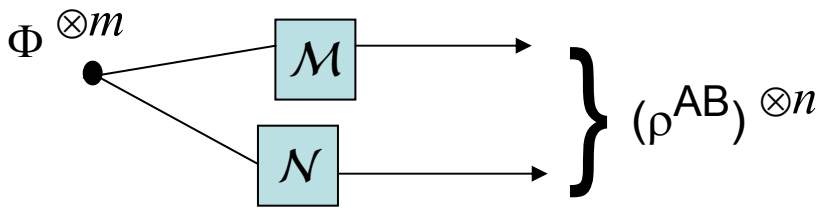
$$C(X,Y) := \min \{ I(XY;W) \mid I(X;Y|W)=0 \}$$

= least rate of a perfectly correlated bipartite source from which many examples of (XY) can be asymptotically prepared by local decoding



Quantum entanglement of purification: $E_{pur}(\rho^{AB})$

= least pure entanglement from which many copies of ρ^{AB} can be asymptotically prepared by local operations and a sublinear amount of communication



The Wyner Common Information may be viewed as the specialization of E_{pur} to the case where the decoders $\mathcal{M}$ and $\mathcal{N}$ are single-letter measurements followed by classical blockwise processing.