# Introduction to Information Theory

## Shannon Entropy and Logical Entropy

David Ellerman

UCR

January 2012

# Charles Bennett on Nature of Information

Information = Distinguishability.

(Using a pencil, a piece of paper can be put into a various states distinguishable at a later time.)

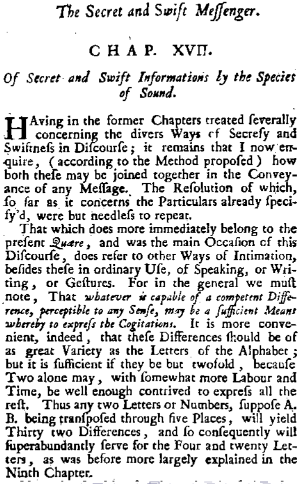
- Information is reducible to bits ( **0** ,**1** )
- Information processing, to reveal implicit truths, can be reduced to logic gates (**NOT**, **AND** )
- bits and gates are *fungible*, independent of physical embodiment, making possible Moore's law

It is natural to assume that information

- can be copied at will without disturbing it
- cannot travel faster than light or backward in time
- can be erased when it is no longer wanted

# John Wilkins, 1641. *Mercury: The Secret and Swift Messenger*.

- "For in the general we must note, That whatever is capable of a competent Difference, perceptible to any Sense, may be a sufficient Means whereby to express the Cogitations." [John Wilkins 1641 quoted in: Gleick 2011, p. 161]



*The Secret and Swift Meſſenger.*

CHAP. XVII.

*Of Secret and Swift Informations by the Species of Sound.*

HAving in the former Chapters treated ſeverally concerning the divers Ways of Secreſy and Swiftneſs in Diſcourſe; it remains that I now enquire, (according to the Method propoſed) how both theſe may be joined together in the Conveyance of any Meſſage. The Reſolution of which, ſo far as it concerns the Particulars already ſpecify'd, were but needleſs to repeat.

That which does more immediately belong to the preſent *Quære*, and was the main Occaſion of this Diſcourſe, does refer to other Ways of Intimation, beſides theſe in ordinary Uſe, of Speaking, or Writing, or Geſtures. For in the general we muſt note, That *whatever is capable of a competent Difference, perceptible to any Senſe, may be a ſufficient Means whereby to expreſs the Cogitations.* It is more convenient, indeed, that theſe Differences ſhould be of as great Variety as the Letters of the Alphabet; but it is ſufficient if they be but twofold, becauſe Two alone may, with ſomewhat more Labour and Time, be well enough contrived to expreſs all the reſt. Thus any two Letters or Numbers, ſuppoſe A. B. being tranſpoſed through five Places, will yield Thirty two Differences, and ſo conſequently will ſuperabundantly ſerve for the Four and twenty Letters, as was before more largely explained in the Ninth Chapter.

# James Gleick on John Wilkins

- Gleick, James 2011. *The Information: A History, A Theory, A Flood*. New York: Pantheon, discovered this stunning 3-century anticipation of idea that information = differences in 1641 (Newton born in 1642).
- "Any difference meant a binary choice. Any binary choice began the expressing of cogitations. Here, in this arcane and anonymous treatise of 1641, the essential idea of information theory poked to the surface of human thought, saw its shadow, and disappeared again for four hundred years." [Gleick 2011, p. 161] (actually 300 years)

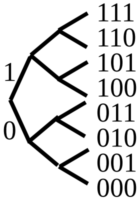
# Overview of Basic Theme: Information = Distinctions

- Two related notions of "information content" or "entropy" of a probability distribution $p = (p_1, ..., p_n)$:
  - Shannon entropy in base 2: $H(p) = H_2(p) = \sum_i p_i \log_2(1/p_i)$, or Shannon entropy that is base-free: $H_m(p) = \prod_i \left(\frac{1}{p_i}\right)^{p_i} = 2^{H_2(p)}$;
  - Logical entropy: $h(p) = \sum_i p_i (1 - p_i) = 1 - \sum_i p_i^2$.
- Logical entropy arises out of partition logic–just as finite probability theory arises out of ordinary subset logic;
- Logical entropy and Shannon entropy (in the base-dependent or base-free versions) are all just different ways to measure the amount of distinctions.

# Interpretation of Shannon entropy

- $H(p) = \sum_i p_i \log_2 \left(\frac{1}{p_i}\right)$ is usually interpreted as the average minimum number of yes-or-no questions needed to distinguish or single-out a chosen element from among $n$ with the probabilities $p = (p_1, ..., p_n)$.
- Example: Game of 20 questions with $2^n$ equipossible choices. Code $2^n$ elements with $n$ binary digits. Ask $n$ binary questions: "Is $i^{th}$ digit a 1?" for $i = 1, ..., n$.

Shannon entropy of $p = \left\{\frac{1}{8}, ..., \frac{1}{8}\right\}$ is $H(p) = \sum_{i=1}^{8} \frac{1}{8} \log_2 \left(\frac{1}{1/8}\right) = 8 \times \frac{1}{8} \times 3 = 3.$

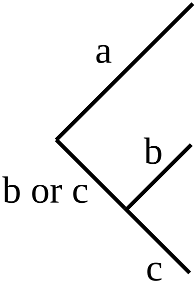



$2^3 = 8$

# Shannon entropy with unequal probs: I

- Now suppose the choices or messages are not equiprobable but that probabilties are still powers of $1/2$. With an alphabet of $a, b, c$, let $p_a = \frac{1}{2}$ and $p_b = \frac{1}{4} = p_c$.
- 1 character messages: efficient minimum number of questions (on average) are:

  - "Is the message "$a$"? If "yes" then finished, and if not, then:
  - "Is the message "$b$"? Either way, message is determined.



# Shannon entropy with unequal probs: II

- The efficient binary code for these messages is just a description of the questions:

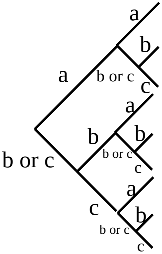
$$a = 1;\; b = 01; c = 00.$$

- Average # questions:

$$\frac{1}{2}\#\left(1\right) + \frac{1}{4}\#\left(01\right) + \frac{1}{4}\#\left(00\right) = \left(\frac{1}{2} \times 1\right) + \left(\frac{1}{4} \times 2\right) + \left(\frac{1}{4} \times 2\right)$$
$$= \frac{3}{2} = \sum p_i \log_2\left(\frac{1}{p_i}\right) = H\left(p\right).$$

# Unequal probabilities: 2 character messages

- 2 character messages:
- $\sum p_i \#(q's) =$ $\frac{2}{4} + \frac{3}{8} + \frac{3}{8} + \frac{3}{8} + \frac{4}{16} +$ $\frac{4}{16} + \frac{3}{8} + \frac{4}{16} + \frac{4}{16} =$ $\frac{48}{16} = 3 = 2H(p)$.
- Average #questions per character = $2H(p)/2 = H(p)$.
- In general, $H(p)$ interpreted as average #questions per character.

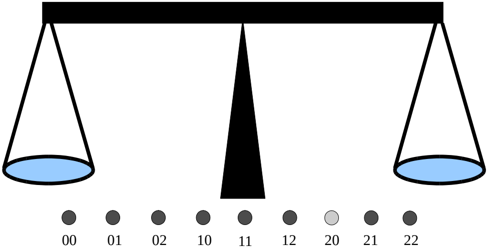


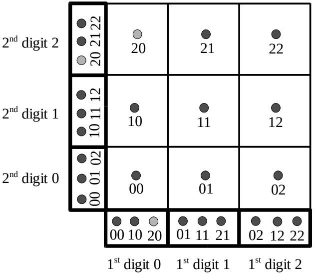

aa = 11 $\#(11)/2^2$
ab = 101 $\#(101)/2^3$
ac = 100 $\#(100)/2^3$
ba = 011 $\#(011)/2^3$
bb = 0101 $\#(0101)/2^4$
bc = 0100 $\#(0100)/2^4$
ca = 001 $\#(001)/2^3$
cb = 0001 $\#(0001)/2^4$
cc = 0000 $\#(0000)/2^4$

# Shannon entropy with a different base: I

- Given $3^n$ identical looking coins, one counterfeit (lighter than others) and a balance scale.
- Find counterfeit coin with $n$ ternary questions:
  - Code the coins in ternary arithmetic so each coin has $n$ ternary digits ("trits").
  - $i^{th}$ question = "What is $i^{th}$ ternary digit of counterfeit coin?"
  - $H_3(p) = \sum_{i=1}^{3^n} \frac{1}{3^n} \log_3\left(\frac{1}{1/3^n}\right) = 3^n \times \frac{1}{3^n} \times n = n$ questions.
  - Asking questions by weighing:
    - Group coins in three piles according to $i^{th}$ ternary digit.
    - Put two piles on balance scale. If one side light, coin is in the group; otherwise in third pile.

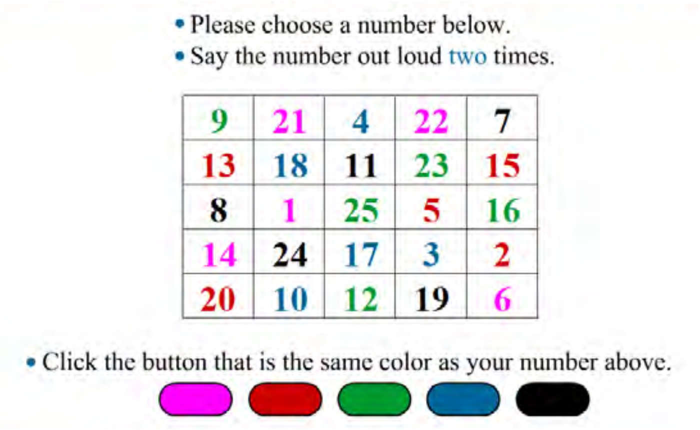
# Shannon entropy with a different base: II




- 2 weighings = 2 questions = $H_3(p)$ where $p = \left(\frac{1}{9}, \ldots, \frac{1}{9}\right)$.

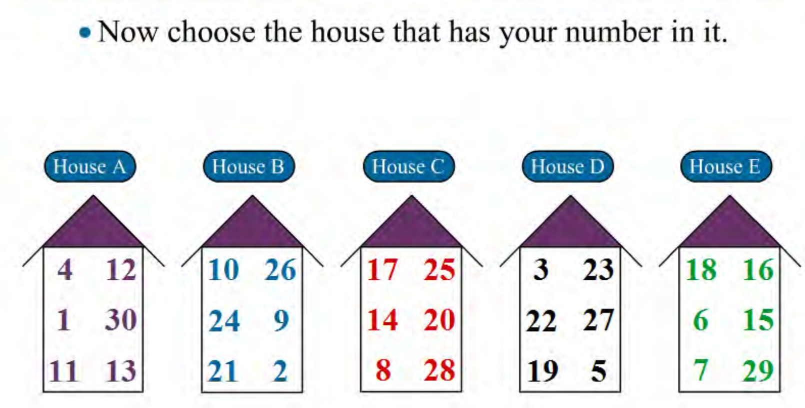
- Go to this web example of $5^2$: http://www.quizyourprofile.com/guessyournumber.swf



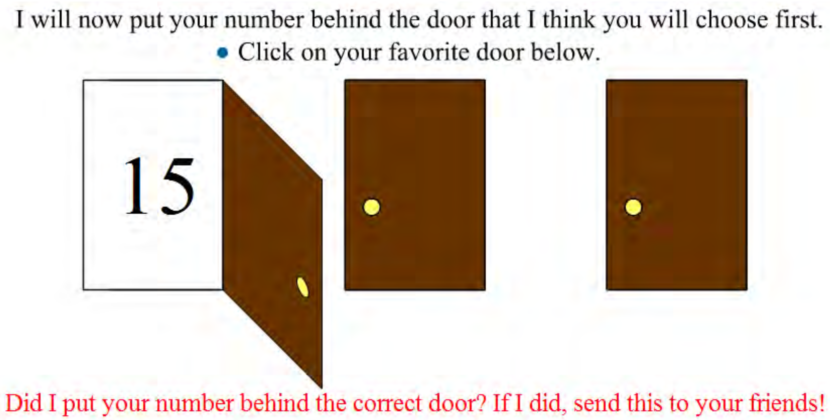

| | | | | |
|---|---|---|---|---|
| 9 | 21 | 4 | 22 | 7 |
| 13 | 18 | 11 | 23 | 15 |
| 8 | 1 | 25 | 5 | 16 |
| 14 | 24 | 17 | 3 | 2 |
| 20 | 10 | 12 | 19 | 6 |

# Web example with base 5: II

- Choosing color is equivalent to choosing one base-5 digit in a two digit number.



- Choosing house is equivalent to choosing another base-5 digit so number is determined.



# Partitions dual to Subsets: I

- Ordinary "propositional" logic is viewed as subset logic–where all operations are viewed as subset operations on subsets of universe $U$ ("propositional" special case where $U$ is one element set 1 with subsets 1 and 0);
- Category-theoretic duality between monomorphisms and epimorphisms:
  - For Sets, it is the duality of subsets of a set and partitions on a set;
  - Duality throughout algebra between subobjects and quotient objects.
- Lattice of subsets $\wp(U)$ (power-set) of $U$ with inclusion order, join = union, meet = intersection, top = $U$, and bottom = $\emptyset$;

# Partitions dual to Subsets: II

- Lattice of partitions $\prod(U)$ on $U$ where for partitions $\pi = \{B\}$ and $\sigma = \{C\}$:
  - refinement ordering: $\sigma \preceq \pi$ if $\forall B \in \pi, \exists C \in \sigma$ with $B \subseteq C$ ($\pi$ refines $\sigma$);
  - join of partitions $\pi \vee \sigma$: blocks are non-empty intersections $B \cap C$, and
  - meet of partitions $\pi \wedge \sigma$: define undirected graph on $U$ with link between $u$ and $u'$ if they are in same block of $\pi$ or $\sigma$. Then connected components of graph are blocks of meet.
  - Top = discrete partition of singleton blocks: $1 = \{\{u\} : u \in U\}$, and bottom = indiscrete partition with one block: $0 = \{U\}$.
  - NB: in combinatorial theory literature, $\prod(U)$ is usually written upside-down with "unrefinement" ordering that reverses join and meet, and top and bottom.

# Representing lattice of partitions in UxU: I

- An *equivalence relation* on $U$ is a reflexive, symmetric, and transitively closed subset $E \subseteq U \times U$.
- Given a partition $\pi = \{B, B', ...\}$, it is the set of equivalence classes of an equivalence relation $\text{indit}(\pi) = \{(u, u') : \exists B \in \pi, u, u' \in B\}$, the *indistinctions* of $\pi$.
- Upside-down lattice $\prod (U)^{op}$ is lattice of equivalence relations on $U$.
- Complement $E^c = (U \times U) - E$ of an equivalence relation is a *partition relation*, i.e., an anti-reflexive, symmetric, and anti-transitive subset.

# Representing lattice of partitions in UxU: II

- Given a partition $\pi$, the partition relation is: $\operatorname{dit}(\pi) = \{(u, u') : \exists B, B' \in \pi, B \neq B', u \in B, u' \in B'\}$, the *distinctions* of $\pi$ where $\operatorname{dit}(\pi) = \operatorname{indit}(\pi)^c$.
- Equivalence relations are *closed subsets* of $U \times U$ with *closure* $\overline{S}$ as the refl.-symm.-trans. closure. Then partition relations are *open subsets* and interior operation is $\operatorname{int}(S) = \overline{(S^c)}^c$. Closure op is not topological since $\overline{S} \cup \overline{T}$ is not nec. closed.
- Lattice of partition relations $\mathcal{O}(U)$ on $U \times U$ is isomorphic to $\prod(U)$, so the partition relations give a representation of $\prod(U)$ with the isomorphism: $\pi \longleftrightarrow dit(\pi)$:
  - $\sigma \preceq \pi$ iff $\operatorname{dit}(\sigma) \subseteq \operatorname{dit}(\pi)$;
  - $\operatorname{dit}(\pi \vee \sigma) = \operatorname{dit}(\pi) \cup \operatorname{dit}(\sigma)$;
  - $\operatorname{dit}(\pi \wedge \sigma) = int(\operatorname{dit}(\pi) \cap \operatorname{dit}(\sigma))$;
  - Top = $U \times U - \Delta$ and bottom = $\emptyset$.

# First table of analogies between subset and partition logic

| | **Subset Logic** | **Partition Logic** |
|---|---|---|
| 'Elements' | Elements $u \in U$ | Distinctions $(u,u') \in (U \times U) - \Delta_U$ |
| All elements | Universe set U | Discrete partition **1** (all dits) |
| No elements | Empty set $\emptyset$ | Indiscrete partition **0** (no dits) |
| Variables in formulas | Subset $S \subseteq U$ | Partition $\pi$ on U |
| Interpretation | $f:S' \to U$ so $Im(S') = S$ defines *property* on U. | $f:U \to R$ so $f^{-1}(R) = \pi$ defines R-valued *attribute* on U. |
| Logical operations | Subset ops $\cup$, $\cap$, $\Rightarrow$,… | Partition ops $\cong$ Interior of subset ops applied to dit-sets. |
| Formula $\Phi(\pi, \sigma,\ldots)$ holds of an element | Element u is in $\Phi(\pi,\sigma,\ldots)$ as a subset. | A dit (u,u') is distinguished by $\Phi(\pi,\sigma,\ldots)$ as a partition. |
| Valid formula $\Phi(\pi,\sigma,\ldots)$ | $\Phi(\pi,\sigma,\ldots) = U$ (top) for any subsets $\pi,\sigma,\ldots$ of any U ($1 \leq \lvert U \rvert$). | $\Phi(\pi,\sigma,\ldots) = \mathbf{1}$ (top = discrete partition) for any partitions $\pi,\sigma,\ldots$ on any U ($2 \leq \lvert U \rvert$). |

# Second table of analogies

| | Finite Prob. Theory | Logical Information Theory |
|---|---|---|
| 'Outcomes' | Elements $u \in U$ finite | Pairs $(u,u') \in U \times U$ finite |
| 'Events' | Subsets $S \subseteq U$ | Partitions $\pi$, i.e., $\mathrm{dit}(\pi) \subseteq U \times U$ |
| Normalized size | *Probability* $\Pr(S) = \|S\|/\|U\|$ = number of elements (normalized). | $h(\pi) = \|\mathrm{dit}(\pi)\|/\|U \times U\|$ = *Logical Entropy* of partition $\pi$ = no. of distinctions (normalized). |
| Equiprobable outcomes | $\Pr(S)$ = probability randomly drawn element is in subset S | $h(\pi)$ = probability randomly drawn pair (w/replacement) is distinguished by partition $\pi$ |
| Generalize to finite prob. distribution | Partition $\pi = \{S_i\}$ with $\Pr(S_i) = p_i$ gives $p = \{p_1,\ldots,p_n\}$ | $h(\pi) = \|\mathrm{dit}(\pi)\|/\|U^2\| = \sum_{i\neq j}\|S_i\|\|S_j\|/\|U^2\| = \sum_{i\neq j} p_i p_j = 1-\sum p_i^2 = h(p)$. |

# Elementary Information Theory

## Shannon entropy and logical entropy

David Ellerman

UCR

January 2012

# Shannon entropy (base 2 and base-free) and logical entropy

For a partition $\pi = \{B\}$ on a finite universe set $U$ with probability $p_B = \frac{|B|}{|U|}$ of a random drawing giving an element of the block $B$:

- Shannon entropy (base 2):
  $H(\pi) = H_2(\pi) = \sum_{B\in\pi} p_B \log_2\left(\frac{1}{p_B}\right)$;
- Shannon entropy (base-free):
  $H_m(\pi) = \prod_{B\in\pi}\left(\frac{1}{p_B}\right)^{p_B} = 2^{H_2(\pi)} = 3^{H_3(\pi)} = e^{H_e(\pi)} = \ldots$;
- Logical entropy: $h(\pi) = \sum_{B\in\pi} p_B(1-p_B) = 1 - \sum_{B\in\pi} p_B^2$.

# Block entropies

Each entropy is an average (arithmetical or geometric) of *block entropies*:

- Shannon base 2 block entropy: $H_2(B) = \log_2\left(\frac{1}{p_B}\right)$ so average is: $H_2(\pi) = \sum_B p_B H_2(B)$;
- Shannon base-free block entropy: $H_m(B) = \frac{1}{p_B}$ so geometrical average is: $H_m(\pi) = \prod_B H_m(B)^{p_B}$;
- Logical block entropy: $h(B) = 1 - p_B$ so average is: $h(\pi) = \sum_B p_B h(B)$.

Mathematical relationship between block entropies:

$$\boxed{h(B) = 1 - \frac{1}{H_m(B)} = 1 - \frac{1}{2^{H(B)}}}.$$

# Mutual information: I

Given two partitions $\pi = \{B\}$ and $\sigma = \{C\}$:

- Think of block entropies $H(B) = \log\left(\frac{1}{p_B}\right)$ (all logs base 2 unless otherwise specified) like a subset in a heuristic "Venn diagram" and same for $H(C) = \log\left(\frac{1}{p_C}\right)$. Block entropies for join $\pi \vee \sigma$ are $H(B \cap C) = \log\left(\frac{1}{p_{B\cap C}}\right)$ are like the unions of the "subsets" in the "Venn diagram." By this heuristics, the block entropies for the *mutual information* $I(B;C)$ are the overlaps in the "Venn diagram" which can be computed as the sum minus the union:

$$H(B) + H(C) - H(B \cap C) =$$
$$\log\left(\frac{1}{p_B}\right) + \log\left(\frac{1}{p_C}\right) - \log\left(\frac{1}{p_{B\cap C}}\right) = \log\left(\frac{p_{B\cap C}}{p_B p_C}\right).$$

# Mutual information: II

Then the average mutual information is:

*Shannon mutual information*: $I(\pi;\sigma) = \sum_{B\in\pi, C\in\sigma} p_{B\cap C} I(B;C)$.

- If information = distinctions, then mutual information = mutual distinctions. Thus for logical entropy, the mutual information $m(\pi;\sigma)$ is obtained by the actual Venn diagram in the closure space $U \times U$:

*Logical mutual information*: $m(\pi;\sigma) = \frac{|\operatorname{dit}(\pi)\cap\operatorname{dit}(\sigma)|}{|U\times U|}$.

- Inclusion-exclusion principle follows from heuristic or actual Venn diagram:
  - $I(\pi;\sigma) = H(\pi) + H(\sigma) - H(\pi\vee\sigma)$ for Shannon entropy.
  - $m(\pi;\sigma) = h(\pi) + h(\sigma) - h(\pi\vee\sigma)$ for logical entropy.

# Stochastically independent partitions: I

- Partitions $\pi$ and $\sigma$ are *(stochastically) independent* if $\forall B \in \pi, C \in \sigma$:

$$p_{B\cap C} = p_B p_C.$$

- For Shannon, one of the main motivations for using the log-version rather than the base-free notion was:

$$\text{If } \pi \text{ and } \sigma \text{ are independent: } H(\pi \vee \sigma) = H(\pi) + H(\sigma)$$
$$\text{so that: } I(\pi;\sigma) = 0.$$

- For Shannon base-free entropy: $\pi, \sigma$ independent implies $H_m(\pi \vee \sigma) = H_m(\pi) H_m(\sigma)$.

# Stochastically independent partitions: II

- Since logical entropy has a direct probabilistic interpretation [$h(\pi)$ = prob. randomly drawing a pair distinguished by $\pi$], we have:

$1 - h(\pi \vee \sigma) =$ prob. drawing a pair not-distinguished by $\pi \vee \sigma$
= prob. pair not-distinguished by $\pi$ and not-distinguished by $\sigma$
= (using independence) prob. pair not-distinguished by $\pi$ times prob. pair not-distinguished by $\sigma$
= $[1 - h(\pi)][1 - h(\sigma)]$ so:

If $\pi$ and $\sigma$ are independent:

$$1 - h(\pi \vee \sigma) = [1 - h(\pi)][1 - h(\sigma)]$$

so that: $m(\pi;\sigma) = h(\pi)\,h(\sigma)$.

# Conditional entropy: I

- Given a block $C \in \sigma$, $\pi = \{B\}$ induces a partition $\{B \cap C\}$ on $C$ with the prob. distribution $p_{B|C} = \frac{p_{B\cap C}}{p_C}$ so we have the Shannon entropy: $H(\pi|C) = \sum_{B\in\pi} p_{B|C} \log\left(\frac{1}{p_{B|C}}\right)$. Then the *Shannon conditional entropy* is defined as the average of these entropies:

$$H(\pi|\sigma) = \sum_{C\in\sigma} p_C H(\pi|C)$$
$$= H(\pi \vee \sigma) - H(\sigma) = H(\pi) - I(\pi;\sigma).$$

- This is interpreted as the information in $\pi$ given $\sigma$ is the information in both minus the information in $\sigma$, which also is the information in $\pi$ minus the mutual information.
- Under independence: $H(\pi|\sigma) = H(\pi)$.

# Conditional entropy: II

- Since information = distinctions, the *logical conditional entropy* is just the (normalized) distinctions of $\pi$ that were not distinctions of $\sigma$:

$$h(\pi|\sigma) = \frac{|\operatorname{dit}(\pi) - \operatorname{dit}(\sigma)|}{|U \times U|}$$
$$= h(\pi \vee \sigma) - h(\sigma) = h(\pi) - m(\pi;\sigma).$$

- The interpretation is the probability that a randomly drawn pair is distinguished by $\pi$ but not by $\sigma$.
- Under independence: $h(\pi|\sigma) = h(\pi)[1 - h(\sigma)]$ = prob. random pair is distinguished by $\pi$ times the prob. random pair is not distinguished by $\sigma$.

# Cross-entropy and divergence: I

Given pdf's $p = (p_1, ..., p_n)$ and $q = (q_1, ..., q_n)$ (instead of two partitions):

- *Shannon cross-entropy* is defined as: $H(p\|q) = \sum_i p_i \log\left(\frac{1}{q_i}\right)$ (which is non-symmetric) where if $p = q$, then $H(p\|q) = H(p)$.
- *Kullback-Leibler divergence* is defined as:

$$D(p\|q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right) = H(p\|q) - H(p).$$

Basic information inequality: $D(p\|q) \geq 0$
with equality iff $\forall i, p_i = q_i$.

# Cross-entropy and divergence: II

- *Logical cross-entropy* has simple motivation: in drawing the pair, draw once according to $p$ and once according to $q$ so that: $h(p\|q) = \sum_i p_i(1-q_i) = 1 - \sum_i p_i q_i$ = prob. drawing a distinction [where, obviously, $h(p\|q) = h(q\|p)$ and if $p = q$, then $h(p\|q) = h(p)$].
- Obvious notion of distance or divergence between two probability distributions is the Euclidean distance (squared) so the *logical divergence* is:

$$d(p\|q) = \sum_i (p_i - q_i)^2 = 2h(p\|q) - h(p) - h(q).$$

Basic information inequality: $d(p\|q) \geq 0$

with equality iff $\forall i, p_i = q_i$.

# Table of analogous formulas

| | Shannon Entropy | Logical Entropy |
|---|---|---|
| Block entropy | $H(B) = \log(1/p_B)$ | $h(B) = 1-p_B$ |
| Entropy | $H(\pi) = \Sigma_B p_B H(B)$ | $h(\pi) = \Sigma_B p_B h(B)$ |
| Mutual Information | $I(\pi;\sigma) = H(\pi)+H(\sigma)-H(\pi\vee\sigma)$ | $m(\pi;\sigma) = h(\pi)+h(\sigma)-h(\pi\vee\sigma)$ |
| Independence | $I(\pi;\sigma) = 0$ | $m(\pi;\sigma) = h(\pi)h(\sigma)$ |
| Conditional entropy | $H(\pi\|\sigma) = H(\pi\vee\sigma)-H(\sigma) = H(\pi)-I(\pi;\sigma)$ | $h(\pi\|\sigma) = h(\pi\vee\sigma)-h(\sigma) = h(\pi)-m(\pi;\sigma)$ |
| Cross entropy | $H(p\|\|q) = \Sigma p_i \log(1/q_i)$ | $h(p\|\|q) = \Sigma p_i(1-q_i)$ |
| Divergence | $D(p\|\|q) = H(p\|\|q)-H(p)$ | $d(p\|\|q) = 2h(p\|\|q) - h(p) - h(q)$ |
| Information Inequality | $D(p\|\|q) \geq 0$ with = iff $p_i = q_i$ for all i. | $d(p\|\|q) \geq 0$ with = iff $p_i = q_i$ for all i. |

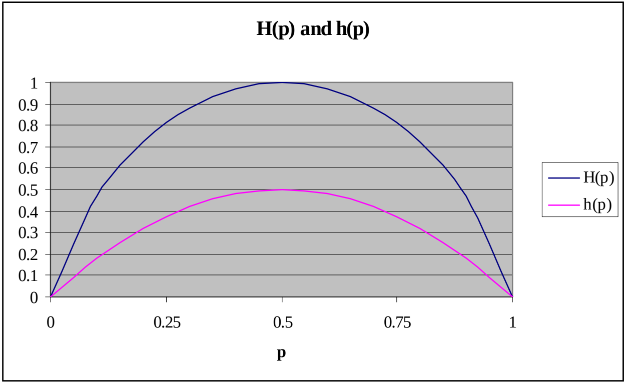
## Special cases of interest: I

- For the indiscrete partition $0 = \{U\}, H(0) = h(0) = 0$.
- For the discrete partition $1 = \{\{u\}\}_{u \in U}$ where $|U| = n$ or equivalently, for $p = (p_1, ..., p_n)$ with $p_i = \frac{1}{n}$:

$$\begin{aligned} H(1) &= H\left(\frac{1}{n}, ..., \frac{1}{n}\right) = \log n \\ H_m(1) &= n \\ h(1) &= h\left(\frac{1}{n}, ..., \frac{1}{n}\right) = 1 - \frac{1}{n} \end{aligned}$$

- Note that: $h(1) = 1 - \frac{1}{n}$ = probability of not drawing the same element twice.

# Special cases of interest: II




For two element distributions $(p, 1-p)$.
Note: $h(p, 1-p) = 2p(1-p)$ = variance of binomial dist. for sampled pairs = prob. not sampling same outcome twice.

# Introduction to density matrices and all that

## Pure states and mixed states

David Ellerman

UCR

January 2012

$H = n$-dimensional Hilbert space:

- Given $m$ (not nec. orthog.) state vectors $|\psi_1\rangle, ..., |\psi_m\rangle$ ($m$ not related to dimension $n$) and a finite probability distribution $p = (p_1, ..., p_m)$, this defines the *density operator*:

$$\rho = \sum_{i=1}^{m} p_i |\psi_i\rangle \langle\psi_i|.$$

- The density operator $\rho$ is said to represent a *pure state* if $\rho^2 = \rho$, i.e., $m = 1$ so $\rho = |\psi\rangle\langle\psi|$ for some state vector $|\psi\rangle$. Otherwise, $\rho$ is said to represent a *mixed state*.
- Motivation: think of a quantum ensemble where proportion $p_i$ of the ensemble is in state $|\psi_i\rangle$. Then the density operator represents all the probabilistic information in the ensemble.
- Nota bene: a pure state is any state which may be a superposition of eigenstates of an observable (don't confuse "mixed" and "superposition").

# Density matrices and traces

- Density operators are Hermitian, $\rho = \rho^\dagger$, and positive semidefinite, $\langle\psi|\rho|\psi\rangle \geq 0$ for any $|\varphi\rangle$.
- Given any orthonormal basis $\left\{\left|\varphi_j\right\rangle\right\}$, a density operator can be represented as an $n \times n$ *density matrix* using that basis with the $i, j$-entry: $\rho_{ij} = \left\langle \varphi_i \left|\rho\right| \varphi_j \right\rangle$.
- The *trace* of a matrix is the sum of its diagonal elements. For any density matrix:

$$\operatorname{tr}(\rho) = \sum_{j=1}^{n} \rho_{jj} = 1.$$

- Recall that the trace of a matrix is invariant under similarity transformations. In particular, if $\rho$ was diagonalized by $S$ to give the diagonal matrix of $\rho$'s eigenvalues, then the eigenvalues are non-negative (positive-definiteness) and their sum is $\operatorname{tr}\left(S\rho S^{-1}\right) = \operatorname{tr}(\rho) = 1$ and thus form a probability distribution.

# Density matrix for a pure state: I

- Let $\rho = |\psi\rangle\langle\psi|$ where $|\psi\rangle = \sum_{j=1}^{n} c_j \left|\varphi_j\right\rangle$ for an orthonormal basis $\left\{\left|\varphi_j\right\rangle\right\}$.

$$\rho = |\psi\rangle\langle\psi| = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \begin{bmatrix} c_1^* & \cdots & c_n^* \end{bmatrix} = \begin{bmatrix} c_1 c_1^* & \cdots & c_1 c_n^* \\ \vdots & \ddots & \vdots \\ c_n c_1^* & \cdots & c_n c_n^* \end{bmatrix} \text{ so}$$

$$\rho_{ij} = c_i c_j^*.$$

- Diagonal entries are $c_i c_i^* = |c_i|^2$ = probabilities of getting $i^{th}$ outcome when measuring $|\psi\rangle$ using observable with eigenvectors $\left\{\left|\varphi_j\right\rangle\right\}$.

# Density matrix for a pure state: II

- Off-diagonal entries $c_i c_j^*$ ($i \neq j$) are interpreted recalling that complex number $c_i$ can be represented in polar form as $c_i = |c_i| \, e^{-i\phi_i}$ so off-diagonal entry is: $c_i c_j^* = |c_i| \, |c_j| \, e^{-i\left(\phi_i - \phi_j\right)}$ which represents the degree of coherence in the superposition state.
- With each pure state $\rho = |\psi\rangle \langle\psi|$, we may associate a mixed state $\hat{\rho}$ that samples with the same probabilities for the basis states $\left|\varphi_j\right\rangle$ as $\rho$ measures: $\hat{\rho} = \sum_{j=1}^{n} c_j c_j^* \left|\varphi_j\right\rangle \left\langle \varphi_j\right|$ = diagonal matrix with diagonal entries $c_j c_j^* = \left|c_j\right|^2$ probabilities.
- $\hat{\rho}$ is the *decohered* $\rho$ that represents the change due to a measurement of $|\psi\rangle$ with an observable with the eigenstates $\left\{\left|\varphi_j\right\rangle\right\}$.

# Other properties of density matrices

- $\operatorname{tr}\left(\rho^2\right) \leq 1$ with equality iff $\rho^2 = \rho$, i.e., $\rho$ is a pure state.
- If $\rho = \sum_{i=1}^{m} p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|$ and $A$ is a Hermitian operator, then the $\rho$ *ensemble average* of $A$ is:

$$\begin{aligned}
[A]_\rho &= \sum_{i=1}^{m} p_i \left\langle \psi_i \left| A \right| \psi_i \right\rangle \\
&= \sum_i p_i \sum_{k=1}^{n} \sum_{j=1}^{n} \left\langle \psi_i \middle| \varphi_k \right\rangle \left\langle \varphi_k \right| A \left| \varphi_j \right\rangle \left\langle \varphi_j \middle| \psi_i \right\rangle \\
&= \sum_{k=1}^{n} \sum_{j=1}^{n} \left[ \sum_i p_i \left\langle \varphi_j | \psi_i \right\rangle \left\langle \psi_i | \varphi_k \right\rangle \right] \left\langle \varphi_k | A | \varphi_j \right\rangle \\
&= \sum_{j,k} \left\langle \varphi_j | \rho | \varphi_k \right\rangle \left\langle \varphi_k | A | \varphi_j \right\rangle = \sum_j \left\langle \varphi_j | \rho A | \varphi_j \right\rangle = \operatorname{tr}\left(\rho A\right).
\end{aligned}$$

- $[A]_\rho = \operatorname{tr}\left(\rho A\right)$ is a strong result with many applications.

Let $\{|1\rangle, |2\rangle, |3\rangle\}$ be an orthonormal basis in three-dimensional Hilbert space.

- Let $|A\rangle = \frac{1}{2}\left(|1\rangle + \sqrt{2}\,|2\rangle + |3\rangle\right)$ be a superposition state, so its pure state density matrix is:

$$\rho = |A\rangle\langle A| = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{\sqrt{2}} \\ \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4}\sqrt{2} & \frac{1}{4} \\ \frac{1}{4}\sqrt{2} & \frac{1}{2} & \frac{1}{4}\sqrt{2} \\ \frac{1}{4} & \frac{1}{4}\sqrt{2} & \frac{1}{4} \end{bmatrix}.$$

- As a pure state density matrix, $\rho^2 = \rho$.

- If measured by an operator with the eigenstates $\{|1\rangle, |2\rangle, |3\rangle\}$ then the diagonal entries of $\rho$ expressed in that basis are the probabilities $p(i)$ of those eigenstates: $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ respectively.
- The measurement makes the transition from the pure to the mixed state: $\rho \to \hat{\rho} =$ "decohered $\rho$".

$$\hat{\rho} = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix}.$$

- Then $\operatorname{tr}\left(\hat{\rho}^2\right) = \sum_{i=1}^{3} p(i)^2 = \frac{1}{16} + \frac{1}{4} + \frac{1}{16} = \frac{3}{8}$ instead of $\operatorname{tr}\left(\rho^2\right) = 1$ for the pure state $\rho$.

# Example 2: From pure to completely mixed states: I

- Consider the equal-amplitude pure state:

$$|\psi\rangle = \frac{1}{\sqrt{3}}|1\rangle + \frac{1}{\sqrt{3}}|2\rangle + \frac{1}{\sqrt{3}}|3\rangle.$$

$$\rho = |\psi\rangle\langle\psi| = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

- If measured by an operator with the eigenstates $\{|1\rangle, |2\rangle, |3\rangle\}$ then the diagonal entries of $\rho$ are the equal probabilities $\frac{1}{3}$ of getting one of the eigenstates.
- The decohered version is:

# Example 2: From pure to completely mixed states: II

$$\hat{\rho} = \frac{1}{3}I = \frac{1}{3}\left(|1\rangle\langle 1| + |2\rangle\langle 2| + |3\rangle\langle 3|\right) = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}.$$

- Such an equiprobable mixed state is called a *completely mixed state*.
- In $n$-dim. Hilbert space, a completely mixed state $\hat{\rho}$ has $\operatorname{tr}\left(\hat{\rho}^2\right) = \frac{1}{n}$ where in this case: $\operatorname{tr}\left(\hat{\rho}^2\right) = 3 \times \frac{1}{9} = \frac{1}{3}$. For $n = 2$, unpolarized light is a completely mixed state.
- In general: $\frac{1}{n} \leq \operatorname{tr}\left(\rho^2\right) \leq 1$ ($n$-dim. space) with the two extremes being completely mixed states and pure states.
- Note similarity: $\operatorname{tr}\left(\rho^2\right) \approx \sum_i p_i^2$ for probability distributions with the two extremes being $\left(\frac{1}{n}, ..., \frac{1}{n}\right)$ and $(0, ..., 0, 1, 0, ..., 0)$.

# Unitary evolution of density matrices: I

- Time evolution of state can be given by unitary operator $U(t, t_0)$ so that:

$$|\psi(t)\rangle = U(t, t_0)\,|\psi(t_0)\rangle.$$

- A density matrix $\rho$ is used to represent a pure or mixed state but it is an operator $\rho : H \to H$ on the Hilbert space so the time evolution of the operator $\rho(t)$ is obtained by the operator that:

1. uses $U(t, t_0)^{-1} = U(t, t_0)^{\dagger}$ to translate state back to time $t_0$,
2. apply the operator $\rho(t_0)$, and
3. use $U(t, t_0)$ to translate the result back to time $t$:

$$\rho(t) = U(t, t_0)\,\rho(t_0)\,U(t, t_0)^{-1} : H \to H.$$

# Unitary evolution of density matrices: II

- Then we have: $\rho(t)^2 = U(t,t_0)\rho(t_0)^2 U(t,t_0)^{-1}$ so if $\rho(t_0)^2 = \rho(t_0)$, then $\rho(t)^2 = \rho(t)$, i.e.,

  unitary evolution always takes pure states to pure states.

- The simple idea of a (projective) measurement is when a pure $\rho = |\psi\rangle\langle\psi|$ is expressed in terms of the orthonormal basis of eigenvectors $\{\varphi_i\}$ for an operator $A$, then the effect of the measurement is $\rho \to \hat{\rho}$, to go from a pure state to the decohered mixed state of the probability-weighted eigenstates.
- This cannot be a unitary evolution since unitary evolutions can only take pure states $\to$ pure states.

# Decomposition of density op. not unique: I

Consider the three Pauli spin matrices:

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}; \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The eigenvectors for each operator are:

$x_+ = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}; x_- = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}; y_+ = \begin{bmatrix} \frac{-i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}; y_- = \begin{bmatrix} \frac{i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix};$

$z_+ = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $z_- = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

The projection operators to the one-dimensional subspaces spanned by these eigenvectors are:

$$P_{x+} = |x_+\rangle \langle x_+| = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

# Decomposition of density op. not unique: II

$$P_{x-} = |x_-\rangle\langle x_-| = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} =$$
$$\begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}$$

$$P_{y+} = |y_+\rangle\langle y_+| = \begin{bmatrix} \frac{-i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2}i \\ \frac{1}{2}i & \frac{1}{2} \end{bmatrix}$$

$$P_{y-} = |y_-\rangle\langle y_-| = \begin{bmatrix} \frac{i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -\frac{1}{2}i\sqrt{2} & \frac{1}{2}\sqrt{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2}i \\ -\frac{1}{2}i & \frac{1}{2} \end{bmatrix}$$

$$P_{z+} = |z_+\rangle\langle z_+| = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$P_{z-} = |z_-\rangle\langle z_-| = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Then we have:

# Decomposition of density op. not unique: III

$$\rho_{unpolarized} = \frac{1}{2}P_{x+} + \frac{1}{2}P_{x-} = \frac{1}{2}P_{y+} + \frac{1}{2}P_{y-} = \frac{1}{2}P_{z+} + \frac{1}{2}P_{z-} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}.$$

- For $|x_\pm\rangle = \frac{1}{\sqrt{2}}|x_+\rangle + \frac{1}{\sqrt{2}}|x_-\rangle$, $\rho_{x\pm} = |x_\pm\rangle\langle x_\pm| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.
- For $|y_\pm\rangle = \frac{1}{\sqrt{2}}|y_+\rangle + \frac{1}{\sqrt{2}}|y_-\rangle$ , $\rho_{y\pm} = |y_\pm\rangle\langle y_\pm| = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.
- For $|z_\pm\rangle = \frac{1}{\sqrt{2}}|z_+\rangle + \frac{1}{\sqrt{2}}|z_-\rangle$, $\rho_{z\pm} = |z_\pm\rangle\langle z_\pm| = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$.

# Tensor products, reduced density matrices, and the measurement problem

David Ellerman

UCR

January 2012

# Tensor products: I

- If a quantum system $A$ is modeled in the Hilbert space $H^A$ and similarly for a system $B$ and the Hilbert space $H^B$, then the *composite system* $AB$ is modeled in the tensor product $H^A \otimes H^B$.
- In general a concept for sets "lifts" to the appropriate vector space concept for quantum mechanics by applying the set concept to a basis set of a vector space, and then generate the corresponding vector space concept.
- Thus the appropriate v.s. concept of "product" of two spaces $V, V'$ for QM is to apply the set concept of product (i.e., the Cartesian product) to two bases for $V$ and $V'$, and then those ordered pairs of basis elements form a basis for a vector space called the *tensor product* $V \otimes V'$.

- Given a basis $\{|u_i\rangle\}$ for $V$ and a basis $\left\{\left|u'_j\right\rangle\right\}$ for $V'$, the set of all ordered pairs $|u_i\rangle \otimes \left|u'_j\right\rangle$ (often denoted as $|u_i\rangle \left|u'_j\right\rangle$ or $\left|u_i, u'_j\right\rangle$) form a basis for $V \otimes V'$.
- Tensor products are *bilinear* and *distributive* in the sense that for any $|v\rangle \in V$ and $|v'\rangle \in V'$:
  1. for any scalar $\alpha$, $\alpha\left(|v\rangle \otimes |v'\rangle\right) = \left(\alpha |v\rangle\right) \otimes |v'\rangle = |v\rangle \otimes \left(\alpha |v'\rangle\right)$;
  2. for any $|v\rangle \in V$ and $|v'_1\rangle, |v'_2\rangle \in V'$,
  $|v\rangle \otimes \left(|v'_1\rangle + |v'_2\rangle\right) = \left(|v\rangle \otimes |v'_1\rangle\right) + \left(|v\rangle \otimes |v'_2\rangle\right)$;
  3. for any $|v_1\rangle, |v_2\rangle \in V$ and $|v'\rangle \in V'$,
  $\left(|v_1\rangle + |v_2\rangle\right) \otimes |v'\rangle = \left(|v_1\rangle \otimes |v'\rangle\right) + \left(|v_2\rangle \otimes |v'\rangle\right)$.
- The *tensor product of operators* on the component spaces is obtained by applying the operators component-wise, i.e.,

$$\text{for } T : V \to V \text{ and } T' : V' \to V',$$
$$(T \otimes T')\left(|v\rangle \otimes |v'\rangle\right) = T\left(|v\rangle\right) \otimes T'\left(|v'\rangle\right).$$

- The *inner product on the tensor product* is defined component-wise on basis elements and extended (bi)linearly to the whole space:

$$\left\langle u_i, u'_j | u_k, u'_l \right\rangle = \langle u_i | u_k \rangle \left\langle u'_j | u'_l \right\rangle.$$

- The *tensor (or Kronecker) product* of an $m \times n$ matrix $A$ and a $p \times q$ matrix $B$ is the $nq \times mp$ matrix $A \otimes B$ obtained by inserting $B$ after each entry $a_{ij}$ of $A$, e.g.,

# Tensor products: IV

$$\text{if } X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \text{ then}$$

$$X \otimes H = \begin{bmatrix} 0H & 1H \\ 1H & 0H \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{bmatrix}.$$

- States of the form $|v\rangle \otimes |v'\rangle \in V \otimes V'$ are called *separated*; other states in the tensor product are *entangled*.

# The measurement problem: I

- A measurement of a quantum system $Q$, represented in $H^Q$, by a measurement apparatus $M$, represented in $H^M$, is modeled by the tensor product $H^Q \otimes H^M$.
- If the Hermitian operator $A : H^Q \to H^Q$, which represents the observable being measured, has the orthonormal eigenstates $|u_1\rangle, ..., |u_n\rangle \in H^Q$, then the idea is to pair or correlate these eigenstates with $n$ orthonormal *indicator states* $|v_1\rangle, ..., |v_n\rangle \in H^M$ in the tensor product.
- The state $|\psi\rangle = \sum_i \alpha_i |u_i\rangle \in H^Q$ is the *initial state* and there is another *initial indicator state* $|v_0\rangle \in H^M$.
- Thus the composite system starts off in the state: $|\psi\rangle \otimes |v_0\rangle$.

# The measurement problem: II

- Taking the quantum system and the measurement apparatus as together being an isolated quantum system, the initial state unitarily evolve according to QM to the entangled state: $\sum_i |u_i\rangle \otimes |v_i\rangle$ (ignoring normalization).
- But that is another superposition state, like the original $|\psi\rangle = \sum_i \alpha_i |u_i\rangle$, whereas the usual notion of a "measurement" is that that system ends up in a specific $|u_i\rangle \otimes |v_i\rangle$ state of the composite system and thus in the eigenstate $|u_i\rangle$ of the system $Q$ having the corresponding eigenvalue $\lambda_i$ as the measured value.
- What causes the "collapse of the wave-packet" or the state reduction to that eigenstate?

# The measurement problem: III

- We considered the system $QM$ represented in $H^Q \otimes H^M$ as being isolated so that it evolved according to the Schrodinger equation, i.e., by a unitary transformation of state.
- If we say the superposition $\sum_i |u_i\rangle \otimes |v_i\rangle$ was collapsed by the intervention of another system $M'$, then assuming the universality of the laws of quantum mechanics, we can consider the isolated composite system $H^Q \otimes H^M \otimes H^{M'}$ and then by the same argument will end up by a unitary transformation in another uncollapsed superposition state: $\sum_i |u_i\rangle \otimes |v_i\rangle \otimes |v_i'\rangle$.
- And so forth in what is called *von Neumann's chain*.

# The measurement problem: IV

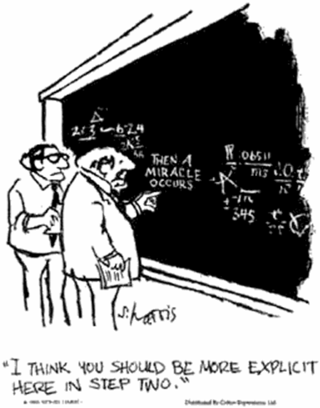
- Since the laws of QM only lead to this chain of ever larger superpositions, Schrodinger tried to show the implausibility of the chain with his famous Schrodinger's cat example.
- Others like Wigner suggested that perhaps it is human consciousness ("reading the dial") that terminates von Neumann's chain, and that led to countless books fully of fuzzy thinking about QM and consciousness. Woo-woo.
- Others like Everett have avoided the whole problem of the collapse of the superposition by assuming that the whole universe splits so that each eigenstate is continued in one of the possibilities. Thus there is splitting of worlds rather than reduction to an eigenstate in the one and only world. The utter silliness of this option (which has its followers)

# The measurement problem: V

shows the extremes to which otherwise-sane physicists are driven by "the measurement problem."

- The standard Copenhagen interpretation tries to simply eschew such questions, but that amounts to postulating a state-reducing property called "macroscopic." At some point along von Neumann's chain (e.g., at the first step), the measurement apparatus is assumed to have the property of being "macroscopic" which means that its indicator states $|v_i\rangle$ cannot be in superposition, and hence the measurement apparatus is in one of the indicator states.
- You ask, "What happened to the laws of QM in the interaction with this 'macroscopic' apparatus? When does the miracle occur?"

# The measurement problem: VI




- The Copenhagen answer is: "Don't ask."

- The mystery deepens when we analyze the measurement problem using density operators.
- We start with the state $|\psi\rangle \otimes |v_0\rangle \in H^Q \otimes H^M$ represented by the pure state density operator $\rho_0$ which unitarily evolves to the state $\sum_i |u_i\rangle \otimes |v_i\rangle$ represented by pure state density operator $\rho = \left(\sum_i |u_i\rangle \otimes |v_i\rangle\right)\left(\sum_i \langle u_i| \otimes \langle v_i|\right)$.
- But what is happening in the component system $H^Q$?
- In general, given a (pure or mixed) state $\rho$ on a tensor product $V \otimes V'$, there is a *reduced density operator* $\rho^V : V \to V$ such that for any observable operator $T : V \to V$,

$$\operatorname{tr}(\rho^V T) = \operatorname{tr}_{V'}\left(\rho\left(T \otimes I\right)\right)$$

# Reduced density operators: II

where $\mathrm{tr}_{V'}()$ is the *partial trace* defined by:

$$\mathrm{tr}_{V'}(|v_1\rangle\langle v_2| \otimes |v_1'\rangle\langle v_2'|) = |v_1\rangle\langle v_2| \, \mathrm{tr}(|v_1'\rangle\langle v_2'|) = |v_1\rangle\langle v_2| \langle v_2'|v_1'\rangle$$

"Taking the partial trace over $V'$".

- The principal fact is that if the pure state on the tensor product is a perfectly correlated "measurement state" $\sum_i \alpha_i |u_i\rangle \otimes |v_i\rangle$ (orthogonal states from both components), then the state represented by the reduced density operator $\rho^V$ is the *mixed* state:

$$\rho^V = \sum_i \alpha_i \alpha_i^* |u_i\rangle\langle u_i|.$$

# Reduced density operators: III

- This is exactly the mixture of probabilistic outcomes one would expect from a measurement on the initial state: $|\psi\rangle = \sum_i \alpha_i |u_i\rangle$.
- Here is where the usual "ignorance interpretation" of mixed states breaks down. Under that interpretation $\rho^V$, the first component system is actually in some state $|u_i\rangle$ with probability $\alpha_i\alpha_i^*$, which due to the entanglement forces the other component into the state $|v_i\rangle$. But then the composite system is in the state $|u_i\rangle \otimes |v_i\rangle$ with probability $\alpha_i\alpha_i^*$ which is a mixed state in contrast to the pure superposition state $\sum_i \alpha_i |u_i\rangle \otimes |v_i\rangle$.

# Reduced density operators: IV

- One reaction in the literature is to simply consider two different types of mixed states. For instance, Bernard D'Espagnat has "proper mixtures" (the usual sort) and "improper mixtures" (reductions of entangled pure state on tensor products), while others call them mixed states of the "first kind" and "second kind." See following Charles Bennett slide.

*Mixed States and Density Matrices*

The quantum states we have been talking about so far, identified with rays in Hilbert space, are called pure states. They represent situations of minial ignorance, where there is nothing more to know about the system. Pure states are fundamental in the sense that the quantum mechanics of any closed system can be completely described as a unitary evolution of pure states, without need of further notions. However, a very useful notion, the mixed state, has been introduced to deal with situations of greater ignorance, in particular

an ensemble $\mathcal{E}$ in which the system in question may be in any of several pure states $\psi_1$, $\psi_2$ ... with probabilities $p_1$, $p_2$ ....

a situation in which the system in question (call it $A$) is part of larger system $AB$, which itself is in an entangled pure state $\Psi(AB)$.

In open systems, a pure state may naturally evolve into a mixed state (which can also be described as a pure state of a larger system comprising the original system and its environment)

# A Bell state as a perfectly correlated measurement state: I

- Consider the Bell basis vector:
  $|\Phi^+\rangle = \frac{1}{\sqrt{2}}\left[|0_A\rangle \otimes |0_B\rangle + |1_A\rangle \otimes |1_B\rangle\right] \in \mathbb{C}^2 \otimes \mathbb{C}^2$.
- The corresponding pure state density operator is:

$$\rho = |\Phi^+\rangle\langle\Phi^+|$$
$$= \frac{1}{2}\left[|0_A\rangle \otimes |0_B\rangle + |1_A\rangle \otimes |1_B\rangle\right]\left[\langle 0_A| \otimes \langle 0_B| + \langle 1_A| \otimes \langle 1_B|\right]$$
$$= \frac{1}{2}\left[\begin{array}{c} (|0_A\rangle \otimes |0_B\rangle)(\langle 0_A| \otimes \langle 0_B|) + (|0_A\rangle \otimes |0_B\rangle)(\langle 1_A| \otimes \langle 1_B|) \\ + (|1_A\rangle \otimes |1_B\rangle)(\langle 0_A| \otimes \langle 0_B|) + (|1_A\rangle \otimes |1_B\rangle)(\langle 1_A| \otimes \langle 1_B|) \end{array}\right]$$
$$= \frac{1}{2}\left[\begin{array}{c} |0_A\rangle\langle 0_A| \otimes |0_B\rangle\langle 0_B| + |0_A\rangle\langle 1_A| \otimes |0_B\rangle\langle 1_B| \\ + |1_A\rangle\langle 0_A| \otimes |1_B\rangle\langle 0_B| + |1_A\rangle\langle 1_A| \otimes |1_B\rangle\langle 1_B| \end{array}\right].$$

- Then the reduced density operator for the first system is:

# A Bell state as a perfectly correlated measurement state: II

$$\rho^A = \frac{1}{2}\begin{bmatrix} |0_A\rangle\langle 0_A| \operatorname{tr}(|0_B\rangle\langle 0_B|) + |0_A\rangle\langle 1_A| \operatorname{tr}(|0_B\rangle\langle 1_B|) \\ + |1_A\rangle\langle 0_A| \operatorname{tr}(|1_B\rangle\langle 0_B|) + |1_A\rangle\langle 1_A| \operatorname{tr}(|1_B\rangle\langle 1_B|) \end{bmatrix}$$
$$= \frac{1}{2}\begin{bmatrix} |0_A\rangle\langle 0_A| \langle 0_B|0_B\rangle + |0_A\rangle\langle 1_A| \langle 1_B|0_B\rangle \\ + |1_A\rangle\langle 0_A| \langle 0_B|1_B\rangle + |1_A\rangle\langle 1_A| \langle 1_B|1_B\rangle \end{bmatrix}$$
$$= \frac{1}{2}\left[|0_A\rangle\langle 0_A| + |1_A\rangle\langle 1_A|\right] = \frac{1}{2}I_A.$$

- The key step is: $\langle 1_B|0_B\rangle = 0 = \langle 0_B|1_B\rangle$ which decoheres the state.
- The reduced density operator is a decohered mixed state, indeed, it is a completely mixed state (like unpolarized light).

# A Bell state as a perfectly correlated measurement state: III

- This mixed state describes the mixed state one would expect from a "wave-packet-collapsing" measurement (with the eigenstates $|0_A\rangle$ and $|1_A\rangle$) on the initial state: $|\psi\rangle = \frac{1}{\sqrt{2}}[|0_A\rangle + |1_A\rangle]$. That pure state density matrix is:

$$\rho_1 = |\psi\rangle\langle\psi| = \frac{1}{\sqrt{2}}[|0_A\rangle + |1_A\rangle]\frac{1}{\sqrt{2}}[\langle 0_A| + \langle 1_A|]$$
$$= \frac{1}{2}[|0_A\rangle\langle 0_A| + |0_A\rangle\langle 1_A| + |1_A\rangle\langle 0_A| + |1_A\rangle\langle 1_A|]$$
$$= \frac{1}{2}\left[\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} + \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} + \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix} + \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}\right]$$
$$= \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}.$$

# A Bell state as a perfectly correlated measurement state: IV

- Thus the corresponding "decohered state" $\hat{\rho}_1$ is obtained by setting all the non-diagonal elements of $\rho_1$ to 0 and the result is the reduced density matrix: $\hat{\rho}_1 = \frac{1}{2}I = \rho^A$.
- Nielsen-Chuang's mention of the decohered version of a density operator $\rho$ is given by the formulas 2.150-2.152 on p. 101 but there is a nasty typo in that they have the same symbol $\rho$ for the decohered version, rather than something like $\hat{\rho}$. Hence they have "incoherent" formulas 2.151-2 with same symbol for different $\rho$'s on the LHS and RHS.
- What happened since we need not depart from unitary evolution to get from some initial state $|\psi\rangle \otimes |v_0\rangle$ to the pure Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}\left[|0_A\rangle \otimes |0_B\rangle + |1_A\rangle \otimes |1_B\rangle\right]$?

# A Bell state as a perfectly correlated measurement state: V

1. The superposed eigenstates of $|\psi\rangle = \frac{1}{\sqrt{2}}[|0_A\rangle + |1_A\rangle]$, represented by the density operator $\rho_1$, were "marked with which-way information" in the composite state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}[|0_A\rangle \otimes |0_B\rangle + |1_A\rangle \otimes |1_B\rangle]$ represented by $\rho$.
2. That is sufficient to have the reduced state to be the incoherent completely mixed state $\rho^A$.
3. Thus instead of non-unitary jump $\rho_1 \to \hat{\rho}_1$ from a pure state to a mixed state, we have the expansion of the $|\psi\rangle$ to form the pure composite state $|\psi\rangle \otimes |v_0\rangle$ which unitarily evolves to the pure "measurement state" $|\Phi^+\rangle = \frac{1}{\sqrt{2}}[|0_A\rangle \otimes |0_B\rangle + |1_A\rangle \otimes |1_B\rangle]$ which, in terms of density operators, has the reduced state $\rho^A = \hat{\rho}_1$.

# Any change in quantum state by embedding in larger Hilbert space and reducing

## The Church of the Larger Hilbert Space

This is the name given by John Smolin to the habit of always thinking of a mixed state as a pure state of some larger system; and of any nonunitary evolution as being embedded in some unitary evolution of a larger system: No one can stop us from thinking this way; and Church members find it satisfying and helpful to their intuition:

This doctrine only makes sense in a quantum context, where because of entanglement a pure whole can have impure parts: Classically; a whole can be no purer than its most impure part.

- This is Bennett and Smolin's play on the Mormon Church which is officially "Church of the Latter Day Saints" or CLDS.

# A Quantum Eraser example of which-way marking: I

- Consider the setup of the two-slit experiment where the superposition state, $|Slit1\rangle + |Slit2\rangle$, evolves to show interference on the wall.

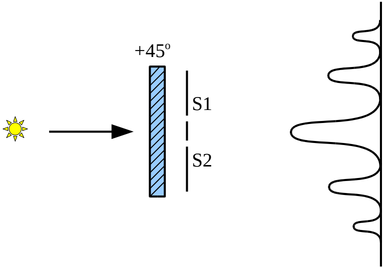



Figure 1: Interference pattern from two-slits

# A Quantum Eraser example of which-way marking: II

- Then horizontal and vertical are inserted in front of the slits which marks slit-eigenstates with which-way polarization information so the perfectly correlated "measurement state" might be represented schematically as: $|Slit1\rangle \otimes |Horiz\rangle + |Slit2\rangle \otimes |Vert\rangle$. This marking suffices to eliminate the interference pattern but it is not a "packet-collapsing" quantum jump since the state is still a pure superposition state.

# A Quantum Eraser example of which-way marking: III

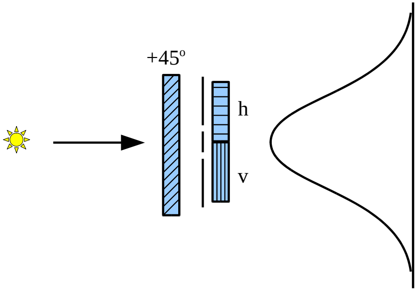



Figure 2: Mush pattern with interference eliminated by which-way markings

# A Quantum Eraser example of which-way marking: IV

- If $P_{\Delta y}$ is the projection operator representing finding a particle in the region $\Delta y$ along the wall, then that probability is:

$$\langle S1 \otimes H + S2 \otimes V | P_{\Delta y} \otimes I | S1 \otimes H + S2 \otimes V \rangle$$
$$= \langle S1 \otimes H + S2 \otimes V | P_{\Delta y} S1 \otimes H + P_{\Delta y} S2 \otimes V \rangle$$
$$= \langle S1 \otimes H | P_{\Delta y} S1 \otimes H \rangle + \langle S1 \otimes H | P_{\Delta y} S2 \otimes V \rangle$$
$$+ \langle S2 \otimes V | P_{\Delta y} S1 \otimes H \rangle + \langle S2 \otimes V | P_{\Delta y} S2 \otimes V \rangle$$
$$= \langle S1 | P_{\Delta y} S1 \rangle \langle H | H \rangle + \langle S1 | P_{\Delta y} S2 \rangle \langle H | V \rangle$$
$$+ \langle S2 | P_{\Delta y} S1 \rangle \langle V | H \rangle + \langle S2 | P_{\Delta y} S2 \rangle \langle V | V \rangle$$
$$= \langle S1 | P_{\Delta y} S1 \rangle + \langle S2 | P_{\Delta y} S2 \rangle$$

= sum of separate slot probabilities.

# A Quantum Eraser example of which-way marking: V

- The key step is how the orthogonal polarization markings decohered the state since $\langle H|V\rangle = 0 = \langle V|H\rangle$ and thus eliminated the interference between the $Slot1$ and $Slot2$ terms.
- The state-reduction occurs only when the evolved superposition state hits the far wall which measures the positional component (i.e., $P_{\Delta y}$) of the composite state and shows decohered non-interference pattern.

# A Quantum Eraser example of which-way marking: VI

- The key point is that in spite of the bad terminology of "which-way" or "which-slit" information, the polarization markings do NOT create a half-half mixture of horizontally polarized photons going through slit 1 and vertically polarized photons going through slit 2. It creates the superposition state $|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle$.
- This can be verified by inserting a $+45°$ polarizer between the two-slit screen and the far wall.

# A Quantum Eraser example of which-way marking: VII

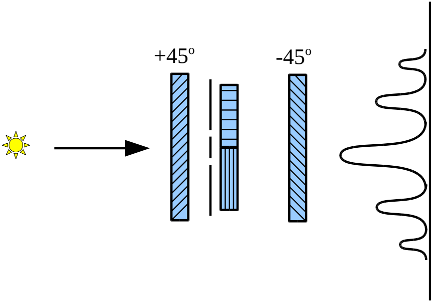



Figure 3: Fringe interference pattern produced by $+45°$ polarizer

# A Quantum Eraser example of which-way marking: VIII

- Each of the horizontal and vertical polarization states can be represented as a superposition of $+45°$ and $-45°$ polarization states. Just as the horizontal polarizer in front of slit 1 threw out the vertical component so we have no $|S1\rangle \otimes |V\rangle$ term in the superposition, so now the $+45°$ polarizer throws out the $-45°$ component of each of the $|H\rangle$ and $|V\rangle$ terms so the state transformation is:

$$|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle$$
$$\rightarrow |S1\rangle \otimes |+45°\rangle + |S2\rangle \otimes |+45°\rangle = (|S1\rangle + |S2\rangle) \otimes |+45°\rangle.$$

# A Quantum Eraser example of which-way marking: IX

- Then at the wall, the positional measurement of the first component is the evolved superposition $|S1\rangle + |S2\rangle$ which again shows an interference pattern. But it is NOT the original interference pattern before any polarizers were inserted since only half the photons (statistically speaking) got through the $+45°$ polarizer. This "shifted" interference pattern is called the *fringe* pattern.
- Alternatively we could insert a $-45°$ polarizer which would transform the state $|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle$ into $(|S1\rangle + |S2\rangle) \otimes |-45°\rangle$ which produces the interference pattern from the "other half" of the photons and which is called the *anti-fringe* pattern.

# A Quantum Eraser example of which-way marking: X

- The all-the-photons sum of the fringe and anti-fringe patterns reproduces the "mush" non-interference pattern of Figure 2.
- This is one of the simplest examples of a quantum eraser experiment.
  1. The insertion of the horizontal and vertical polarizers marks the photons with "which-slot" information that eliminates the interference pattern.
  2. The insertion of the, say, $+45°$ polarizer "erases" the which-slot information so an interference pattern reappears.

# A Quantum Eraser example of which-way marking: XI

- But there is a **mistaken interpretation** of the quantum eraser experiment that leads one to infer that there is retrocausality. Woo-woo. The incorrect reasoning is as follows:

  1. The insertion of the horizontal and vertical polarizers causes each photon to be reduced to either a horizontally polarized photon going through slit 1 or a vertically polarized photon going through slit 2.
  2. The insertion of the $+45°$ polarizer erases that which-slot information so interference reappears which means that the photon had to "go through both slits."

# A Quantum Eraser example of which-way marking: XII

3. Hence the delayed choice to insert or not insert the $+45^\circ$ polarizer–after the photons have traversed the screen–retrocauses the photons to either go through both slits or to only go through one slit or the other.

- Hence we see the importance of realizing that prior to inserting the $+45^\circ$ polarizer, the photons were in the superposition state $|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle$, not a half-half mixture of the reduced states $|S1\rangle \otimes |H\rangle$ or $|S2\rangle \otimes |V\rangle$.

# A Quantum Eraser example of which-way marking: XIII

- The **proof** that the system was not in that mixture is obtained by inserting the $+45°$ polarizer which yields the (fringe) interference pattern. If a photon had been, say, in the state $|S1\rangle \otimes |H\rangle$ then, with 50% probability, the photon would have passed through the filter in the state $|S1\rangle \otimes |+45°\rangle$, but that would not yield any interference pattern at the wall since their was no contribution from slit 2. And similarly if a photon in the state $|S2\rangle \otimes |V\rangle$ hits the $+45°$ polarizer.

# A Quantum Eraser example of which-way marking: XIV

- The fact that the insertion of the $+45°$ polarizer yielded interference proved that the incident photons were in a superposition state $|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle$ which, in turn, means there was no "going through one slit or the other" in case the $+45°$ polarizer had not been inserted.
- Thus a correct interpretation of the quantum eraser experiment removes any inference of retrocausality and fully accounts for the experimentally verified facts given in the figures. See the full treatment on my website:

  http://www.ellerman.org/a-common-fallacy/.

# Set version of reduced state description: I

- One way to better understand part of QM math using vector spaces is to see the set-analogue just using sets–prior to "lifting" it to vector spaces. The bridge from sets to vector spaces is the vector spaces over 2, $\mathbb{Z}_2^n$.
- Without here giving the whole "sets-to-vector-spaces" lifting program, we will just give enough to get a better understanding of the reduced mixtures.
- The set-analogue of a vector or *pure state* in a vector space is a *subset* of a set $U$ (see the $\mathbb{Z}_2^n$ bridge where a vector "is" just a subset). If the vector space is a Hilbert space, then the set-analogue $U$ has a probability distribution $\{\Pr(u) : u \in U\}$ over its elements.

# Set version of reduced state description: II

- The set-analogue of a *mixed-state* is just a set of *subsets with probabilities* assigned to them like $S_1, ..., S_n \subseteq U$ with corresponding probability distribution $\{\Pr(S_i)\}$.
- The set-analogue of the *tensor product* of two vector spaces is the *direct product* $U_A \times U_B$ of two sets (always finite dimension spaces and finite sets). If the vector spaces are Hilbert spaces, then we may assume a joint probability distribution on the product $\{\Pr(a, b) : a \in U_A, b \in U_B\}$.
- The set-analogue of a *separated state* in a tensor product is a *product subset* $S_A \times S_B \subseteq U_A \times U_B$ for some subsets $S_A \subseteq U_A$ and $S_B \subseteq U_B$. If a subset of order pairs from $U_A \times U_B$ cannot be expressed in this way, then it is "*entangled*."

# Set version of reduced state description: III

- Given a pure state $\rho$ on $H_A \otimes H_B$, there is the reduced mixture $\rho_A$ on $H_A$. For the set-analogue, given a "pure" subset $S \subseteq U_A \times U_B$ (which is a "trivial mixture with just one subset with probability 1"), then for any element $b \in H_B$ that appears in the ordered pair $(a,b) \in S$,
  - Define the subset $S_A^{(b)} = \{a \in U_A : (a,b) \in S\}$ and
  - Assign it the marginal probability of $b$ (suitably normalized), i.e., the probability $\Pr\left(S_A^{(b)}\right) = \sum\left\{\Pr(a,b) : a \in S_A^{(b)}\right\} / \sum_{(a,b)\in S} \Pr(a,b)$.
  - If the same subset of $U_A$ appears multiple times, they can be formally added by just adding the probabilties assigned to that subset so it only appears once. If $S_A^{(b)} = S_A^{(b')}$, then $\Pr\left(S_A^{(b,b')}\right) = \Pr\left(S_A^{(b)}\right) + \Pr\left(S_A^{(b')}\right)$ is assigned to that subset denoted $S_A^{(b,b')}$.

# Set version of reduced state description: IV

- This defines reduced mixture $S_A$ on $U_A$ which consists of the subsets $S_A^{(b,\ldots,b')}$ with the probabilities $\Pr\left(S_A^{(b,\ldots,b')}\right)$.

- Example 1: if $S = S_A \times S_B$ is a separated subset, then the reduced mixture on $U_A$ is in fact the subset $S_A$ considered as a trivial mixture with probability 1 assigned to it.
- Example 2: Nondegenerate measurement states
  - In the Hilbert space case, if $\rho$ comes from a perfectly correlated state $|\psi\rangle = \sum \alpha_i |a_i\rangle \otimes |b_i\rangle$ (where $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$ are orthonormal bases of $H_A$ and $H_B$), then $\rho_A$ is the reduced mixture of the states $|a_i\rangle$ with the probabilities $\alpha_i \alpha_i^* = \Pr(a_i)$.

# Set version of reduced state description: V

- In the set case, if $S$ is the graph of an injective function $f : U_A \to U_B$ given by $a_i \longmapsto b_i$ with the already-conditionalized probabilities $\Pr(a_i, b_i)$ assigned to the pairs in the graph, then the reduced mixture on $U_A$ is just the discrete partition $f^{-1} = \{\{a_i\}\}_{a_i \in U_A}$ with the probabilities $\Pr(a_i, b_i)$ assigned to the singleton subsets $\{a_i\}$.

- Example 3: In the general case of degenerate measurements, take $S$ as the graph of any function $f : U_A \to U_B$ and the reduced mixture is the partition $f^{-1}$ on $U_A$ with the probabilities assigned to the blocks:

$$\Pr\left(f^{-1}(b)\right) = \sum_{f(a)=b} \Pr(a,b) \; / \; \sum_{(a,b)\in graph(f)} \Pr(a,b).$$

# Taking mystery out of CLHS Eucharist

- Making distinctions and defining partitions. One way of making distinctions is joining a partition onto the given distinctions, like getting a binary question answered. Another way is mapping elements to other elements already distinct so those mapped to distinct elements are distinguished; those mapped to same element are in same block of inverse-image partition. Thus a given partition (in the codomain) induces a partition on the domain by the inverse-image operation.
- Measurement (degenerate or nondegenerate) is one example where mapping given by ordered pairs [basis elements $a \otimes b$ in the tensor product] and the already-distinguished states in the codomain are the indicator states of the measurement apparatus.

# Schmidt Decomposition

David Ellerman

UCR

January 2012

# Schmidt decomposition

- We have seen the special properties of the perfectly correlated marked superpositions $|\Psi\rangle = \sum_{i=1}^{n} \alpha_i |\varphi_i\rangle \otimes |\psi_i\rangle$ in a tensor product $H_A \otimes H_B$. The Schmidt decomposition shows that any pure state in $H_A \otimes H_B$ can be put into this form as:

$$|\Psi\rangle = \sum_i \lambda_i |\varphi_i\rangle \otimes |\psi_i\rangle.$$

- The *Schmidt coefficients* $\lambda_i$ are non-negative reals with $\sum_i \lambda_i^2 = 1$ and the states $\{|\varphi_i\rangle\}$ and $\{|\psi_i\rangle\}$ are orthonormal in their respective spaces.
- Then $|\Psi\rangle$ is a separated state iff only one Schmidt coefficient $\lambda_i = 1$ and the rest are 0; otherwise the state is entangled. The state is said to be *maximally entangled* if all the Schmidt coefficients are equal.

# Proof using singular value decomposition: I

- From linear algebra, we have that for any complex matrix $a = [a_{jk}]$, there are unitary matrices $u, v$ and a diagonal matrix $d$ of non-negative reals such that $a = udv$.
- Assuming $H_A$ and $H_B$ are of dimension $n$, a general pure state of $H_A \otimes H_B$ has the form $|\Psi\rangle = \sum_{jk} a_{jk} |j\rangle \otimes |k\rangle$ for some orthonormal bases $\{|j\rangle\}$ and $\{|k\rangle\}$ respectively.
- Then we can use the singular value decomposition of the $[a]$ where all the matrices are $n \times n$:

$$|\Psi\rangle = \sum_{ijk} u_{ji} d_{ii} v_{ik} |j\rangle \otimes |k\rangle.$$

- Then we can define: $|i_A\rangle = \sum_j u_{ji} |j\rangle$ and $|i_B\rangle = \sum_k v_{ik} |k\rangle$ and $\lambda_i = d_{ii}$ and then we have:

# Proof using singular value decomposition: II

$$|\Psi\rangle = \sum_i \lambda_i |i_A\rangle \otimes |i_B\rangle.$$

Schmidt decomposition

- Since the $|i_A\rangle$ result from a unitary transformation of the orthonormal basis $\{|j\rangle\}$ and the $|i_B\rangle$ similarly result from a unitary transformation of $\{|k\rangle\}$, they are also orthonormal bases.

# Alternative proof without assuming SVD: I

- We again start with a pure state $|\Psi\rangle$ in $H_A \otimes H_B$ and then we take the reduced density operator $\rho^A = \mathrm{tr}_B\left(|\Psi\rangle\langle\Psi|\right)$.
- As a positive semidefinite operator on $H_A$, we can express it in terms of its eigenvector projections with non-negative real coefficients: $\rho^A = \sum_{i=1}^{n} \lambda_i^2 |\varphi_i\rangle\langle\varphi_i|$ so $\{|\varphi_i\rangle\}$ is an orthonormal basis for $H_A$.
- Take any orthonormal basis $\left\{\left|\psi_j'\right\rangle\right\}_{j=1}^{m}$ for $H_B$ and then expand the original state $|\Psi\rangle$ in terms of the basis $\varphi_i \otimes \psi_j'$:

$$|\Psi\rangle = \sum_{i,j} \left\langle \Psi | \varphi_i \otimes \psi_j' \right\rangle \left| \varphi_i \otimes \psi_j' \right\rangle.$$

# Alternative proof without assuming SVD: II

- Taking the summation over the $\left|\psi'_j\right\rangle$ with the $\left\langle \Psi|\varphi_i \otimes \psi'_j\right\rangle$ coefficients, we have the vectors in $H_B$, $\left|\psi''_i\right\rangle = \sum_j \left\langle \Psi|\varphi_i \otimes \psi'_j\right\rangle \left|\psi'_j\right\rangle$ with the property that:

$$|\Psi\rangle = \sum_{i=1}^{n} |\varphi_i\rangle \otimes \left|\psi''_i\right\rangle.$$

- But the $\left|\psi''_i\right\rangle$ may not be normalized, so using the defining characteristic of the reduced density operator $\rho^A$: for any operator $T$ on $H_A$,

$$\sum_i \lambda_i^2 \langle \varphi_i|T\varphi_i\rangle = \langle T\rangle = \mathrm{tr}\left(\rho^A T\right) = \langle \Psi|\left(T \otimes I_B\right)\Psi\rangle = \sum_{i,k} \langle \varphi_i|T\varphi_k\rangle \left\langle \psi''_i|\psi''_k\right\rangle.$$

# Alternative proof without assuming SVD: III

- But since this holds for *any* operator $T$, the equation must hold term by term so that:

$$\langle \psi_i'' | \psi_k'' \rangle = \lambda_i^2 \delta_{ik}.$$

- Thus for $\lambda_i > 0$, define $|\psi_i\rangle = \frac{1}{\lambda_i} |\psi_i''\rangle$ so the $\{|\psi_i\rangle\}$ are both normalized and orthogonal. Then we have:

$$|\Psi\rangle = \sum_i \lambda_i |\varphi_i\rangle \otimes |\psi_i\rangle$$

Schmidt decomposition.

- We have seen the progression: $\rho \xrightarrow{CLHS} \varrho \xrightarrow{red.} \varrho_1$ starting with a pure $\rho$. The Schmidt decomposition allows us to start with any mixed state $\rho^A$ on $H_A$ and then to define a pure state $\varrho$ on $H^A \otimes H^A$ so that the reduced density matrix on the first component is $\varrho_1 = \rho^A$.
- Given any mixed state $\rho^A$ on $H_A$, we, as above, can express it as: $\rho^A = \sum_i \lambda_i^2 |\varphi_i\rangle \langle\varphi_i|$ for non-negative reals $\lambda_i$ with $\sum_i \lambda_i^2 = 1$ and orthonormal $\{|\varphi_i\rangle\}$.
- Then $|\Psi\rangle = \sum_i \lambda_i |\varphi_i\rangle \otimes |\varphi_i\rangle$ is a pure state on $H_A \otimes H_A$ called its *purification* so that for $\varrho = |\Psi\rangle \langle\Psi|$, $\varrho_1 = \rho^A$.
- Thus we always have:

$$\rho^A \xrightarrow{CLHS} \varrho \xrightarrow{red.} \varrho_1 = \rho^A.$$

- Consider the example on $\mathbb{C}^2 \otimes \mathbb{C}^2$:

$$|\Psi\rangle = \frac{1}{\sqrt{3}}\left[|0_A\rangle \otimes |0_B\rangle + |0_A\rangle \otimes |1_B\rangle + |1_A\rangle \otimes |0_B\rangle\right].$$

- Thus the $2 \times 2$ matrix is: $a = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & 0 \end{bmatrix}$.
- Using a computational program, the SVD is: $a = udv =$

$$\begin{bmatrix} \sqrt{\frac{2}{5-\sqrt{5}}} & -\sqrt{\frac{2}{5+\sqrt{5}}} \\ \sqrt{\frac{2}{5+\sqrt{5}}} & \sqrt{\frac{2}{5-\sqrt{5}}} \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{6}\sqrt{5}+\frac{1}{2}} & 0 \\ 0 & \sqrt{\frac{1}{2}-\frac{1}{6}\sqrt{5}} \end{bmatrix}$$
$$\begin{bmatrix} \sqrt{\frac{2}{5-\sqrt{5}}} & \sqrt{\frac{2}{5+\sqrt{5}}} \\ \sqrt{\frac{2}{5+\sqrt{5}}} & -\sqrt{\frac{2}{5-\sqrt{5}}} \end{bmatrix}.$$

# Example of Schmidt decomposition: II

- In the $|0_A\rangle$,$|1_A\rangle$, the two Schmidt basis vectors for the first component $\mathbb{C}^2$ are the two columns of $u$.
- In the $|0_B\rangle$ , $|1_B\rangle$ basis, the two Schmidt basis vectors for the second component $\mathbb{C}^2$ are two rows of $v$ transposed as columns.
- Hence the Schmidt decomposition is:

$$|\Psi\rangle = \sqrt{\tfrac{1}{6}\sqrt{5}+\tfrac{1}{2}}\begin{bmatrix}\sqrt{\frac{2}{5-\sqrt{5}}}\\ \sqrt{\frac{2}{5+\sqrt{5}}}\end{bmatrix}\otimes\begin{bmatrix}\sqrt{\frac{2}{5-\sqrt{5}}}\\ \sqrt{\frac{2}{5+\sqrt{5}}}\end{bmatrix}$$
$$+\sqrt{\tfrac{1}{2}-\tfrac{1}{6}\sqrt{5}}\begin{bmatrix}-\sqrt{\frac{2}{5+\sqrt{5}}}\\ \sqrt{\frac{2}{5-\sqrt{5}}}\end{bmatrix}\otimes\begin{bmatrix}\sqrt{\frac{2}{5+\sqrt{5}}}\\ -\sqrt{\frac{2}{5-\sqrt{5}}}\end{bmatrix}.$$

# Example of Schmidt decomposition: III

- To check it, let's compute the coefficient of $|0_A\rangle \otimes |0_B\rangle$:

$$\sqrt{\tfrac{1}{6}\sqrt{5}+\tfrac{1}{2}}\left(\frac{2}{5-\sqrt{5}}\right) - \sqrt{\tfrac{1}{2}-\tfrac{1}{6}\sqrt{5}}\left(\frac{2}{5+\sqrt{5}}\right) = \frac{1}{\sqrt{3}}.\checkmark$$

- Coefficient of $|0_A\rangle \otimes |1_B\rangle$:

$$\sqrt{\tfrac{1}{6}\sqrt{5}+\tfrac{1}{2}}\left(\sqrt{\frac{2}{5-\sqrt{5}}}\right)\left(\sqrt{\frac{2}{5+\sqrt{5}}}\right) +$$
$$\sqrt{\tfrac{1}{2}-\tfrac{1}{6}\sqrt{5}}\left(-\sqrt{\frac{2}{5+\sqrt{5}}}\right)\left(-\sqrt{\frac{2}{5-\sqrt{5}}}\right) = \frac{1}{\sqrt{3}}.\checkmark$$

- Coefficient of $|1_A\rangle \otimes |0_B\rangle$:

# Example of Schmidt decomposition: IV

$$\sqrt{\tfrac{1}{6}\sqrt{5}+\tfrac{1}{2}}\left(\sqrt{\tfrac{2}{5+\sqrt{5}}}\right)\left(\sqrt{\tfrac{2}{5-\sqrt{5}}}\right)+$$
$$\sqrt{\tfrac{1}{2}-\tfrac{1}{6}\sqrt{5}}\left(\sqrt{\tfrac{2}{5-\sqrt{5}}}\right)\left(\sqrt{\tfrac{2}{5+\sqrt{5}}}\right)=\frac{1}{\sqrt{3}}.\checkmark$$

- Coefficient of $|1_A\rangle \otimes |1_B\rangle$:

$$\sqrt{\tfrac{1}{6}\sqrt{5}+\tfrac{1}{2}}\left(\tfrac{2}{5+\sqrt{5}}\right)-\sqrt{\tfrac{1}{2}-\tfrac{1}{6}\sqrt{5}}\left(\tfrac{2}{5-\sqrt{5}}\right)=0.\checkmark$$

# Purification example: I

- We start with a mixed state of $\mathbb{C}^2$ which is:

$$\tfrac{1}{3} \text{ of } |\psi_1\rangle = \tfrac{1}{\sqrt{2}}\left(|0_A\rangle + |1_A\rangle\right) \text{ and } \tfrac{2}{3} \text{ of } |\psi_2\rangle = |0_A\rangle.$$

- Hence its density matrix in the usual coordinates is:

$$\begin{aligned}
\rho^A &= \tfrac{1}{3}|\psi_1\rangle\langle\psi_1| + \tfrac{2}{3}|\psi_2\rangle\langle\psi_2| \\
&= \tfrac{1}{3}\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix} + \tfrac{2}{3}\begin{bmatrix}1\\0\end{bmatrix}\begin{bmatrix}1 & 0\end{bmatrix} \\
&= \tfrac{1}{3}\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix} + \tfrac{2}{3}\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} = \begin{bmatrix}\frac{5}{6} & \frac{1}{6}\\ \frac{1}{6} & \frac{1}{6}\end{bmatrix}.
\end{aligned}$$

- The orthonormal eigenvectors and their eigenvalues are:

# Purification example: II

$$|\varphi_1\rangle = \frac{1}{\sqrt{10-4\sqrt{5}}}\begin{bmatrix} 2-\sqrt{5} \\ 1 \end{bmatrix} \text{ with } \lambda_1 = \frac{3-\sqrt{5}}{6}$$
$$|\varphi_2\rangle = \frac{1}{\sqrt{10+4\sqrt{5}}}\begin{bmatrix} 2+\sqrt{5} \\ 1 \end{bmatrix} \text{ with } \lambda_2 = \frac{3+\sqrt{5}}{6}.$$

- These give the orthonormal decomposition of the density matrix since:

$$\begin{aligned}
&\lambda_1 |\varphi_1\rangle \langle\varphi_1| + \lambda_2 |\varphi_2\rangle \langle\varphi_2| \\
&= \frac{3-\sqrt{5}}{6}\frac{1}{10-4\sqrt{5}}\begin{bmatrix} 2-\sqrt{5} \\ 1 \end{bmatrix}\begin{bmatrix} 2-\sqrt{5} & 1 \end{bmatrix} \\
&+ \frac{3+\sqrt{5}}{6}\frac{1}{10+4\sqrt{5}}\begin{bmatrix} 2+\sqrt{5} \\ 1 \end{bmatrix}\begin{bmatrix} 2+\sqrt{5} & 1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{5}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} \end{bmatrix} = \rho^A. \checkmark
\end{aligned}$$

# Purification example: III

- The purification is then the pure state of $\mathbb{C}^2 \otimes \mathbb{C}^2$:

$$|\Psi\rangle = \sqrt{\lambda_1}\,|\varphi_1\rangle \otimes |\varphi_1\rangle + \sqrt{\lambda_2}\,|\varphi_2\rangle \otimes |\varphi_2\rangle.$$

- The density matrix is $\varrho = |\Psi\rangle\langle\Psi|$ and the reduced density matrix over the first component is:

$$\varrho_1 = \lambda_1\,|\varphi_1\rangle\langle\varphi_1| + \lambda_2\,|\varphi_2\rangle\langle\varphi_2| = \rho^A.\checkmark$$

# Two-Slit Quantum Eraser Example

David Ellerman

UCR

January 2012

# Quantum eraser example before markings: I

- Consider the setup of the two-slit experiment where the superposition state, $\frac{1}{\sqrt{2}}(|S1\rangle + |S2\rangle)$, evolves to show interference on the wall.
- If we put a $+45^\circ$ polarizer in front of the slits to control the incoming polarization, then we can represent the system after the polarizer as a tensor product with the second component giving the polarization state. The evolving state after the two slits is the superposition:

$$\frac{1}{\sqrt{2}}(|S1\rangle \otimes |45^\circ\rangle + |S2\rangle \otimes |45^\circ\rangle).$$

# Quantum eraser example before markings: II

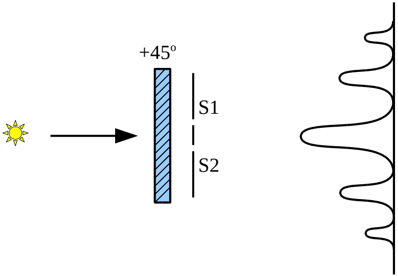



Figure 1: Interference pattern from two-slits

# Simultaneous insertion of H,V polarizers: I

- Then horizontal and vertical polarizers are simultaneous inserted behind the $S1$ and $S2$ slits respectively.
- This will change the evolving state to: $\frac{1}{\sqrt{2}}\left(|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle\right)$ but since these new polarizers involve some measurements, not just unitary evolution, it may be helpful to go through the calculation in some detail.
- The state that "hits" the $H, V$ polarizers is:

$$\frac{1}{\sqrt{2}}\left(|S1\rangle \otimes |45^\circ\rangle + |S2\rangle \otimes |45^\circ\rangle\right).$$

- The $45^\circ$ polarization state can be resolved by inserting the identity operator $I = |H\rangle\langle H| + |V\rangle\langle V|$ to get:

# Simultaneous insertion of H,V polarizers: II

$$\left|45^\circ\right\rangle = \left[\left|H\right\rangle\left\langle H\right| + \left|V\right\rangle\left\langle V\right|\right]\left|45^\circ\right\rangle = \left\langle H|45^\circ\right\rangle\left|H\right\rangle + \left\langle V|45^\circ\right\rangle\left|V\right\rangle = \frac{1}{\sqrt{2}}\left[\left|H\right\rangle + \left|V\right\rangle\right].$$

- Substituting this for $\left|45^\circ\right\rangle$, we have the state that hits the $H, V$ polarizers as:

$$\frac{1}{\sqrt{2}}\left(\left|S1\right\rangle \otimes \left|45^\circ\right\rangle + \left|S2\right\rangle \otimes \left|45^\circ\right\rangle\right)$$
$$= \frac{1}{\sqrt{2}}\left(\left|S1\right\rangle \otimes \frac{1}{\sqrt{2}}\left[\left|H\right\rangle + \left|V\right\rangle\right] + \left|S2\right\rangle \otimes \frac{1}{\sqrt{2}}\left[\left|H\right\rangle + \left|V\right\rangle\right]\right)$$
$$= \frac{1}{2}\left[\left|S1\right\rangle \otimes \left|H\right\rangle + \left|S1\right\rangle \otimes \left|V\right\rangle + \left|S2\right\rangle \otimes \left|H\right\rangle + \left|S2\right\rangle \otimes \left|V\right\rangle\right]$$

which can be regrouped in two parts as:

$$= \frac{1}{2}\left[\left|S1\right\rangle \otimes \left|H\right\rangle + \left|S2\right\rangle \otimes \left|V\right\rangle\right] + \frac{1}{2}\left[\left|S1\right\rangle \otimes \left|V\right\rangle + \left|S2\right\rangle \otimes \left|H\right\rangle\right].$$

# Simultaneous insertion of H,V polarizers: III

- Then the $H, V$ polarizers are making a degenerate measurement that give the first state $|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle$ with probability $\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2}$.
- The other state $|S1\rangle \otimes |V\rangle + |S2\rangle \otimes |H\rangle$ is obtained with the same probability, and it is blocked by the polarizers.
- Thus with probability $\frac{1}{2}$, the state that evolves is the state (after being normalized):

$$\frac{1}{\sqrt{2}}\left[|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle\right].$$

- Logically, we should get the same result if we insert the $H$ and $V$ polarizers sequentially.

# Sequential insertion of H,V polarizers: I

- Suppose that we imposed the $H$ and $V$ polarizers one at a time in a sequence. We start by just putting the $H$ polarizers after slit 1. We have the same state evolving after the two slits but a different grouping for the degenerate measurement.

$$\frac{1}{2}\left[|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle\right] + \frac{1}{2}\left[|S1\rangle \otimes |V\rangle\right].$$

- Then with probability $\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{3}{4}$ the measurement yields the result $|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle$ and with probability $\frac{1}{4}$ we get $|S1\rangle \otimes |V\rangle$. Since the latter state is blocked by the $H$ filter at $S1$, the normalized state that continues is:

# Sequential insertion of H,V polarizers: II

$$\frac{1}{\sqrt{3}}\left[|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle\right].$$

- Then we insert the $V$ polarizer so that it only effects the $S2$ portion and do another degenerate measurement with the grouping:

$$\frac{1}{\sqrt{3}}\left[|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle\right] + \frac{1}{\sqrt{3}}\left[|S2\rangle \otimes |H\rangle\right].$$

- With probability $\left(\frac{1}{\sqrt{3}}\right)^2 + \left(\frac{1}{\sqrt{3}}\right)^2 = \frac{2}{3}$ we get $|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle$ and with probability $\left(\frac{1}{\sqrt{3}}\right)^2 = \frac{1}{3}$ we get $|S2\rangle \otimes |H\rangle$ which is the blocked state.
- Hence with probability $\frac{2}{3}$ we get, after the second polarizer, the previous normalized state:

# Sequential insertion of H,V polarizers: III

$$\frac{1}{\sqrt{2}}\left[|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle\right].$$

- Combining the probabilities from the sequential $H$ and $V$ polarizers, we get the above state with the probability: $\frac{3}{4} \times \frac{2}{3} = \frac{1}{2}$ exactly as when the $H, V$ polarizers are inserted simultaneously rather than sequentially.

# Interference removed by H,V polarizer markings: I

- If $P_{\Delta y}$ is the projection operator representing finding a particle in the region $\Delta y$ along the wall, then that probability in the state $\frac{1}{\sqrt{2}}\left[|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle\right]$ is:

$$\frac{1}{2}\left\langle S1 \otimes H + S2 \otimes V \middle| P_{\Delta y} \otimes I \middle| S1 \otimes H + S2 \otimes V \right\rangle$$
$$= \frac{1}{2}\left\langle S1 \otimes H + S2 \otimes V \middle| P_{\Delta y} S1 \otimes H + P_{\Delta y} S2 \otimes V \right\rangle$$
$$= \frac{1}{2}\left[\left\langle S1 \otimes H \middle| P_{\Delta y} S1 \otimes H \right\rangle + \left\langle S1 \otimes H \middle| P_{\Delta y} S2 \otimes V \right\rangle\right.$$
$$\left. + \left\langle S2 \otimes V \middle| P_{\Delta y} S1 \otimes H \right\rangle + \left\langle S2 \otimes V \middle| P_{\Delta y} S2 \otimes V \right\rangle\right]$$
$$= \frac{1}{2}\left[\left\langle S1 \middle| P_{\Delta y} S1 \right\rangle \langle H|H\rangle + \left\langle S1 \middle| P_{\Delta y} S2 \right\rangle \langle H|V\rangle\right.$$
$$\left. + \left\langle S2 \middle| P_{\Delta y} S1 \right\rangle \langle V|H\rangle + \left\langle S2 \middle| P_{\Delta y} S2 \right\rangle \langle V|V\rangle\right]$$
$$= \frac{1}{2}\left[\left\langle S1 \middle| P_{\Delta y} S1 \right\rangle + \left\langle S2 \middle| P_{\Delta y} S2 \right\rangle\right]$$

= average of separate slot probabilities.

# Interference removed by H,V polarizer markings: II

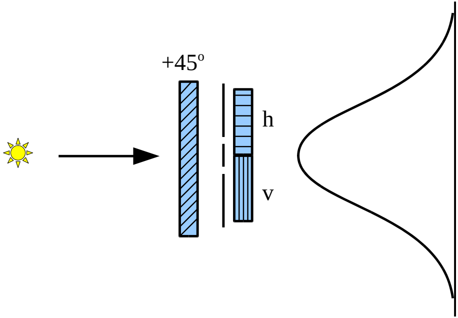



Figure 2: Mush pattern with interference eliminated by which-way markings

# Interference removed by H,V polarizer markings: III

- The key step is how the orthogonal polarization markings decohered the state since $\langle H|V\rangle = 0 = \langle V|H\rangle$ and thus eliminated the interference between the $S1$ and $S2$ terms.
- The state-reduction occurs only when the evolved superposition state hits the far wall which measures the positional component (i.e., $P_{\Delta y}$) of the composite state and shows the non-interference pattern.

# "Erasing" the markings: I

- The key point is that in spite of the bad terminology of "which-way" or "which-slit" information, the polarization markings do NOT create a half-half mixture of horizontally polarized photons going through slit 1 and vertically polarized photons going through slit 2. It creates the (incoherent) *superposition* state $\frac{1}{\sqrt{2}}\left[|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle\right]$.
- This can be verified by inserting a $+45°$ polarizer between the two-slit screen and the far wall.

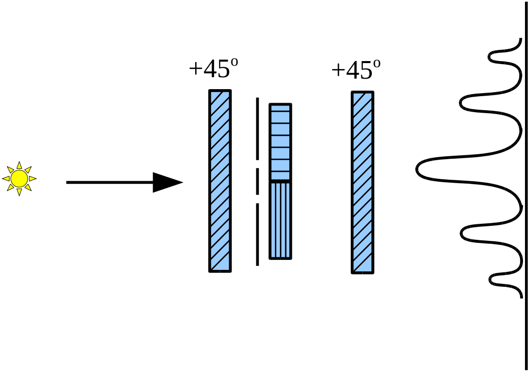



Figure 3: Fringe interference pattern produced by $+45°$ polarizer

# "Erasing" the markings: III

- Each of the horizontal and vertical polarization states can be represented as a superposition of $+45^\circ$ and $-45^\circ$ polarization states. Just as the horizontal polarizer in front of slit 1 threw out the vertical component so we have no $|S1\rangle \otimes |V\rangle$ term in the superposition, so now the $+45^\circ$ polarizer throws out the $-45^\circ$ component of each of the $|H\rangle$ and $|V\rangle$ terms so the state transformation is:

$$\frac{1}{\sqrt{2}}\left[|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle\right]$$
$$\rightarrow \frac{1}{\sqrt{2}}\left[|S1\rangle \otimes |{+45^\circ}\rangle + |S2\rangle \otimes |{+45^\circ}\rangle\right] =$$
$$\frac{1}{\sqrt{2}}\left(|S1\rangle + |S2\rangle\right) \otimes |{+45^\circ}\rangle.$$

- It might be useful to again go through the calculation in some detail.

# "Erasing" the markings: IV

1. $|H\rangle = (|+45^\circ\rangle\langle+45^\circ| + |-45^\circ\rangle\langle-45^\circ|)\,|H\rangle = \langle+45^\circ|H\rangle\,|+45^\circ\rangle + \langle-45^\circ|H\rangle\,|-45^\circ\rangle$ and since a horizontal vector at $0^\circ$ is the sum of the $+45^\circ$ vector and the $-45^\circ$ vector, $\langle+45^\circ|H\rangle = \langle-45^\circ|H\rangle = \frac{1}{\sqrt{2}}$ so that: $|H\rangle = \frac{1}{\sqrt{2}}\left[|+45^\circ\rangle + |-45^\circ\rangle\right]$.
2. $|V\rangle = (|+45^\circ\rangle\langle+45^\circ| + |-45^\circ\rangle\langle-45^\circ|)\,|V\rangle = \langle+45^\circ|V\rangle\,|+45^\circ\rangle + \langle-45^\circ|V\rangle\,|-45^\circ\rangle$ and since a vertical vector at $90^\circ$ is the sum of the $+45^\circ$ vector and the negative of the $-45^\circ$ vector, $\langle+45^\circ|V\rangle = \frac{1}{\sqrt{2}}$ and $\langle-45^\circ|V\rangle = -\frac{1}{\sqrt{2}}$ so that: $|V\rangle = \frac{1}{\sqrt{2}}\left[|+45^\circ\rangle - |-45^\circ\rangle\right]$.

- Hence making the substitutions gives:

$$\frac{1}{\sqrt{2}}\left[|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle\right]$$

$$= \frac{1}{\sqrt{2}}\left[\begin{array}{c} |S1\rangle \otimes \frac{1}{\sqrt{2}}\left[|+45^\circ\rangle + |-45^\circ\rangle\right] \\ +|S2\rangle \otimes \frac{1}{\sqrt{2}}\left[|+45^\circ\rangle - |-45^\circ\rangle\right] \end{array}\right].$$

# "Erasing" the markings: V

- We then regroup the terms according to the measurement being made by the 45° polarizer:

$$= \frac{1}{\sqrt{2}} \left[ \begin{array}{c} \frac{1}{\sqrt{2}} \left[ |S1\rangle \otimes |+45^\circ\rangle + |S2\rangle \otimes |+45^\circ\rangle \right] \\ + \frac{1}{\sqrt{2}} \left[ |S1\rangle \otimes |+45^\circ\rangle - |S2\rangle \otimes |-45^\circ\rangle \right] \end{array} \right]$$

$$= \frac{1}{2} \left( |S1\rangle + |S2\rangle \right) \otimes |+45^\circ\rangle + \frac{1}{2} \left( |S1\rangle - |S2\rangle \right) \otimes |-45^\circ\rangle .$$

- Then with probability $\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2}$, the $+45^\circ$ polarization measure passes the state $\left( |S1\rangle + |S2\rangle \right) \otimes |+45^\circ\rangle$ and blocks the state $\left( |S1\rangle - |S2\rangle \right) \otimes |-45^\circ\rangle$. Hence the normalized state that evolves is: $\frac{1}{\sqrt{2}} \left( |S1\rangle + |S2\rangle \right) \otimes |+45^\circ\rangle$, as indicated above.

# "Erasing" the markings: VI

- Then at the wall, the positional measurement $P_{\Delta y}$ of the first component is the evolved superposition $|S1\rangle + |S2\rangle$ which again shows an interference pattern. But it is not the same as the original interference pattern before $H, V$ or $+45°$ polarizers were inserted. This "shifted" interference pattern is called the *fringe* pattern of figure 3.
- Alternatively we could insert a $-45°$ polarizer which would transform the state $\frac{1}{\sqrt{2}}\left[|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle\right]$ into $\frac{1}{\sqrt{2}}\left(|S1\rangle + |S2\rangle\right) \otimes |-45°\rangle$ which produces the interference pattern from the "other half" of the photons and which is called the *anti-fringe* pattern.
- The all-the-photons sum of the fringe and anti-fringe patterns reproduces the "mush" non-interference pattern of Figure 2.

# "Erasing" the markings: VII

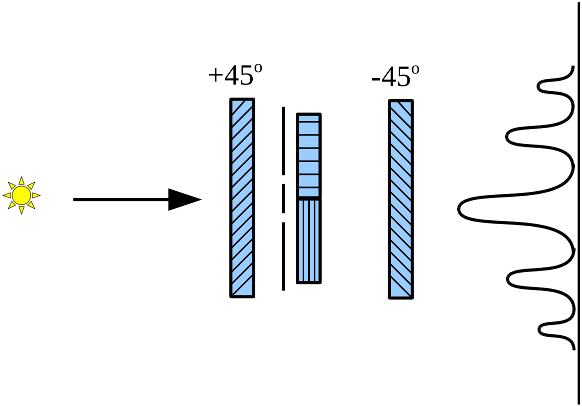



Figure 4: Anti-fringe interference pattern produced by $-45°$ polarizer

# Interpreting the Quantum Eraser: I

- This is one of the simplest examples of a quantum eraser experiment.
  1. The insertion of the horizontal and vertical polarizers marks the photons with "which-slot" information that eliminates the interference pattern.
  2. The insertion of a $+45°$ or $-45°$ polarizer "erases" the which-slot information so an interference pattern reappears.
- But there is a **mistaken interpretation** of the quantum eraser experiment that leads one to infer that there is *retrocausality*. *Woo-woo*. The incorrect reasoning is as follows:

# Interpreting the Quantum Eraser: II

1. The markings by insertion of the horizontal and vertical polarizers creates the half-half *mixture* where each photon is reduced to either a horizontally polarized photon going through slit 1 or a vertically polarized photon going through slit 2. Hence the photon "goes through one slit or the other." [Fail!]
2. The insertion of the $+45°$ polarizer erases that which-slot information so interference reappears which means that the photon had to "go through both slits."
3. Hence the delayed choice to insert or not insert the $+45°$ polarizer–after the photons have traversed the screen and $H, V$ polarizers–*retrocauses* the photons to either:
    - go through both slits, or
    - to only go through one slit or the other.

# Interpreting the Quantum Eraser: III

- Now we can see the importance of realizing that prior to inserting the $+45°$ polarizer, the photons were in the *superposition* state $\frac{1}{\sqrt{2}}\left[|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle\right]$, not a half-half mixture of the reduced states $|S1\rangle \otimes |H\rangle$ or $|S2\rangle \otimes |V\rangle$.
- The **proof** that the system was not in that mixture is obtained by inserting the $+45°$ polarizer which yields the (fringe) interference pattern.
  1. If a photon had been, say, in the state $|S1\rangle \otimes |H\rangle$ then, with 50% probability, the photon would have passed through the filter in the state $|S1\rangle \otimes |+45°\rangle$, but that would not yield any interference pattern at the wall since their was no contribution from slit 2.

# Interpreting the Quantum Eraser: IV

2. And similarly if a photon in the state $|S2\rangle \otimes |V\rangle$ hits the $+45°$ polarizer.

- The fact that the insertion of the $+45°$ polarizer yielded interference proved that the incident photons were in a superposition state $\frac{1}{\sqrt{2}}[|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle]$ which, in turn, means there was no "going through one slit or the other" in case the $+45°$ polarizer had not been inserted.
- Thus a correct interpretation of the quantum eraser experiment removes any inference of *retrocausality* and fully accounts for the experimentally verified facts given in the figures. See the treatment on my mathblog:

http://www.mathblog.ellerman.org/2011/11/a-common-qm-fallacy/.

# The Scully Maser Eraser

## A Non-Retrocausal Analysis

David Ellerman

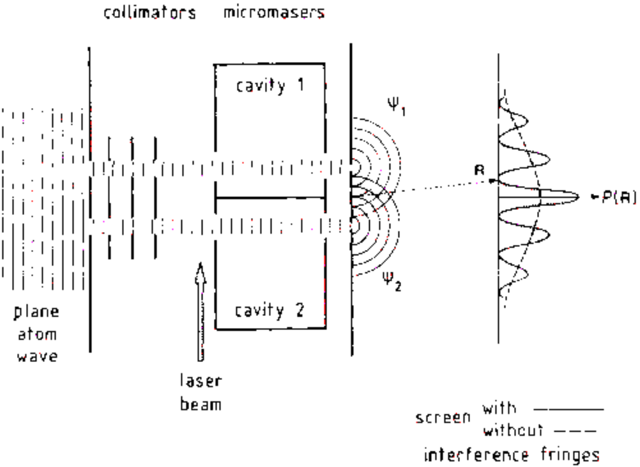
UCR

February 2012

# Scully's Maser Quantum Eraser: I

- The Scully maser eraser [Scully et al. 1991. Quantum optical tests of complementarity. *Nature*. 351 (May 9, 1991)] follows the same logic as the simpler eraser based on polarizers *except* that it allows the eraser to be applied after the hit at the wall–which *appears* to make a stronger case for retrocausality. The Walborn et al. quantum eraser model is, as they explicitly state, an optical realization of Scully's suggested model using masers (Walborn et al. 2002. Double-slit quantum eraser. *Physical Review A*. 65 (3)).
- We will present the formulas for maser case parallel to the previous formulas for the simple polarizer model, while allowing for the retro-application of the eraser.

# Scully's Maser Quantum Eraser: II




Scully Figure 3

# Scully's Maser Quantum Eraser: III

- Start with Figure 3 but first suppose the laser and maser are not there so we only have a two-slit superposition state $\Psi(r) = \frac{1}{\sqrt{2}}[\psi_1(r) + \psi_2(r)]$, Scully's formula (2) which is like the formula $\frac{1}{\sqrt{2}}[|S1\rangle + |S2\rangle]$.
- The probability of a particle falling at $R$ is given by Scully's formula (3):

$$P(R) = \frac{1}{2}\left[|\psi_1(R)|^2 + |\psi_2(R)|^2 + \psi_1(R)^* \psi_2(R) + \psi_2(R)^* \psi_1(R)\right]$$

which is his version of the polarizer-model formula:

$$\frac{1}{2}[\langle S1|P_{\Delta y}S1\rangle + \langle S2|P_{\Delta y}S2\rangle + \langle S1|P_{\Delta y}S2\rangle + \langle S2|P_{\Delta y}S1\rangle].$$

# Scully's Maser Quantum Eraser: IV

- Then in Figure 3, Scully introduces the laser (but not the maser cavities) which excites the incoming atoms so he moves up to the tensor product of positional space ("centre-of-mass coordinates") and the internal state of the atom. Then his formula (4) is (where I have taken the freedom to always insert the tensor product symbol $\otimes$):

$$\Psi(r) = \frac{1}{\sqrt{2}}\left[\psi_1(r) + \psi_2(r)\right] \otimes |i\rangle$$

which is analogous to the polarizer-model formula after introducing the first 45° polarizer:

$$\frac{1}{\sqrt{2}}\left[|S1\rangle + |S2\rangle\right] \otimes |45^\circ\rangle.$$

- This again yields an interference pattern as indicated by Scully's formula (5):

$$P(R) = \frac{1}{2}\left[|\psi_1|^2 + |\psi_2|^2 + \psi_1{}^*\psi_2 + \psi_2{}^*\psi_1\right]\langle i|i\rangle$$

where the analogous formula would be:

$$\frac{1}{2}[\langle S1|P_{\Delta y}S1\rangle + \langle S2|P_{\Delta y}S2\rangle + \langle S1|P_{\Delta y}S2\rangle + \langle S2|P_{\Delta y}S1\rangle]\langle 45^\circ|45^\circ\rangle.$$

# Adding the which-way markers: I

- Scully then puts in two maser cavities in front of the two slits as in his Figure 3 analogous to putting the $H, V$ polarizers just after the slits (they could have been in front). The top cavity is 1 and the bottom cavity is 2. The idea is that as an atom passes through a cavity it may emit a photon so $|1_1 0_2\rangle$ represents the state of 1 photon in cavity 1 and 0 photons in cavity 2, and similarly for $|0_1 1_2\rangle$ representing a photon in the second cavity. This again expands the Hilbert space with the two photon-in-cavity states which serve as the which-way markings. The atom's state is excited to $|b\rangle$. Hence after a particle passes through the masers and slits, its superposition state is Scully's formula (6):

# Adding the which-way markers: II

$$\Psi(r) = \frac{1}{\sqrt{2}}\left[\psi_1(r) \otimes |1_1 0_2\rangle + \psi_2(r) \otimes |0_1 1_2\rangle\right] \otimes |b\rangle$$

which, aside from the extra $|b\rangle$ state, is Scully's version of polarization-model's post-markings formula:

$$\frac{1}{\sqrt{2}}\left(|S1\rangle \otimes |H\rangle + |S2\rangle \otimes |V\rangle\right).$$

- That markings-step was key in both arguments and it is the source of some confusion that later leads to inferences of retrocausality.
- The common-sense assumption is that the atom has to go through one maser chamber or the other so there is a tell-tale photon in one or the other and we just don't know which. This mental imagery of the tell-tale photon being, premeasurement, in one cavity or the other is wrong.

# Adding the which-way markers: III

- Similarly it might be thought that marking the slits with the different polarizers caused each photon to be either horizontally polarized or vertically polarized.
- I n either case, the system would then be in a $50, 50$ mixture.
- But it is NOT a mixture. In both cases, the system is in a *superposition* state, e.g., Scully's formula (6).
- In Scully's case, there is an entangled state where "slit 1 & photon in cavity 1" is *superposed* with "slit 2 & photon in cavity 2".
- In the polarization model, there is the entangled state where "slit 1 & H" is *superposed* with "slit 2 & V".
- Then in either case, we redo the probability calculations and we find in both cases that the which-way markings suffice to eliminate the cross-terms and the interference.

# Adding the which-way markers: IV

- In Scully's model, the probability formula (7) is:

$$P(R) = \frac{1}{2}\left[|\psi_1|^2 + |\psi_2|^2 + \psi_1^*\psi_2 \langle 1_1 0_2 | 0_1 1_2 \rangle + \psi_2^*\psi_1 \langle 0_1 1_2 | 1_1 0_2 \rangle\right] \langle b | b \rangle$$

where since $\langle 1_1 0_2 | 0_1 1_2 \rangle = 0 = \langle 0_1 1_2 | 1_1 0_2 \rangle$, the interference terms drop out so we get the no-fringes formula (8):

$$P(R) = \frac{1}{2}\left[|\psi_1|^2 + |\psi_2|^2\right].$$

- In the polarization model, the probability formula is:

$$\frac{1}{2}[\langle S1 | P_{\Delta y} S1 \rangle \langle H | H \rangle + \langle S2 | P_{\Delta y} S2 \rangle \langle V | V \rangle + \langle S1 | P_{\Delta y} S2 \rangle \langle H | V \rangle + \langle S2 | P_{\Delta y} S1 \rangle \langle V | H \rangle]$$

# Adding the which-way markers: V

where since $\langle H|V\rangle = 0 = \langle V|H\rangle$, the interference terms drop out so we get the no-fringes formula:

$$\frac{1}{2}\left[\left\langle S1|P_{\Delta y}S1\right\rangle + \left\langle S2|P_{\Delta y}S2\right\rangle\right].$$

# Introducing the quantum eraser: I

- Scully then introduces the "eraser" element that is analogous to introducing either a $+45°$ or $-45°$ polarizer before the wall. But the details are quite different.
- In the positional coordinates, Scully introduces a change of basis to:
  - Symmetric state: $\psi_{+}(r) = \frac{1}{\sqrt{2}}[\psi_1(r) + \psi_2(r)]$ which is analogous to $\frac{1}{\sqrt{2}}[|S1\rangle + |S2\rangle]$ in the polarization model, and
  - Antisymmetric state: $\psi_{-}(r) = \frac{1}{\sqrt{2}}[\psi_1(r) - \psi_2(r)]$ which is analogous to $\frac{1}{\sqrt{2}}[|S1\rangle - |S2\rangle]$.
- Scully also introduces a change of basis for the maser states:
  - $|+\rangle = \frac{1}{\sqrt{2}}[|1_1 0_2\rangle + |0_1 1_2\rangle]$ which is analogous to $|+45°\rangle = \frac{1}{\sqrt{2}}[|H\rangle + |V\rangle]$ and

# Introducing the quantum eraser: II

- $|-\rangle = \frac{1}{\sqrt{2}}\left[|1_1 0_2\rangle - |0_1 1_2\rangle\right]$ which is analogous to $|-45°\rangle = \frac{1}{\sqrt{2}}\left[|H\rangle - |V\rangle\right]$.

- Then shutters and a detector are placed between the chambers as in Figure 5. When the shutters are closed, there is no "erasure" as when no final $+45°$ or $-45°$ polarizers are inserted in the polarization model and the mush pattern is observed.

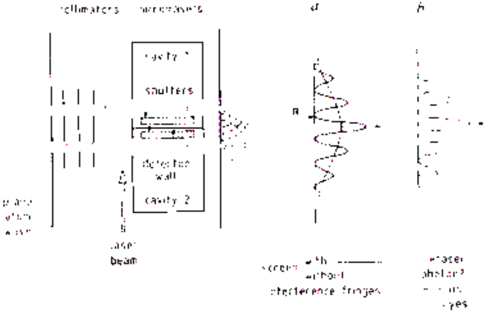



FIG. 5 a. Quantum erasure configuration in which electro-optic shutters separate microwave photons in two cavities from the thin-film semiconductor (detector wall) which absorbs microwave photons and acts as a photodetector. b. Density of particles on the screen depending upon whether a photocount is observed in the detector wall ('yes') or not ('no'), demonstrating that correlations between the event on the screen and the eraser photocount are necessary to retrieve the interference pattern.

# Introducing the quantum eraser: IV

- When the shutters are open, then the detector is a measurement of the $|+\rangle$ or $|-\rangle$ states.
    - The detector state $|e\rangle$ is the excited state that registers $|+\rangle$;
    - The detector state $|d\rangle$ is the de-excited state registering $|-\rangle$.
- Prior to opening the shutters, the detector is in the neutral state which is also $|d\rangle$. Then adding another component to the composite system to represent the detector, the formula (6)

$$\Psi(r) = \frac{1}{\sqrt{2}}\left[\psi_1(r) \otimes |1_1 0_2\rangle + \psi_2(r) \otimes |0_1 1_2\rangle\right] \otimes |b\rangle$$

is transformed in the new bases and with the new component into Scully's formula (12):

$$\Psi(r) = \frac{1}{\sqrt{2}}\left[\psi_+(r) \otimes |+\rangle + \psi_-(r) \otimes |-\rangle\right] \otimes |b\rangle \otimes |d\rangle.$$

- When the shutters are open, so the maser chambers interact with the detector, then the new state is given by Scully's formula (13):

$$\Psi(r) = \frac{1}{\sqrt{2}}\left[\psi_{+}(r) \otimes |0_1 0_2\rangle \otimes |e\rangle + \psi_{-}(r) \otimes |-\rangle \otimes |d\rangle\right] \otimes |b\rangle.$$

- If there is no measurement to collapse to the $|e\rangle$ portion or the $|d\rangle$ portion of the entangled state (i.e., shutters closed), then the probability distribution at the wall is the usual mush pattern with no interference.
- On an atom-by-atom basis, we can first make a positional measurement by observing a hit at the wall and then make another measurement by opening the shutters and observing the detector:

# Introducing the quantum eraser: VI

- If the detector state is $|e\rangle$, the hit on the wall is labeled "yes-atom."
- If the detector state is $|d\rangle$, the hit on the wall is labeled "no-atom."

- After recording much data, the yes-atoms will show the fringe interference pattern of formula (15):

$$P_e(R) = \frac{1}{2}\left[|\psi_1(R)|^2 + |\psi_2(R)|^2\right] + \mathrm{Re}(\psi_1{}^*(R)\,\psi_2(R)).$$

In the polarizer model, inserting the $+45°$polarizer gives the state $\frac{1}{\sqrt{2}}[|S1\rangle + |S2\rangle] \otimes |+45°\rangle$ so the corresponding probability formula is: (using $\langle 45°|45°\rangle = 1$)

$$\frac{1}{2}[\langle S1|P_{\Delta y}S1\rangle + \langle S2|P_{\Delta y}S2\rangle] + \mathrm{Re}\left(\langle S1|P_{\Delta y}S2\rangle\right).$$

# Introducing the quantum eraser: VII

- In both cases, the fringe is like the original interference pattern.
- After recording much data, the no-atoms will show the antifringe interference pattern of formula (16):

$$P_d(R) = \frac{1}{2}\left[\left|\psi_1(R)\right|^2 + \left|\psi_2(R)\right|^2\right] - \mathrm{Re}(\psi_1{}^*(R)\,\psi_2(R)).$$

In the polarizer model, inserting the $-45°$ gives the state $\frac{1}{\sqrt{2}}(|S1\rangle - |S2\rangle) \otimes |-45°\rangle$ so the probability calculation for the anti-fringe pattern is:

# Introducing the quantum eraser: VIII

$$\frac{1}{2}\left\langle (S1-S2)\otimes -45^\circ \middle| P_{\Delta y}\otimes I \middle| (S1-S2)\otimes -45^\circ \right\rangle$$
$$= \frac{1}{2}\left\langle (S1-S2)\otimes -45^\circ \middle| \left(P_{\Delta y}S1 - P_{\Delta y}S2\right)\otimes -45^\circ \right\rangle$$
$$= \frac{1}{2}\left[\left\langle S1|P_{\Delta y}S1\right\rangle - \left\langle S1|P_{\Delta y}S2\right\rangle - \left\langle S2|P_{\Delta y}S1\right\rangle + \left\langle S2|P_{\Delta y}S2\right\rangle\right]\left\langle -45^\circ | -45^\circ\right\rangle$$
$$= \frac{1}{2}\left[\left\langle S1|P_{\Delta y}S1\right\rangle + \left\langle S2|P_{\Delta y}S2\right\rangle\right] - \mathrm{Re}\left(\left\langle S2|P_{\Delta y}S1\right\rangle\right).$$

[analogue of (16) above]

- The sum of the fringe and antifringe patterns gives the original mush pattern in both models.

# The extra Scully-Walborn mystery

- In the polarizer model, the hit at the wall was after either a $\pm 45°$ polarizer inserted so we can easily visualize each filter picking out one pattern or the other out of the mush.
- But in the maser model, the analogue to putting in a $\pm 45°$ filter is the measurement of the detector which can happen AFTER the hit at the wall. Isn't that retrocausality?
- How does an atom know where to land if only the future event of the detector registering a "yes" or "no" determines if it is in the fringe or antifringe pattern? Thus the detector event seems to retrocause the atom to be in one pattern or the other.
- Doesn't this show that the Scully maser model (and the isomorphic Walborn model) exhibit genuine retrocausality, unlike the simpler polarizer model (where the $\pm 45°$ polarizer does the fringe-antifringe filtering *before* the hit at the wall)?

# Order of measuring two components does not matter: I

- The answer to this additional puzzle in the Scully and Walborn models lies in seeing that the time-ordering of the measurements does not alter the final probability distribution–as was pointed out by Bram Gaasbeek in a recent 2010 paper: *Demystifying the Delayed Choice Experiments*. [quant-ph] arXiv:1007.3977v1.
- The irrelevance of time-order of these measurements is the QM version of the probability theory result that given a joint distribution $\Pr(X, Y)$ over random variables $X, Y$, one can arrive at the same probability $\Pr(X = x_0, Y = y_0)$ by first sampling $Y$ to get $y_0$ with probability $\Pr(Y = y_0) = \sum_x \Pr(X = x, Y = y_0)$, and taking the

# Order of measuring two components does not matter: II

probability of getting $X = x_0$ conditional on $Y = y_0$:
$\Pr(X = x_0 | Y = y_0) = \Pr(X = x_0, Y = y_0) / \Pr(Y = y_0)$ or the reverse sequence. Either way, the result is the same:

$$\begin{aligned}
\Pr(X = x_0, Y = y_0) &= \Pr(X = x_0 | Y = y_0) \Pr(Y = y_0) \\
&= \Pr(Y = y_0 | X = x_0) \Pr(X = x_0) .
\end{aligned}$$

# Order of measuring two components does not matter: III

- Gaasbeek gives the quantum version: given a state $|\psi\rangle = \sum_{ij} \alpha_{ij} |i\rangle \otimes |j\rangle$ in a two-component system, the probability that a measurement on the first component yields $|i_0\rangle$ is $\Pr(i = i_0) = \sum_j |\alpha_{i_0 j}|^2$ and similarly $\Pr(j = j_0) = \sum_i |\alpha_{ij_0}|^2$ for a measurement on the second component.
- If we first measure the second component and get $|j_0\rangle$, then the state collapses as:

$$|\psi\rangle \to |\psi'\rangle = \sum_i \alpha_{ij_0} |i\rangle \otimes |j_0\rangle \,/\, \sum_i |\alpha_{ij_0}|^2 .$$

Then starting in the state $|\psi'\rangle$, the probability of a first component measurement giving $|i_0\rangle$ is:

# Order of measuring two components does not matter: IV

$$\Pr\left(i = i_0 | j = j_0\right) = \left|\alpha_{i_0 j_0}\right|^2 / \sum_i \left|\alpha_{i j_0}\right|^2.$$

- If we perform the measurements in the opposite order, then we find;

$$\Pr\left(j = j_0 | i = i_0\right) = \left|\alpha_{i_0 j_0}\right|^2 / \sum_j \left|\alpha_{i_0 j}\right|^2$$

so that:

$$\begin{aligned} \Pr\left(i = i_0, j = j_0\right) &= \Pr\left(i = i_0 | j = j_0\right) \Pr\left(j = j_0\right) \\ &= \Pr\left(j = j_0 | i = i_0\right) \Pr\left(i = i_0\right). \end{aligned}$$

# Order of measuring two components does not matter: V

- While this result is simple and well-known, Gaasbeek's contribution is to use it to eliminate the last bit of retrocausal mystery out of the Scully maser or Walborn optical models. Applying this result to the Scully model, it means that the probability of the joint event of hitting the wall in region $r$ and getting an excited reading $|e\rangle$ is the same regardless of the order in which we took the measurements. Thus instead of reading the detector after the atom hit the wall at some $r$, we could have read the detector while the atom was in flight before it hit the wall without changing the statistical correlations.

- With either order of doing the measurements, what counts are the correlations:

$$|e\rangle \longleftrightarrow \text{"yes" hit at wall = fringe, and}$$
$$|d\rangle \longleftrightarrow \text{"no" hit at wall = antifringe.}$$

- The key formula is Scully's formula (13):

$$\Psi(r) = \frac{1}{\sqrt{2}}\left[\psi_{+}(r)\otimes|0_1 0_2\rangle\otimes|e\rangle + \psi_{-}(r)\otimes|-\rangle\otimes|d\rangle\right]\otimes|b\rangle.$$

# Explaining the mystery: II

The two either-order events are $|e\rangle$ , $|d\rangle$ measurements and hit-the-wall at $r$ (for different $r$'s) measurements. The detector states are not entangled with specific positions $r$ (so the atom does not jump from a hit in the fringe pattern to an antifringe hit or vice-versa). It is entangled with the symmetric $\psi_+(r)$ or antisymmetric $\psi_-(r)$ states–which give the fringe and antifringe distributions respectively.

- Correlating the $|e\rangle$ readings, atom by atom, with the hits, i.e., the yes-atoms, will single out hits following the $\psi_+(r)$ state's fringe-distribution, which is analogous to "correlating" the photons that went through the $+45°$ filter with the evolved $\frac{1}{\sqrt{2}}[|S1\rangle + |S2\rangle]$ distribution.

- Correlating the $|d\rangle$ readings with the no-atoms hits singles out the hits following the $\psi_{-}(r)$ state's antifringe-distribution, which is analogous to "correlating" the photons that went through the $-45°$ filter with the evolved $\frac{1}{\sqrt{2}}[|S1\rangle - |S2\rangle]$ distribution.
- Thus one should exorcise any mental imagery of a detector reading retrocausing the wall hit to be in the appropriate pattern. What counts to build up the interference patterns is the joint probabilities:

$$Pr[detector = e, hit = r \text{ according to } \psi_{+}(r)], \text{ and}$$
$$Pr[detector = d, hit = r \text{ according to } \psi_{-}(r)]$$

and those probabilities are the same regardless of the order of measurement.

# Explaining the mystery: IV

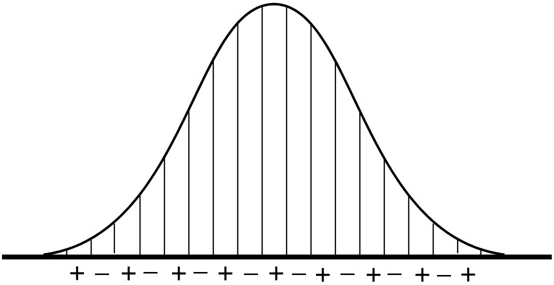
- Neither Scully nor Walborn developed the formula that would show the entanglement between the *direct* position measurement of $r$ and the detector states $|e\rangle$ and $|d\rangle$–only the formula (13) entanglement between the detector states and the distributions given by $\psi_+(r)$ and $\psi_-(r)$.
- But formula (13) can be used to tell the corrections between $r$ and detector states.
- If we *first* take the detector measurement while the atom was in flight, then, say, a $|d\rangle$ reading would, via the entanglement, cause the later hit-probabilities to be according to $\psi_-(r)$, and that would accordingly increase the probability of getting a peak-$r$ in the $\psi_-(r)$ anti-fringe pattern and decrease the probability of a valley-$r$ in the $\psi_+(r)$ fringe pattern.

# Explaining the mystery: V

- By the result that the order of measurement does not matter, if the atom first hits at a position $r$ that is in a valley of the fringe distribution and at a peak of the anti-fringe distribution, then, via the entanglement, that will accordingly change the probabilities of the later detector measurement giving $|e\rangle$ (less probability) or $|d\rangle$ (more probability).
- Thus Scully's procedure of marking the hits "yes" or "no" according to the detector readings $|e\rangle$ or $|d\rangle$ will *tend* to respectively pick out two statistical patterns of the fringe and antifringe.

# Explaining the mystery: VI

- There is another way to make the point clear. Instead of the statistical patterns, fringe and antifringe, suppose we simplify to a rigid separation of probability slices so that the sum of the $+$ slices gives the probability of the $\psi_+$ state and the sum of the $-$ slices gives the probability of the $\psi_-$ state.



# Explaining the mystery: VII

- The $r$-dependence is transformed into a two-way possibility of getting $\psi_+$ or $\psi_-$ so Scully's formula (13) becomes: $\Psi = \frac{1}{\sqrt{2}}\left[\psi_+ \otimes |0_1 0_2\rangle \otimes |e\rangle + \psi_- \otimes |-\rangle \otimes |d\rangle\right] \otimes |b\rangle$.
- Then we have a rigid entanglement $\psi_+ \otimes |e\rangle + \psi_- \otimes |d\rangle$ and the order of measuring the detector state $|e\rangle$, $|d\rangle$ or the $\psi_\pm$ state does not matter.
- If a $\psi_+$ was first recorded, then the entanglement would not only change the probability but would ensure the later detector reading of $|e\rangle$, and vice-versa.
- If a $\psi_-$ was first recorded, then the entanglement would ensure the later detector reading of $|d\rangle$, and vice-versa.
- In the actual model, these rigid connections are replaced by the probability distributions $\psi_+(r)$ and $\psi_-(r)$.

# Scully resolution of the "Jaynes Paradox": I

- It is important to note that Scully et al. point out that their proposed quantum eraser does not involve the retrocausality that occasioned the remarkable rant by Jaynes:

"By applying or not applying the eraser mechanism before measuring the state of the microwave cavities we can, at will, force the atomic beam into either: (1) a state with a known path, and no possibility of interference effects in any subsequent measurement; (2) a state with both $\psi_1$ and $\psi_2$ present with a known relative phase. Interference effects are then not only observable, but predictable. And we can decide which to do after the interaction is over and the atom is far from the cavities, so there can be no thought of any physical influence on the atom's centre-of-mass wavefunction!"

# Scully resolution of the "Jaynes Paradox": II

"...I say that [present quantum theory] constitutes a violent irrationality, that somewhere in this theory the distinction between reality and our knowledge of reality has become lost, and the result has more the character of medieval necromancy than of science." [Edwin Jaynes, quoted in Scully et al. 1991, p. 114] ["necromancy" = "a method of divination through invocation of the dead" (Webster)]

- But Scully et al. point out that their model does NOT involve any such retrocausality (not to mention, necromancy) since by correlating the "yes"-atoms with the $|e\rangle$ readings and the "no"-atoms with the $|d\rangle$ readings, they can statistically bring out the fringe or antifringe patterns.
- "In this way, we have resolved the 'Jaynes Paradox.'" [Scully et al. 1991, p. 115]

# Quantum entropies

David Ellerman

UCR

February 2012

# Review of Shannon and Logical Entropies:

- Assume distributions $\{p_x\}$ and $\{q_x\}$ are over same indices. When given joint distribution $\Pr(X = x, Y = y) = p_{xy}$, then $p_x = \sum_y p_{xy}$ and $p_y = \sum_x p_{xy}$ are the marginals.

| | Shannon Entropy | Logical Entropy |
|---|---|---|
| Entropy | $H(p_x) = \Sigma_x p_x \log(1/p_x)$ | $h(p_x) = \Sigma_x p_x(1-p_x)$ |
| Uniform 1/n | $\log(n)$ | $1-1/n$ |
| Cross entropy | $H(p_x \Vert q_x) = \Sigma p_x \log(1/q_x)$ | $h(p_x \Vert q_x) = \Sigma p_x(1-q_x)$ |
| Divergence | $D(p_x \Vert q_x) = \Sigma_x p_x \log(p_x/q_x)$ | $d(p_x \Vert q_x) = \Sigma_x(p_x - q_x)^2$ |
| Information Ineq. | $D(p\Vert q) \geq 0$ with = iff $p_i = q_i$ all i. | $d(p\Vert q) \geq 0$ with = iff $p_i = q_i$ all i. |
| Joint entropy | $H(X,Y) = \Sigma_{xy} p_{xy} \log(1/p_{xy})$ | $h(X,Y) = \Sigma_{xy} p_{xy}(1-p_{xy})$ |
| Mutual info. | $H(X:Y) = H(X)+H(Y)-H(X,Y)$ | $m(X,Y) = h(X)+h(Y)-h(X,Y)$ |
| Independence | $H(X,Y) = H(X) + H(Y)$ | $m(X,Y) = h(X)h(Y)$ |

# von Neumann and Logical Entropies:

- Let $\rho$ and $\sigma$ be mixed states.

| | von Neumann Entropy | Logical Entropy |
|---|---|---|
| Entropy | $S(\rho) = -\mathrm{tr}(\rho\log(\rho))$ | $h(\rho) = \mathrm{tr}(\rho(1-\rho)) = 1-\mathrm{tr}(\rho^2)$ |
| $\rho = \Sigma r_i \vert i\rangle\langle i\vert$ | $S(\rho) = \Sigma r_i \log(1/r_i)$ | $h(\rho) = 1-\Sigma r_i^2$ |
| $\sigma = \Sigma s_j \vert j\rangle\langle j\vert$ | $S(\sigma) = \Sigma s_j \log(1/s_j)$ | $h(\sigma) = 1-\Sigma s_j^2$ |
| Pure $\rho = \vert\psi\rangle\langle\psi\vert$ | $S(\rho) = 0$ | $h(\rho) = 0$ |
| Compl. mixed I/n | $\log(n)$ | $1-1/n$ |
| Divergence | $S(\rho\Vert\sigma) = \mathrm{tr}[\rho\log(\rho)-\rho\log(\sigma)]$ | $d(\rho\Vert\sigma) = \mathrm{tr}[(\rho-\sigma)^2]$ |
| Information Ineq. | $S(\rho\Vert\sigma) \geq 0$ | $d(\rho\Vert\sigma) \geq 0$ |
| Tensor product | $S(\rho\otimes\sigma) = S(\rho) + S(\sigma)$ | $h(\rho\otimes\sigma) = h(\rho)[1-h(\sigma)]+h(\sigma)$ |

# Interpretation of logical entropy: I

- The interpretation of the *classical logical entropy* $h(p) = 1 - \sum_i p_i^2$ of a probability distribution $p = \{p_i\}$ is the probability of drawing a distinction $i \neq i'$ in two independent samplings of the distribution.
- The interpretation of the *quantum logical entropy* $h(\rho) = 1 - \mathrm{tr}\left(\rho^2\right)$ of a mixed state $\rho = \sum_i r_i |i\rangle \langle i|$ (orthogonal decomposition) is the probability of getting distinct eigenstates $|i\rangle \neq |i'\rangle$ in two independent measurements of $\rho$ (using the $\{|i\rangle\}$ measurement basis), i.e., the total distinction probability.

# Interpretation of logical entropy: II

- The interpretation can be expressed without using the orthogonal decomposition so we start with $\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i|$. Then $\rho$ represented in any basis $\{|m\rangle\}$ has the entries: $\rho_{mm'} = \sum_i p_i \langle m|\psi_i\rangle \langle\psi_i|m'\rangle$ which can be interpreted as the amplitude for $m$ to be indistinct from $m'$, the $m, m'$ *indistinction amplitude*. Then the $m^{th}$ diagonal element of $\rho^2$ is:

$$\left(\rho^2\right)_{mm} = \sum_{m'=1}^{d} \rho_{mm'}\rho_{m'm} = \left(\rho_{mm}\right)^2 + \sum_{m' \neq m} \left|\rho_{mm'}\right|^2.$$

- Note that every $\left|\rho_{mm'}\right|^2$ for $\rho_{mm'}$ an entry in $\rho$ is included just once in:

# Interpretation of logical entropy: III

$$\mathrm{tr}\left(\rho^2\right) = \sum_m \left(\rho^2\right)_{mm} = \sum_m \rho_{mm}^2 + \sum_m \sum_{m' \neq m} \left|\rho_{mm'}\right|^2$$
$$\mathrm{tr}\left(\rho^2\right) = \sum_{m,m'} \left|\rho_{mm'}\right|^2$$
sum of *indistinction probabilities*.

- Thus the quantum logical entropy is again:

$$h\left(\rho\right) = 1 - \mathrm{tr}\left(\rho^2\right)$$
sum of *distinction probabilities*.

# Interpreting coherence terms as indistinction amplitudes: I

- Consider any *pure* state $\rho = |\psi\rangle\langle\psi|$. The general three-dimensional case is illustrative:

$$|\psi\rangle = \alpha_1 |1\rangle + \alpha_2 |2\rangle + \alpha_3 |3\rangle \text{ so: } \rho = \begin{bmatrix} \alpha_1\alpha_1^* & \alpha_1\alpha_2^* & \alpha_1\alpha_3^* \\ \alpha_2\alpha_1^* & \alpha_2\alpha_2^* & \alpha_2\alpha_3^* \\ \alpha_3\alpha_1^* & \alpha_3\alpha_2^* & \alpha_3\alpha_3^* \end{bmatrix}.$$

- The diagonal term $\rho_m = \alpha_m\alpha_m^*$ is the probability that a $\{|m\rangle\}$-basis measurement of the state $|\psi\rangle$ will result in the eigenstate $|m\rangle$.
- The probability $\rho_m\rho_{m'} = \alpha_m\alpha_m^*\alpha_{m'}\alpha_{m'}^*$ is the probability that two independent measurements would result in the pair of eigenstates $(|m\rangle, |m'\rangle)$.

# Interpreting coherence terms as indistinction amplitudes: II

- The off-diagonal *coherence term* $\rho_{mm'}$ of $\rho$ is the amplitude $\alpha_m \alpha_{m'}^* = \langle m|\psi\rangle \langle\psi|m'\rangle$ for $\psi$ to superpose or be indistinct between $m$ and $m'$, whose corresponding probability

$|\rho_{mm'}|^2 = \rho_{mm'}\rho_{m'm} = \alpha_m\alpha_{m'}^*\alpha_{m'}\alpha_m^* = \alpha_m\alpha_m^*\alpha_{m'}\alpha_{m'}^* = \rho_m\rho_{m'}$ probability of two measurements giving pair $(|m\rangle, |m'\rangle)$.

- Now in the pure state $\rho = |\psi\rangle\langle\psi|$, no distinctions or measurements have been made yet, so all the amplitudes giving pair-probabilities are interpreted as indistinction amplitudes.
- In a pure state, since there are no distinctions, all the indistinction probabilities sum to 1 so:

# Interpreting coherence terms as indistinction amplitudes: III

- $\operatorname{tr}(\rho^2) = \sum_{m,m'} |\rho_{mm'}|^2 = 1$, and
- $h(\rho) = 1 - \operatorname{tr}(\rho^2) = 0$ for any pure state $\rho$.

- Thus the coherence terms $\rho_{mm'}$ in any pure density matrix $\rho = |\psi\rangle\langle\psi|$ can be interpreted as the amplitudes for the indistinction probabilities $\rho_m \rho_{m'}$ (the $\rho_m = \rho_{mm}$ being the diagonal entries).
- In a general mixed state density matrix $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, that interpretation in each pure state $|\psi_i\rangle\langle\psi_i|$ is weighted by a probability $p_i$. The general entries $\rho_{mm'}$ are thus weighted-amplitudes and the corresponding indistinction probabilities give:

# Interpreting coherence terms as indistinction amplitudes: IV

$$\operatorname{tr}\left(\rho^2\right) = \sum_{m,m'} \left|\rho_{mm'}\right|^2$$

Sum of indistinction probabilities.

- But distinctions have to be made for a pure state to give a mixed state so the sum of the indistinction probabilities will not in general be one, and the complementary sum of distinction probabilities is the logical entropy:

$$h(\rho) = 1 - \operatorname{tr}\left(\rho^2\right)$$

Total of distinction probabilities.

# Classical to quantum logical entropy: I

- Given an index set $U = \{1, 2, 3\}$ with the probabilities $p = \{p_1, p_2, p_3\}$, the probability of drawing the ordered pair $(i, j)$ in a pair of independent draws is $p_i p_j$.
- Given a partition $\pi$ on $U$, the logical entropy $h(\pi)$ of the partition is probability of drawing a distinction, which is $1-$ probability of drawing an indistinction.

1. If $\pi = \{U\} = \mathbf{0}$, the indiscrete or blob partition, then any drawn pair is an indistinction so the total indistinction probability is 1, and the logical entropy is $h(\mathbf{0}) = 0$.
2. If $\pi = \{\{1\}, \{2\}, \{3\}\} = \mathbf{1}$, the discrete partition, then only the diagonal pairs $(i, i)$ are indistinctions to the total indistinction probability is $p_1^2 + p_2^2 + p_3^2$ and the logical entropy is $h(\mathbf{1}) = 1 - \sum_i p_i^2$. In the equiprobable case, $h(\mathbf{1}) = 1 - \left(\frac{1}{9} + \frac{1}{9} + \frac{1}{9}\right) = 1 - \frac{1}{3} = \frac{2}{3}$.

# Classical to quantum logical entropy: II

To make the QM connection clearest, we construct the corresponding quantum example:

- Instead of the set $U = \{1, 2, 3\}$, we start with an orthonormal basis set $\{|1\rangle, |2\rangle, |3\rangle\}$ where each basis element $|i\rangle$ has an associated amplitude $\sqrt{p_i}$.

1. In the state $|\psi\rangle$ where each basis vector is superposed with its amplitude $|\psi\rangle = \sqrt{p_1}\,|1\rangle + \sqrt{p_2}\,|2\rangle + \sqrt{p_3}\,|3\rangle$, the pure state density matrix is:

$$\rho = |\psi\rangle\langle\psi| = \begin{bmatrix} p_1 & \sqrt{p_1}\sqrt{p_2} & \sqrt{p_1}\sqrt{p_3} \\ \sqrt{p_2}\sqrt{p_1} & p_2 & \sqrt{p_2}\sqrt{p_3} \\ \sqrt{p_3}\sqrt{p_1} & \sqrt{p_3}\sqrt{p_2} & p_3 \end{bmatrix}.$$

# Classical to quantum logical entropy: III

The sum of the indistinction probabilities corresponding to the indistinction amplitudes is:

$\mathrm{tr}\left(\rho^2\right) = \sum_{i,j} \left|\rho_{ij}\right|^2 = \sum_{i,j} \left(\sqrt{p_i p_j}\sqrt{p_j p_i}\right) = \sum_{i,j} p_i p_j = 1$ so
$h\left(\rho\right) = 1 - \mathrm{tr}\left(\rho_2\right) = 0.$

2. In a nondegenerate measurement of $\rho$, we get the eigenstate $|i\rangle$ with probability $\left(\sqrt{p_i}\right)^2 = p_i$, so the measurement results can be described as the mixed state $p_1 |1\rangle\langle 1| + p_2 |2\rangle\langle 2| + p_3 |3\rangle\langle 3|$ with the density matrix: $\rho' = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & p_2 & 0 \\ 0 & 0 & p_3 \end{bmatrix}$. Since the measurement was nondegenerate ("discrete"), the only indistinction probabilities are the diagonal terms $p_1^2 + p_2^2 + p_3^2$ so the

# Classical to quantum logical entropy: IV

logical entropy is: $h(\rho') = 1 - (p_1^2 + p_2^2 + p_3^2)$, which in the equi-amplitude completely mixed case is:
$h(\rho') = 1 - \frac{1}{3} = \frac{2}{3}$.

- A nondegenerate measurement distinguishes between the eigenstates $|i\rangle$ so it converts all the off-diagonal coherence terms, which represented indistinction probabilities in the pure state, into distinction probabilities in the mixed state giving the measurement results. All those off-diagonal coherence terms in the pure state $\rho$ became 0 due to the decohering measurement in $\rho'$:

$$\rho = \begin{bmatrix} p_1 & \sqrt{p_1 p_2} & \sqrt{p_1 p_3} \\ \sqrt{p_2 p_1} & p_2 & \sqrt{p_2 p_3} \\ \sqrt{p_3 p_1} & \sqrt{p_3 p_2} & p_3 \end{bmatrix} \xrightarrow{meas.} \rho' = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & p_2 & 0 \\ 0 & 0 & p_3 \end{bmatrix}.$$

# Classical to quantum logical entropy: V

- The logical entropy $h(\rho')$ of the measured state is precisely the sum of the distinction probabilities resulting from those "disappeared" or "zeroed" off-diagonal coherence terms:

$$h(\rho') = \sum_{i \neq j} \left( \sqrt{p_i p_j} \right)^2 = 1 - \sum_i p_i^2 = 1 - \operatorname{tr}(\rho').$$

# Example of degenerate measurement: I

- Let's return to the same set example $p = \{p_1, p_2, p_3\}$ as the probabilities for $U = \{1, 2, 3\}$.
- Instead of seeing a measurement going from the undifferentiated blob $\{1, 2, 3\}$ to the discrete partition $\{\{1\}, \{2\}, \{3\}\}$, let's consider a "degenerate measurement" that goes from the blob only to the non-discrete partition $\pi' = \{\{1\}, \{2, 3\}\}$. This partition has four distinctions $(1, 2), (1, 3), (2, 1)$, and $(3, 1)$ with the total probability:

$$h(\pi') = 2p_1p_2 + 2p_1p_3$$

which could also be seen as $1-$ sum of probs for the five remaining indistinctions $(1, 1), (2, 2), (2, 3), (3, 2)$ ,and $(3, 3)$.

# Example of degenerate measurement: II

- In the quantum version, the measurement only yields the mixture of $|1\rangle$ with probability $p_1$ and $\frac{1}{\sqrt{2}}(|2\rangle + |3\rangle)$ with probability $p_2 + p_3$. This gives the mixed state density matrix:

$$\rho'' = p_1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + (p_2 + p_3) \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$= \begin{bmatrix} p_1 & 0 & 0 \\ 0 & \frac{p_2+p_3}{2} & \frac{p_2+p_3}{2} \\ 0 & \frac{p_2+p_3}{2} & \frac{p_2+p_3}{2} \end{bmatrix}.$$

# Example of degenerate measurement: III

Note the two non-zero off-diagonal coherence terms representing the indistinction amplitude of superposing $|2\rangle$ and $|3\rangle$. Note also the four zero off-diagonal terms representing the fact that $|1\rangle$ was distinguished from $|2\rangle$ and $|3\rangle$, which corresponds to the four pairs $(1,2)$, $(1,3)$, $(2,1)$, and $(3,1)$ that went from being indistinctions to distinctions in the set case.

- Thus the degenerate measurement has the effect:

$$\rho = \begin{bmatrix} p_1 & \sqrt{p_1 p_2} & \sqrt{p_1 p_3} \\ \sqrt{p_2 p_1} & p_2 & \sqrt{p_2 p_3} \\ \sqrt{p_3 p_1} & \sqrt{p_3 p_2} & p_3 \end{bmatrix} \xrightarrow{meas.} \rho'' = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & \frac{p_2+p_3}{2} & \frac{p_2+p_3}{2} \\ 0 & \frac{p_2+p_3}{2} & \frac{p_2+p_3}{2} \end{bmatrix}.$$

# Example of degenerate measurement: IV

- The sum of the indistinction probabilities is $p_1^2 + 4 \times \left(\frac{p_2+p_3}{2}\right)^2 = p_1^2 + (p_2+p_3)^2$ so the distinction probabilities are:

$$h(\rho'') = 1 - \left[p_1^2 + (p_2+p_3)^2\right]$$
$$= 1 - p_1^2 - \left(p_2^2 + 2p_2p_3 + p_3^2\right) = 2p_1p_2 + 2p_1p_3.$$

- The four new distinctions in the set case are here represented by the four disappeared or zeroed coherence terms which give the total new distinction probabilities of:

$$\left(\sqrt{p_1p_2}\right)^2 + \left(\sqrt{p_1p_3}\right)^2 + \left(\sqrt{p_2p_1}\right)^2 + \left(\sqrt{p_3p_1}\right)^2$$
$$= 2p_1p_2 + 2p_1p_3 = h(\rho'').\checkmark$$

# Modeling measurement in general: I

- Measurement (projective) makes distinctions and thus increases information so it should increase the entropies.
- How does a (projective) measurement change a mixed state $\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i|$. Let $\{|m\rangle\}$ be the orthonormal measurement basis with the projection matrices $P_m = |m\rangle \langle m|$ where $\sum_m P_m = I$.
- Then a measurement will, with probability $p_i$, start with a state $|\psi_i\rangle$ and will result in the state $|m\rangle$ with probability $|\langle m|\psi_i\rangle|^2$ so the total probability of getting the state $|m\rangle$ is $\sum_i p_i |\langle m|\psi_i\rangle|^2$.
- Hence the mixed state $\rho'$ giving the measurement outcomes weighed by their probabilities is: $\rho' = \sum_m \sum_i p_i |\langle m|\psi_i\rangle|^2 |m\rangle \langle m|$.

# Modeling measurement in general: II

- But

$$\sum_m P_m \rho P_m = \sum_m |m\rangle \langle m| \sum_i p_i |\psi_i\rangle \langle \psi_i| |m\rangle \langle m|$$
$$= \sum_m \sum_i p_i \langle m|\psi_i\rangle \langle \psi_i|m\rangle |m\rangle \langle m| = \rho'.$$

- Thus the effect of the $m$-basis measurement is:

$$\rho = \begin{bmatrix} \rho_{11} & \cdots & \rho_{1d} \\ \vdots & \ddots & \vdots \\ \rho_{d1} & \cdots & \rho_{dd} \end{bmatrix} \xrightarrow{meas.} \rho' = \begin{bmatrix} \rho_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \rho_{dd} \end{bmatrix}.$$

# Measurement increases vN entropy

- From information inequality: $0 \leq S(\rho' || \rho) = -S(\rho) - \operatorname{tr}(\rho \log \rho')$ so it would be sufficient to show that $-\operatorname{tr}(\rho \log \rho') = S(\rho')$.
- Using $\sum_m P_m = I$, $P_m^2 = P_m$, and $\operatorname{tr}(AB) = \operatorname{tr}(BA)$;

$$-\operatorname{tr}(\rho \log \rho') = -\operatorname{tr}\left(\sum_m P_m \rho \log \rho\right) = -\operatorname{tr}\left(\sum_m P_m \rho \log \rho' P_m\right)$$

and $\rho' P_m = P_m \rho P_m = P_m \rho'$ so $P_m$ commutes with $\rho'$ and thus with $\log \rho'$, so

$$-\operatorname{tr}(\rho \log \rho') = -\operatorname{tr}\left(\sum_m P_m \rho P_m \log \rho'\right) = S(\rho').$$

- Proof gives no insight as to why measurement increases vN entropy (in addition to no interpretation to vN entropy).

# Measurement increases logical entropy: I

- Let $\rho = \begin{bmatrix} \rho_{11} & \cdots & \rho_{1d} \\ \vdots & \ddots & \vdots \\ \rho_{d1} & \cdots & \rho_{dd} \end{bmatrix}$ be the representation of $\rho$ in the measurement basis. Then the off-diagonal terms $\rho_{mm'}$ for $m \neq m'$ represent the coherence, i.e., the amplitude for superposition (indistinction) between $m$ and $m'$.
- Measurement decoheres, i.e.,

$$\rho \xrightarrow{meas.} \rho' = \sum_m P_m \rho P_m = \begin{bmatrix} \rho_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \rho_{dd} \end{bmatrix}.$$

- Logical entropy after measurement:

# Measurement increases logical entropy: II

$$h(\rho') = 1 - \operatorname{tr}\left(\rho'^2\right) = 1 - \sum_m \rho_{mm}^2.$$

- Logical entropy before measurement ($\rho$ not nec. pure):

$$h(\rho) = 1 - \operatorname{tr}\left(\rho^2\right) = 1 - \sum_m \left(\rho^2\right)_{mm} = 1 - \sum_m \sum_{m'} \rho_{mm'}\rho_{m'm}$$
$$= 1 - \sum_m \rho_{mm}^2 - \sum_{m \neq m'} \rho_{mm'}\rho_{m'm} = h(\rho') - \sum_{m \neq m'} \left|\rho_{mm'}\right|^2$$

so

$$h(\rho') - h(\rho) = \sum_{m \neq m'} \left|\rho_{mm'}\right|^2.$$

Increase in entropy = sum of new distinction probs resulting from disappeared off-diagonal coherence terms.

- Coherence terms give amplitude for keeping eigenvectors indistinct in a superposition. Measurement makes the distinctions that takes away that coherence.
- Logical entropy records precisely that loss of the coherence terms in the measurement.

# Cross-entropy, divergence, and related concepts

David Ellerman

UCR

March 2012

# Logical Cross-Entropy and Cross-Fidelity

- Given two probability distributions $\{p_x\}$ and $\{q_x\}$ with the same indices, the (classical) *logical cross-entropy* is:

$$h(p_x||q_x) = \sum_x p_x [1 - q_x] = 1 - \sum_x p_x q_x.$$

- The interpretation of the logical cross-entropy of two distributions is the probability of drawing distinct indices ("distinction probability") if one draw is according to $p_x$ and the other draw according to $q_x$. Note: $h(p_x||p_x) = h(p_x)$.
- The complementary notion might be defined as the (classical) *logical cross-fidelity*:

$$f(p_x||q_x) = \sum_x p_x q_x = 1 - h(p_x||q_x).$$

- It is the probability of drawing the same index ("indistinction probability") with one draw according to $p_x$ and the other according to $q_x$.

# Quantum logical cross-entropy and cross-fidelity: I

- Given mixed states $\rho$ and $\sigma$, the *quantum logical cross-entropy* is:

$$h\left(\rho||\sigma\right) = 1 - \mathrm{tr}\left(\rho\sigma\right).$$

- And the complementary notion is the *quantum logical cross-fidelity (purity* or *cross-coherence?)*:

$$f\left(\rho||\sigma\right) = \mathrm{tr}\left(\rho\sigma\right) = 1 - h\left(\rho||\sigma\right).$$

- If $\rho = \sum p_k \left|k\right\rangle \left\langle k\right|$ and $\sigma = \sum_k q_k \left|k\right\rangle \left\langle k\right|$ share an orthonormal basis, then $\mathrm{tr}\left(\rho\sigma\right) = \sum_k p_k q_k$ and we are back in the classical case.

# Quantum logical cross-entropy and cross-fidelity: II

- In general, $\rho$ and $\sigma$ have orthogonal decompositions $\rho = \sum_i p_i |i\rangle \langle i|$ and $\sigma = \sum_j q_j |j\rangle \langle j|$ and then:

$$f(\rho||\sigma) = \operatorname{tr}(\rho\sigma) = \langle\sigma\rangle_\rho = \sum_i p_i \langle i|\sigma|i\rangle$$
$$= \sum_i p_i \left\langle i \right| \sum_j q_j |j\rangle \langle j| \left. i \right\rangle = \sum_{i,j} p_i q_j |\langle i|j\rangle|^2.$$

- For the probability interpretation, we consider the direct and indirect ways of getting $|i\rangle$ twice:
   1. Direct draw: the mixed state $\rho$ gives an ordinary probability distribution $\{p_i\}$ over the states $\{|i\rangle\}$, so in the direct draw, we get a specific $|i\rangle$ with probability $p_i$.

# Quantum logical cross-entropy and cross-fidelity: III

2. Indirect draw: the mixed state $\sigma$ similarly gives a draw of a basis state $|j\rangle$ with probability $q_j$, and then a quantum measurement in the $\{|i\rangle\}$ basis gives the state $|i\rangle$ with probability $|\langle i|j\rangle|^2$ so the total probability of getting $|i\rangle$ by this indirect method is $\sum_j q_j |\langle i|j\rangle|^2$.
3. Thus the probability of getting the same $|i\rangle$ in both draws is the **indistinction probability**:

$$f(\rho||\sigma) = \mathrm{tr}(\rho\sigma) = \sum_{i,j} p_i q_j |\langle i|j\rangle|^2$$

and the **distinction probability** is:

$$h(\rho||\sigma) = 1 - \mathrm{tr}(\rho\sigma) = 1 - f(\rho||\sigma).$$

# Example: I

- Consider the following states in $\mathbb{C}^2$ associated with spin states ($\pm z$ spin basis):

$$\rho = \frac{1}{3}\left|x_+\right\rangle\left\langle x_+\right| + \frac{2}{3}\left|y_-\right\rangle\left\langle y_-\right| = \begin{bmatrix} \frac{1}{2} & \frac{1+2i}{6} \\ \frac{1-2i}{6} & \frac{1}{2} \end{bmatrix}$$

$$\sigma = P_{x-} = \left|x_-\right\rangle\left\langle x_-\right| = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}$$

- Then the product and the trace are computed as follows:

$$\rho\sigma = \begin{bmatrix} \frac{1}{2} & \frac{1+2i}{6} \\ \frac{1-2i}{6} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 1-i & -1+i \\ -1-i & 1+i \end{bmatrix}.$$

# Example: II

The product the other way is: $\sigma\rho = \frac{1}{6}\begin{bmatrix} 1+i & -1+i \\ -1-i & 1-i \end{bmatrix}$ so the trace is the same:

$$\operatorname{tr}(\rho\sigma) = \tfrac{1}{3} \text{ and } h(\rho||\sigma) = \tfrac{2}{3}.$$

- We now work through the interpretation. Since $\sigma = |x_-\rangle\langle x_-|$ is pure, $q_1 = 1$ so we only need to compute the two ways to get that state $|x_-\rangle$ from the two states that make up $\rho$. There is a probability $\frac{1}{3}$ of getting $|x_+\rangle$ but $\langle x_-|x_+\rangle = 0$ so there is no contribution there. There is a probability $\frac{2}{3}$ of getting $|y_-\rangle$ and:

# Example: III

$$\langle x\_|y\_\rangle = x\_^{\dagger}y\_ = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \frac{i}{2} - \frac{1}{2}$$

$$\langle x\_|y\_\rangle \langle y\_|x\_\rangle = \left(\frac{i}{2} - \frac{1}{2}\right)\left(-\frac{i}{2} - \frac{1}{2}\right) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} = |\langle x\_|y\_\rangle|^2$$

so we have:

$$\mathrm{tr}\,(\rho\sigma) = \frac{1}{3}\,|\langle x\_|x_+\rangle|^2 + \frac{2}{3}\,|\langle x_-|y\_\rangle|^2 = 0 + \frac{2}{3}\frac{1}{2} = \frac{1}{3}.\checkmark$$

# "Criticism" of fidelity measure: I

- Richard Jozsa [1994. Fidelity for mixed quantum states. *Journal of Modern Optics*. 41 (12)] notes that if $\rho = \frac{I}{N}$ (completely mixed state), then $\operatorname{tr}\left(\frac{I}{N}\sigma\right) = \frac{1}{N}\operatorname{tr}(\sigma) = \frac{1}{N}$ regardless of $\sigma$ so $\operatorname{tr}(\rho\sigma)$ "is unsatisfactory as a measure of fidelity."
- This assumes that the purpose of "cross-fidelity" [or "purity" or "cross-coherence" or whatever] is to distinguish states, but that is the role of **divergence**.
- This aspect of the complement of cross-entropy is there even in the classical case. Given two probabilities distributions, an arbitrary one $\{p_i\}$ and the uniform distribution $\left\{\frac{1}{N}\right\}$ over the same indices, what is the indistinction probability if the first draw is according to $\{p_i\}$ and the second draw according to the uniform distribution?

# "Criticism" of fidelity measure: II

- No matter what was drawn on the first draw according to $\{p_i\}$, the probability of getting that same index on the second draw is $\frac{1}{N}$.
- Totally incoherent state is the "dominant gene" w.r.t. cross-fidelity, purity, or cross-coherence (or whatever it is called).

- For mixed states $\rho, \sigma$, the *quantum logical divergence* is:

$$d(\rho||\sigma) = \operatorname{tr}\left[(\rho - \sigma)^2\right].$$

- The Hermitian operator $\rho - \sigma$ can be unitarily diagonalized as $\rho - \sigma = UDU^{\dagger}$ and then the diagonal matrix $D$ has the Jordan decomposition $D = D^{+} - D^{-}$ as the difference of two positive matrices with orthogonal support. Thus

$$\rho - \sigma = U\left(D^{+} - D^{-}\right)U^{\dagger} = UD^{+}U^{\dagger} - UD^{-}U^{\dagger} = P - Q$$

is the difference of two positive operators $P, Q$ of orthogonal support.

$$d(\rho||\sigma) = \operatorname{tr}\left[(\rho-\sigma)^2\right] = \operatorname{tr}\left[(P-Q)^2\right] = \\ \operatorname{tr}\left(P^2\right) - 2\operatorname{tr}(PQ) + \operatorname{tr}\left(Q^2\right)$$

where $PQ = 0$ since they have orthogonal support and thus $\operatorname{tr}(PQ) = 0$ so we have:

$$\boxed{d(\rho||\sigma) \geq 0}$$

Quantum information inequality.

- Equivalent formulas are immediate:

$$\begin{aligned} d(\rho||\sigma) &= \operatorname{tr}\left(\rho^2\right) + \operatorname{tr}\left(\sigma^2\right) - 2\operatorname{tr}(\rho\sigma) \\ &= [1-h(\rho)] + [1-h(\sigma)] - 2f(\rho||\sigma) \\ &= 2h(\rho||\sigma) - h(\rho) - h(\sigma). \end{aligned}$$

- Hence the information inequality gives:

$$h\left(\rho||\sigma\right) \geq \frac{h(\rho)+h(\sigma)}{2}$$

Cross-entropy $\geq$ average entropy.

- An inequality for the entropy of the average $h(\frac{\rho+\sigma}{2})$ is:

$$4h\left(\frac{\rho+\sigma}{2}\right) - 2\left[h\left(\rho\right) + h\left(\sigma\right)\right] = \operatorname{tr}\left(\rho^2\right) + \operatorname{tr}\left(\sigma^2\right) - 2\operatorname{tr}\left(\rho\sigma\right) = d\left(\rho||\sigma\right)$$

so the information inequality also gives:

$$h\left(\frac{\rho+\sigma}{2}\right) \geq \frac{h(\rho)+h(\sigma)}{2}$$

"Mixing increases logical entropy"

- Interpretation of divergence:
  1. $h(\rho||\sigma)$ = distinction probability in the direct/indirect or "mixed" measurements;
  2. $h(\rho)$ = distinction probability with the "straight" measurements both in the $\{|i\rangle\}$ basis;
  3. $h(\sigma)$ = distinction probability with both measurements in the $\{|j\rangle\}$ basis.
- Hence the interpretation of the divergence,

$$d(\rho||\sigma) = [h(\rho||\sigma) - h(\rho)] + [h(\rho||\sigma) - h(\sigma)],$$

is the **total excess distinction probability** of the two mixed measurements over the two straight measurements.

- In the case that worried Jozsa where we want to measure the divergence between the completely mixed state $\rho = \frac{I}{N}$ and any state $\sigma$, $h\left(\frac{I}{N}||\sigma\right) = h\left(\frac{I}{N}\right)$ so the term $\left[h\left(\rho||\sigma\right) - h\left(\rho\right)\right]$ drops out and thus:

$$d\left(\frac{I}{N}||\sigma\right) = h\left(\frac{I}{N}\right) - h\left(\sigma\right) = 1 - \frac{1}{N} - h\left(\sigma\right) = \text{tr}\left(\sigma^2\right) - \frac{1}{N}$$

which is just the difference in the distinction probability for the completely mixed state and the $\sigma$ state (or the difference the other way around between the indistinction probabilities).

# Square root of logical divergence = Euclidean metric in matrix space: I

- Using an idea suggested by John DePillis and others, one can treat $n \times n$ matrices in $n$-dimensional Hilbert space as being the vectors in $n \times n$-dimensional Hilbert space.
- An inner product is defined on $n \times n$-dimensional space by:

$$\langle B|A\rangle \equiv \operatorname{tr}\left(AB^{\dagger}\right) \text{ for any two } n \times n \text{ matrices.}$$

- In any Hilbert space, the *Cauchy-Schwarz inequality* is (N&C, p. 68):

$$|\langle B|A\rangle|^2 \leq \langle A|A\rangle \langle B|B\rangle.$$

# Square root of logical divergence = Euclidean metric in matrix space: II

- If $A, B$ are Hermitian matrices, i.e., $A = A^\dagger$ and $B = B^\dagger$, like density matrices $\rho, \sigma$, then we have: $[\mathrm{tr}\,(\rho\sigma)]^2 \leq \mathrm{tr}\,(\rho^2)\,\mathrm{tr}\,(\sigma^2)$ where, incidentally, $[1 - h\,(\rho||\sigma)]^2 = [\mathrm{tr}\,(\rho\sigma)]^2$ and $\mathrm{tr}\,(\rho^2)\,\mathrm{tr}\,(\sigma^2) = [1 - h\,(\rho)]\,[1 - h\,(\sigma)] = 1 - h\,(\rho \otimes \sigma) = \mathrm{tr}\left[(\rho \otimes \sigma)^2\right]$.
- For $\sqrt{d\,(\rho||\sigma)} = \sqrt{\mathrm{tr}\left[(\rho - \sigma)^2\right]}$ to be a *metric*, we need:

1. $\sqrt{\mathrm{tr}\left[(\rho - \sigma)^2\right]} \geq 0$ (non-negativity);
2. $\sqrt{\mathrm{tr}\left[(\rho - \sigma)^2\right]} = 0$ if and only if $\rho = \sigma$;

# Square root of logical divergence = Euclidean metric in matrix space: III

3. $\sqrt{\operatorname{tr}\left[(\rho-\sigma)^2\right]} = \sqrt{\operatorname{tr}\left[(\sigma-\rho)^2\right]}$ (symmetry);
4. $\sqrt{\operatorname{tr}\left[(\rho-\tau)^2\right]} \le \sqrt{\operatorname{tr}\left[(\rho-\sigma)^2\right]} + \sqrt{\operatorname{tr}\left[(\sigma-\tau)^2\right]}$ (triangle inequality).

- Only the triangle inequality needs a proof.

Since $\rho-\tau = (\rho-\sigma)+(\sigma-\tau)$,

$$\operatorname{tr}\left((\rho-\tau)^2\right) =$$
$$\operatorname{tr}\left[(\rho-\sigma)^2\right] + 2\operatorname{tr}\left[(\rho-\sigma)(\sigma-\tau)\right] + \operatorname{tr}\left[(\sigma-\tau)^2\right].$$

# Square root of logical divergence = Euclidean metric in matrix space: IV

By the Cauchy-Schwarz inequality,
$\left[\operatorname{tr}\left[(\rho-\sigma)(\sigma-\tau)\right]\right]^2 \leq \operatorname{tr}\left[(\rho-\sigma)^2\right]\operatorname{tr}\left[(\sigma-\tau)^2\right]$ so taking the square root of each side:

$$\operatorname{tr}\left[(\rho-\sigma)(\sigma-\tau)\right] \leq \sqrt{\operatorname{tr}\left[(\rho-\sigma)^2\right]}\sqrt{\operatorname{tr}\left[(\sigma-\tau)^2\right]}.$$

Substituting in the middle term of the expansion for $\operatorname{tr}\left((\rho-\tau)^2\right)$ gives:

# Square root of logical divergence = Euclidean metric in matrix space: V

$$\operatorname{tr}\left((\rho-\tau)^2\right) \leq$$

$$\operatorname{tr}\left[(\rho-\sigma)^2\right] + 2\sqrt{\operatorname{tr}\left[(\rho-\sigma)^2\right]}\sqrt{\operatorname{tr}\left[(\sigma-\tau)^2\right]} + \operatorname{tr}\left[(\sigma-\tau)^2\right]$$

$$= \left[\sqrt{\operatorname{tr}\left[(\rho-\sigma)^2\right]} + \sqrt{\operatorname{tr}\left[(\sigma-\tau)^2\right]}\right]^2$$

so taking the square root of each side yields:

$$\sqrt{\operatorname{tr}\left((\rho-\tau)^2\right)} \leq \sqrt{\operatorname{tr}\left[(\rho-\sigma)^2\right]} + \sqrt{\operatorname{tr}\left[(\sigma-\tau)^2\right]},$$

the triangle inequality for the Euclidean metric in $n \times n$-dimensional Hilbert space. ■

# Limit cases

- The limits for quantum logical cross-entropy and cross-fidelity are complementary and since both are interpreted as distinction or indistinction probabilities, they are between 0 and 1:

$$0 \leq h(\rho||\sigma) \leq 1 \text{ and } 1 \geq f(\rho||\sigma) \geq 0$$
$$\text{left equality iff } \rho = \sigma = \text{pure}$$
$$\text{right equality iff orthogonal support.}$$

- The limits for the quantum logical divergence are:

$$0 \leq d(\rho||\sigma) \leq 2$$
$$\text{left equality iff } \rho = \sigma$$
$$\text{right equality iff } \rho, \sigma \text{ pure and orthogonal, i.e., } \rho\sigma = 0.$$

# Tensor product and entropies: I

**Theorem (Joint entropy theorem (N&C, p. 513))**

*Suppose $p_i$ are probabilities, $|i\rangle$ are orthogonal states for the system A, and $\rho_i$ is any set of density operators for another system B, then*

$$S\left(\sum_i p_i |i\rangle\langle i| \otimes \rho_i\right) = H(p_i) + \sum_i p_i S(\rho_i);$$
$$h\left(\sum_i p_i |i\rangle\langle i| \otimes \rho_i\right) = h(p_i) + \sum_i p_i^2 h(\rho_i).$$

- For any mixed states $\rho$ and $\sigma$:

# Tensor product and entropies: II

$$S(\rho \otimes \sigma) = S(\rho) + S(\sigma)$$

$$\begin{aligned} h(\rho \otimes \sigma) &= h(\rho)[1 - h(\sigma)] + h(\sigma) \\ &= h(\rho) + h(\sigma)[1 - h(\rho)] \\ &= h(\rho) + h(\sigma) - h(\rho)h(\sigma) \end{aligned}$$

- Interpretation of $h(\rho \otimes \sigma)$ = distinction probability of $\rho$ *times* indistinction probability of $\sigma$ plus distinction probability for $\sigma$.
- Special cases:

# Tensor product and entropies: III

$$S\left(\rho \otimes |\psi\rangle\langle\psi|\right) = S(\rho) \text{ for any pure state } \sigma = |\psi\rangle\langle\psi|$$
$$h\left(\rho \otimes |\psi\rangle\langle\psi|\right) = h(\rho) \text{ for any pure state } \sigma = |\psi\rangle\langle\psi|$$

Tensoring with a zero entropy state adds nothing to entropy

–

$$S\left(\rho \otimes \tfrac{I}{N}\right) = S(\rho) + \log N$$
$$h\left(\rho \otimes \tfrac{I}{N}\right) = \frac{h(\rho)}{N} + 1 - \frac{1}{N}$$

Tensoring with max entropy state

–

$$S\left(\tfrac{I}{N} \otimes \tfrac{I}{N}\right) = 2\log N$$
$$h\left(\tfrac{I}{N} \otimes \tfrac{I}{N}\right) = 1 - \frac{1}{N^2}$$

Max entropy state tensor-squared

# "Bad" definitions for vN entropy: I

- A joint probability distribution $p_{xy} = p(x,y)$ $= \Pr(X = x, Y = y)$ is defined on the *direct product* $X \times Y$, and the classical joint entropies are defined using that distribution: $H(X,Y) = \sum_{x,y} p_{xy} \log\left(\frac{1}{p_{xy}}\right)$ and $h(X,Y) = \sum_{xy} p_{xy}(1 - p_{xy})$.
- The tensor product of vector spaces is substantially different than the direct product (since it allows superposition), and yet the "vN joint entropy" is defined as if it were the quantum generalization.
- Given a composite system $AB$ represented by $H_A \otimes H_B$ for component systems $A$ and $B$, and given a density operator $\rho^{AB}$ on the tensor product, the *vN joint entropy* is defined (N&C, p. 514) as:

# "Bad" definitions for vN entropy: II

$$S(A,B) = -\operatorname{tr}(\rho^{AB} \log(\rho^{AB})).$$

- To make matters worse, where $S(A) = S(\rho^A)$ and $S(B) = S(\rho^B)$ are the vN entropies of the reduced density operators, then the *conditional vN entropy* and the *mutual vN information* are simply defined by formulas analogous to the classical case (where classically they at least had some motivation):

$$S(A|B) =_{df} S(A,B) - S(B)$$
$$S(A:B) =_{df} S(A) + S(B) - S(A,B).$$

A bit of trouble with Shannon's explanation of conditional entropy:

# "Bad" definitions for vN entropy: III

- Given a joint distribution $p_{xy}$, a conditional probability distribution is $p\left(Y=y|x\right)=\frac{p_{xy}}{p_x}$ where $p_x=\sum_y p_{xy}$, and thus there is a Shannon entropy $H\left(p\left(Y|x\right)\right)=H\left(Y|x\right)$ of that probability distribution.
- Shannon then defines the *conditional entropy* $H\left(Y|X\right)$ as the average of these entropies of the conditional distributions:

$$H\left(Y|X\right)=\sum_x p_x H\left(Y|x\right).$$

- After this motivated definition of the conditional entropy, then it is a theorem (not a definition) that:

$$H\left(Y|X\right)=H\left(X,Y\right)-H\left(X\right)=H\left(Y\right)-H\left(X:Y\right).$$

# "Bad" definitions for vN entropy: IV

- Since the Shannon conditional entropy is a non-negative sum of non-negative entropies (N&C, p. 514),

$$H(X) \leq H(X, Y).$$

- N&C explain this as "surely we cannot be more uncertain about the state of $X$ than we are about the joint state of $X$ and $Y$." (Ibid.)
- Since the Shannon mutual information $H(X : Y)$ is also always non-negative, we also have:

$$H(Y|X) \leq H(Y).$$

# "Bad" definitions for vN entropy: V

- Shannon explains this with a similar remark: "The uncertainty about $Y$ is never increased by knowledge of $X$." [quoted in: Uffink, Jos. *Measures of Uncertainty and the Uncertainty Principle*, University of Utrecht dissertation, 1990, p. 82 (on the web)].
- The intuition behind this explanation is wrong as is shown by Uffink's example:

$$[p_{xy}] = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} .98 & 0 \\ .01 & .01 \end{bmatrix}.$$

- The joint entropy is $H(X,Y) = 0.16$ and the entropy of the marginal distribution $p_y$ is $H(Y) = 0.08$. The conditional distribution $p(Y|X=2) = (0.5, 0.5)$ so that $H(Y|x=2) = 1$ and thus:

# "Bad" definitions for vN entropy: VI

$$H(Y|x=2) = 1 \nleq 0.08 = H(Y).$$

- Thus the explanation that "The uncertainty about $Y$ is never increased by knowledge of $X$" is clearly wrong, but Shannon's formula for conditional entropy is an average of the entropies of the conditional distributions.
  - The other conditional distribution $p(Y|x=1) = (1,0)$ with the entropy $H(Y|x=1) = 0$.
  - Hence the conditional entropy is:

$$H(Y|X) = 0.98 \times 0 + 0.01 \times 1 = .01 \leq 0.08 = H(Y).$$

- While there are some "problems" with the explanations of Shannon's conditional entropy, the *definition* of "vN conditional entropy" is shameless:

# "Bad" definitions for vN entropy: VII

$$S(A|B) =_{df} S(A,B) - S(B).$$

- As if to emphasize the lack of interpretation of the "vN conditional entropy" and the shameless formula-mongering, they go on to show that "vN conditional entropy" can be negative!
- Take a combined two qubit system $AB$ in the pure state $\frac{1}{\sqrt{2}}[|00\rangle + |11\rangle]$ so that the reduced density operator for $B$ (and $A$) is $\frac{I}{2}$ so that:

$$S(A|B) = S(A,B) - S(B) = 0 - 1 = -1!$$

# "Bad" definitions for vN entropy: VIII

- Instead of taking this as definitive evidence that the formula-mongering has taken a wrong turn, they say "intuition fails for quantum states." Woo-woo.
- They even derive the "result" that since separated pure states $|\psi\rangle\langle\psi| \otimes |\varphi\rangle\langle\varphi|$ have components in pure states (so all have zero entropy), we have that: pure states $|AB\rangle$ are entangled iff the conditional entropy $S(B|A) < 0$. Entanglement (woo-woo) means negative (conditional) entropy! More woo-woo.
- IMHO, this is another case (like retrocausality) where people have lost track of what is reasonable, and just accept the "result" as more quantum weirdness.

# Miscellany

David Ellerman

UCR

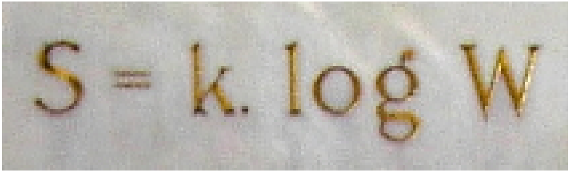
March 2012

# Statistical Mechanics entropy and Shannon entropy: I

- There is a constant meme in Shannon's information theory that his entropy $H(p) = \sum_i p_i \ln\left(\frac{1}{p_i}\right)$ (where I have used natural logs rather than base 2 logs) has the same **functional form** as entropy in statistical mechanics.
- However, the connection is only via a numerical approximation, the Stirling approximation, where only if the first two terms in the Stirling approximation are used, then the Shannon formula is obtained.
- The first two terms in the Stirling approximation for $\ln(N!)$ are: $\ln(N!) \approx N(\ln(N) - 1)$. The first three terms in the Stirling approximation are:
$\ln(N!) \approx N(\ln(N) - 1) + \frac{1}{2}\ln(2\pi N)$.

# Statistical Mechanics entropy and Shannon entropy: II

- If we consider a partition on a finite $U$ with $|U| = N$, with $n$ blocks of size $N_1, ..., N_n$, then the number of ways of distributing the individuals in these $n$ boxes with those numbers $N_i$ in the $i^{th}$ box is: $W = \frac{N!}{N_1! \times ... \times N_n!}$. The normalized natural log of $W$, $\frac{1}{N} \ln(W)$ is one form of entropy in statistical mechanics.
- On Boltzmann's gravestone:



# Statistical Mechanics entropy and Shannon entropy: III

- The entropy formula can then be developed using the first two terms in the Stirling approximation.

$$
\begin{gathered}
S=\frac{1}{N}\ln\left(W\right)=\frac{1}{N}\ln\left(\frac{N!}{N_1!\times\ldots\times N_n!}\right)=\frac{1}{N}\left[\ln(N!)-\sum_i\ln(N_i!)\right]\\
\approx\frac{1}{N}\left[N\left[\ln\left(N\right)-1\right]-\sum_i N_i\left[\ln\left(N_i\right)-1\right]\right]\\
=\frac{1}{N}\left[N\ln(N)-\sum N_i\ln(N_i)\right]=\frac{1}{N}\left[\sum N_i\ln\left(N\right)-\sum N_i\ln\left(N_i\right)\right]\\
=\sum\frac{N_i}{N}\ln\left(\frac{1}{N_i/N}\right)=\sum p_i\ln\left(\frac{1}{p_i}\right)=H\left(p\right)
\end{gathered}
$$

where $p_i=\frac{N_i}{N}$.

# Statistical Mechanics entropy and Shannon entropy: IV

- The Stirling approximation is an excellent *numerical* approximation for large $N$ (e.g., in statistical mechanics). But the common meme is not that Shannon's entropy formula is a good numerical approximation for entropy in statistical mechanics, but that it has the *same functional form*. That is simply *false* in view of the use of Stirling's approximation in the above derivation.
- The point can be emphasized by using the three-term Stirling approximation to get an even better numerical approximation.

# Statistical Mechanics entropy and Shannon entropy: V

$$\frac{1}{N}\ln\left(W\right)=\frac{1}{N}\ln\left(\frac{N!}{N_1!\times\ldots\times N_n!}\right)=\frac{1}{N}\left[\ln(N!)-\sum_i\ln(N_i!)\right]$$
$$\approx$$
$$\frac{1}{N}\left[N\left[\ln\left(N\right)-1\right]+\frac{1}{2}\ln\left(2\pi N\right)-\sum_i\left\{N_i\left[\ln\left(N_i\right)-1\right]-\frac{1}{2}\ln\left(2\pi N_i\right)\right\}\right]$$
$$=\frac{1}{N}\left[N\ln(N)+\frac{1}{2}\ln\left(2\pi N\right)-\sum\left\{N_i\ln\left(N_i\right)-\frac{1}{2}\ln\left(2\pi N_i\right)\right\}\right]$$
$$=$$
$$\frac{1}{N}\left[\sum_i N_i\ln(N)-\sum N_i\ln\left(N_i\right)\right]+\frac{1}{N}\left[\frac{1}{2}\ln\left(2\pi N\right)-\sum\frac{1}{2}\ln\left(2\pi N_i\right)\right]$$
$$=\left[\sum_i\frac{N_i}{N}\ln\left(\frac{1}{N_i/N}\right)\right]+\frac{1}{N}\frac{1}{2}\left[\ln\left(2\pi N\right)-\ln\left(\left(2\pi^n\right)\Pi N_i\right)\right]$$
$$=H\left(p\right)+\frac{1}{2N}\ln\left(\frac{2\pi N}{\left(2\pi\right)^n\Pi N_i}\right)=H\left(p\right)+\frac{1}{2N}\ln\left(\frac{2\pi N^n}{\left(2\pi\right)^n\Pi p_i}\right).$$

# Statistical Mechanics entropy and Shannon entropy: VI

- Thus the expression $H(p) + \frac{1}{2N}\ln\left(\frac{2\pi N^n}{(2\pi)^n \Pi p_i}\right)$ is an even *better* approximation to the entropy $\frac{1}{N}\ln(W)$ than $H(p)$. If anyone really thinks the Shannon functional form is "justified" by the connection to entropy in statistical mechanics, then they are welcome to redo information theory with the even "better" entropy formula: $H(p) + \frac{1}{2N}\ln\left(\frac{2\pi N^n}{(2\pi)^n \Pi p_i}\right)$.
- Thus any justification of the functional form of Shannon's entropy formula should not be done by waving one's hand in the direction of statistical mechanics.

# Entropy invariance under trace-preserving transformations

- Both the vN quantum entropy and the quantum logical entropy are defined using the trace of density operators, so those notions are invariant under similarity transformation (including unitary transformations) and indeed under any trace-preserving transformation.
- The logical cross-entropy $h(\rho||\sigma) = 1 - \mathrm{tr}(\rho\sigma)$ is also invariant for the same reason. For instance, under the unitary evolution $U(t_0, t_1)$, $\rho \to U\rho U^\dagger$ and $\sigma \to U\sigma U^\dagger$ so $\rho\sigma \to U\rho U^\dagger U\sigma U^\dagger = U\rho\sigma U^\dagger$ so

$$h\left(U\rho U^\dagger||U\sigma U^\dagger\right) = h(\rho||\sigma).$$

- Since the divergence $d(\rho||\sigma) = 2h(\rho||\sigma) - h(\rho) - h(\sigma)$, we also have:

$$d\left(U\rho U^\dagger||U\sigma U^\dagger\right) = d(\rho||\sigma).$$

# Positive semidefiniteness of matrices: I

- Any density operator or matrix $\rho$ is a *positive semidefinite* operator or matrix which means that $x^{\dagger}\rho x \geq 0$ for all vectors $x$, i.e., all its eigenvalues are non-negative.
- Another characterization of positive semidefiniteness uses the notion of a principal minor.
- One must be careful to distinguish between "successive principal minors" and "principal minors" in general.
    1. A *principal minor of order k* is the determinant of any square submatrix whose diagonal is along the main diagonal of the matrix, and where "submatrix" allows any permutation of indices or, equivalently interchanging the $i^{th}$ and $j^{th}$ rows and the $i^{th}$ and $j^{th}$ columns.

# Positive semidefiniteness of matrices: II

2. The *successive principal minors* are the principal minors of orders $1, \ldots, n$ starting in the NW corner of the matrix without any interchanging of rows and columns.

$$\left|\begin{array}{ccc} p_1 & \rho_{12} & \rho_{13} \\ \rho_{21} & p_2 & \rho_{23} \\ \rho_{31} & \rho_{32} & p_3 \end{array}\right|.$$

Successive principal minors of $\rho$

- While positive definiteness can be characterized by all the successive principal minors being positive, one cannot similarly characterize positive *semi*definiteness as all the successive principal minors being non-negative. For instance,

# Positive semidefiniteness of matrices: III

$$\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$$

has all the successive principal minors being non-negative, but it is not positive semidefinite (in fact it is negative semidefinite). Hence we need to strengthen the non-negativity condition to *all* principal minors.

- A matrix $\rho$ is positive semidefinite if and only if all principal minors are non-negative.

# Positive semidefiniteness of matrices: IV

- In the previous counterexample, the principal minor $[-1]$ has a negative determinant, so it fails this stronger condition. The condition referring to all principal minors could be equivalently stated in terms of the successive principal minors if we allow the arbitrary interchanges of the same rows and columns. For instance, interchanging the first and second rows and columns moves the $-1$ up to become the first successive principal minor.

$$\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \xrightarrow{\text{row}} \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} \xrightarrow{\text{col}} \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}.$$

# Connecting classical and quantum logical entropy: I

- Any density operator $\rho$ can be represented as a density matrix (e.g., $n = 3$):

$$\rho = \begin{bmatrix} p_1 & \rho_{12} & \rho_{13} \\ \rho_{21} & p_2 & \rho_{23} \\ \rho_{31} & \rho_{32} & p_3 \end{bmatrix}$$

in any orthonormal basis $M = \{|m_i\rangle\}$ where $p_i = \rho_{ii}$.

- The logical entropy $h(\rho)$ is defined as: $h(\rho) = 1 - \operatorname{tr}(\rho^2)$.
- It was previously shown that:

$$\operatorname{tr}(\rho^2) = \sum_{i,j} \left|\rho_{ij}\right|^2 = \sum_i p_i^2 + 2\sum_{i<j} \left|\rho_{ij}\right|^2,$$

# Connecting classical and quantum logical entropy: II

i.e., the trace of $\rho^2$ is just sum of probabilities $\left|\rho_{ij}\right|^2$ associated with all the (amplitudes) $\rho_{ij}$ entries in $\rho$.

- But the 1 in $h(\rho) = 1 - \operatorname{tr}(\rho^2)$ can be expanded as:

$$1 = \left(\sum_i p_i\right)\left(\sum_j p_j\right) = \sum_{i,j} p_i p_j = \sum_i p_i^2 + 2\sum_{i<j} p_i p_j.$$

- Hence we have:

$$\begin{aligned} h(\rho) &= 1 - \operatorname{tr}(\rho^2) \\ &= \left[\sum_i p_i^2 + 2\sum_{i<j} p_i p_j\right] - \left[\sum_i p_i^2 + 2\sum_{i<j} \left|\rho_{ij}\right|^2\right] \\ &= 2\sum_{i<j}\left[p_i p_j - \left|\rho_{ij}^2\right|\right]. \end{aligned}$$

# Connecting classical and quantum logical entropy: III

- That is the characterization of the logical entropy as the sum of the terms $p_i p_j - \left|\rho_{ij}^2\right|$ for any $i \neq j$ where since $\rho$ is Hermitian, we can just double the sum of those terms from the upper triangular section where $i < j$.
- When measuring $\rho$ using the measurement basis $M$, $p_i$ is the probability of getting the result $|m_i\rangle = |i\rangle$. If $\rho$ was a diagonal matrix in the $M$ basis with all $\rho_{ij} = 0$ for $i \neq j$, then we have essentially a classical discrete partition $\{\{|1\rangle\}, ..., |n\rangle\}$ or $\{\{1\}, ..., \{n\}\}$ with the logical entropy $1 - \sum_i p_i^2 = 2\sum_{i<j} p_i p_j$.

# Connecting classical and quantum logical entropy: IV

- But even classically the elements $i$ might be in bigger-than-singleton blocks and then the terms $p_i p_j$ are counted only when $i$ and $j$ are in different blocks. That is, if $i, j$ were in the same block, then we might say that their "*indistinction amplitude*" was $\sqrt{p_i p_j}$ and their *indistinction probability* was $\left|\sqrt{p_i p_j}\right|^2 = p_i p_j$ which would have to be subtracted off from the $p_i p_j$ that appeared in the sum $2\sum_{i<j} p_i p_j$ for the discrete partition to account for the fact that now $i, j$ are in the same block. Thus that $i, j$ term in the sum becomes $p_i p_j - \left|\sqrt{p_i p_j}\right|^2 = 0$ as the *net distinction probability* associated with $i, j$ and that is how it "drops out" of the sum when $i, j$ are in the same block.

# Connecting classical and quantum logical entropy: V

- This interpretation carries over to the quantum case where:
  1. $\rho_{ij}$ is the *indistinction amplitude;*
  2. $\left|\rho_{ij}\right|^2$ is the *indistinction probability;*
  3. $p_i p_j - |\rho_{ij}|^2$ is the *net distinction probability* for the pair $|i\rangle$ and $|j\rangle$; and
  4. $h(\rho) = 2\sum_{i<j}\left[p_i p_j - \left|\rho_{ij}\right|^2\right]$ is the *total of the net distinction probabilities* for the pairs of basis states $|i\rangle$ and $|j\rangle$.

# Connecting classical and quantum logical entropy: VI

- Thus we not only have a simple interpretation of logical entropy in the quantum case that directly generalizes the classical case, we can use the associated concepts like "indistinction amplitude" to interpret the entries $\rho_{ij}$ in the density matrix itself. Previously $\rho_{ij}$ was seen as an indicator of the "coherence" between the basis states $|i\rangle$ and $|j\rangle$ in the mixed state $\rho$.

# Seeing classical case through quantum lens: I

- We can retro-engineer the classical case using some of the fancy concepts from the quantum case–like density matrices.
- Let's take the "classical case" as having a finite set of points $U = \{1, 2, ..., n\}$ where each point $i$ has the probability $p_i$.
- If the partition is the discrete one, $\mathbf{1} = \{\{1\}, ..., \{n\}\}$, then the logical entropy of that partition is just the logical entropy of the probability distribution $p = \{p_1, ..., p_n\}$, i.e.,

$$h(\mathbf{1}) = h(p) = 1 - \sum_i p_i^2 = 2\sum_{i<j} p_i p_j.$$

- The "classical" density matrix corresponding to this discrete case is the $n \times n$ diagonal matrix with the $p_i$ along the diagonal.

# Seeing classical case through quantum lens: II

- But when the elements are grouped together in larger blocks, then some pairs $(i, j)$ of indices that were distinctions in the discrete case now become indistinctions since they are in the same block. Hence the off-diagonal term corresponding to the indistinction pairs goes from 0 to $\sqrt{p_i p_j}$ as the indistinction amplitude that gives the indistinction probability $p_i p_j$.
- If, for instance, $U = \{1, 2, 3\}$ and the elements $2, 3$ are grouped together in a block of a non-discrete partition $\pi = \{\{1\}, \{2, 3\}\}$, then the associated density matrix is:

$$\rho = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & p_2 & \sqrt{p_2 p_3} \\ 0 & \sqrt{p_2 p_3} & p_3 \end{bmatrix}$$

# Seeing classical case through quantum lens: III

and the logical entropy is:

$$h(\pi) = 2\sum_{i<j}\left[p_i p_j - \left|\rho_{ij}\right|^2\right]$$
$$= 2\left[(p_1 p_2 - 0) + (p_1 p_3 - 0) + \left(p_2 p_3 - \left|\sqrt{p_2 p_3}\right|^2\right)\right]$$
$$= 2p_1 p_2 + 2p_1 p_3.$$

- Since each off-diagonal term $\sqrt{p_i p_j}$ wipes out the $p_i p_j$ term in the sum for logical entropy, the terms that survive correspond to the off-diagonal zero terms.
- If $U = \{1, 2, 3, 4\}$ and the partition $\pi = \{\{1, 3\}, \{2, 4\}\}$, then the density matrix is:

# Seeing classical case through quantum lens: IV

$$\rho = \begin{bmatrix} p_1 & 0 & \sqrt{p_1 p_3} & 0 \\ 0 & p_2 & 0 & \sqrt{p_2 p_4} \\ \sqrt{p_1 p_3} & 0 & p_3 & 0 \\ 0 & \sqrt{p_2 p_4} & 0 & p_4 \end{bmatrix}$$

and the logical entropy is:

$$h(\pi) = 2\sum_{i<j}\left[p_i p_j - \left|\rho_{ij}\right|^2\right]$$
$$= 2p_1p_2 + 2p_1p_4 + 2p_2p_3 + 2p_2p_4.$$

# Seeing classical case through quantum lens: V

- Note that by interchanging rows and columns, each classical density matrix representing a partition can be turned into a block-diagonal matrix where the "blocks" of the matrix correspond to the blocks of the partition.
- In the special case of equiprobable points $p_i = \frac{1}{n}$, then we can factor a $\frac{1}{n}$ scalar outside of the classical density matrix. The remaining matrix then just has $0, 1$ entries and it is precisely the incidence matrix of the reflexive, symmetric, and transitive equivalence relation for the partition.
- For instance in the last example:

# Seeing classical case through quantum lens: VI

$$\rho = \begin{bmatrix} \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

where the $0,1$ matrix is the incidence matrix for the equivalence relation corresponding to the partition $\{\{1,3\},\{2,4\}\}$. That binary equivalence relation is exactly the set of indistinctions or "indits" of the partition, so the $1'$s in the $0,1$ matrix occur where there are indistinctions (all with "amplitude" $\frac{1}{4}$ in this case).

- By seeing classical case through a quantum lens, we can better understand the quantum case by keeping in mind the classical precursor.

# Seeing classical case through quantum lens: VII

- Classically, a pair $(i, j)$ is either a distinction of a partition (so $\rho_{ij} = 0$) or the pair is an indistinction (so $\rho_{ij} = \sqrt{p_i p_j}$). It is a yes-or-no business. In terms of the coherence language, the elements $i, j$ are either totally coherent or indistinct (i.e., in the same block) or totally decoherent or distinct (in distinct blocks).
- In the quantum case, a state $\rho$ can have the basis states $|i\rangle$ and $|j\rangle$ as being partially indistinct–as indicated by the indistinction amplitude $\rho_{ij}$–and thus partially distinct so the net distinction probability $p_i p_j - \left|\rho_{ij}\right|^2$ can be anywhere between the classical limits of 0 and $p_i p_j$.

# Another characterization of logical entropy: I

- We know that $h(\rho) = 1 - \operatorname{tr}(\rho^2) \geq 0$ since $\frac{1}{n} \leq \operatorname{tr}(\rho^2) \leq 1$.
- But in the expansion:

$$h(\rho) = \sum_{i<j} \left[p_i p_j - \left|\rho_{ij}\right|^2\right],$$

we have not shown that each term $p_i p_j - \left|\rho_{ij}\right|^2$ is non-negative.

- The slick proof of this is that since $\rho$ is positive semi-definite, all principal minors of any order $k$ are non-negative, and the principal minors of order 2 are:

$$\begin{vmatrix} p_i & \rho_{ij} \\ \rho_{ij}^* & p_j \end{vmatrix} = p_i p_j - \left|\rho_{ij}\right|^2 \geq 0.$$

# Another characterization of logical entropy: II

- Moreover, we thus have a new characterization of the logical entropy as the sum of all the principal minors of order 2:

$$h(\rho) = \sum_{i \neq j} \begin{vmatrix} p_i & \rho_{ij} \\ \rho_{ji} & p_j \end{vmatrix}.$$

# Miscellany II

David Ellerman

UCR

March 2012

# Projective measurement and generalized "measurement": I

- The usual sort of measurement in QM determined by a Hermitian operator is given by a set of projection operators $\{P_m\}$ such that $\sum_m P_m = I$ and $P_m P_{m'} = P_m \delta_{mm'}$, and it is now given the retronym, *"projective" measurement*. The probability of getting the result $m$ is $\mathrm{tr}(P_m \rho)$ and the post-measurement state is $\rho_m = P_m \rho P_m / \mathrm{tr}(P_m \rho)$.
- There are other quantum operation called *generalized "measurements"* given by a set of "measurement" operators $\{M_m\}$ such that $\sum_m M_m^\dagger M_m = I$ (but no orthogonality condition). The probability of getting the result $m$ is $\mathrm{tr}\left[M_m^\dagger M_m \rho\right]$ and the post-measurement state is $\rho_m = M_m \rho M_m^\dagger / \mathrm{tr}\left[M_m^\dagger M_m \rho\right]$.

# Projective measurement and generalized "measurement": II

- For projective measurement, entropy increases (or remains the same) for both vN and logical entropy.
- But for so-called generalized "measurement", entropy might **decrease** for both types of entropy.

# Example of entropy-decreasing "measurement": I

- Let $M_1 = |0\rangle\langle 0|$ and $M_2 = |0\rangle\langle 1|$. Then as matrices in the basis $\{|0\rangle, |1\rangle\}$,

$$M_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } M_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

so that

$$M_1^\dagger M_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$
$$M_2^\dagger M_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

# Example of entropy-decreasing "measurement": II

and thus $\sum_m M_m^\dagger M_m = I$ as required. Then for any:

$$\rho = \begin{bmatrix} p_1 & \rho_{12} \\ \rho_{21} & p_2 \end{bmatrix}$$

the result of the generalized "measurement" is
$\hat{\rho} = M_1 \rho M_1^\dagger + M_2 \rho M_2^\dagger$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} p_1 & \rho_{12} \\ \rho_{21} & p_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} p_1 & \rho_{12} \\ \rho_{21} & p_2 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$
$$\begin{bmatrix} p_1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} p_2 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} p_1 + p_2 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

which is a pure state of entropy 0 with either notion of entropy.

# Example of entropy-decreasing "measurement": III

- Yet the initial state $\rho$ could be any state such as $\rho = I/2$ which has positive entropy $S(\rho) = \log 2 = 1$ or $h(\rho) = 1 - \frac{1}{2} = \frac{1}{2}$.
- Hence the so-called "generalized measurement" decreased entropy.
- Both notions of entropy are related to the notion of distinctions, and any notion of measurement worth the name is about making distinctions. Hence these general quantum operations called "generalized measurements" are not well-named and that has caused much confusion.

# Example of entropy-decreasing "measurement": IV

- Since $M_1^\dagger M_1 = P_1$ and $M_2^\dagger M_2 = P_2$ are both projection matrices, the resultant state of a projective measurement using those projections is with probability $\mathrm{tr}(P_1\rho) = p_1$ the state:

$$\rho_1 = P_1\rho P_1 / \mathrm{tr}(P_1\rho) = \frac{1}{p_1}\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\begin{bmatrix}p_1 & \rho_{12}\\\rho_{21} & p_2\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} =$$

$$\frac{1}{p_1}\begin{bmatrix}p_1 & 0\\0 & 0\end{bmatrix} = \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$$

and with probability $\mathrm{tr}(P_2\rho) = p_2$, the resultant state is:

$$\rho_2 = P_2\rho P_2 / \mathrm{tr}(P_2\rho) = \frac{1}{p_2}\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}p_1 & \rho_{12}\\\rho_{21} & p_2\end{bmatrix}\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix} =$$

$$\frac{1}{p_2}\begin{bmatrix}0 & 0\\0 & p_2\end{bmatrix} = \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}.$$

# Example of entropy-decreasing "measurement": V

- Hence the total resulting state is:

$$\hat{\rho} = p_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + p_2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix}.$$

- For the probability distribution $p = \{p_1, p_2\}$,

$$H(p) = S(\hat{\rho}) = \sum_i p_i \log\left(\frac{1}{p_i}\right) \geq S(\rho), \text{ and}$$

$$h(p) = h(\hat{\rho}) = 1 - \sum_i p_i^2 = 2p_1 p_2 \geq h(\rho) = 2\left[p_1 p_2 - |\rho_{12}|^2\right].$$

- Both type of entropy increase ("increase" always includes staying the same) under projective measurements.

# Example of entropy-decreasing "measurement": VI

- Using the Church of the Larger Hilbert Space, the generalized "measurement" is turned into a projective measurement by embedding it in a larger Hilbert space.
- In our example, take the qubit space to be $H_A$ and tensor it with $H_B$ with the basis $\{|0_B\rangle, |1_B\rangle\}$.
- Then without going through all the details, the two results of the generalized measurement are now "marked" by the ancillia basis to become orthogonal so the measurement becomes projective.
- That is, the results of the measurement are:

$$\begin{bmatrix} p_1 & 0 \\ 0 & 0 \end{bmatrix} \otimes |0_B\rangle\langle 0_B| \text{ and } \begin{bmatrix} p_2 & 0 \\ 0 & 0 \end{bmatrix} \otimes |1_B\rangle\langle 1_B|$$

# Example of entropy-decreasing "measurement": VII

so the resultant state in the larger Hilbert space is:

$$\hat{\rho} = \begin{bmatrix} p_1 & 0 & 0 & 0 \\ 0 & p_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

which is a *mixed* state.

# Example of entropy-decreasing "measurement": VIII

- Moreover the entropy is now $h(\hat{\rho}) = 1 - \sum_i p_i^2 = 2p_1p_2$ whereas the entropy before the embedding and measurement was $h(\rho) = 2\left[p_1p_2 - |\rho_{12}|^2\right]$ so the entropy now *increased* by the amount $2|\rho_{12}|^2$, as in the case when a projective measurement was made in the first place without going through the embedding in the larger Hilbert space.

# Projective measurement decreases logical divergence: I

- The logical entropy of each state and the logical divergence between states stays constant under unitary transformation, so the question is what happens when a projective measurement takes place so that

$$\rho \to \hat{\rho} = \sum_m P_m \rho P_m \text{ and } \sigma \to \hat{\sigma} = \sum_m P_m \sigma P_m .$$

- We have already seen that projective measurement increases logical entropy (as well as vN entropy), i.e.,

$$h(\rho) \leq h(\hat{\rho}) \text{ under projective measurement.}$$

# Projective measurement decreases logical divergence: II

- Hence the question is: what happens to the logical divergence between states under projective measurement?

**Theorem**

*For projective measurement:* $d\left(\hat{\rho}||\hat{\sigma}\right) \leq d\left(\rho||\sigma\right)$.

Proof: $(\hat{\rho}-\hat{\sigma})^2 = \left(\sum_m P_m\rho P_m - \sum_m P_m\sigma P_m\right)^2$
$= \left(\sum_m P_m(\rho-\sigma)P_m\right)^2$
$= \sum_{m,m'}\left[P_m\left(\rho-\sigma\right)P_m\right]\left[P_{m'}\left(\rho-\sigma\right)P_{m'}\right]$
$= \sum_m P_m\left(\rho-\sigma\right)P_mP_m\left(\rho-\sigma\right)P_m$
$= \sum_m P_m\left(\rho-\sigma\right)P_m\left(\rho-\sigma\right)P_m$.
Hence $d\left(\hat{\rho}||\hat{\sigma}\right) = \operatorname{tr}\left[(\hat{\rho}-\hat{\sigma})^2\right]$

# Projective measurement decreases logical divergence: III

$$
\begin{aligned}
&= \operatorname{tr}\left[\sum_m P_m(\rho-\sigma) P_m(\rho-\sigma) P_m\right] \\
&= \sum_m \operatorname{tr}\left[P_m(\rho-\sigma) P_m(\rho-\sigma) P_m\right] \\
&\leq \sum_m \operatorname{tr}\left[P_m(\rho-\sigma)(\rho-\sigma) P_m\right] \\
&= \sum_m \operatorname{tr}\left[P_m P_m(\rho-\sigma)^2\right] \\
&= \operatorname{tr}\left[\sum_m P_m(\rho-\sigma)^2\right] \\
&= \operatorname{tr}\left[(\rho-\sigma)^2\right] \\
&= d(\rho||\sigma). \ \square
\end{aligned}
$$

# vN quantum relative entropy: I

- The vN quantum relative entropy $S(\rho||\sigma)$ seems to have a similar role to the quantum logical divergence $d(\rho||\sigma)$ in that both satisfy the basic inequalities, $S(\rho||\sigma) \geq 0$ and $d(\rho||\sigma) \geq 0$ with equality iff $\rho = \sigma$.
- Since $\sqrt{d(\rho||\sigma)}$ is a metric and $d(\rho||\sigma)$ decreases under projective measurement, it is natural to ask if $S(\rho||\sigma)$ has the same properties.
- Firstly, $S(\rho||\sigma)$ is not symmetric so $\sqrt{S(\rho||\sigma)}$ fails to be a metric for that simple reason.
- Hence we symmetrize it: $S^*(\rho||\sigma) = \frac{1}{2}[S(\rho||\sigma) + S(\sigma||\rho)]$ and then ask the same question about the symmetric version.

# vN quantum relative entropy: II

- At his point, I have neither proofs nor counterexamples that $\sqrt{S^*(\rho||\sigma)}$ is a metric or that $S(\rho||\sigma)$ or $S^*(\rho||\sigma)$ decreases under projective measurements, but I suspect that if those were true, then surely N&C would have mentioned it.

# Mixed States Entropy Formula

David Ellerman

UCR

March 2012

# Basic mixed state entropy formula: I

- In general, any mixed state $\rho$ can be expressed as the probability mixture of mixed states:

$$\rho = \sum_k q_k \rho_k$$

General mixed state representation

- Any $\rho$ can also be expressed as the probability mixture of pure states

$$\rho = \sum_k q_k \left|\psi_k\right\rangle \left\langle\psi_k\right|$$

Pure state representation

# Basic mixed state entropy formula: II

- One particular way to express a mixed state as a probability sum of pure states is its orthogonal decomposition where the probabilities $p_i$ are the real non-negative eigenvalues of $\rho$.

$$\rho = \sum_i p_i \left|i\right\rangle \left\langle i\right|$$

Orthogonal pure state representation

- Section 11.3.6 (p. 518) of N&C is entitled: "The entropy of a mixture of quantum states" where they refer to vN entropy. For a general mixed state $\rho = \sum_k q_i \rho_k$, the best relation there seems to be for vN entropy is the inequality:

$$\sum_k q_k S(\rho_k) \leq S\left(\sum_k q_k \rho_k\right) \leq \sum_k q_k S(\rho_k) + H(q_i).$$

# Basic mixed state entropy formula: III

- For logical entropy, however, there is mixed state master *equation* that gives the entropy $h(\rho)$ of a mixed state as the *same* mixture of the entropies "within" the states plus the weighted average of the divergences "between" the $\rho_k$ states.

**Theorem (Entropies and divergence form)**

*Given any representation $\rho = \sum_k q_k \rho_k$, the logical entropy of $\rho$ is:*

$$h\left(\sum_k q_k \rho_k\right) = \sum_k q_k h(\rho_k) + \frac{1}{2}\sum_{j,k} q_j q_k d\left(\rho_j || \rho_k\right).$$

Proof:
$h(\rho) = h\left(\sum_k q_k \rho_k\right)$

# Basic mixed state entropy formula: IV

$= 1 - \mathrm{tr}\left[\left(\sum_k q_k \rho_k\right)^2\right]$

$= 1 - \mathrm{tr}\left[\sum_k q_k^2 \rho_k^2 + 2\sum_{j<k} q_j q_k \rho_j \rho_k\right]$

$= 1 - \sum_k q_k^2 \,\mathrm{tr}\left(\rho_k^2\right) - 2\sum_{j<k} q_j q_k \,\mathrm{tr}\left(\rho_j \rho_k\right)$

$= \left[\sum_k q_k^2 + 2\sum_{j<k} q_j q_k\right] - \sum_k q_k^2 \,\mathrm{tr}\left(\rho_k^2\right) - 2\sum_{j<k} q_j q_k \,\mathrm{tr}\left(\rho_j \rho_k\right)$

$= \sum_k q_k^2 \left[1 - \mathrm{tr}\left(\rho_k^2\right)\right] + 2\sum_{j<k} q_j q_k \left[1 - \mathrm{tr}\left(\rho_j \rho_k\right)\right]$

$= \sum_k q_k^2 h\left(\rho_k\right) + 2\sum_{j<k} q_j q_k h\left(\rho_j || \rho_k\right)$ [this is used later in the cross-entropies version]

$= \sum_k q_k q_k h\left(\rho_k\right) + \sum_{j<k} q_j q_k 2h\left(\rho_j || \rho_k\right)$

$= \sum_k q_k q_k h\left(\rho_k\right) + \sum_{j<k} q_j q_k \left[d\left(\rho_j || \rho_k\right) + h\left(\rho_j\right) + h\left(\rho_k\right)\right]$

# Basic mixed state entropy formula: V

$= \sum_k q_k q_k h(\rho_k) + \sum_{j<k} q_j q_k d\left(\rho_j || \rho_k\right) + \sum_{j<k} q_j q_k h\left(\rho_j\right) + \sum_{j<k} q_j q_k h(\rho_k)$

$= \sum_k q_k q_k h(\rho_k) + \sum_{j<k} q_j q_k d\left(\rho_j || \rho_k\right) + \sum_{k<j} q_j q_k h(\rho_k) + \sum_{j<k} q_j q_k h(\rho_k)$

$= \sum_{k,j} q_k q_j h(\rho_k) + \sum_{j<k} q_j q_k d\left(\rho_j || \rho_k\right)$

$= \sum_k q_k h(\rho_k) + \sum_{j<k} q_j q_k d\left(\rho_j || \rho_k\right)$

$= \sum_k q_k h(\rho_k) + \frac{1}{2} \sum_{j,k} q_j q_k d\left(\rho_j || \rho_k\right) . \square$

# Basic mixed state entropy formula: VI

## Remark

- *There is some tension here between defining "divergence" as I did,* $d\left(\rho_j||\rho_k\right) = \mathrm{tr}\left[\left(\rho_j - \rho_k\right)^2\right]$ *so it is the Euclidean distance squared or defining it as half that so that the* $\frac{1}{2}$ *can be left out of the above formula.*
- *These formulas were derived in biostatistics with great generality in: Rao, C. R. 1982. Diversity and Dissimilarity Coefficients: A Unified Approach. Theoretical Population Biology. 21: 24-43. Rao calls (a more general form of) the logical entropy, the "diversity coefficient" and he uses half-divergence as the "dissimilarity coefficient." This reinforces the theme that "entropy" is about distinctions, differences, and decoherence which in biostatistics are diversity and dissimilarity.*

# Other mixed state entropy formulas: I

## Corollary (Cross-entropy form)

*Given any representation* $\rho = \sum_k q_k \rho_k$, *the logical entropy of* $\rho$ *is:*

$$h(\rho) = \sum_k q_k^2 h(\rho_k) + \sum_{j \neq k} q_j q_k h\left(\rho_j || \rho_k\right) = \sum_{j,k} q_j q_k h\left(\rho_j || \rho_k\right).$$

## Proof.

Picking up at a line in the above proof:

$h(\rho) = \sum_k q_k^2 h(\rho_k) + 2 \sum_{j<k} q_j q_k h\left(\rho_j || \rho_k\right)$

$= \sum_k q_k^2 h(\rho_k) + \sum_{j \neq k} q_j q_k h\left(\rho_j || \rho_k\right)$

$= \sum_{j,k} q_j q_k h\left(\rho_j || \rho_k\right)$. $\square$

# Other mixed state entropy formulas: II

- The cross-entropy version of the master formula is formally like the ANOVA formula that gives the variance of a weighted set of populations as a weighted average of the variances "within" each population plus a weighted average of the covariances "between" the populations. The two formulas are:

$$
\begin{aligned}
h\left(\sum_k q_k \rho_k\right) &= \sum_k q_k^2 h\left(\rho_k\right) + \sum_{j \neq k} q_j q_k h\left(\rho_j || \rho_k\right) \\
Var\left(\sum_i a_i X_i\right) &= \sum_i a_i^2 Var\left(X_i\right) + \sum_{i \neq j} a_i a_j Cov\left(X_i, X_j\right).
\end{aligned}
$$

# Other mixed state entropy formulas: III

- Since $\operatorname{tr}\left(\rho^2\right)$ is just the complement of cross-entropy, i.e., $\operatorname{tr}\left(\rho_j\rho_j\right) = 1 - h\left(\rho_j||\rho_j\right)$, we immediately have a trace version of the formula (which is easy to prove directly).

## Corollary (Trace form)

*Given any representation* $\rho = \sum_k q_k\rho_k$*:*

$$\operatorname{tr}\left(\rho^2\right) = \sum_{j,k} q_j q_k \operatorname{tr}\left(\rho_j\rho_k\right).$$

- Since $\operatorname{tr}\left[\rho\sigma\right]$ and $h\left(\rho||\sigma\right)$ are complements, one might ask: "Why not just work with trace formulas rather than (logical) entropy formulas?" The answer is that we are trying to develop a theory of quantum *information* where:

# Other mixed state entropy formulas: IV

- *Information* is about: distinctions, discernibility, distinguishability, discrimination, diversity, dissimilarity, divergence, decoherence, and the like.
- There could be a complementary trace-based theory about "lack-of-information" or ignorance where:
- *"Ignorance"* is about indistinction, indiscernibility, indistinguishability, lack-of-diversity, similarity, convergence, coherence, and the like .
- The same choice occurred in the development of partition logic where one could work with partition relations or their complements, equivalence relations. To develop the analogies with ordinary logic, it was key to work with partition relations, although all formulas have a dual form in terms of equivalence relations.

# Special cases: I

- Each special type of representation of a mixed state can be plugged into the general formula to derive a special formula.
- Any mixed state can be expressed as a mixture of *pure* states (in many different ways).

## Corollary (Mixture of pure states)

*Given any representation $\rho = \sum_k q_k |\psi_k\rangle \langle\psi_k|$ in terms of pure states, the logical entropy of $\rho$ is:*

$$h(\rho) = \sum_{j,k} q_j q_k \left[1 - \left|\left\langle \psi_j | \psi_k \right\rangle\right|^2\right]$$

*(where $\left|\left\langle \psi_j | \psi_k \right\rangle\right|^2$ can be interpreted as an "indistinction probability").*

**Proof.**

Using the cross-entropy form of the theorem:

$$h(\rho) = \sum_{j,k} q_j q_k h\left(\left|\psi_j\right\rangle\left\langle\psi_j\right| || \left|\psi_k\right\rangle\left\langle\psi_k\right|\right)$$

$$= \sum_{j,k} q_j q_k \left[1 - \operatorname{tr}\left(\left|\psi_j\right\rangle\left\langle\psi_j\right| \left|\psi_k\right\rangle\left\langle\psi_k\right|\right)\right]$$ where

$$\operatorname{tr}\left(\left|\psi_j\right\rangle\left\langle\psi_j\right| \left|\psi_k\right\rangle\left\langle\psi_k\right|\right) = \left\langle\psi_j|\psi_k\right\rangle \operatorname{tr}\left[\left|\psi_j\right\rangle\left\langle\psi_k\right|\right] = \left|\left\langle\psi_j|\psi_k\right\rangle\right|^2.$$

□

## Corollary (Trace form for pure states)

*Given any representation* $\rho = \sum_k q_k |\psi_k\rangle \langle\psi_k|$ *in terms of pure states:*

$$\operatorname{tr}\left(\rho^2\right) = 1 - \sum_{j \neq k} q_j q_k \left[1 - \left|\left\langle\psi_j|\psi_k\right\rangle\right|^2\right].$$

## Remark

*Does* $\operatorname{tr}\left(\rho^2\right) = 1$ *imply that* $\rho$ *is a pure state? By the above formula, since all the* $q_j > 0$ *and* $0 \leq \left|\left\langle \psi_j | \psi_k \right\rangle\right|^2 \leq 1$*,* $\operatorname{tr}\left(\rho^2\right) = 1$ *implies* $\left|\left\langle \psi_j | \psi_k \right\rangle\right|^2 = 1$ *for all* $\psi_j$ *and* $\psi_k$*. This means the* $|\psi_k\rangle$ *are vectors that differ at most in an absolute phase factor* $e^{i\varphi_k}$ *(different* $\varphi_k$ *for different* $k$*). Then any* $\rho = \sum_k q_k |\psi_k\rangle\langle\psi_k|$ *has* $\left|\left\langle \psi_j | \psi_k \right\rangle\right|^2 = 1$ *for any* $j, k$ *so that* $\operatorname{tr}\left(\rho^2\right) = 1$*. But all the projection matrices* $|\psi_k\rangle\langle\psi_k|$ *are the same* $|\psi\rangle\langle\psi|$ *(since the phases cancel out) so that* $\rho = \sum_k q_k |\psi_k\rangle\langle\psi_k| = |\psi\rangle\langle\psi|$ *which is a pure state.*

# Special cases: VI

- It should be recalled that there is no connection between the dimension of the Hilbert space and the number of $\rho_k$'s involved in a representation $\rho = \sum_k q_k \rho_k$ using mixed states or $\rho = \sum_k q_k |\psi_k\rangle \langle\psi_k|$ using pure states.
- Another special case is the orthogonal decomposition $\rho = \sum_i p_i |i\rangle \langle i|$ where the states $|i\rangle$ are orthonormal. Then the entropy formula is: $h(\rho) = \sum_{i,j} p_i p_j \left[1 - |\langle i|j\rangle|^2\right]$ where $\langle i|j\rangle = \delta_{ij}$ so:

$$h(\rho) = \sum_{i,j} p_i p_j \left[1 - \delta_{ij}\right] = \sum_{i \neq j} p_i p_j = 1 - \sum_i p_i^2.$$

Entropy of orthogonal decomposition is classical logical entropy

### Remark

*We have not discussed logical entropy in the continuous case, but the formula* $h(\rho) = \sum_{i,j} p_i p_j \left[1 - \delta_{ij}\right]$ *indicates one approach. Given a continuous probability distribution* $P(x)$*, the logical entropy of the probability distribution is:*
$h(P) = \int (1 - \delta(x_1 - x_2)) P(x_1) P(x_2) \, dx_1 dx_2$ *where* $\delta(x_1 - x_2)$ *is the Dirac delta function. But* $h(P) = 1 - \int P(x)^2 \, dx$ *is much simpler.*

- Another special case is the Schmidt decomposition of a pure state $|\psi\rangle$ on $H^A \otimes H^B$ which is: $|\psi\rangle = \sum_i \sqrt{p_i} \, |i_A\rangle \otimes |i_B\rangle$ where the $i_A$ and $i_B$ are orthonormal states of the two systems $H_A$ and $H_B$. Then the reduced density matrices on $H^A$ and $H^B$ are:

$\rho^A = \sum_i p_i \left| i_A \right\rangle \left\langle i_A \right|$ and $\rho^B = \sum_i p_i \left| i_B \right\rangle \left\langle i_B \right|$.

- Now we can apply the formula:

$$h\left(\rho^A\right) = \sum_{j \neq k} \left[ p_j p_k - p_j p_k \left| \langle j_A | k_A \rangle \right|^2 \right] = \sum_{j \neq k} p_j p_k = 1 - p_i^2.$$

- Since the result depends on the $p_i$'s, the logical entropy of $\rho^B$ is the same.
- If $|\psi\rangle$ is not only a pure state but a separated (or product) state, then $|\psi\rangle = |i_A\rangle \otimes |i_B\rangle$, $p_1 = 1$, and $h\left(\rho^A\right) = 0 = h\left(\rho^B\right)$. Hence:

$|\psi\rangle$ is entangled iff $h\left(\rho^A\right) = h\left(\rho^B\right) > 0$
(similarly for vN entropy).

- The joint entropy theorem for vN entropy was proven in the N&C book (p. 513). The theorem for logical entropy was previously stated but not proven. It can be proven by adapting the proof for vN entropy that uses the orthogonal decomposition but it can also be derived using the mixed state entropy formula.

## Lemma

*If $p = \{p_i\}$ is a probability distribution and the states $\rho_i$ have orthogonal support, then:*

$$h\left(\sum_j p_j \rho_j\right) = h(p) + \sum_j p_j^2 h\left(\rho_j\right).$$

**Proof.**

Using the cross-entropy version of the formula:

$$h(\rho) = \sum_{j,k} p_j p_k h\left(\rho_j || \rho_k\right) = \sum_{j,k} p_j p_k - \sum_{j,k} p_j p_k \operatorname{tr}\left(\rho_j \rho_k\right)$$
$$= \sum_j p_j^2 + \sum_{j \neq k} p_j p_k - \sum_j p_j^2 \operatorname{tr}\left(\rho_j^2\right) = h(p) + \sum_j p_j^2 \left[1 - \operatorname{tr}\left(\rho_j^2\right)\right]$$
$$= h(p) + \sum_j p_j^2 h\left(\rho_j\right). \qquad \square$$

**Theorem (Joint entropy theorem)**

*Suppose* $p = \{p_i\}$ *are probabilities,* $|i\rangle$ *are orthogonal states for system A, and* $\{\rho_i\}$ *are any set of density operators for system B. If* $\rho = \sum_i p_i |i\rangle \langle i| \otimes \rho_i$*, then:*

$$h\left(\sum_i p_i |i\rangle \langle i| \otimes \rho_i\right) = h(p) + \sum_i p_i^2 h(\rho_i).$$

# Joint entropy theorem: IV

**Proof.**

The states $|i\rangle\langle i| \otimes \rho_i$ have orthogonal support so the lemma gives $h(\rho) = h(p) + \sum_i p_i^2 h(|i\rangle\langle i| \otimes \rho_i)$. It was previous shown that for any states $\sigma, \tau$, $h(\sigma \otimes \tau) = h(\sigma) + h(\tau) - h(\sigma) h(\tau)$ and $|i\rangle\langle i|$ is a pure state so $h(|i\rangle\langle i|) = 0$ and thus $h(|i\rangle\langle i| \otimes \rho_i) = h(\rho_i)$. $\square$

# End of seminar

For other writings on quantum mechanics, see my website:
www.ellerman.org/category/qm/