# PHYSICS 491: Symmetry and Quantum Information

Michael Walter, Stanford University

Spring 2017

## Abstract

This course gives an introduction to quantum information theory, targeted to advanced undergraduates and graduate students that have taken a quantum mechanics class on the level of Physics 230. We use symmetries as a guiding principle to study the fundamental features of quantum mechanics and their exploitation for quantum information processing tasks.

## Acknowledgements

I would like to thank Sepehr Nezami for his careful proofreading of the problem sets. Thanks also to Prakash Mudholka, Grant Salton and Zhaoyou Wang for pointing out typos in these lecture notes.

Last updated: November 12, 2018.

# Contents

**PHYSICS 491: Symmetry and Quantum Information** April 4, 2017

# Quantum correlations, non-local games, rigidity

*Lecture 1* *Michael Walter, Stanford University*

*These lecture notes are not proof-read and are offered for your convenience only. They include additional detail and references to supplementary reading material. I would be grateful if you email me about any mistakes and typos that you find.*

Quantum mechanics can seem quite strange at times! We have phenomena such as superpositions ($|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$), entanglement ($|\phi\rangle_{AB} \neq |\phi\rangle_A \otimes |\phi\rangle_B$), incompatible measurements ($[X, Y] \neq 0$), etc. This "strangeness" manifests itself through the *correlations* predicted by quantum mechanics. A modern perspective of studying and comparing correlations is through the notions of a *nonlocal game*. You have met nonlocal games already in Physics 230, but we will discuss some interesting new aspects that you may not have seen before.

## 1.1 Nonlocal games

In a *nonlocal game*, we imagine that a number of *players* play against a *referee*. The referee hands them *questions* and the players reply with appropriate *answers* that win them the game. The players' goal is to collaborate and maximize their chances of winning. Before the game, the players meet and may agree upon a joint strategy – but then they move far apart from each other and cannot communicate with each other while the game is being played (this can be ensured by the laws of special relativity). The point then is the following: *Since the players are constrained by the laws of physics, we can concoct games where players utilizing a quantum strategy may have an advantage.* This way of reasoning about quantum correlations is eminently operational and quantitative, as we will see in the following.

The *GHZ (Greenberger-Horne-Zeilinger) game* is a famous example of a nonlocal game due to Mermin (1990); cf. Greenberger et al. (1990). Figure 1 illustrates the setup of the GHZ game. It involves three players – Alice, Bob, and Charlie. Each receives as questions a bit $x, y, z \in \{0, 1\}$ and their answers are likewise bits $a, b, c \in \{0, 1\}$. They win the game if the sum of their answers modulo 2 is as follows:

| $x$ | $y$ | $z$ | $a \oplus b \oplus c$ |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 1 |
| 1 | 0 | 1 | 1 |
| 0 | 1 | 1 | 1 |

Note that not all bit strings $xyz$ are questions that the referee asks. The winning condition can be succintly stated as follows: $a \oplus b \oplus c = x \vee y \vee z$. We write $\oplus$ for addition modulo 2 and $\vee$ for the logical OR. Those of you that have taken the Physics 230 final are already familiar with the rules of this game.

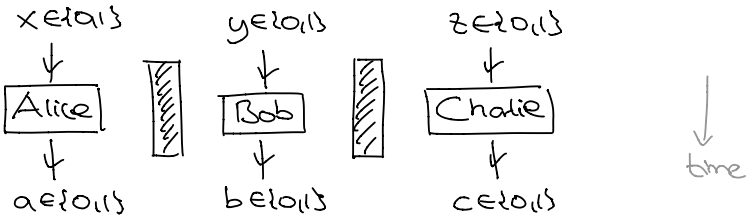



Figure 1: Setup of the three-player GHZ game. The winning condition is that $a \oplus b \oplus c = x \vee y \vee z$.

## Classical strategies

It is easy to see that the GHZ game cannot be won if the players' strategies are described by a "local" and "realistic" theory. As in Physics 230, "local" means that each player's answer does only depend on its immediate surroundings, and "realistic" means that the theory must assign a pre-existing value to every possible measurement before the measurement is made. In our case, "measurements" correspond to "questions" and "outcomes" to "answers". Thus in a local and realistic theory we assume that

$$a = a(x), \quad b = b(y), \quad c = c(z).$$

When we say that the players may jointly agree on a strategy before the game is being played, we mean that they may select "question-answer functions" $a$, $b$, $c$ in a correlated way. For example, when the players meet before the game is being played, they could flip a coin, resulting in some random $\lambda \in \{0, 1\}$, and agree on the strategy $a(x) = x \oplus \lambda$, $b(y) = y \oplus \lambda$, $c(z) = z \oplus \lambda$. Thus, in mathematical terms, the functions $a$,$b$,$c$ can be correlated random variables. Equivalently, we could say that $\lambda$ is a "hidden variable", with some probability distribution $p_\lambda(0) = p_\lambda(1) = 1/2$, and consider $a = a(x, \lambda)$ as a deterministic function of both the input and the hidden variable. You will discuss this point of view in problem 1.1. If the players strategy can be described by classical mechanics then the above would provide an adequate model. Thus, strategies of this form are usually referred to as *local hidden variable strategies* or simply as *classical strategies*.

Suppose now for sake of finding a contradiction that Alice, Bob, and Charlie can win the GHZ game perfectly. Then,

$$\begin{aligned} 1 &= 0 \oplus 1 \oplus 1 \oplus 1 \\ &= (a(0) \oplus b(0) \oplus c(0)) \oplus (a(1) \oplus b(1) \oplus c(0)) \oplus (a(1) \oplus b(0) \oplus c(1)) \oplus (a(0) \oplus b(1) \oplus c(1)) = 0. \end{aligned}$$

The last equality holds because $a(x) \oplus a(x) \equiv 0$ etc., whatever the value of $a(x)$. This is a contradiction! We conclude that there is no perfect classical winning variable strategy for the GHZ game. Suppose, e.g., that the referee selects each possible question $xyz$ with equal probability 1/4. Then the game can be won with probability at most

$$p_{\text{win,cl}} = 3/4.$$

This winning probability can be achieved by, e.g., the trivial strategy $a(x) = b(y) = c(z) \equiv 1$.

## Quantum strategies

In a *quantum strategy*, we imagine that the three players are described by quantum mechanics. Thus they start out by sharing an arbitrary joint state $|\psi\rangle_{ABC} \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$, where $\mathcal{H}_A$ is the Hilbert

space describing a quantum system in Alice' possession, etc., and upon receiving their questions $x, y, z \in \{0,1\}$ they will measure corresponding observables $A_x$, $B_y$, $C_z$ on their respective Hilbert spaces. While it might not be immediately obvious, any classical strategy is also a quantum strategy, as you will show in problem 1.1.

It will be convenient to take the eigenvalues (i.e., measurement outcomes) of the observables to be in $\{\pm 1\}$ rather than in $\{0,1\}$. Provided the outcome of Alice's measurement of $A_x$ is $(-1)^a$, she sends back $a$ as the answer, etc. In this case, the eigenvalues of $A_x \otimes B_y \otimes C_z$ are $(-1)^{a+b+c} = (-1)^{a \oplus b \oplus c}$, that is, they correspond precisely to the the sum modulo two of the answers. Thus, a perfect quantum strategy is one where

$$\begin{aligned}
(A_0 \otimes B_0 \otimes C_0)\,|\psi\rangle_{ABC} &= +\,|\psi\rangle_{ABC}\,,\\
(A_1 \otimes B_1 \otimes C_0)\,|\psi\rangle_{ABC} &= -\,|\psi\rangle_{ABC}\,,\\
(A_1 \otimes B_0 \otimes C_1)\,|\psi\rangle_{ABC} &= -\,|\psi\rangle_{ABC}\,,\\
(A_0 \otimes B_1 \otimes C_1)\,|\psi\rangle_{ABC} &= -\,|\psi\rangle_{ABC}\,,
\end{aligned} \tag{1.1}$$

In problem 1.1 you will verify that, more generally,

$$p_{\text{win},q} = \frac{1}{2} + \frac{1}{8}\,\langle\psi_{ABC}|A_0 \otimes B_0 \otimes C_0 - A_1 \otimes B_1 \otimes C_0 - A_1 \otimes B_0 \otimes C_1 - A_0 \otimes B_1 \otimes C_1|\psi_{ABC}\rangle$$

is the probability of winning the GHZ game (for uniform choice of questions $xyz$).

Remarkably, there is a quantum strategy for the GHZ game that allows the players to win the game *every single time* (i.e., $p_{\text{win,q}} = 1$). Following Watrous (2006), the players share the three-qubit state

$$|\Gamma\rangle_{ABC} = \frac{1}{2}\,(|000\rangle - |110\rangle - |101\rangle - |011\rangle) \in \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2, \tag{1.2}$$

where we imagine that the first qubit is in Alice's possession, the second in Bob's, and the third in Charlie's. Upon receiving $x = 0$, Alice measures the observable $A_0 = Z = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ on her qubit, while upon receiving $x = 1$ she measures the observable $A_1 = X = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Bob and Charlie perform exactly the same strategy on their qubits. To see that this quantum strategy wins the GHZ game every single time, we only need to verify (1.1). Indeed:

$$\begin{aligned}
(Z \otimes Z \otimes Z)\,|\Gamma\rangle_{ABC} &= |\Gamma\rangle_{ABC}\,,\\
(X \otimes X \otimes Z)\,|\Gamma\rangle_{ABC} &= \frac{1}{2}\,(|110\rangle - |000\rangle - (-1)\,|011\rangle - (-1)\,|101\rangle) = -\,|\Gamma\rangle_{ABC}\,,
\end{aligned}$$

and similarly $(X \otimes Z \otimes X)\,|\Gamma\rangle_{ABC} = (Z \otimes X \otimes X)\,|\Gamma\rangle_{ABC} = -\,|\Gamma\rangle_{ABC}$.

This shows that in a precise quantitative sense, quantum mechanics enables much stronger "non-local correlations" than what is possible using a local realistic theory.

**Exercise.** *This looks different from what you remember from the Physics 230 exam! It is a fun exercise to relate the strategy above to the one you remember from the Physics 230 exam.*

## Device-independent quantum cryptography

When the three players perform the optimal strategy described above then not only do their answers satisfy the winning condition but their answers are in fact completely *random*, subject only to the constraint that $a \oplus b \oplus c$ must sum to the desired value $x \vee y \vee z$. In particular, $a, b \in \{0,1\}$ are two independent random bits. You can easily verify this by inspection: E.g., for $x = y = z = 0$,

Alice, Bob, and Charlie each measure their local $Z$ observable. The eigenvectors are $|abc\rangle$ and so it is clear from eq. (1.2) that we obtain $abc \in \{000, 110, 101, 011\}$ with equal probability 1/4. The randomness obtained in this way is also *private* in the following sense: Suppose that apart from Alice, Bob, Charlie, there is also an evil eavesdropper (Evan) who would like to learn about the random bits generated in this way. Their joint state will be described by a pure state $|\psi\rangle_{ABCE}$ (we may assume that this is a pure state – just hand all other systems to the eavesdropper; this will only give him more power). If Alice, Bob, and Charlie indeed share the state in eq. (1.2) (or for that matter *any* pure state) then it must be the case that $|\psi\rangle_{ABCE} = |\Gamma\rangle_{ABC} \otimes |\psi\rangle_E$. You will show this in problem 2.2. This means that Evan is completely decoupled from Alice, Bob and Charlie's state, and it follows that the random bits $a$ and $b$ are completely uncorrelated from the $E$ system. All these means that the players' answers can be used to generate private randomess – the referee simply locks Alice, Bob, and Charlie (best thought of as quantum devices) into his laboratory, ensures that they cannot communicate, and interrogates them with questions. But of course, the referee cannot in general trust Alice, Bob, and Charlie to actually play the strategy above! So this observation might seem not very useful at first glance...

*However, what if the optimal strategy for winning the GHZ game was actually unique?* In this case, the referee could test Alice, Bob, and Charlie with randomly selected questions and check that they pass the test every time. After a while, the referee might be confident that the players are in fact able to win the GHZ game every time. But then, by uniqueness of the winning strategy, the referee should in fact know the precise strategy that Alice, Bob, and Charlie are pursuing! The referee in this case would *not* have to put any trust in Alice, Bob, Charlie – they would prove their worth by winning the GHZ game every time around. This remarkable idea for generating private random bits was first proposed by Colbeck (2009). (Note that we need private random bits in the first place to generate the random questions – thus this protocol proposes to achieve a task known as *randomness expansion*. Private random bits cannot be generated without an initial seed of random bits.) The argument sketched so far is of course not rigorous at all: ignoring questions of robustness, we need to take into account that Alice, Bob, Charlie may not behave the same way every time we play the game, may have a (quantum) memory, etc.

However, these challenges can be circumvented and secure randomness expansion protocols using completely untrusted devices do exist (see, e.g., Miller and Shi (2014) and the review Acín and Masanes (2016))! This general line of research is known as *device-independent quantum cryptography* (Mayers and Yao, 1998), since it does not rely on assumptions on the inner workings of the devices involved, but only on their observed correlations. Other applications of include device-independent quantum key distribution (Vazirani and Vidick, 2014) and the command of an adversarial quantum system (Reichardt et al., 2013).

## 1.2 Rigidity of the GHZ game

For the remainder of the lecture, we will content ourselves with showing that the winning strategy for the GHZ game is indeed essentially unique (Colbeck and Kent, 2011). We say that the GHZ game is *rigid* – or that it is a *self-test* for the state (1.2).

**Remark.** *The CHSH game which you might remember from Physics 230 is likewise rigid; see Tsirel'son (1987), Summers and Werner (1987), Popescu and Rohrlich (1992), McKague et al. (2012), Reichardt et al. (2013). (Here, optimal quantum winning probability is* $1/2 + 1/2\sqrt{2} < 1$*!) Robust rigidity results for general XOR games are contained in Slofstra (2011), Miller and Shi (2013),*

*Ostrev (2015). Rigidity is also closely related to the question of how much entanglement is needed to win a nonlocal game (e.g., Slofstra, 2011). Surveying some of these results would make for great (but challenging) course projects.*

To prove the rigidity result, we first observe that in the three-qubit strategy discussed above, the state $|\Gamma\rangle_{ABC}$ is already uniquely determined by the measurement operators: Indeed, any eigenvector of $Z\otimes Z\otimes Z$ is necessarily of the form $\alpha|000\rangle+\beta|110\rangle+\gamma|101\rangle+\delta|011\rangle$, and the other three conditions are only satisfied if $\alpha=-\beta=-\gamma=-\delta=1/2$, up to overall phase.

Let us now consider a general optimal strategy given by operators $A_x$, $B_y$, $C_z$ with $A_x^2=\mathbb{1}$ etc. and a state $|\psi\rangle_{ABC}\in\mathcal{H}_A\otimes\mathcal{H}_B\otimes\mathcal{H}_C$ such that eq. (1.1) are satisfied. The basic strategy to prove the rigidity theorem will be to uncover some *hidden symmetries* in the problem to reduce to the case of three qubits:

**Claim 1.1.** *In any optimal strategy, the observables must anticommute: "$\{A_0,A_1\}=0$, $\{B_0,B_1\}=0$, $\{C_0,C_1\}=0$" (see below for fine-print).*

We will prove this claim later, but let us see first see how this allows us to identify three qubits.

## How to find a qubit?

Consider, e.g., the pair of observables $A_0,A_1$. They satisfy $A_0^2=A_1^2=\mathbb{1}$ and $\{A_0,A_1\}=0$. Hence, $A_2=-\frac{i}{2}[A_0,A_1]=-iA_0A_1=iA_1A_0$ is such that

$$[A_1,A_2]=A_1A_2-A_2A_1=iA_1A_1A_0+iA_0A_1A_1=2iA_0,$$

and similarly $[A_2,A_0]=2iA_1$. This means that $A_0,A_1,A_2$ transform like the Pauli matrices $X,Y,Z$! It follows that the Hilbert space decomposes into irreducible representations of $SU(2)$:

$$\mathcal{H}_A=V_{j_1}\oplus V_{j_2}\oplus\cdots=\bigoplus_{j=0,1/2,1,\ldots}V_j\otimes\mathbb{C}^{m_j},$$

where $m_j$ counts the number of times the spin-$j$ representation $V_j$ appears in $\mathcal{H}_A$. We claim that, since $\{A_0,A_1\}=0$, this representation of $SU(2)$ has to be $j=1/2$! Indeed, $A_2^2=-iA_0A_1iA_1A_0=1$ and so

$$\frac{1}{4}(A_0^2+A_1^2+A_2^2)=\frac{3}{4}=\frac{1}{2}(\frac{1}{2}+1)$$

acts by a scalar. Comparing with $j(j+1)$ we find that $j=1/2$ (cf. remark 5.4).

Therefore, $\mathcal{H}_A\cong\mathbb{C}^2\otimes\mathcal{H}_{A'}$, where $\mathcal{H}_{A'}$ is some auxiliary Hilbert space of dimension $m_{1/2}$, and $A_0,A_1$ act by

$$Z\otimes\mathbb{1},X\otimes\mathbb{1}.$$

**Exercise.** *Can you find an argument that avoids using the representation theory of $SU(2)$?*

The same argument works for Bob and Charlie's pairs of observables. Thus the total Hilbert space decomposes as

$$\mathcal{H}_A\otimes\mathcal{H}_B\otimes\mathcal{H}_C\cong(\mathbb{C}^2\otimes\mathbb{C}^2\otimes\mathbb{C}^2)\otimes(\mathcal{H}_{A'}\otimes\mathcal{H}_{B'}\otimes\mathcal{H}_{C'})$$

and the measurement operators act as in the three-qubit solution. We saw above that in the three-qubit solution the state is uniquely determined by the measurement operators. Thus,

$$|\psi\rangle_{ABC}=|\Gamma\rangle\otimes|\gamma\rangle_{A'B'C'},$$

where $|\Gamma\rangle$ is the three-qubit state from eq. (1.2) and $|\gamma\rangle_{A'B'C'}$ some auxiliary state (which is irrelevant because the observables do not act on it). This is the desired rigidity result.

## Anticommutations from correlations (proof of the claim)

We first note that the optimality condition eq. (1.1) can be written as

$$\begin{aligned} A_0 |\psi\rangle &= +B_0 C_0 |\psi\rangle \\ A_0 |\psi\rangle &= -B_1 C_1 |\psi\rangle \\ A_1 |\psi\rangle &= -B_1 C_0 |\psi\rangle \\ A_1 |\psi\rangle &= -B_0 C_1 |\psi\rangle . \end{aligned}$$

Here and in the following we write $A_0$ instead of $A_0 \otimes \mathbb{1}_B \otimes \mathbb{1}_C$ to make the formulas more transparent. From the first two and last two equations, respectively,

$$\begin{aligned} A_0 |\psi\rangle &= +\frac{1}{2} (B_0 C_0 - B_1 C_1) |\psi\rangle \\ A_1 |\psi\rangle &= -\frac{1}{2} (B_1 C_0 + B_0 C_1) |\psi\rangle \end{aligned}$$

Hence,

$$\begin{aligned} A_0 A_1 |\psi\rangle &= -\frac{1}{4} (B_1 C_0 + B_0 C_1)(B_0 C_0 - B_1 C_1) |\psi\rangle = -\frac{1}{4} (B_1 B_0 - C_0 C_1 + C_1 C_0 - B_0 B_1) |\psi\rangle , \\ A_1 A_0 |\psi\rangle &= -\frac{1}{4} (B_0 C_0 - B_1 C_1)(B_1 C_0 + B_0 C_1) |\psi\rangle = -\frac{1}{4} (B_0 B_1 - C_1 C_0 + C_0 C_1 - B_1 B_0) |\psi\rangle \end{aligned}$$

and so

$$\{A_0, A_1\} |\psi\rangle = 0.$$

How can we show that $\{A_0, A_1\} = 0$?

This is in fact not exactly true – hence the "quotes" in claim 1.1. But what is true is that $\{A_0, A_1\} = 0$ on a subspace $\tilde{\mathcal{H}}_A$ of $\mathcal{H}_A$ such that $|\psi\rangle_{ABC} \in \tilde{\mathcal{H}}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$. Indeed, we can expand

$$|\psi\rangle_{ABC} = \sum_i s_i |e_i\rangle_A \otimes |f_i\rangle_{BC}$$

where the $|e_i\rangle$ and $|f_i\rangle$ are orthonormal and $s_i > 0$. If there are $\dim \tilde{\mathcal{H}}_A$ terms then the $|e_i\rangle$ form a complete basis of $\mathcal{H}_A$ and so $\{A_0, A_1\} |\psi\rangle = 0$ implies that $\{A_0, A_1\} = 0$. Otherwise, we can restrict to the subspace $\tilde{\mathcal{H}}_A := \mathrm{span}\{|e_i\rangle_A\}$ – this is called the *Schmidt decomposition* and we will discuss it in more detail in a future lecture. In the latter case, $|\psi\rangle_{ABC} \in \tilde{\mathcal{H}}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$, the operators $A_x$ are block diagonal with respect to $\tilde{\mathcal{H}}_A \oplus \tilde{\mathcal{H}}_a^{\perp}$, and $\{A_0, A_1\} = 0$ on $\tilde{\mathcal{H}}_A$. We can proceed likewise for $B_y$ and $C_z$.

## Statement of the rigidity theorem

What have we proved? In mathematical terms, we have established the following theorem:

**Theorem 1.2** (Rigidity for the GHZ game)**.** *Consider an optimal strategy for the GHZ game given by operators $A_x$, $B_y$, $C_z$ with $A_x^2 = \mathbb{1}_A$ etc. and a state $|\psi\rangle_{ABC} \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$. Then there exist isometries $V_A : \mathbb{C}^2 \otimes \mathcal{H}_{A'} \to \mathcal{H}_A$, $V_B : \mathbb{C}^2 \otimes \mathcal{H}_{B'} \to \mathcal{H}_B$, $V_C : \mathbb{C}^2 \otimes \mathcal{H}_{C'} \to \mathcal{H}_C$ such that*

*(i)* $|\psi\rangle_{ABC} = (V_A \otimes V_B \otimes V_C)(|\Gamma\rangle \otimes |\gamma\rangle)$ *for some* $|\gamma\rangle \in \mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}'C$.

*(ii)* $V_A^\dagger A_0 V_A = Z \otimes \mathbb{1}_{A'}$, $V_A^\dagger A_1 V_A = X \otimes \mathbb{1}_{A'}$, *and similarly for* $B_y$ *and* $C_z$.

In the coming lectures, we will revisit many of the techniques used above in a more systematic way. I would suggest that you come back to this lecture at the end of the term – at this point you should be well equipped to write up a complete proof of theorem 1.2.

### Outlook

There are many further aspects of nonlocal games related to what we discussed in this lecture. For example, how do winning probabilities and optimal strategies behave when one plays many instances of a game – either in multiple rounds (sequentially) or even at the same time (in parallel)? It is clear that if $p$ is the optimal winning probability for a single instance then for $n$ instances the winning probability is at least $p^n$ – but we might be able to do better by using strategies that exploit correlations or entanglement in a clever way! Indeed, the maximal classical winning probability for a single instance of the CHSH game is $3/4$ – while for two instances it is $10/16 > 9/16 = (3/4)^2$ (Barrett et al., 2002). On the other hand, it is proved in Cleve et al. (2007) not only for the CHSH game but for arbitrary *XOR games* (games where the winning condition only depends on the sum modulo two of the answers, $a \oplus b \oplus \dots$) that the optimal quantum winning probability for $n$ instances is equal to $p^n$ – this is known as a perfect *parallel repetition theorem*. Surveying some of the papers in this area could also make for a good course project.

## Bibliography


N. David Mermin. Quantum mysteries revisited, *American Journal of Physics*, 58(8):731–734, pages 731–734, 1990.

Daniel M. Greenberger, Michael A. Horne, Abner Shimony, and Anton Zeilinger. Bell's theorem without inequalities, *American Journal of Physics*, 58(12):1131–1143, pages 1131–1143, 1990.

John Watrous. Bell inequalities and nonlocality. 2006. URL `https://cs.uwaterloo.ca/~watrous/CPSC519/LectureNotes/20.pdf`.

Roger Colbeck. *Quantum and relativistic protocols for secure multi-party computation.* PhD thesis, 2009. arXiv:0911.3814.

Carl A. Miller and Yaoyun Shi. Universal security for randomness expansion from the spot-checking protocol. 2014. arXiv:1411.6608.

Antonio Acín and Lluis Masanes. Certified randomness in quantum physics, *Nature*, 540(7632): 213–219, pages 213–219, 2016.

Dominic Mayers and Andrew Yao. Quantum cryptography with imperfect apparatus. In *Foundations of Computer Science, 1998. Proceedings. 39th Annual Symposium on*, pages 503–509. IEEE, 1998. arXiv:quant-ph/9809039.

Umesh Vazirani and Thomas Vidick. Fully device-independent quantum key distribution, *Physical Review Letters*, 113(14):140501, page 140501, 2014.

Ben W. Reichardt, Falk Unger, and Umesh Vazirani. Classical command of quantum systems, *Nature*, 496(7446):456–460, pages 456–460, 2013. arXiv:1209.0448.

Roger Colbeck and Adrian Kent. Private randomness expansion with untrusted devices, *Journal of Physics A: Mathematical and Theoretical*, 44(9):095305, page 095305, 2011. arXiv:1011.4474.

Boris S. Tsirel'son. Quantum analogues of the Bell inequalities. the case of two spatially separated domains, *Journal of mathematical sciences*, 36(4):557–570, pages 557–570, 1987.

Stephen J. Summers and Reinhard Werner. Maximal violation of Bell's inequalities is generic in quantum field theory, *Communications in Mathematical Physics*, 110(2):247–259, pages 247–259, 1987.

Sandu Popescu and Daniel Rohrlich. Which states violate Bell's inequality maximally?, *Physics Letters A*, 169(6):411–414, pages 411–414, 1992.

Matthew McKague, Tzyh Haur Yang, and Valerio Scarani. Robust self-testing of the singlet, *Journal of Physics A: Mathematical and Theoretical*, 45(45):455304, page 455304, 2012.

William Slofstra. Lower bounds on the entanglement needed to play xor non-local games, *Journal of Mathematical Physics*, 52(10):102202, page 102202, 2011.

Carl A. Miller and Yaoyun Shi. Optimal robust self-testing by binary nonlocal XOR games. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 22. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2013. arXiv:1207.1819.

Dimiter Ostrev. The structure of optimal and nearly-optimal quantum strategies for non-local XOR games. 2015. arXiv:1506.00607.

Jonathan Barrett, Daniel Collins, Lucien Hardy, Adrian Kent, and Sandu Popescu. Quantum nonlocality, bell inequalities, and the memory loophole, *Physical Review A*, 66(4):042111, page 042111, 2002. arXiv:quant-ph/0205016.

Richard Cleve, William Slofstra, Falk Unger, and Sarvagya Upadhyay. Perfect parallel repetition theorem for quantum XOR proof systems. In *Computational Complexity, 2007. CCC'07. Twenty-Second Annual IEEE Conference on*, pages 109–114. IEEE, 2007. arXiv:quant-ph/0608146.

**PHYSICS 491: Symmetry and Quantum Information** April 6, 2017

# Measurements, symmetric subspace, pure state estimation

*Lecture 2* *Michael Walter, Stanford University*

*These lecture notes are not proof-read and are offered for your convenience only. They include additional detail and references to supplementary reading material. I would be grateful if you email me about any mistakes and typos that you find.*

Today we will talk about measurements in quantum mechanics and discuss the problem of estimating an unknown pure state.

## 2.1 Generalized measurements

From your quantum mechanics class you know that observable in quantum mechanics are modeled by Hermitian operators $X$. Let $X = \sum_{x\in\Omega} x P_x$ denote the spectral decomposition of an observable, i.e., $P_x$ denotes the projector onto the eigenspace corresponding to an eigenvalue $x \in \Omega$. Thus we can repackage $X$ in terms the collection of projections $P_x$, labeled by the possible measurement outcomes $x \in \Omega$. This is convenient for two reasons: First, the probability of outcome $x$ in state $|\psi\rangle$ is given by by the *Born rule*:

$$\Pr(\text{outcome } x) = \langle\psi|P_x|\psi\rangle\,, \tag{2.1}$$

which is naturally expressed in terms of the projections $P_x$. Second, this formalism allows us to consider more general sets of outcomes $\Omega$ that are not necessarily real numbers. Instead of using observables, we will therefore often prefer to work with the collection of operators $\{P_x\}_{x\in\Omega}$. We call $\{P_x\}_{x\in\Omega}$ a *projective measurement*. Mathematically, it is specified by operators $P_x$ such that (i) $P_x \geq 0$, (ii) $\sum_x P_x = \mathbb{1}$, and (iii) $P_x P_y = \delta_{xy} P_x$.

Can we think of more general measurement schemes? Suppose we couple our system $A$ to an auxiliary system $B$ that is initialized in a fixed state:

$$|\psi\rangle \mapsto |\psi\rangle_A \otimes |0\rangle_B$$

We then apply an arbitrary projective measurement on the joint system, modelled by some $\{P_{AB,x}\}$. The subscript $AB$ reminds us that we are applying a projective measurement on the full system. See fig. 2 for illustration. Then the Born rule eq. (2.1) says that

$$\Pr(\text{outcome } x) = (\langle\psi|_A \otimes \langle 0|_B)\, P_{AB,x}\, (|\psi\rangle_A \otimes |0\rangle_B) = \langle\psi_A| \left( \underbrace{(\mathbb{1}_A \otimes \langle 0|_B)\, P_{AB,x}\, (\mathbb{1}_A \otimes |0\rangle_B)}_{=:Q_x} \right) |\psi_A\rangle\,,$$

where we have introduce new operators $Q_x$ on $\mathcal{H}_A$. These operators have the property that (i) $Q_x \geq 0$ and (ii) $\sum_x Q_x = \mathbb{1}_A$.

We say call any collection of operators $\{Q_x\}$ satisfying (i) and (ii) a *generalized measurement* or a *POVM measurement* (POVM is short for positive-operator valued measure). The $Q_x$ are called *POVM elements*. As we saw above, the Born rule for POVM measurements takes the familiar form

$$\Pr(\text{outcome } x) = \langle\psi|Q_x|\psi\rangle\,. \tag{2.2}$$

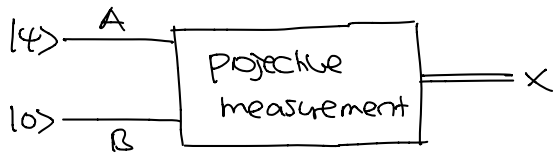



Figure 2: A generalized measurement implemented by coupling the system $A$ to an auxiliary system $B$ initialized in a fixed state $|0\rangle_B$ and performing a projective measurement on the joint system.

A *binary POVM measurement*, i.e., one that has precisely two outcomes, has the form $\{Q, \mathbb{1} - Q\}$ and is therefore specified by a single POVM element $0 \leq Q \leq \mathbb{1}$.

**Remark.** *In problem 1.4, you will show any POVM can be implemented in the fashion described above. An alternative way of thinking about a POVM measurement is the following: After coupling to an auxiliary system $B$, we apply a unitary $U_{AB}$ and then perform a projective measurement on the auxiliary system. This fits nicely with our intuitive model of measuring a quantum system – we couple it to an apparatus $B$, apply an interacting unitary time evolution, and read off the result at the apparatus.*

While eqs. (2.1) and (2.2) look identical, POVM measurements are truely more general than projective measurements. This is because while the projections $P_x$ are necessarily orthogonal, $P_x P_y = \delta_{xy} P_x$, this does not need to be the case for the $Q_x$.

**Example.** *The four operators* $\frac{1}{2}|0\rangle\langle 0|$, $\frac{1}{2}|1\rangle\langle 1|$, $\frac{1}{2}|+\rangle\langle +|$, $\frac{1}{2}|-\rangle\langle -|$ *make up a POVM with four possible outcomes. It can be thought of performing either a projective measurement in the basis* $|0\rangle$,$|1\rangle$ *or in the basis* $|+\rangle$,$|-\rangle$*, with 50% probability each.*

**Example 2.1.** *Another example is the POVM that consists of the three (mutually non-orthogonal) operators* $\{\frac{2}{3}|0\rangle\langle 0|, \frac{2}{3}|\alpha^+\rangle\langle\alpha^+|, \frac{2}{3}|\alpha^-\rangle\langle\alpha^-|\}$*, where* $|\alpha^\pm\rangle = \frac{1}{2}|0\rangle \pm \frac{\sqrt{3}}{2}|1\rangle$*. Indeed, it is easily verified that*

$$\frac{2}{3}|0\rangle\langle 0| + \frac{2}{3}|\alpha^+\rangle\langle\alpha^+| + \frac{2}{3}|\alpha^-\rangle\langle\alpha^-| = \mathbb{1}.$$

*Unlike the previous example, this POVM cannot be decomposed in an interesting way.*

In problem 1.3 you will study a state discrimination scenario where POVM measurements outperform projective measurements.

## Continuous POVMs

How can we generalize the concept of a POVM measurement to an infinite set of outcomes $\Omega$ (e.g., the set of all real numbers $\mathbb{R}$, the set of all quantum states, ...)? Let us assume that the space of outcomes $\Omega$ carries some measure $dx$. Then the conditions on $\{Q_x\}_{x\in\Omega}$ to be a POVM measurement are as follows, (i) $Q_x \geq 0$, as before, and (ii) $\int_\Omega dx Q_x = \mathbb{1}$, and Born's rule now states that

$$p(x) = \langle\psi|Q_x|\psi\rangle$$

is now the *probability density* of the outcome distribution. In other words, probabilities and expectation values can be computed as follows:

$$\Pr(\text{outcome } x \in S) = \int_S dx \, \langle\psi|Q_x|\psi\rangle \,,$$
$$E\,[f(x)] = \int dx \, \langle\psi|Q_x|\psi\rangle \, f(x). \tag{2.3}$$

We sometimes say that $\{Q_x\}$ is a *continuous POVM*.

**Remark.** *This is the most general kind of POVM measurement on a finite-dimensional Hilbert space. In infinite dimensions, one needs a more mathematically sophisticated concept – positive operator-valued measures – which is where the term "POVM" originated (e.g., Holevo, 2011).*

*You might be concerned whether we need an infinite-dimensional auxiliary Hilbert space in order to implement a POVMs with infinitely many outcomes. Interestingly, any continuous POVM on a finite-dimensional Hilbert space can be implemented by performing a discrete POVM chosen at random from a continuous probability distribution (Chiribella et al., 2007). This paper could make for a good course project.*

### Today's goal: State estimation

Suppose we are given a quantum system and we would like to learn about the underlying quantum state $|\psi\rangle$. Is there a measurement that gives us a classical description "$\psi$" of the state $|\psi\rangle$? Clearly, this cannot be done perfectly – since otherwise we could first perform this measurement and then prepare the state from its classical description multiple times, thereby achieving the impossible task of *cloning*:

$$|\psi\rangle \mapsto \text{“}\psi\text{”} \mapsto |\psi\rangle \otimes |\psi\rangle \,.$$

On the other hand, suppose that we are not given just one copy of a state, but in fact many copies $|\psi\rangle^{\otimes n}$. Note that $\langle\psi^{\otimes n}|\phi^{\otimes n}\rangle = \langle\psi|\phi\rangle^n$, so if two states are not equal then they rapidly become orthogonal as $n$ becomes large – suggesting that we can distinguish them arbitrarily well. Of course, since $\langle\psi|\phi\rangle$ can be arbitrarily close to one this is not yet a completely rigorous argument. But note that in this case the states are essentially the same, and so we make only a small error by conflating them. Thus it seems plausible that we can achieve the following task, known as *pure state estimation*:

We want to design a continuous POVM $\{Q_{\hat{\psi}}\}$ on $(\mathbb{C}^d)^{\otimes n}$, labeled by the pure states on $\mathbb{C}^d$, such when we measure on $|\psi\rangle^{\otimes n}$ we obtain an outcome $\hat{\psi}$ that is "close" to $\psi$ (on average, or even with high probability).

To solve this problem and come up with a good measurement for estimating pure states, we need to talk about the symmetries inherent in this problem: If $|\psi\rangle \in \mathbb{C}^d$ then not only is $|\psi\rangle^{\otimes n} \in (\mathbb{C}^d)^{\otimes n}$, but $|\psi\rangle^{\otimes n}$ is invariant under permuting the subsystems. Let's make this a bit more precise.

## 2.2 Symmetric subspace

Let $S_n$ denote the *symmetric group* on $n$ symbols. Its elements are permutations $\pi\colon\{1,\ldots,n\} \to \{1,\ldots,n\}$. Thus, $S_n$ has $n!$ elements. This is a *group*, meaning that products and inverses are again contained in $S_n$. For any $\pi \in S_n$, we can define an operator $R_\pi$ on the $n$-fold tensor power $(\mathbb{C}^d)^{\otimes n}$ in the following way:

$$R_\pi \, |\psi_1\rangle \otimes \ldots \otimes |\psi_n\rangle = |\psi_{\pi^{-1}(1)}\rangle \otimes \ldots \otimes |\psi_{\pi^{-1}(n)}\rangle$$

It is clear that

$$R_1 = \mathbb{1}, \quad R_\tau R_\pi = R_{\tau\pi} \tag{2.4}$$

Indeed, the latter is guaranteed by our judicious use of inverses:

$$\begin{aligned} R_\tau R_\pi |\psi_1\rangle \otimes \ldots \otimes |\psi_n\rangle &= R_\tau |\psi_{\pi^{-1}(1)}\rangle \otimes \ldots \otimes |\psi_{\pi^{-1}(n)}\rangle \\ &= R_\tau |\psi_{\pi^{-1}(1)}\rangle \otimes \ldots \otimes |\psi_{\pi^{-1}(n)}\rangle \\ &= |\psi_{\pi^{-1}(\tau^{-1}(1))}\rangle \otimes \ldots \otimes |\psi_{\pi^{-1}(\tau^{-1}(n))}\rangle \\ &= |\psi_{(\tau\pi)^{-1}(1)}\rangle \otimes \ldots \otimes |\psi_{(\tau\pi)^{-1}(n)}\rangle \\ &= R_{\tau\pi} |\psi_1\rangle \otimes \ldots \otimes |\psi_n\rangle . \end{aligned}$$

Equation (2.4) says that the map $\pi \mapsto R_\pi$ turns $(\mathbb{C}^d)^{\otimes n}$ into a *representation* of the symmetric group $S_n$.

Let us return to the vectors $|\psi\rangle^{\otimes n}$. Clearly, they have the property that $R_\pi |\psi\rangle^{\otimes n} = |\psi\rangle^{\otimes n}$ for all $\pi$. That is, $|\psi\rangle^{\otimes n}$ are elements of the *symmetric subspace*

$$\mathrm{Sym}^n(\mathbb{C}^d) = \{|\Phi\rangle \in (\mathbb{C}^d)^{\otimes n} : R_\pi |\Phi\rangle = |\Phi\rangle\}.$$

The symmetric subspace is also known as the $n$-particle sector of the bosonic Fock space for $d$ modes.

Given an arbitrary vector $|\Phi\rangle \in (\mathbb{C}^d)^{\otimes n}$, we can always symmetrize it to obtain a vector in the symmetric subspace. Indeed, let us define the *symmetrizer*

$$\Pi_n = \frac{1}{n!} \sum_{\pi \in S_n} R_\pi$$

This operator is the projector on the symmetric subspace. Let's verify this: (i) If $|\Phi\rangle$ is in the symmetric subspace then $\Pi_n |\Phi\rangle = |\Phi\rangle$:

$$\Pi_n |\Phi\rangle = \frac{1}{n!} \sum_{\pi \in S_n} R_\pi |\Phi\rangle = \frac{1}{n!} \sum_{\pi \in S_n} |\Phi\rangle = |\Phi\rangle .$$

(ii) For any vector $|\Phi\rangle \in (\mathbb{C}^d)^{\otimes n}$, the vector $|\tilde{\Phi}\rangle = \Pi_n |\Phi\rangle$ is in the symmetric subspace:

$$R_\tau |\tilde{\Phi}\rangle = R_\tau(\Pi_n |\Phi\rangle) = R_\tau \frac{1}{n!} \sum_{\pi \in S_n} R_\pi |\Phi\rangle = \frac{1}{n!} \sum_{\pi \in S_n} R_{\tau\pi} |\Phi\rangle = \frac{1}{n!} \sum_{\pi' \in S_n} R_{\pi'} |\Phi\rangle = \Pi_n |\Phi\rangle = |\tilde{\Phi}\rangle .$$

Here, we used that as $\pi$ ranges over all permutations, so does $\pi' = \tau\pi$ (indeed, we obtain any $\pi'$ exactly from $\pi = \tau^{-1}\pi'$).

In particular, we can obtain a basis of the symmetric subspace by taking a basis $|i\rangle$ of $\mathbb{C}^d$, considering a tensor product basis element $|i_1, \ldots, i_n\rangle$, and symmetrizing. The result does not depend on the order of the elements, but only on the number of times $t_i = \#\{i_k = i - 1\}$. Thus $\mathrm{Sym}^n(\mathbb{C}^d)$ has the *occupation number basis*

$$\|t_1, \ldots, t_d\rangle\!\rangle \propto \Pi_n(|1\rangle^{\otimes t_1} \otimes \ldots \otimes |d\rangle^{\otimes t_d}), \tag{2.5}$$

where $t_i \geq 0$ and $\sum_i t_i = n$.

**Example** (n=2,d=2)**.** *A basis of* $\mathrm{Sym}^2(\mathbb{C}^2)$ *is given by*

$$\|2,0\rangle\!\rangle = |00\rangle\,, \quad \|1,1\rangle\!\rangle = \frac{1}{\sqrt{2}}\left(|10\rangle + |01\rangle\right), \quad \|0,2\rangle\!\rangle = |11\rangle\,.$$

*Note that we can complete this to a basis of* $\mathbb{C}^2 \otimes \mathbb{C}^2$ *by adding the* antisymmetric *singlet state* $(|10\rangle - |01\rangle)/\sqrt{2}$*. It is true more generally that* $(\mathbb{C}^d)^{\otimes 2} = \mathrm{Sym}^2(\mathbb{C}^d) \oplus \bigwedge^2(\mathbb{C}^d)$*.*

In general, there are $\binom{n+d-1}{n}$ such basis vectors and therefore

$$\dim \mathrm{Sym}^n(\mathbb{C}^d) = \mathrm{tr}\,\Pi_n = \binom{n+d-1}{n} = \frac{(n+d-1)!}{n!(d-1)!}.$$

## A resolution of the identity for the symmetric subspace

The reason why we studied the symmetric subspace is that it contains the states $|\psi\rangle^{\otimes n}$ that arise in our estimation problem. Not every vector in $\mathrm{Sym}^n(\mathbb{C}^d)$ is of this form – for example, $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$ isn't. Moreover, the $|\psi\rangle^{\otimes n}$ are not orthogonal. Nevertheless, we have the following alternative formula for the projection onto the symmetric subspace:

$$\Pi_n' = \binom{n+d-1}{n} \int d\psi\, |\psi\rangle^{\otimes n} \langle\psi|^{\otimes n}\,. \tag{2.6}$$

The integral requires some explanation: We integrate over all unit vectors $|\psi\rangle \in \mathbb{C}^d$, and the measure $d\psi$ is the unique probability measure that is invariant under the unitary group $U(d)$. That is, expectation values do not change when we substitute $|\psi\rangle \mapsto U|\psi\rangle$, where $U$ is a unitary $d \times d$ matrix. Sometimes this measure is called the *Haar measure*. (Concretely, we can think of the $|\psi\rangle$ as unit vectors in $S^{2d-1}$ and the Haar measure can be realized as the unique rotation invariant measure on that sphere.) (Mathematically speaking, I am somewhat conflating the vectors $|\psi\rangle$ and the pure states $|\psi\rangle\langle\psi|$ – but if this concerns you then you know how to fix it!) For example, the invariance property immediately implies the following:

$$\Pi_n' = U^{\otimes n}\Pi_n' U^{\dagger,\otimes n}, \quad \text{or} \quad U^{\otimes n}\Pi_n' = \Pi_n' U^{\otimes n} \tag{2.7}$$

One way of interpreting eq. (2.6) is that the vectors $|\psi\rangle^{\otimes n}$ form an "overcomplete basis" of the symmetric subspace. Indeed, if $|\Phi\rangle$ is an arbitrary vector then

$$|\Phi\rangle = \Pi_n |\Phi\rangle = \binom{d+n-1}{n} \int d\psi\, |\psi\rangle^{\otimes n} \langle\psi^{\otimes n}|\Psi\rangle = \int d\psi\, c_\psi(\Psi)\, |\psi\rangle^{\otimes n}\,,$$

where $c_\psi(\Psi) = \binom{d+n-1}{n} \langle\psi^{\otimes n}\|\Psi\rangle$. This means that we can write $|\Phi\rangle$ as a linear combination of the states $|\psi\rangle^{\otimes n}$.

Another way to interpret eq. (2.6), though, is that it shows that

$$Q_{\hat\psi} = \binom{d+n-1}{n} |\hat\psi\rangle^{\otimes n} \langle\hat\psi|^{\otimes n} \tag{2.8}$$

defines a continuous POVM $\{Q_{\hat\psi}\}$ on the symmetric subspace! It is this so-called *uniform POVM* that we will use to solve our estimation problem!

## 2.3 Pure state estimation

We will now solve the problem of pure state estimation (cf. Chiribella, 2010, Brandao et al., 2016, Harrow, 2013). Recall that we are given $n$ copies of some $|\psi\rangle^{\otimes n}$. To obtain a good estimate, we want to measure the uniform POVM (2.8).

How do we quantify the goodness of this strategy? There are several options, but the one that is most natural in the present context is to consider the overlap squared, $|\langle\psi|\hat{\psi}\rangle|^2$, between estimate and true state. We will in fact look at a slightly more general figure of merit, namely $|\langle\psi|\hat{\psi}\rangle|^{2k}$ for some fixed $k > 0$, since this is just as easy and we will use it in Tuesday's lecture.

**Remark.** *If $k > 1$ then this is a more stringent figure of merit since unequal states become more orthogonal in this way: $|\langle\psi|\hat{\psi}\rangle|^{2k} < |\langle\psi|\hat{\psi}\rangle|^2$.*

**Remark.** *The overlap has a good operational meaning: In problem 1.2, you will show that two quantum states with overlap close to one are indeed almost indistinguishable by any possible measurement.*

Let us compute the expected value of $|\langle\psi|\hat{\psi}\rangle|^{2k}$ (the average is over the measurement outcome $\hat{\psi}$):

$$\begin{aligned}
E\left[|\langle\psi|\hat{\psi}\rangle|^{2k}\right] &= \int d\hat{\psi}\ \langle\psi^{\otimes n}|Q_{\hat{\psi}}|\psi^{\otimes n}\rangle\ |\langle\psi|\hat{\psi}\rangle|^{2k} \\
&= \binom{n+d-1}{n} \int d\hat{\psi}\, |\langle\psi|\hat{\psi}\rangle|^{2(k+n)} \\
&= \binom{n+d-1}{n} \langle\psi^{\otimes(k+n)}| \left( \int d\hat{\psi}\, |\hat{\psi}\rangle^{\otimes(k+n)}\, \langle\hat{\psi}|^{\otimes(k+n)} \right) |\psi^{\otimes(k+n)}\rangle \\
&= \binom{n+d-1}{n} \binom{n+k+d-1}{n+k}^{-1} \langle\psi^{\otimes(k+n)}|\Pi_{n+k}|\psi^{\otimes(k+n)}\rangle \\
&= \binom{n+d-1}{n} \binom{n+k+d-1}{n+k}^{-1} \\
&= \frac{(n+d-1)!}{n!} \frac{(n+k)!}{(n+k+d-1)!} = \frac{(n+d-1)\dots(n+1)}{(n+k+d-1)\dots(n+k+1)} \\
&\geq \left(\frac{n+1}{n+k+1}\right)^{d-1} = \left(1 - \frac{k}{n+k+1}\right)^{d-1} \\
&\geq 1 - \frac{k(d-1)}{n+k+1} \geq 1 - \frac{kd}{n}.
\end{aligned} \tag{2.9}$$

The first equality holds because $\langle\psi^{\otimes n}|Q_{\hat{\psi}}|\psi^{\otimes n}\rangle$ is the probability density of the measurement outcome $\hat{\psi}$, as we know from eq. (2.3). For the second equality, we plugged in the definition of the POVM element eq. (2.8). The third is just some simple manipulation using linearity of the integral, and the fourth follows by plugging in the formula for the projector onto the symmetric subspace $\mathrm{Sym}^{n+k}(\mathbb{C}^d)$. The rest are some simple inequalities that I explained in class.

Success! We have shown that the uniform POVM (2.8) gives us a very good estimate of $|\psi\rangle$ as soon as $n \gg d$ (if we measure its goodness by the overlap squared, corresponding to $k = 1$).

**Remark.** *Later in this course we will learn how to go beyond the symmetric subspace and solve the state estimation problem for general, not necessarily pure quantum states (lecture 7).*

## Bibliography

Alexander S Holevo. *Probabilistic and statistical aspects of quantum theory*, volume 1. Springer Science & Business Media, 2011.

Giulio Chiribella, Giacomo Mauro D'Ariano, and Dirk Schlingemann. How continuous quantum measurements in finite dimensions are actually discrete, *Physical Review Letters*, 98(19):190403, page 190403, 2007. arXiv:quant-ph/0702068.

Giulio Chiribella. On quantum estimation, quantum cloning and finite quantum de finetti theorems. In *Conference on Quantum Computation, Communication, and Cryptography*, pages 9–25. Springer, 2010.

Fernando GSL Brandao, Matthias Christandl, Aram W Harrow, and Michael Walter. The Mathematics of Entanglement. 2016. arXiv:1604.01790.

Aram W Harrow. The church of the symmetric subspace. 2013. arXiv:1308.6595.

**PHYSICS 491: Symmetry and Quantum Information** April 11, 2017

# Representation theory, density operators, partial trace

*Lecture 3* *Michael Walter, Stanford University*

*These lecture notes are not proof-read and are offered for your convenience only. They include additional detail and references to supplementary reading material. I would be grateful if you email me about any mistakes and typos that you find.*

In this lecture, we'll revisit some fundamentals: First, we discuss representation theory more systematically and prove the "resolution of the identity" formula (2.6) from last lecture. Then we recall the notion of a density operator and discuss the partial trace, which allows us to define the quantum state of subsystems.

## 3.1 Representation theory primer

A *(finite-dimensional unitary) representation* of a group $G$ is given by (i) a (finite-dimensional) Hilbert space $\mathcal{H}$, and (ii) unitary operators $R_g$ on $\mathcal{H}$ for every group element $g \in G$ such that the following two laws are satisfied:

$$R_1 = \mathbb{1}, \quad R_{gh} = R_g R_h$$

Every group has a *trivial representation*, given by identity operators $R_g = \mathbb{1}_{\mathcal{H}}$ acting on a one-dimensional space $\mathcal{H}$. We will often simply speak of "the representation $H$", but we always have associated operators $R_g$ in mind. All representations that we will ever study in this course will be unitary and finite-dimensional.

A useful way of understanding a representation is to decompose it into smaller building blocks. Suppose that $\tilde{\mathcal{H}} \subseteq \mathcal{H}$ is an *invariant subspace*, i.e., a subspace such that $R_g \tilde{\mathcal{H}} \subseteq \tilde{\mathcal{H}}$ for all $g \in G$. Then, the orthogonal complement $\tilde{\mathcal{H}}^\perp$ is also an invariant subspace! Indeed, if $|\phi\rangle \in \tilde{\mathcal{H}}^\perp$ then, for all $|\psi\rangle \in \tilde{\mathcal{H}}$,

$$\langle \psi | R_g | \phi \rangle = \langle R_g^\dagger \psi | \phi \rangle = 0,$$

since $R_g^\dagger |\psi\rangle \in \tilde{\mathcal{H}}$; this shows that $R_g |\phi\rangle \in \tilde{\mathcal{H}}^\perp$. As a consequence, the operators $R_g$ are block diagonal with respect to the decomposition $\mathcal{H} = \tilde{\mathcal{H}} \oplus \tilde{\mathcal{H}}^\perp$, i.e.,

$$R_g = \begin{pmatrix} \tilde{R}_g & 0 \\ 0 & \tilde{R}_g^\perp \end{pmatrix}.$$

Note that the block $\tilde{R}_g$ is a representation on $\tilde{\mathcal{H}}$ and the block $\tilde{R}_g^\perp$ is a representation on $\tilde{\mathcal{H}}^\perp$. Thus we have successfully decomposed the given representation into two "smaller" representations. We can apply the same reasoning separately to $\tilde{\mathcal{H}}$ and $\tilde{\mathcal{H}}^\perp$, and continue until we arrive at a decomposition

$$\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \ldots \oplus \mathcal{H}_m \tag{3.1}$$

that cannot be refined further. That is, the building blocks $\mathcal{H}_j$ have no interesting invariant subspaces (i.e., the only invariant subspaces are $\mathcal{H}_j$ itself and $\{0\}$, neither of which allow us to decompose further). We call such representations $\mathcal{H}_j$ *irreducible representations* – or *"irreps"*.

How can we compare different representations? An *intertwiner* $J: \mathcal{H} \to \mathcal{H}'$ is a map such that

$$JR_g = R'_g J$$

(hence the name). If there exists an *invertible* intertwiner $J$ then we say that the two representations $\mathcal{H}$ and $\mathcal{H}'$ are *equivalent*, and write $\mathcal{H} \cong \mathcal{H}'$. This invertible intertwiner can always be chosen to be a unitary operator, and we will always assume that all invertible intertwiners under consideration are unitary operators. Note that in this case we have

$$JR_g J^{-1} = JR_g J^\dagger = R'_g$$

so the operators $\{R_g\}$ and $\{R'_g\}$ differ only by an overall "base change". We will use the notation $\mathcal{H} \cong \mathcal{H}'$ and $R_g \cong R'_g$.

**Example.** *An example that you all know well is the group* $\mathrm{SU}(2)$ *of unitary* $2 \times 2$*-matrices with unit determinant, which arises in the study of rotational symmetries of quantum systems. Up to equivalence, its irreducible representations are labeled by their* spin

$$j \in \{0, \frac{1}{2}, 1, \frac{3}{2}, \dots\}.$$

*E.g.,* $V_0$ *is the one-dimensional trivial representation (also called the* singlet*),* $V_{1/2} \cong \mathbb{C}^2$*,* $V_1$ *is the* triplet *representation, etc. We used the decomposition of* $\mathrm{SU}(2)$*-representations into irreducibles briefly in section 1.2 to find a qubit, and will revisit it in greater detail in a later lecture.*

**Example 3.1.** *The permutation group* $S_3$ *has three irreducible representations (up to equivalence):*

*(i) The trivial representation* $W_{\ydiagram{3}} = \mathbb{C}|0\rangle$*, with* $R_\pi|0\rangle = |0\rangle$*.*

*(ii) The sign representation* $W_{\ydiagram{1,1,1}} = \mathbb{C}|0\rangle$*, with* $R_\pi|0\rangle = \operatorname{sign}\pi\,|0\rangle$*.*

*Here* $\operatorname{sign}\pi$ *denotes the* sign *of a permutation* $\pi \in S_n$*, defined to be* $-1$ *for transpositions ("swaps")* $i \leftrightarrow j$*. It is extended to arbitrary permutations by the requirement that* $\operatorname{sign}\pi\tau = (\operatorname{sign}\pi)(\operatorname{sign}\tau)$*. (This assignment is well-defined, as you may verify, e.g., in the special case* $S_3$*.)*

*Now consider the representation* $\mathcal{H} = \mathbb{C}^3$*, with* $R_\pi|i\rangle = |\pi(i)\rangle$*. It is not itself irreducible. However:*

*(iii) The invariant subspace*

$$W_{\ydiagram{2,1}} = \{\alpha|0\rangle + \beta|1\rangle + \gamma|2\rangle : \alpha + \beta + \gamma = 0\} \subseteq \mathbb{C}^3$$

*is a two-dimensional irreducible representation of* $S_3$*.*

*Its orthogonal complement is* $W_{\ydiagram{2,1}}^\perp = \mathbb{C}(|0\rangle + |1\rangle + |2\rangle) \cong W_{\ydiagram{3}}$*. Hence:*

$$\mathbb{C}^3 \cong W_{\ydiagram{2,1}} \oplus W_{\ydiagram{3}}$$

*The curious labeling of the irreps will become more clear when we discuss Schur-Weyl duality (see remark 7.2).*

An important tool for us is the following mathematical result, known as *Schur's lemma*.

**Lemma 3.2** (Schur). *Let $J: \mathcal{H} \to \mathcal{H}'$ be an intertwiner between* irreducible *representations $R_g$, $R'_g$.*

*(i) Either $J$ is invertible (and hence $\mathcal{H} \cong \mathcal{H}'$) or $J = 0$.*

*(ii) If $\mathcal{H} = \mathcal{H}'$ and $R_g = R'_g$ then $J \propto \mathbb{1}_{\mathcal{H}}$ (i.e., any self-intertwiner is necessarily a multiple of the identity operator).*

Schur's lemma shows that intertwiners between irreducible representations are rigidly determined. In particular, there are no nonzero intertwiners between inequivalent irreducible representations. We will not prove this result – you are encouraged to look it up in your favorite textbook (e.g., Fulton and Harris, 2013) – but we will profitably use it many times in this class.

## Normal forms of representations

Now suppose that someone handed us a list of irreducible representations of a group $G$. Let us write $V_j$ for the Hilbert space, $R_g^{(j)}$ for the operators, and $j$ runs over some index set $J$ that labels the different irreps. We assume that the list is *complete* (i.e., that any other irreducible representation is equivalent to some $V_j$) and that it is *irredundant* (i.e., that $V_j \ncong V_{j'}$ if $j \neq j'$). We just saw two such lists for $G = \mathrm{SU}(2)$ and $G = S_3$, respectively.

Then, if $\mathcal{H}$ is an arbitrary representation of $G$, we can first decompose as in eq. (3.1). Since each $\mathcal{H}_k$ in eq. (3.1) is irreducible, it must be equivalent to some $V_j$ – say $\mathcal{H}_k \cong V_{j_k}$. Thus:

$$\mathcal{H} \cong V_{j_1} \oplus \ldots \oplus V_{j_m} \tag{3.2}$$

Suppose that $n_j$ is the number of times that $V_j$ appears in this list, i.e., $n_j = \#\{k : j_k = j\}$. Let us reorder (3.2) according to the different values of $j$:

$$\mathcal{H} \cong \bigoplus_{j \in J} \underbrace{V_j \oplus \ldots \oplus V_j}_{n_j \text{ times}} \tag{3.3}$$

The numbers $n_j$ are uniquely determined – as a consequence of Schur's lemma! They fully characterize the representation $\mathcal{H}$, up to equivalence. A useful alternative way of writing down the decomposition (3.3) is as follows:

$$\mathcal{H} \cong \bigoplus_{j \in J} V_j \otimes \mathbb{C}^{n_j}, \tag{3.4}$$

where $G$ acts on the right-hand side by the block-diagonal matrices

$$\bigoplus_k R_g^{(j)} \otimes \mathbb{1}_{n_j}.$$

(We use the notation $\oplus$ to stress that they are block diagonal with respect to the direct sum decomposition of the Hilbert space that they act on, i.e., eq. (3.4).) We may think of eq. (3.3) or eq. (3.4) as a "normal form" of the representation $\mathcal{H}$.

**Remark.** *The fact that unitary representation $\mathcal{H}$ can be brought into a normal form is completely analogous to how, e.g., a unitary or Hermitian matrix can always be diagonalized.*

Representation theory tells us about the list of irreducible representations for a given group $G$ and how to determine the decomposition (3.3) or (3.4) of a representation into its irreducible pieces (in particular, how to calculate the numbers $n_j$).

### Proof of the resolution of the identity for the symmetric subspace

Schur's lemma allows us to at last deduce eq. (2.6). To see this, we first observe that the space

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

is not only a representation of $S_n$, as discussed in section 2.2, but also of the unitary group $U(d)$. Its elements are the unitary $d \times d$-matrices $U$, and its representation on $(\mathbb{C}^d)^{\otimes n}$ is defined as follows:

$$T_U = (U \otimes \ldots \otimes U) = U^{\otimes n}$$

Next week, we will learn much more about the way $(\mathbb{C}^d)^{\otimes n}$ decomposes with respect to the groups $S_n$ and $U(d)$. For today, we only note that the two group actions commute:

$$R_\pi T_U = T_U R_\pi, \quad \text{or} \quad [R_\pi, T_U] = 0. \tag{3.5}$$

Let us verify this explicitly:

$$\begin{aligned} &R_\pi T_U(|\psi_1\rangle \otimes \ldots \otimes |\psi_1\rangle) = R_\pi(U|\psi_1\rangle \otimes \ldots \otimes U|\psi_1\rangle) \\ &= U|\psi_{\pi^{-1}(1)}\rangle \otimes \ldots \otimes U|\psi_{\pi^{-1}(n)}\rangle = T_U R_\pi(|\psi_1\rangle \otimes \ldots \otimes |\psi_1\rangle). \end{aligned}$$

Equation (3.5) implies at once that the symmetric subspace $\mathrm{Sym}^n(\mathbb{C}^d)$ is an invariant subspace for $U(d)$. Indeed, if $|\Phi\rangle \in \mathrm{Sym}^n(\mathbb{C}^d)$ then $R_\pi(T_U|\Phi\rangle) = T_U(R_\pi|\Phi\rangle) = T_U|\Phi\rangle$ and so $T_U|\Phi\rangle \in \mathrm{Sym}^n(\mathbb{C}^d)$.

*Importantly, the symmetric subspace is in fact an irreducible representation of $U(d)$.* You will show this in problem 2.3. It is now easy to see that the operator $\Pi'_n$ defined in eq. (2.6) is equal to the projector onto the symmetric subspace. First, note that eq. (2.7) asserts precisely that $\Pi'_n$ is a self-intertwiner, i.e., $T_U \Pi'_n = \Pi'_n T_U$ (this follows from the invariance of the integral under substituting $|\psi\rangle \mapsto U|\psi\rangle$). Second, note that $\Pi'_n$ is supported only on the symmetric subspace. We may therefore safely think of $\Pi'_n$ as an operator from $\mathrm{Sym}^n(\mathbb{C}^d)$ to $\mathrm{Sym}^n(\mathbb{C}^d)$. But since the symmetric subspace is irreducible, Schur's lemma tells us that $\Pi'_n$ must be proportional to the identity operator on $\mathrm{Sym}^n(\mathbb{C}^d)$, i.e., to $\Pi_n$. Since moreover

$$\operatorname{tr}\Pi'_n = \binom{n+d-1}{n} \int d\psi \underbrace{\operatorname{tr}\left[|\psi\rangle^{\otimes n}\langle\psi|^{\otimes n}\right]}_{=1} = \binom{n+d-1}{n} = \operatorname{tr}\Pi_n,$$

we conclude that $\Pi_n = \Pi'_n$.

## 3.2 Density operators and mixed states

Before we proceed with entanglement and symmetries, let us talk a bit about ensembles of quantum states. Many of you know density operators and partial traces, but I hope this might be a good reminder for everyone.

Suppose that $\{p_i, |\psi_i\rangle\}$ is an ensemble of quantum states on some Hilbert space $\mathcal{H}$, i.e., we have the state $|\psi_i\rangle$ with probability $p_i$. If $X$ is an observable then we can compute its expectation value by

$$\langle X\rangle = \sum_i p_i \langle\psi_i|X|\psi_i\rangle = \sum_i p_i \operatorname{tr}\left[|\psi_i\rangle\langle\psi_i| X\right] = \operatorname{tr}\left[\underbrace{\sum_i p_i |\psi_i\rangle\langle\psi_i|}_{=:\rho} X\right] = \operatorname{tr}[\rho X].$$

The operator $\rho$ is called a *density operator*, or a *density matrix*, or simply a *quantum state* on $\mathcal{H}$. It safisfies $\rho \geq 0$ and $\operatorname{tr}\rho = 1$, and any such operator arises from some ensemble of quantum states (think of the spectral decomposition!). The *Born rule* for density operators reads

$$\Pr(\text{outcome } x) = \operatorname{tr}[\rho Q_x],$$

as follows from our preceding calculation.

If $\rho = |\psi\rangle\langle\psi|$ then we say that it is a *pure state* (and it is not uncommon to simply write $\rho = \psi$ in this case). Otherwise, $\rho$ is called a *mixed state* (but we will often be sloppy and say "mixed state" when we really should say "density operator"). Note that $\rho$ is pure if and only if $\operatorname{rk}\rho = 1$, or if $\rho^2 = \rho$, or if the eigenvalue spectrum is $\{1, 0\}$.

**Example 3.3** (Warning!)**.** *In general the ensemble that determines a density operator is* not *unique. E.g.,* $\tau = \mathbb{1}/2$ *can be written in an infinite number of ways:*

$$\tau = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) = \frac{1}{2}(|+\rangle\langle +| + |-\rangle\langle -|) = \dots.$$

The states $\tau_{\mathcal{H}} = \mathbb{1}_{\mathcal{H}}/\dim\mathcal{H}$ are known as *maximally mixed states*. They are the analogues of uniform distributions in probability theory.

More generally, if $p(x_1, \dots, x_n)$ is a probability distribution then we may consider the ensemble $\{p(x_1, \dots, x_n), |x_1\rangle \otimes \dots \otimes |x_n\rangle\}$. The corresponding density operator is

$$\rho_{X_1,\dots,X_n} = \sum_{x_1,\dots,x_n} p(x_1, \dots, x_n)\, |x_1\rangle\langle x_1| \otimes \dots \otimes |x_n\rangle\langle x_n| \tag{3.6}$$

and we call such a state a *classical state*. If all probabilities $p(x_1, \dots, x_n)$ are the same then $\rho_{X_1,\dots,X_n}$ is a maximally mixed state, $\rho = \tau$. In a later problem set, you will explore more generally how classical probability theory can be embedded into quantum mechanics.

In quantum physics, density operators arise in a number of places: As statistical ensembles (e.g., Gibbs states in statistical quantum physics), when describing noisy sources, . . .– but importantly, also when describing the state of subsystems, as we will discuss in the following.

### Reduced density matrices and partial trace

Suppose that $\rho_{AB}$ is a quantum state on $\mathcal{H}_A \otimes \mathcal{H}_B$ and $X_A$ an observable on $\mathcal{H}_A$. The axioms of quantum mechanics tell us $X_A \otimes \mathbb{1}_B$ is the appropriate observable on the joint system $\mathcal{H}_A \otimes \mathcal{H}_B$. Let's calculate the expectation value of this observable in the state $\rho_{AB}$:

$$\begin{aligned}
\langle X_A\rangle &= \operatorname{tr}[\rho_{AB}(X_A \otimes \mathbb{1}_B)] = \sum_{a,b} \langle a,b|\rho_{AB}(X_A \otimes \mathbb{1}_B)|a,b\rangle \\
&= \sum_{a,b} (\langle a| \otimes \langle b|)|\rho_{AB}(X_A \otimes \mathbb{1}_B)(|a\rangle \otimes |b\rangle) \\
&= \sum_{a,b} \langle a|\,(\mathbb{1}_A \otimes \langle b|)\rho_{AB}(X_A \otimes \mathbb{1}_B)(\mathbb{1}_A \otimes |b\rangle)\,|a\rangle \\
&= \sum_{a,b} \langle a|\,(\mathbb{1}_A \otimes \langle b|)\,\rho_{AB}\,(\mathbb{1}_A \otimes |b\rangle)\,X_A|a\rangle \\
&= \sum_{a} \langle a|\underbrace{\sum_{b} (\mathbb{1}_A \otimes \langle b|)\,\rho_{AB}\,(\mathbb{1}_A \otimes |b\rangle)}_{=:\operatorname{tr}_B[\rho_{AB}]}\,X_A|a\rangle
\end{aligned}$$

The operation $\mathrm{tr}_B$ just introduced is called the *partial trace* over $B$. If $\rho_{AB}$ is a quantum state, then $\mathrm{tr}_B[\rho_{AB}]$ is called the *reduced density operator* or the *reduced density matrix* $\rho_A$ of $\rho_{AB}$. We will often denote it by $\rho_A = \mathrm{tr}_B[\rho_{AB}]$ (even though this can at times seem ambiguous). Dually, $\rho_{AB}$ is said to be an *extension* of $\rho_A$. By construction,

$$\mathrm{tr}[\rho_{AB}(X_A \otimes \mathbb{1}_B)] = \mathrm{tr}[\rho_A X_A], \tag{3.7}$$

and so the reduced density operator $\rho_A$ contains all information necessary to evaluate observables on $A$. It therefore faithfully describes the state of the subsystem $A$.

We can also compute partial traces of operator that are not quantum states: If $M_{AB}$ is an arbitrary operator on $\mathcal{H}_A \otimes \mathcal{H}_B$ then its partial trace over $B$ is defined just as before,

$$\mathrm{tr}_B[M_{AB}] = \sum_b (\mathbb{1}_A \otimes \langle b|)\, M_{AB}\, (\mathbb{1}_A \otimes |b\rangle)\,.$$

(However, if $M_{AB}$ is not a state then we will *never* denote this partial trace by $M_A$.)

The following useful rule tells us how to compute partial traces of tensor product operators $M_A \otimes N_B$ and justifies the term "partial trace":

$$\mathrm{tr}_B[M_A \otimes N_B] = M_A \,\mathrm{tr}[N_B] \tag{3.8}$$

It follows directly from the definition:

$$\mathrm{tr}_B[M_A \otimes N_B] = \sum_b (\mathbb{1}_A \otimes \langle b|)\,(M_A \otimes N_B)\,(\mathbb{1}_A \otimes |b\rangle) = M_A \sum_b \langle b|N_B|b\rangle = M_A \,\mathrm{tr}[N_B].$$

Other useful properties are

- $\mathrm{tr}_B[(M_A \otimes \mathbb{1}_B)X_{AB}(M'_A \otimes \mathbb{1}_B)] = M_A \,\mathrm{tr}_B[O_{AB}]M'_B$,
- $\mathrm{tr}_B[(\mathbb{1} \otimes M_B)O_{AB}] = \mathrm{tr}_B[O_{AB}(\mathbb{1} \otimes M_B)]$.

**Remark.** *A useful convention that you will often find in the literature is that tensor products with the identity operator are omitted. E.g., instead of $X_A \otimes \mathbb{1}_B$ we would write $X_A$, since the subscripts already convey the necessary information. Thus, instead of eqs.* (3.7) *and* (3.8) *we would write*

$$\mathrm{tr}[\rho_{AB}X_A] = \mathrm{tr}[\rho_A X_A],$$
$$\mathrm{tr}_B[M_A N_B] = M_A \,\mathrm{tr}[N_B]$$

*which is arguably easier to read.*

**Example** (Warning!)**.** *Even if $\rho_{AB}$ is a pure state, $\rho_A$ can be mixed. For example, consider the maximally entangled state $|\psi\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Then,*

$$\begin{aligned}
\rho_{AB} = |\psi\rangle\langle\psi|_{AB} &= \frac{1}{2}\,(|00\rangle + |11\rangle)\,(\langle 00| + \langle 11|) \\
&= \frac{1}{2}\,(|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11|) \\
&= \frac{1}{2}\,(|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |1\rangle\langle 0| \otimes |1\rangle\langle 0| + |0\rangle\langle 1| \otimes |0\rangle\langle 1| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|)\,,
\end{aligned}$$

*and so, using eq.* (3.8)*,*

$$\rho_A = \mathrm{tr}_B[|\psi\rangle\langle\psi|_{AB}] = \frac{1}{2}\,(|0\rangle\langle 0| + |1\rangle\langle 1|)\,.$$

*Thus $\rho_A$ is a mixed state – in fact, the maximally mixed state $\tau_A$ introduced previously in example 3.3.*

The preceding example was not an accident. Every pure state $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ has a so-called *Schmidt decomposition*

$$|\psi\rangle_{AB} = \sum_i s_i |e_i\rangle_A \otimes |f_i\rangle_B ,$$

where $s_i > 0$ and the $|e_i\rangle_A$ and $|f_i\rangle_B$ are sets of orthonormal vectors in $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Note:

$$\rho_A = \sum_i s_i^2 |e_i\rangle \langle e_i|_A \quad \text{and} \quad \rho_B = \sum_i s_i^2 |f_i\rangle \langle f_i|_B .$$

*Thus the eigenvalues of the reduced density matrices are directly related to the cofficients $s_i$.*

The Schmidt decomposition is a very important tool that we already briefly met in the fine-print of lecture 1. For one, it helps us to understand entanglement in pure states: E.g., if $|\psi\rangle_{AB} = |\psi\rangle_A \otimes |\psi\rangle_B$ is a product state then the reduced density matrices are pure. Conversely, if the reduced density matrices of a pure state $|\psi\rangle_{AB}$ are mixed then this is a signature of entanglement. You will discuss this in more detail on problem 2.1. (This also justifies why quantities such as *entanglement entropies* that some of you might already know might be good entanglement measures (only) for pure states.)

We mention two last important facts that you will prove in problem 2.2:

(i) Any mixed state $\rho_A$ has a *purification*: That is, there exists a pure state $|\psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, with $\mathcal{H}_B$ an auxiliary Hilbert space, such that

$$\rho_A = \mathrm{tr}_B[|\psi_{AB}\rangle \langle \psi_{AB}|].$$

**Remark.** *This justifies why in lecture 1 we were allowed to only consider quantum strategies that involved pure states and observables. At the expense of adding an auxiliary Hilbert space, we can always replace mixed states by pure states and generalized measurements by measurements of observables (you proved the latter in problem 1.4).*

(ii) If $\rho_A = |\psi\rangle_A \langle\psi|_A$ is pure then any extension $\rho_{AB}$ is necessarily a product, i.e., $\rho_{AB} = \rho_A \otimes \rho_B$ – whether $\rho_{AB}$ is pure or mixed. We already mentioned this fact when discussing the privacy of random bits in lecture 1.

## Bibliography

William Fulton and Joe Harris. *Representation theory: a first course*, volume 129. Springer Science & Business Media, 2013.

**PHYSICS 491: Symmetry and Quantum Information** April 13, 2017

# Mixed state entanglement, monogamy of entanglement

*Lecture 4* *Michael Walter, Stanford University*

*These lecture notes are not proof-read and are offered for your convenience only. They include additional detail and references to supplementary reading material. I would be grateful if you email me about any mistakes and typos that you find.*

Monogamy of entanglement is the idea that if two systems are strongly entangled then each of them cannot be entangled very much with other systems. For example, suppose that

$$\rho_{AB} = |\Psi\rangle\langle\Psi|_{AB}$$

where $|\Psi\rangle_{AB}$ is in a pure state – say, a maximally entangled state. Since $\rho_{AB}$ is pure, any extension $\rho_{ABC}$ must factorize,

$$\rho_{ABC} = \rho_{AB} \otimes \rho_C,$$

as we discussed at the end of lecture 3. Thus $A$ and $B$ are both completely uncorrelated with $C$ (fig. 3). In particular, $\rho_{AC} = \rho_A \otimes \rho_C$ and $\rho_{BC} = \rho_B \otimes \rho_C$ are product states.

**Remark.** *While correct, the above analysis should perhaps be taken with a grain of salt. Since it only relied on $\rho_{AB}$ being in a pure state, it is also applicable to, say, $\psi_{AB} = |0\rangle_A \otimes |0\rangle_B$ – which is a product state, not an entangled state! Nevertheless, the conclusion remains that also in this case $\rho_{AC}$ and $\rho_{BC}$ have to product states. However, this is a consequence of $\rho_A = |0\rangle\langle 0|_A$ and $\rho_B = |0\rangle\langle 0|_B$ being pure, not of entanglement between $A$ and $B$.*

Does monogamy hold more generally and can it be made quantitative? Indeed this is possible – and we will see that symmetry is the key.

## 4.1 Mixed state entanglement

First, though, we will have to talk about what it means for a quantum state to be entangled. For pure states $|\psi\rangle_{AB}$, the answer is simple: A state is entangled if and only if is *not* a tensor product,

$$|\psi\rangle_{AB} \neq |\psi\rangle_A \otimes |\psi\rangle_B .$$

For mixed states, however, there are non-product quantum states that should nevertheless not be considered entangled.

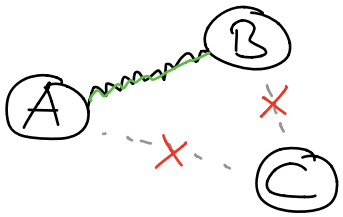



Figure 3: Illustration of monogamy of entanglement.

**Example 4.1** (Classical joint distributions)**.** *Let $p(x, y)$ be a probability distribution of two random variables. Following* (3.6)*, we construct a corresponding density operator*

$$\rho_{AB} = \sum_{x,y} p(x,y) \left|xy\right\rangle_{AB} \left\langle xy\right|_{AB} = \sum_{x,y} p(x,y) \left|x\right\rangle \left\langle x\right|_A \otimes \left|y\right\rangle \left\langle y\right|_B .$$

*In general, $\rho_{AB}$ is not a product state (indeed, this is only the case if the random variables are statistically independent). Yet this corresponds to classical correlations, not to quantum entanglement. For example, if Alice and Bob know the outcome of a fair coin flip, their state would be described by the density operator*

$$\rho_{AB} = \frac{1}{2} \left( \left|00\right\rangle \left\langle 00\right|_{AB} + \left|11\right\rangle \left\langle 11\right|_{AB} \right),$$

*that is not of product form.*

This suggests the following general definition: We say that a quantum state $\rho_{AB}$ is *entangled* if it is *not* a mixture of product states:

$$\rho_{AB} \neq \sum_i p_i \rho_A^{(i)} \otimes \rho_B^{(i)}. \tag{4.1}$$

Here, $\{p_i\}$ is an arbitrary probability distribution and the $\rho_A^{(i)}$ and $\rho_B^{(i)}$. We say that states of the right-hand side form are *separable*, or simply *unentangled*. If $\rho_{AB} = \left|\psi\right\rangle \left\langle\psi\right|_{AB}$ is a pure state then it is separable exactly if it is a product, $\left|\psi\right\rangle_{AB} = \left|\psi\right\rangle_A \otimes \left|\psi\right\rangle_B$.

**Remark.** *There are more separable states than the classical states in example 4.1. This is because we do not demand the operators $\{\rho_A^{(i)}\}$ and $\{\rho_B^{(i)}\}$ in eq.* (4.1) *are orthogonal.*

Separable states have a pleasant operational interpretation. They are the largest class of quantum states $\sigma_{AB}$ that can be created by Alice and Bob in their laboratories if allow Alice and Bob to perform arbitrary quantum operations in their laboratory but restrict their communication with each other to be classical.

Let us denote the set of all density operators on $\mathcal{H}_A \otimes \mathcal{H}_B$ by

$$Q_{AB} = \{\rho_{AB} \geq 0, \operatorname{tr} \rho_{AB} = 1\}$$

and the subset of separable states by

$$SEP_{AB} = \{\rho_{AB} \text{separable}\}.$$

Both sets are *convex*. As a consequence of $SEP_{AB}$ being convex, it can be fully characterized by separating hyperplanes, i.e., hyperplanes that contain all separable state on one side (fig. 4). These hyperplanes gives rise to *entanglement witness* – one-sided tests that can be used to certify that a state is entangled. You will explore them in problem 2.4.

Yet, it is unfortunately a difficult problem to decide if a mixed state is entangled or not. In fact, the problem of deciding whether a given quantum state $\rho_{AB}$ is separable is *NP-hard*. This implies that we are unlikely to ever find an efficient (polynomial-time) algorithm. In practice, the situation is less bleak since we have ways of testing whe a quantum state is approximately separable (see below).

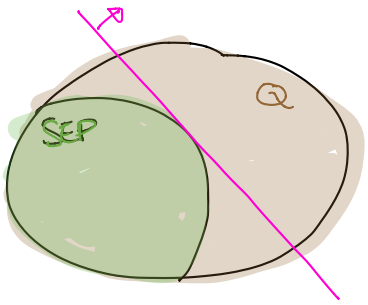



Figure 4: The set of separable states $SEP$ is a convex subset of the set of all quantum states $Q$. Hyperplanes (such as the pink one) that contain all separable states on one side give rise to entanglement witnesses.

## 4.2 Monogamy and symmetry

We are now ready to study the monogamy of entanglement in more detail. We will consider two situations where we would expect monogamy to play a role:

### De Finetti theorem

First, consider a permutation-symmetric state

$$|\Psi\rangle_{A_1\dots A_n} \in \mathrm{Sym}^n(\mathbb{C}^d).$$

Note that all the reduced density matrices $\rho_{A_iA_j}$ are the same. Thus, every pair of particles is entangled equally, and so we would expect that by monogamy they therefore are not entangled "very much" (fig. 5, (a)).

The *quantum de Finetti theorem* (König and Renner, 2005) asserts that our expectation is indeed correct:

$$\rho_{A_1\dots A_k} \approx \int d\psi\, p(\psi)\, |\psi\rangle^{\otimes k} \langle\psi|^{\otimes k} \tag{4.2}$$

as long as $k \ll (n-k)/d$. Here, $p(\psi)$ is some probability density over the set of pure states that depends on the state $\rho$. In particular, $\rho_{A_1A_2}$ is approximately a mixture of product states for large $n$.

**Example** (Warning)**.** *The GHZ state* $|\gamma\rangle_{A_1A_2A_3} = (|000\rangle + |111\rangle)/\sqrt{2}$ *is a state in the symmetric subspace* $\mathrm{Sym}^3(\mathbb{C}^2)$*. Note that, e.g., the first particle is maximally entangled with the other two – so clearly it is* not *true that permutation symmetric states are unentangled. However, if we look at the reduced state of two particles then we find*

$$\rho_{A_1A_2} = \frac{1}{2}\left(|00\rangle\langle 00| + |11\rangle\langle 11|\right) = \frac{1}{2}|0\rangle^{\otimes 2}\langle 0|^{\otimes 2} + \frac{1}{2}|1\rangle^{\otimes 2}\langle 1|^{\otimes 2}.$$

*Note that* $\rho_{A_1A_2}$ *is a* mixture *of product states. This shows that the partial trace is indeed necessary.*

Permutation symmetric states arise naturally in *mean-field systems*. The ground state $|E_0\rangle$ of a mean-field Hamiltonian $H = \sum_{1\le i<j\le n} h_{ij}$ is necessarily in the symmetric subspace – provided that the ground space is nondegenerate and that $n$ is larger than the single-particle Hilbert space. Thus, the de Finetti theorem shows that, locally, ground states of mean field systems look like mixtures of product states – a property that is highly useful for their analysis. For example, it allows us to use the density $p(\psi)$ as a variational ansatz.


Figure 5: (a) In a permutation symmetric state, any pair of particles is entangled in the same way and should therefore not be entangled very much. (b) Similarly, if Alice is entangled with many Bobs in the same way then she is not entangled very much with each of them.

### Extendibility hierarchy

A closely related situation is the following: Suppose that $\rho_{AB}$ is a quantum state that has an extension $\rho_{AB_1\ldots B_n}$ such that

$$\rho_{AB_i} = \rho_{AB} \qquad (\forall i, j)$$

(fig. 5, (b)). We say that $\rho_{AB}$ has an *n-extension*. Thus $A$ is equally entangled with all $B_i$ and so we would expect that $\rho_{AB}$ is not entangled "very much". Indeed, it is true that, for large $n$,

$$\rho_{AB} \approx \sum_i p_i \rho_A^{(i)} \otimes \rho_B^{(i)},$$

i.e., $\rho_{AB}$ is again approximately a mixture of product states.

In contrast to situation (1), however, there is no longer a symmetry requirement between $A$ and $B$, i.e., this reasoning applies to general states $\rho_{AB}$. It turns out that one in this way obtains a hierarchy of efficient approximates test for separability (Doherty et al., 2002, 2004). Indeed, as you will discuss in problem 2.5, if a state $\rho_{AB}$ is $n$-extendible then it is $O(1/n)$-close to being a separable state (fig. 6).

## 4.3 The trace distance between quantum states

Before we proceed, we should make more precise what we meant when we wrote "$\approx$" above. Let $\rho$ and $\sigma$ be two density operators on some Hilbert space $\mathcal{H}$. We define their *trace distance* to be

$$T(\rho, \sigma) := \max_{0 \leq Q \leq \mathbb{1}_{\mathcal{H}}} \mathrm{tr}[Q(\rho - \sigma)].$$

The trace distance is a metric, and so in particular satisfies the triangle inequality. It has the following alternative expression

$$T(\rho, \sigma) = \frac{1}{2} \|\rho - \sigma\|_1,$$

where we used the *1-norm*, which for general Hermitian operators $\Delta$ with spectral decomposition $\Delta = \sum_i \lambda_i |e_i\rangle \langle e_i|$ is defined by $\|\Delta\|_1 = \sum_i |\lambda_i|$. The trace distance has a natural operational interpretation in terms of the optimal probability of distinguishing $\rho$ and $\sigma$ by a POVM measurement. You discussed the trace distance in problem 1.3 in the special case of pure states, but the above conclusions hold

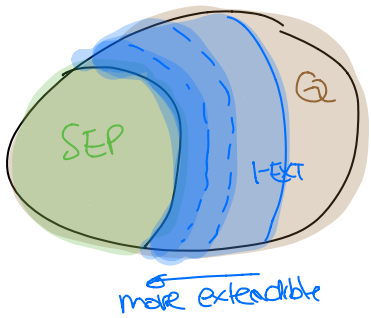



Figure 6: The extendibility hierarchy: If a state is $n$ extendible then it is $O(1/n)$-close to being separable.

in general. There, you also proved that, for pure states $\rho = |\phi\rangle\langle\phi|$ and $\sigma = |\psi\rangle\langle\psi|$, the trace distance and overlap are related by the following formula:

$$T(\rho,\sigma) = \sqrt{1 - |\langle\phi|\psi\rangle|^2} \tag{4.3}$$

**Remark.** *If $X$ is an arbitrary observable then*

$$|\mathrm{tr}[H\rho] - \mathrm{tr}[H\sigma]| \leq 2T(\rho,\sigma)\|H\|_\infty, \tag{4.4}$$

*where $\|H\|_\infty$ denotes the* operator norm *of $H$, defined as the maximal absolute value of all eigenvalues of $H$. Indeed, we can always write $H = Q - Q' <$ where $0 \leq Q, Q' \leq \|H\|_\infty$, and so*

$$|\mathrm{tr}[H\rho] - \mathrm{tr}[H\sigma]| \leq |\mathrm{tr}[Q\rho] - \mathrm{tr}[Q\sigma]| + |\mathrm{tr}[Q'\rho] - \mathrm{tr}[Q'\sigma]| \leq 2\|H\|_\infty T(\rho,\sigma).$$

*Equation* (4.4) *quantifies the difference in expectation values for states with small trace distance. (Note that this gap gap can be arbitrarily large since we can always rescale our observable. This is reflected by the factor $\|H\|_\infty$.)rescale our observables.*

## 4.4 The quantum de Finetti theorem

We will now prove the de Finetti theorem (4.2), following Brandao et al. (2016). Let

$$|\Phi\rangle_{A_1\ldots A_n} \in \mathrm{Sym}^n(\mathbb{C}^d),$$

where $n$ is the number of particles and $d$ the dimension of the single-particle Hilbert space.

The basic idea is the following: Suppose that we measure with the uniform POVM (2.8) on the last $n-k$ systems of $\rho = |\Phi\rangle\langle\Phi|$. Then, if the measurement outcome is some $|\psi\rangle$, we would expect that the first $k$ systems are likewise in the state $|\psi\rangle^{\otimes k}$, at least on average, since the overall state is permutation symmetric among all $n$ subsystems.

Let us try to implement this idea. Since $|\Phi\rangle \in \mathrm{Sym}^n(\mathbb{C}^d)$, it is in particular symmetric under permutations of the last $n-k$ subsystems. Hence, $|\Phi\rangle = (\mathbb{1}_k \otimes \Pi_{n-k})|\Phi\rangle$, and so

$$\begin{aligned}
\rho_{A_1\ldots A_k} &= \mathrm{tr}_{A_{k+1}\ldots A_n}[|\Phi\rangle\langle\Phi|] = \mathrm{tr}_{A_{k+1}\ldots A_n}[(\mathbb{1}_k \otimes \Pi_{n-k})|\Phi\rangle\langle\Phi|] \\
&= \binom{n-k+d-1}{n-k} \int d\psi\, (\mathbb{1}_k \otimes \langle\psi|^{\otimes(n-k)})|\Phi\rangle\langle\Phi|(\mathbb{1}_k \otimes |\psi\rangle^{\otimes(n-k)}) \\
&= \int d\psi p(\psi)|V_\psi\rangle\langle V_\psi|.
\end{aligned}$$

In the second to last step, we have inserted the resolution of identity (2.6), and in the last step, we have introduced introduced unit vectors $|V_\psi\rangle$ and numbers $p(\psi) \geq 0$ such that

$$\sqrt{p(\psi)}\,|V_\psi\rangle = \binom{n-k+d-1}{n-k}^{1/2} (\mathbb{1}_k \otimes \langle\psi|^{\otimes(n-k)})\,|\Phi\rangle\,. \tag{4.5}$$

Note that $p(\psi)$ is a probability density. Indeed, $\int d\psi\, p(\psi) = \operatorname{tr}\rho = 1$, since the overall state is normalized. We would now like to prove that

$$\rho_{A_1\ldots A_k} = \int d\psi\, p(\psi)\; |V_\psi\rangle\langle V_\psi| \approx \int d\psi\, p(\psi)\; |\psi\rangle^{\otimes k}\langle\psi|^{\otimes k} =: \widetilde{\rho}_{A_1\ldots A_k}, \tag{4.6}$$

based on the intuition expressed above that on average the post-measurement states $|V_\psi\rangle$ are close to $|\psi\rangle^{\otimes k}$. Let us first consider the average overlap:

$$\begin{aligned}
&\int d\psi\, p(\psi)\, |\langle V_\psi|\psi^{\otimes k}\rangle|^2 = \int d\psi\, p(\psi)\; \langle V_\psi|\psi^{\otimes k}\rangle\langle\psi^{\otimes k}|V_\psi\rangle \\
&= \binom{n-k+d-1}{n-k} \int d\psi\; \langle\Phi|\psi^{\otimes n}\rangle\langle\psi^{\otimes n}|\,|\Phi\rangle\rangle = \binom{n-k+d-1}{n-k}\binom{n+d-1}{n}^{-1} \underbrace{\langle\Phi|\Pi_n|\Phi\rangle}_{=1} \\
&= \binom{n-k+d-1}{n-k}\binom{n+d-1}{n}^{-1} \geq 1 - \frac{kd}{n-k}.
\end{aligned}$$

In the second step, we inserted the definition of $|V_\psi\rangle$ from eq. (4.5). And the last inequality is precisely (2.9), since there we bounded precisely the ratio of binomial coefficients that we are interested in here (with $n \mapsto n+k$).

It remains to show that the two states $\rho$ and $\widetilde{\rho}$ in eq. (4.6) are close in trace distance. Indeed,

$$\begin{aligned}
&T(\rho_{A_1\ldots A_k}, \widetilde{\rho}_{A_1\ldots A_k}) \leq \int d\psi\, p(\psi) T(|V_\psi\rangle\langle V_\psi|, |\psi\rangle^{\otimes k}\langle\psi|^{\otimes k}) = \int d\psi\, \sqrt{1 - |\langle V_\psi|\psi^{\otimes k}\rangle|^2} \\
&\leq \sqrt{\int d\psi\, \left(1 - |\langle V_\psi|\psi^{\otimes k}\rangle|^2\right)} = \sqrt{1 - \int d\psi\, |\langle V_\psi|\psi^{\otimes k}\rangle|^2} \leq \sqrt{\frac{kd}{n-k}}.
\end{aligned}$$

Here, we first applied the triangle inequality, then we used the relationship between trace distance and fidelity for pure states in eq. (1.2), and the next inequality i Jensen's inequality for the square root function, which is concave. Thus we have proved the de Finetti theorem (4.2):

$$\rho_{A_1\ldots A_k} \approx \int d(\psi)\; |\psi\rangle^{\otimes k}\langle\psi|^{\otimes k}$$

up to error $\sqrt{kd/(n-k)}$ in trace distance. Explicitly, the density $p(\psi)$ that we used in our proof is given by $\langle\Phi|\mathbb{1}_k \otimes Q_\psi|\Phi\rangle$, where $\{Q_\psi\}$ is the uniform POVM (2.8).

### Beyond the symmetric subspace

Our intuition behind the de Finetti theorem only relied on the fact that the reduced density matrices were all the same. But this is a feature that states on the symmetric subspace share with arbitrary *permutation-invariant* states, i.e., states that satisfy

$$[R_\pi, \rho_{A_1\ldots A_n}] = 0, \quad \text{or} \quad R_\pi \rho_{A_1\ldots A_n} = \rho_{A_1\ldots A_n} R_\pi$$

for all $\pi \in S_n$. Examples of permutation-invariant states are states on the *anti*symmetric subspace, or tensor powers of mixed states such as $\rho^{\otimes n}$, which we will study in more detail next week.

A useful fact is that any permutation-invariant state $\rho_{A_1 \dots A_n}$ has a purification on a symmetric subspace: That is, there exists a pure state $|\Phi\rangle_{(A_1 B_1) \dots (A_n B_n)} \in \mathrm{Sym}^n(\mathcal{H}_A \otimes \mathcal{H}_B)$, where $\mathcal{H}_B$ is some auxiliary space, such that $\rho_{(A_1 B_1) \dots (A_n B_n)} = |\Phi\rangle\langle\Phi|$ is an extension of $\rho_{A_1 \dots A_n}$. The auxiliary space $\mathcal{H}_B$ can be chosen of the same dimension as $\mathcal{H}_A$.

If we apply the de Finetti theorem to such a purification, we find that

$$\rho_{(A_1 B_1) \dots (A_k B_k)} \approx \int d\psi_{AB}\, p(\psi_{AB})\, |\psi\rangle_{AB}^{\otimes k}\, \langle\psi|_{AB}^{\otimes k}$$

up to error $d^2 k/(n-k)$, since now the single-particle Hilbert space has dimension $\dim \mathcal{H}_A \otimes \mathcal{H}_B = d^2$. If we take a partial trace over the $B$ systems, we obtain a mixture of product states (which can now be mixed):

$$\rho_{A_1 \dots A_k} \approx \int d\psi_{AB}\, p(\psi_{AB})\, \mathrm{tr}_B[|\psi\rangle\langle\psi|_{AB}]^{\otimes k}$$

Moreover, the trace distance never increases when we take the partial trace. Thus we have proved the following: If $\rho_{A_1 \dots A_n}$ is a permutation-invariant state on $(\mathbb{C}^d)^{\otimes k}$ then its reduced density matrices can be approximated by mixtures of product states

$$\rho_{A_1 \dots A_k} \approx \int d\mu(\rho)\, \rho^{\otimes k}$$

up to error $d^2/(n-k)$ in trace distance. Here, $d\mu$ is some probability measure on the space of mixed states that depends on the state $\rho$.

Nowadays, there are many variants of the de Finetti theorem that quantify the monogamy of entanglement in interesting and useful ways. Surveying some of them could make for an interesting course project.

## Bibliography


Robert König and Renato Renner. A de finetti representation for finite symmetric quantum states, *Journal of Mathematical physics*, 46(12):122108, page 122108, 2005. arXiv:quant-ph/0410229.

Andrew C Doherty, Pablo A Parrilo, and Federico M Spedalieri. Distinguishing separable and entangled states, *Physical Review Letters*, 88(18):187904, page 187904, 2002. arXiv:quant-ph/0112007.

Andrew C Doherty, Pablo A Parrilo, and Federico M Spedalieri. Complete family of separability criteria, *Physical Review A*, 69(2):022308, page 022308, 2004. arXiv:quant-ph/0308032.

Fernando GSL Brandao, Matthias Christandl, Aram W Harrow, and Michael Walter. The Mathematics of Entanglement. 2016. arXiv:1604.01790.

**PHYSICS 491: Symmetry and Quantum Information** April 18, 2017

Shannon theory, data compression, spectrum estimation

*Lecture 5* *Michael Walter, Stanford University*

*These lecture notes are not proof-read and are offered for your convenience only. They include additional detail and references to supplementary reading material. I would be grateful if you email me about any mistakes and typos that you find.*

## 5.1 A first glance at information theory: data compression

Imagine that Alice has acquired a biased coin, with heads coming up with $p = 75\%$ probability. She is excited about her purchase and wants to let Bob know about the result of her coin flips. If the flips the coin once, how many bits does she need to communicate the result to Bob? Clearly, she needs at least one bit. Otherwise, since both outcomes are possible, she would make an error 25% of the time.

Now suppose that Alice flips her coin not only once, but a large number of times – say $n$ times. She would still like to communicate the results of her coin flips to Bob. Clearly, Alice could send over one bit immediately after each coin flip. Can she do better by waiting and looking at the whole sequence of coin flips? If we assume that her coin flips are *independent* then we would expect that heads will come up $j \approx pn$ times for large enough $n$. This suggests the following compression scheme:

- If the number of coin flips $j$ is not within $(p \pm \varepsilon)n$, Alice gives up and signals failure.
- Otherwise, she sends $j$ over to Bob, as well as the index $i$ of her particular sequence of coin flips in a list $\mathcal{L}_j$ that contains all possible coin flips with $j$ heads and $n - j$ tails.

If our two protagonists have agreed beforehand on the lists $\mathcal{L}_j$ (you might call them a *codebook*), then Bob will have no trouble decoding the sequence of coin flips – he merely looks up the $i$-th entry in the list $\mathcal{L}_j$. Note that, for any fixed $\varepsilon > 0$, the probability of failure in the first step is arbitrarily small – this is a consequence of the strong law of large numbers.

**Remark.** *If failure is not an option, Alice may instead send the uncompressed sequence of coin flips instead of giving up. This leads to a similar analysis and will be left as an exercise.*

What is the compression rate of this protocol? To send $j$, we need roughly $(\log n)/n$ bits per coin flip, which is negligible for large $n$.[1] How many sequences are there with $j$ heads and $n - j$ tails? This is given by the binomial coefficient $\binom{n}{j}$. Thus, to communicate the index $i \in \{1, \dots, \binom{n}{j}\}$, Alice needs to send roughly $\frac{1}{n} \log \binom{n}{j}$ bits per coin flip. To estimate this rate, we note that for any $x \in [0, 1]$,

$$x^j (1-x)^{n-j} \binom{n}{j} \leq (x + (1-x))^n = 1$$

[1] Here and throughout the rest of these lecture notes, log denotes the logarithm to the base two.

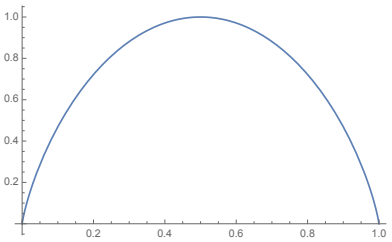



Figure 7: The binary entropy function $h(p)$ defined in eq. (5.2).

and hence, choosing $x = \frac{j}{n}$,

$$\frac{1}{n}\log\binom{n}{j} \leq -\frac{j}{n}\log\frac{j}{n} - (1-\frac{j}{n})\log(1-\frac{j}{n}). \tag{5.1}$$

Since $\frac{j}{n} \approx p$, the right-hand side is approximately equal to the *binary (Shannon) entropy*

$$h(p) := -p\log p - (1-p)\log(1-p). \tag{5.2}$$

See fig. 7 for a plot of the binary entropy function.

In total, the protocol sketched above will achieve a compression rate of roughly $h(p) \leq 1$ bits per coin flip. E.g., $h(75\%) = 0.81$ – so Alice achieve savings of roughly of 19%. We can get arbitrarily close to $h(p)$ by decreasing $\varepsilon$, at the expense of $n$ having to become larger and larger for the probability of failure to vanish. It is not hard to see the compression rate $h(p)$ is optimal. This is Shannon's famous *noiseless coding theorem* – it is called "noiseless" since we assume that the communication line from Alice to Bob is perfect.

The coin flip example illustrates the traditional core principles of information theory, or *Shannon theory*: We are interested in finding *optimal asymptotic rates* for information processing tasks such as compression (the task that you have just solved), information transmission over noisy channels, etc. *Quantum information theory* has very analogous goals – except that now we are dealing with *quantum information* rather than classical information. At a fundamental level, this means that we are interested in the asymptotic behavior of a large number of independent copies of a quantum state $\rho$, i.e., in $\rho^{\otimes n}$ for large $n$ (the so-called *i.i.d.* limit).

**Example 5.1** (Warning)**.** *If $\rho = |\psi\rangle\langle\psi|$ is a pure state then $\rho^{\otimes n} = |\psi\rangle^{\otimes n}\langle\psi|^{\otimes n}$ is an operator on the symmetric subspace. We explored this quite extensively in lectures 2 to 4. However, as soon as $\rho$ is a mixed state, $\rho^{\otimes n}$ is no longer supported purely on the symmetric subspace. A simple example is the maximally mixed state $\tau = \mathbb{1}/d$. Clearly, $\tau^{\otimes n} = \mathbb{1}/d^n$ is supported on all of $(\mathbb{C}^d)^{\otimes n}$. Thus we need to develop new techniques.*

**Remark 5.2.** *In recent years, there has been an increased interest in understanding optimal information processing rates in non-asymptotic scenarios. This is largely beyond the scope of these lectures, although we might have a brief glance at these ideas in the last week of class.*

## 5.2 Spectrum estimation

Today, we will start developing the appropriate machinary for working with independent copies of a quantum state, $\rho^{\otimes n}$. A popular approach that you will find in many textbooks is to work in the

eigenbasis of $\rho$ in order to turn the quantum problem into a classical problem (e.g., Nielsen and Chuang, 2002, Wilde, 2013). *In this class we will pursue a different, and arguably more "invariant" route.* What this means exactly will become clear over the coming lectures, but the practical advantage of exploiting all available symmetries will be that we are naturally led to *universal protocols* that work not only for a single state $\rho$ but for whole classes of states (e.g., all states $\rho$ with the same eigenvalues).

When we discussed the symmetric subspace, our motivation was to solve an estimation problem, namely, the estimation an unknown pure state $|\psi\rangle$ given $n$ copies $|\psi\rangle^{\otimes n}$. Today, we will again be interested in an estimation problem: We would like to estimate the eigenvalues of an unknown density operator $\rho$, given $n$ copies $\rho^{\otimes n}$. That is, if $p_1 \geq \cdots \geq p_d$ denote the eigenvalues of $\rho$ then we would like to define a measurement $\{Q_{\hat{p}}\}$ such that, when we measure on $\rho^{\otimes n}$, we obtain an outcome such that $\hat{p} \approx p$. This task is known as the *spectrum estimation* problem (Keyl and Werner, 2001). It is an easier problem than estimating the full density operator $\rho$, and it allows us to focus on the key difference between pure and mixed states – their eigenvalue spectrum. We will spend the rest of today's lecture and part of lecture 6 solving the spectrum estimation problem.

The tools that we will develop in the course of solving this problem will be prove useful for working with asymptotic quantum information more generally. In lecture 7, we will use them to compress quantum information and we will also sketch how one can estimate the entire unknown quantum state $\rho$ from $\rho^{\otimes n}$, thereby solving the task of quantum states estimation of mixed state, also known as *quantum state tomography.*

## Symmetries of the spectrum estimation problem

If $\rho$ is a quantum state on $\mathbb{C}^d$ then the state $\rho^{\otimes n}$ is a quantum state on $(\mathbb{C}^d)^{\otimes n}$. As discussed in section 3.1, this space is a representation for two groups: (i) the permutation group $S_n$, with representation operators $R_\pi$, and (ii) the unitary group $U(d)$, with representation operators $T_U = U^{\otimes n}$. The operator $\rho^{\otimes n}$ is *permutation-invariant* as defined last time, i.e., it commutes with permutations, $[R_\pi, \rho^{\otimes n}] = 0$ for all $\pi \in S_n$. We may explicitly verify this on a product basis:

$$\begin{aligned} R_\pi \rho^{\otimes n} |x_1, \ldots, x_n\rangle &= R_\pi \left(\rho |x_1\rangle \otimes \ldots \otimes \rho |x_n\rangle\right) = \rho |x_{\pi^{-1}}\rangle \otimes \ldots \otimes \rho |x_{\pi^{-1}}\rangle \\ &= \rho^{\otimes n}\left(|x_{\pi^{-1}}\rangle \otimes \ldots \otimes |x_{\pi^{-1}}\rangle\right) = \rho^{\otimes n} R_\pi |x_1, \ldots, x_n\rangle . \end{aligned}$$

On the other hand, $\rho^{\otimes n}$ does *not* commute with the action of the unitary group: Instead,

$$U^{\otimes n} \rho^{\otimes n} U^{\dagger, \otimes n} = (U \rho U^\dagger)^{\otimes n}$$

which amounts to replacing $\rho \mapsto U \rho U^\dagger$. *This operation changes the eigenbasis, but leaves the eigenvalues the same.* In other words, while the permutation symmetry is a symmetry of the state, the unitary symmetry is a symmetry of the problem that we are trying to solve! This suggests that both symmetries should play an important role, and it prompts us to investigate the representation $(\mathbb{C}^d)^{\otimes n}$ more closely.

**Example 5.3** (Warmup)**.** *Suppose we are just given two copies of the unknown quantum state, i.e., $\rho^{\otimes 2}$. This is a density operator on*

$$(\mathbb{C}^d)^{\otimes 2} = \mathrm{Sym}^2(\mathbb{C}^d) \oplus \bigwedge\nolimits^2(\mathbb{C}^d).$$

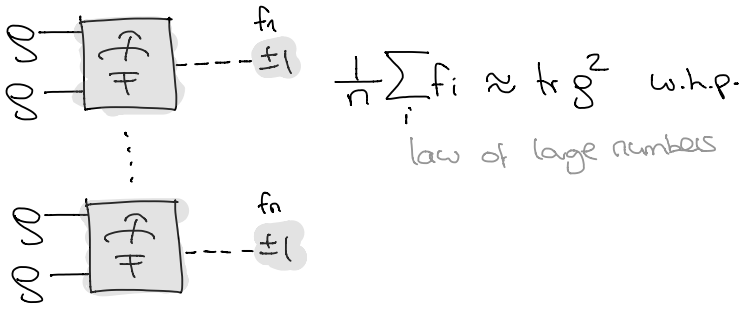



Figure 8: By measuring the swap operator on independent copies of $\rho^{\otimes 2}$, we can estimate the purity $\operatorname{tr} \rho^2$ of the quantum state.

*Both the symmetric and the antisymmetric subspace are irreducible representations (you show this in problem 2.3 for the symmetric subspace; the antisymmetric subspace can be treated completely analogously). The permutation group $S_2$ has just two elements, the identity permutation and the nontrivial permutation $\pi = 1 \leftrightarrow 2$. The corresponding operator is known as the* swap operator

$$F = R_\pi = \sum_{a,b} |a,b\rangle \langle b,a| .$$

*It commutes both with the representation of $U(d)$ as well as the one of $S_2$ (any operator commutes with itself and with the identity matrix). Thus, $F$ is an observable of exactly the kind that we are looking for. Its eigenvalues are $+1$ on the symmetric subspace and $-1$ on the antisymmetric subspace. In problem 2.1, you show the following "swap trick":*

$$\langle F \rangle = \operatorname{tr} \rho^{\otimes 2} F = \operatorname{tr} \rho^2.$$

*The quantity $\operatorname{tr} \rho^2$ is called the* purity *of $\rho$, since it is equal to 1 only if the state $\rho$ is a pure state. (It is closely related to* Rényi-2 entropy $S_2(\rho) = -\log \operatorname{tr} \rho^2$ *that you study in problem 2.1.) The important point though is that if $\rho$ has eigenvalues $r_1 \geq \cdots \geq r_d$ then*

$$\operatorname{tr} \rho^2 = \sum_k r_k^2,$$

*and hence already this simple measurement allows us to learn something about the eigenvalues of $\rho$.*

*Just to be perfectly clear: When measuring the observable $F$ on $\rho^{\otimes 2}$, the measurement outcome is either $\pm 1$. Only when repeated many times on independent copies of $\rho^{\otimes 2}$ will these signs average to $\operatorname{tr} \rho^2$ (fig. 8).*

For qubits, $d = 2$, example 5.3 provides a complete solution (since $p_1 + p_2 = 1$, there is only a single unknown, which can be determined from $\operatorname{tr} \rho^2 = p_1^2 + p_2^2$). In the following, we will discuss a different solution which fully exploits the symmetries of the problem and generalizes readily to any $d$. The protocol is due to Keyl and Werner (2001) and we will follow the proof strategy of Christandl and Mitchison (2006). It will prove to be an important building block for several quantum information applications that we will discuss in the remainder of this course.

## Towards a solution of the spectrum estimation problem

We start by decomposing the Hilbert space of $n$ qubits into irreducible representations of $\mathrm{SU}(2)$. The answer can be written ni the form:

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_j V_j \otimes \mathbb{C}^{m(n,j)}, \tag{5.3}$$

where $V_j$ denotes the irreducible representation of $\mathrm{SU}(2)$ with spin $j$ and $m(n,j)$ are the multiplicities that we need to determine. That is, for any $U \in \mathrm{SU}(2)$ we have that

$$U^{\otimes n} \cong \bigoplus_j T_U^{(j)} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}} = \left[\begin{array}{c|c|c} T_U^{(0)} \otimes \mathbb{1}_{\mathbb{C}^{m(n,0)}} & & \\ \hline & T_U^{(1/2)} \otimes \mathbb{1}_{\mathbb{C}^{m(n,1/2)}} & \\ \hline & & \ddots \end{array}\right]. \tag{5.4}$$

Here we write $T_U^{(j)}$ for the representation operators of the spin-$j$ representation.

Recall that we are looking for a measurement that commutes with both the action of $\mathrm{SU}(2)$ and $S_n$. The projection operator $P_j$ onto a direct summand in eq. (5.3) seems like a plausible candidate. It measures the total spin – generalizing example 5.3. By design, $P_j$ commutes with the action of the unitary group. Indeed, in view of eq. (5.4) it clearly commutes with $U \in \mathrm{SU}(2)$, and any element of $U(2)$ can be written in the form $e^{i\phi}U$ where $U \in \mathrm{SU}(2)$.

Does $P_j$ also commute with the action of $S_n$? Yes, this follows from $[R_\pi, U^{\otimes n}] = 0$ and Schur's lemma, as you will verify in problem 3.5. We have found the desired candidate measurement!

In the remainder of today's lecture, we will start analyzing the projective measurement $\{P_j\}$. That is, we would like to bound the probabilities

$$\Pr(\text{outcome j}) = \mathrm{tr}\left[\rho^{\otimes n} P_j\right]. \tag{5.5}$$

Note that these probabilities remain unchanged if we substitute $\rho \mapsto U\rho U^\dagger$, as $P_j$ commutes with $U^{\otimes n}$. Since we can always diagonalize $\rho$ by a unitary there is thus no harm in assuming that $\rho$ is already a diagonal matrix

$$\rho = \begin{pmatrix} p & \\ & 1-p \end{pmatrix} \tag{5.6}$$

with $p \geq 1-p$, i.e., $p \in [\frac{1}{2}, 1]$. Our goal will be to show that (5.5) is exponentially small in $n$ most of the time – except when we can obtain a good estimate of the spectrum from $j$ (we will later see that $\hat{p} := \frac{1}{2} + \frac{j}{n} \approx p$ provides such an estimate).

How would we go about analyzing eq. (5.5)? The idea is that $\rho^{\otimes n}$ looks just like the representation operators $U^{\otimes n}$ – except that $\rho$ is almost never a unitary matrix! To go beyond unitaries, we need to talk about some more representation theory.

## Representation theory of $\mathrm{SU}(2)$ and $\mathrm{SL}(2)$

As we have already used several times in this course, the irreducible representations of $\mathrm{SU}(2)$ are labeled by their spin $j \in \{0, \frac{1}{2}, 1, \frac{3}{2}, \dots\}$. We denote the spin-$j$ irrep by $V_j$ and its representation operators by $T_U^{(j)}$. The representation $V_j$ is of dimension $2j+1$.

**Remark 5.4.** *In your quantum mechanics class, you have probably analyzed the representation theory of* $\mathrm{SU}(2)$ *by considering its "generators": For any traceless Hermitian matrix* $H$, $U = \exp(iH)$ *is in* $\mathrm{SU}(2)$. *Given a representation* $\tilde{\mathcal{H}}$ *of* $\mathrm{SU}(2)$ *with representation operators* $\tilde{T}_U$, *we can define*

$$\widetilde{H} = \frac{1}{i}\frac{d}{dt}\Big|_{=0}\tilde{T}_{\exp(itH)}.$$

*Sometimes this is called the representation of the Lie algebra of* $\mathrm{SU}(2)$ *(though technically speaking the Lie algebra of* $\mathrm{SU}(2)$ *consists of the* anti*hermitian traceless matrices). Note that the assignment* $H \mapsto \widetilde{H}$ *is linear. Since the real vector space of traceless Hermitian matrices is spanned by the Pauli operators X,Y,Z (the "generators"), we can fully understand the representation* $\tilde{\mathcal{H}}$ *by considering the operators* $\widetilde{X}, \widetilde{Y}, \widetilde{Z}$.

*In your quantum mechanics class, you likely followed this approach to analyze the irreducible representations of* $\mathrm{SU}(2)$. *For example, you might remember that* $V_j$ *has a basis* $|j,m\rangle$, *where* $m = -j, \ldots, j-1, j$, *such that*

$$\widetilde{Z}\,|j,m\rangle = 2m\,|j,m\rangle\,.$$

*Moreover,*

$$\widetilde{Q} = (\widetilde{X})^2 + (\widetilde{Y})^2 + (\widetilde{Z})^2 = 4j(j+\frac{1}{2})\mathbb{1}_{V_j}.$$

*The operator* $\widetilde{Q}$ *is called the* quadratic Casimir operator *of* $\mathrm{SU}(2)$, *and we used the fact that it acts by a scalar on each irreducible representation of* $\mathrm{SU}(2)$ *in lecture 1 to find a qubit .*

In the previous lectures, we used to great effect that the symmetric subspace is irreducible – and you will show this in problem 2.3 by following precisely the strategy outlined in the preceding remark. This means that $\mathrm{Sym}^n(\mathbb{C}^2)$ ought to be one of the spin-$j$ irreps. It is very easy to see that $j = \frac{n}{2}$, and we record this important fact:

$$V_j \cong \mathrm{Sym}^{2j}(\mathbb{C}^2). \tag{5.7}$$

It gives us a very simple way of realizing the spin-$j$ representation concretely, as will be prove useful in just a momenet.

An important fact that was perhaps never explicitly spelled out in your quantum mechanics class is the following: Any unitary representation of $\mathrm{SU}(2)$ can be extended to a (holomorphic, non-unitary) representation of the group $\mathrm{SL}(2)$ in a unique way. For example, our representation $T_U = U^{\otimes n}$ of $\mathrm{SU}(2)$ on $(\mathbb{C}^d)^{\otimes n}$ can be extended to $T_g = g^{\otimes n}$ for $g \in \mathrm{SL}(2)$. We can also restrict this action to the symmetric subspace. Since we can define the spin-$j$ representation using the symmetric subspace (eq. (5.7)), we can likewise define $T_g^{(j)}$ for any $g \in \mathrm{SL}(2)$. Thus, for any $g \in \mathrm{SL}(2)$, eq. (5.3) reads

$$g^{\otimes n} \cong \bigoplus_j T_g^{(j)} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}}. \tag{5.8}$$

**Remark.** *A general way of defining the extension from* $\mathrm{SU}(2)$ *to* $\mathrm{SL}(2)$ *is as follows: In remark 5.4, we defined* $\widetilde{H}$ *for Hermitian matrices we can safely extend it by linearity to arbitrary complex traceless matrices* $M$. *But then* $\exp(M)$ *is an arbitrary matrix in* $\mathrm{SL}(2)$ *and this allows us to extend an arbitrary unitary representation of* $\mathrm{SU}(2)$ *to* $\mathrm{SL}(2)$*: For* $g = \exp(M)$, *define* $R_g := \exp(\widetilde{M})$. *It is not hard to see that a subspace is invariant for* $\mathrm{SU}(2)$ *iff it is invariant for the operators* $\widetilde{H}$ *iff it is invariant for the operators* $\widetilde{M}$ *iff it is invariant for* $\mathrm{SL}(2)$. *This can be used to argue that the finite-dimensional representation theory of* $\mathrm{SU}(2)$ *and of* $\mathrm{SL}(2)$ *is completely identical.*

### Bounding the probability distribution

Why is this important? We are interested in understanding the operator $\rho^{\otimes n}$ on $(\mathbb{C}^2)^{\otimes n}$. Suppose that our density matrix $\rho$ has no zero eigenvalues. Then it is is invertible and

$$\tilde{\rho} := \rho/\sqrt{\det \rho}$$

is an element in the group $\mathrm{SL}(2)$, and we can interpret $\tilde{\rho}^{\otimes n}$ as the corresponding representation operator on $(\mathbb{C}^2)^{\otimes n}$! By eq. (5.8), it follows that

$$\rho^{\otimes n} = (\det \rho)^{n/2}\, \tilde{\rho}^{\otimes n} \cong (\det \rho)^{n/2} \bigoplus_j T^{(j)}_{\tilde{\rho}} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}} = \bigoplus_j \underbrace{(\det \rho)^{n/2}\, T^{(j)}_{\tilde{\rho}}}_{=:T^{(n,j)}_{\rho}} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}} \tag{5.9}$$

By continuity, this equation can be extended to all $\rho \geq 0$.

**Remark.** *Since any operator $X$ can be infinitesimally perturbed to become invertible, we can use the same strategy to analyze $X^{\otimes n}$ for arbitrary operators $X$ on $\mathbb{C}^2$.*

As a consequence of eq. (5.9), our desired probability (5.5) reads

$$\operatorname{tr}\left[P_j \rho^{\otimes n}\right] = \operatorname{tr}\left[T^{(n,j)}_{\rho} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}}\right] = (\det \rho)^{n/2} \operatorname{tr}\left[T^{(j)}_{\tilde{\rho}} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}}\right] = m(n,j)(\det \rho)^{n/2} \operatorname{tr}\left[T^{(j)}_{\tilde{\rho}}\right].$$

How can we compute the right-hand side trace? By eq. (5.7) we can simply compute the trace of $\tilde{\rho}^{\otimes 2j}$ on the symmetric subspace:

$$\operatorname{tr}\left[T^{(j)}_{\tilde{\rho}}\right] = \sum_{k=0}^{2j} \langle\!\langle k \| \tilde{\rho}^{\otimes 2j} \| k \rangle\!\rangle = (\det \rho)^{-j} \sum_{k=0}^{2j} \langle\!\langle k \| \rho^{\otimes 2j} \| k \rangle\!\rangle = (\det \rho)^{-j} \sum_{k=0}^{2j} p^k (1-p)^{2j-k} \leq (\det \rho)^{-j} (2j+1) p^{2j}.$$

Here, we compute the trace in the occupation number basis

$$\| k \rangle\!\rangle \propto |0\rangle^{\otimes k} |1\rangle^{\otimes (2j-k)} + \text{permutations}$$

of the symmetric subspace (see eq. (2.5) and problem 2.3). In the third step, we used that $\rho$ is diagonal, and in the last step we bounded each summand by $p^{2j}$ using that $p \geq 1-p$ (see eq. (5.6)). Thus:

$$\begin{aligned} &\operatorname{tr}\left[T^{(n,j)}_{\rho}\right] = (\det \rho)^{n/2} \operatorname{tr}\left[T^{(j)}_{\tilde{\rho}}\right] \leq (2j+1)(\det \rho)^{n/2-j} p^{2j} = (2j+1) p^{\frac{n}{2}+j} (1-p)^{\frac{n}{2}-j} \\ &= (2j+1) 2^{n\left[\left(\frac{1}{2}+\frac{j}{n}\right)\log p + \left(\frac{1}{2}-\frac{j}{n}\right)\log(1-p)\right]} = (2j+1) 2^{n\left[\hat{p}\log p + (1-\hat{p})\log(1-p)\right]}. \end{aligned} \tag{5.10}$$

where we have defined $\hat{p} := \frac{1}{2} + \frac{j}{n}$. If we plug this back into the preceding equation then we obtain

$$\operatorname{tr}\left[P_j \rho^{\otimes n}\right] \leq (2j+1) m(n,j) 2^{n\left[\hat{p}\log p + (1-\hat{p})\log(1-p)\right]}.$$

This already looks quite suggestively as if the eigenvalue $p$ has something to do with $\hat{p}$!

However, we still need to determine the multiplicities $m(n,j)$. We will do this next time – it will allow us to solve the spectrum estimation problem completely. We will then put the tools developed into a more general context and use them to tackle a number of important applications.

## Bibliography

Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.

Mark M Wilde. *Quantum information theory*. Cambridge University Press, 2013.

Michael Keyl and Reinhard F Werner. Estimating the spectrum of a density operator, *Physical Review A*, 64(5):052311, page 052311, 2001. arXiv:quant-ph/0102027.

Matthias Christandl and Graeme Mitchison. The spectra of quantum states and the Kronecker coefficients of the symmetric group, *Communications in Mathematical Physics*, 261(3):789–797, pages 789–797, 2006. arXiv:quant-ph/0409016.

**PHYSICS 491: Symmetry and Quantum Information** April 20, 2017

Solution of the spectrum estimation problem

*Lecture 6* *Michael Walter, Stanford University*

*These lecture notes are not proof-read and are offered for your convenience only. They include additional detail and references to supplementary reading material. I would be grateful if you email me about any mistakes and typos that you find.*

## 6.1 Solution of the spectrum estimation problem

Last time we started discussing the *spectrum estimation* problem for qubits. Given $\rho^{\otimes n}$, where $\rho$ had eigenvalues $p \geq 1-p$, we wanted to design a measurement that tells us information about $p \in [\frac{1}{2}, 1]$. For this, we considered the decomposition of $(\mathbb{C}^2)^{\otimes n}$ into irreducible representations for $\mathrm{SU}(2)$:

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_j V_j \otimes \mathbb{C}^{m(n,j)} \tag{6.1}$$

and defined $P_j$ as the projector on the spin-$j$ summand. We were led to these projectors because we were looking for a measurement that respected all the symmetries: the unitary invariance of the spectrum of $\rho$ as well the permutation invariance of $\rho^{\otimes n}$. In fact, $P_j$ is the most fine-grained measurement that commutes with $U^{\otimes n}$ and with $R_\pi$ (problem 3.5). Hoping that $\{P_j\}$ might prove to be a good measurement for solving the spectrum estimation problem, we started to calculate the probability

$$\Pr(\text{outcome } j) = \operatorname{tr}\left[P_j \rho^{\otimes n}\right] = ? \tag{6.2}$$

We will now finish this calculation. Our goal will be to show that this probability is exponentially small in $n$, unless

$$\hat{p} := \frac{1}{2} + \frac{j}{n} \approx p.$$

Thus we will find that the measurement outcome $j$ will lead to a good estimate $\hat{p} \approx p$ with very high probability.

The key idea to calculating (6.2) was to extend both $(\mathbb{C}^2)^{\otimes n}$ as well as the spin-$j$ representations $V_j$ from $\mathrm{SU}(2)$ to $\mathrm{SL}(2)$ (see eq. (5.8)). Using that $\rho/\sqrt{\det \rho}$ is an element in $\mathrm{SL}(2)$, we found that

$$\rho^{\otimes n} \cong (\det \rho)^{n/2} \bigoplus_j T^{(j)}_{\rho/\sqrt{\det \rho}} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}} = \bigoplus_j \underbrace{(\det \rho)^{n/2}\, T^{(j)}_{\rho/\sqrt{\det \rho}}}_{=: T^{(n,j)}_\rho} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}} \tag{6.3}$$

(eq. (5.9)) for arbitrary density operators $\rho$. It followed that:

$$\operatorname{tr}\left[P_j \rho^{\otimes n}\right] = m(n,j) \operatorname{tr}\left[T^{(n,j)}_\rho\right]. \tag{6.4}$$

Last time, we calculated the right-hand side trace but not the multiplicities $m(n,j)$. For this, we will recall one last fact that you learned in your quantum mechanics class when studying the total

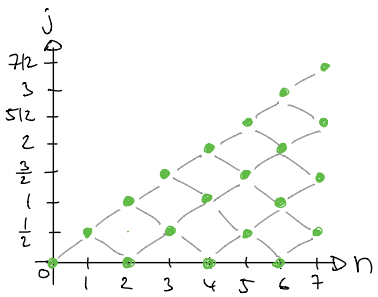


Figure 9: By iterating the Clebsch-Gordan decomposition for $V_{1/2} \otimes V_{1/2} \otimes \ldots$, we obtain a decomposition of $(\mathbb{C}^2)^{\otimes n}$ into irreducible representations of SU(2).

angular momentum. Given two irreducible representations $V_{j_1}$ and $V_{j_2}$, we can consider their tensor product $V_{j_1} \otimes V_{j_2}$. This is a representation of SU(2), with $U$ acting by $T_U^{(j_1)} \otimes T_U^{(j_2)}$. In general this representation is not irreducible and so it can be decomposed it into irreducibles. The famous *Clebsch-Gordan rule* tells us what that this decomposition look as follows:

$$V_{j_1} \otimes V_{j_2} \cong V_{j_1+j_2} \oplus V_{j_1+j_2-1} \oplus \ldots \oplus V_{|j_1-j_2|}$$

In particular, for $j_2 = \frac{1}{2}$, we have

$$V_j \otimes V_{1/2} = \begin{cases} V_{j+1/2} \oplus V_{j-1/2} & \text{if } j > 0 \\ V_{1/2} & \text{if } j = 0 \end{cases}. \tag{6.5}$$

Since a single qubit is nothing but a spin-$1/2$ representation, this allows us to decompose $(\mathbb{C}^2)^{\otimes n}$ by successively applying the Clebsch-Gordan rule (6.5):

$$\begin{aligned} (\mathbb{C}^2)^{\otimes 1} &\cong V_{1/2} \\ (\mathbb{C}^2)^{\otimes 2} &\cong V_{1/2} \otimes V_{1/2} = V_1 \oplus V_0 \\ (\mathbb{C}^2)^{\otimes 3} &\cong (V_1 \oplus V_0) \otimes V_{1/2} = V_{3/2} \oplus \big(V_{1/2} \oplus V_{1/2}\big) \oplus V_0 \\ &\vdots \end{aligned}$$

This process is visualized in fig. 9 and the general result is as follows: The multiplicity $m(n, j)$ of $V_j$ in $(\mathbb{C}^2)^{\otimes n}$ is precisely equal to the number of paths from $(0, 0)$ to $(n, j)$ in fig. 9.

How can we estimate the number of paths? Any path can be specified by a number of $n$ "ups" and "downs". The number of "ups" $u$ must satisfy $(u - (n - u))/2 = u - n/2 = j$ in order to end up at $(n, j)$. Thus there are at most $\binom{n}{\frac{n}{2}+j}$ such paths. (This is only an upper bound because paths that go below zero are invalid.) As a consequence of eq. (5.1), this means that

$$m(n, j) \leq \binom{n}{\frac{n}{2} + j} \leq 2^{nh(\hat{p})}, \tag{6.6}$$

where we recall the binary Shannon entropy

$$h(\hat{p}) = -\hat{p} \log \hat{p} - (1 - \hat{p}) \log(1 - \hat{p})$$

from the compression of coin flips in section 5.1. Thus the multiplicites $m(n,j)$ grow at most exponentially, with exponent is given by precisely by the binary entropy. Note that, as a consequence

$$\operatorname{rk} P_j = (\dim V_j)\, m(n,j) \leq (2j+1)2^{nh(\hat{p})} \leq (n+1)2^{nh(\hat{p})}. \tag{6.7}$$

This fact will prove important later for information theoretic applications.

**Remark.** *More generally, given two representations $\mathcal{H}$ and $\mathcal{H}'$ of some group $G$, we can always consider their tensor product $\mathcal{H} \otimes \mathcal{H}'$ as a representation of the group $G$, with representation operators $T_g \otimes T_{g'}$. Note that this is precisely the same notation as used in eq.* (6.1) *if we think of $\mathbb{C}^{m(n,j)}$ as an $m(n,j)$-dimensional trivial representation of* SU(2).

The other ingredient in eq. (6.4) is the trace of the operator $T_\rho^{(n,j)}$. Last time, we computed the following upper bound (eq. (5.10)):

$$\operatorname{tr}\left[T_\rho^{(n,j)}\right] \leq (2j+1)p^{\frac{n}{2}+j}(1-p)^{\frac{n}{2}-j}$$

We can rewrite this as follows,

$$\begin{aligned}
\operatorname{tr}\left[T_\rho^{(n,j)}\right] &\leq (2j+1)2^{n\left[\hat{p}\log p + (1-\hat{p})\log(1-p)\right]} \leq (2j+1)2^{-n\left[\hat{p}\log\frac{1}{p} + (1-\hat{p})\log\frac{1}{1-p}\right]} \\
&= (2j+1)2^{-n\left[-\hat{p}\log\hat{p} - (1-\hat{p})\log(1-\hat{p}) + \hat{p}\log\frac{\hat{p}}{p} + (1-\hat{p})\log\frac{1-\hat{p}}{1-p}\right]} \\
&\leq (2j+1)2^{-n\left[h(\hat{p}) + \delta(\hat{p}\|p)\right]},
\end{aligned} \tag{6.8}$$

where we have introduced the *binary relative entropy*

$$\delta(\hat{p}\|p) = \hat{p}\log\frac{\hat{p}}{p} + (1-\hat{p})\log\frac{1-\hat{p}}{1-p}.$$

**Remark.** *The relative entropy is an important quantity in information theory and statistics. Note that it is* not *symmetric under exchanging $p \leftrightarrow \hat{p}$.*

What is the purpose of this rewriting? If we plug eqs. (6.6) and (6.8) into eq. (6.4) we obtain the following result:

$$\Pr(\text{outcome } j) = \operatorname{tr}\left[P_j \rho^{\otimes n}\right] \leq (2j+1)2^{-n\delta(\hat{p}\|p)} \tag{6.9}$$

The point now is that the relative entropy is a distance measure between probability distributions: It is nonnegative and $\delta(\hat{p}\|p) = 0$ if and only if $p = \hat{p}$. More quantitatively, it satisfies the following inequality, a special case of the so-called *Pinsker's inequality* (problem 3.4):

$$\delta(\hat{p}\|p) \geq \frac{2}{\ln 2}(\hat{p}-p)^2 \tag{6.10}$$

Thus, unless $\hat{p} \approx$, the probability in eq. (6.9) is exponentially small. This at last allows us to solve the spectrum estimation problem for qubits:

Given $\rho^{\otimes n}$, perform a total spin measurement in the state $\rho^{\otimes n}$ using the projective measurement $\{P_j\}$. Upon outcome $j$, estimate that the maximal eigenvalue of the state $\rho$ is $\hat{p} = \frac{1}{2} + \frac{j}{n}$. Then,

$$\Pr(|\hat{p}-p| > \varepsilon) = \sum_{j \text{ s.th. } |\hat{p}-p|>\varepsilon} \operatorname{tr}\left[P_j \rho^{\otimes n}\right] \leq (n+1)^2 2^{-\frac{2}{\ln 2} n\varepsilon^2}, \tag{6.11}$$

where we have used that eqs. (6.9) and (6.10), that $2j+1 \leq n+1$, and that the sum runs certainly over no more than $n+1$ values of $j$. This means that $\hat{p} \approx p$ with very high probability.

In lecture 10, we will discuss how to implement the spectrum estimation measurement concretely by a quantum circuit. Spectrum estimation has been realized experimentally in Beverland et al. (2016).

## 6.2 Towards quantum data compression

There is another interpretation of what we have just achieved. For fixed $\varepsilon > 0$, consider the projection operator

$$\tilde{P}_n = \sum_{j \text{ s.th. } |\hat{p}-p|<\varepsilon} P_j$$

on all sectors $j$ in eq. (6.1) for which $|\hat{p} - p| < \varepsilon$. Equation (6.11) asserts precisely that

$$\operatorname{tr}\left[\tilde{P}_n \rho^{\otimes n}\right] \to 1 \tag{6.12}$$

as $n \to \infty$, which in turn implies that

$$\tilde{P}_n \rho^{\otimes n} \tilde{P}_n \approx \rho^{\otimes n}.$$

This means that if we perform a measurement $\{\tilde{P}_n, \mathbb{1} - \tilde{P}_n\}$ on $\rho^{\otimes n}$ then, for large $n$, this measurement will proceed with very high probability and leave the state $\rho^{\otimes n}$ almost unchanged. We will call the subspace $\tilde{\mathcal{H}}_n$ that $\tilde{P}_n$ projects on a *typical subspace* for $\rho^{\otimes n}$ (although we caution that the traditional definition is somewhat different).

Since the binary entropy is continuous,

$$|\hat{p} - p| < \varepsilon \Rightarrow |h(\hat{p}) - h(p)| < \delta(\varepsilon)$$

for some function $\delta$ such that $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$. (To obtain a more quantitive bound, you could use Fannes' inequality that your derive in problem 3.4.) In view of eq. (6.7), this implies that the subspace that $\tilde{P}_n$ projects on has dimension no larger than

$$\dim \tilde{\mathcal{H}}_n \leq (n+1)^2 2^{n(h(p)+\delta(\varepsilon))}. \tag{6.13}$$

Thus, the post-measurement state is supported on a possibly much smaller subspace of roughly $n(h(p) + \delta)$ qubits.

Let us end with a word of caution: In the coin flip example in section 5.1, the purpose of the compression scheme was to communicate Alice' actual sequence of coin flips to Bob – *not* for Bob to flip its own biased coin. The latter would only reproduce the probability distribution of the biased coin, but not the actual sequence of coin flips observed by Alice! In the same way, the purpose of a quantum compression scheme is *not* simply to produce the quantum state $\rho^{\otimes n}$ at Bob's side.

In fact, compression protocols are usually designed for known information sources. In the coin flip example, this means that Bob already knows the parameter $p$ of the coin and could flip his own biased coin with no communication required at all. (Since its quantum analogue is the eigenvalue spectrum of $\rho$, you might in fact be concerned that spectrum estimation solves a problem that is completely irrelevant to compression.)

Next time, we will carefully define what it means to compress quantum information and see that the properties in eqs. (6.12) and (6.13) above are nevertheless precisely the properties required to solve the problem.

# Bibliography

Michael E Beverland, Jeongwan Haah, Gorjan Alagic, Gretchen K Campbell, Ana Maria Rey, and Alexey V Gorshkov. Spectrum estimation of density operators with alkaline-earth atoms. 2016. arXiv:1608.02045.

**PHYSICS 491: Symmetry and Quantum Information** April 25, 2017

# Schur-Weyl duality, quantum data compression, tomography

*Lecture 7* *Michael Walter, Stanford University*

*These lecture notes are not proof-read and are offered for your convenience only. They include additional detail and references to supplementary reading material. I would be grateful if you email me about any mistakes and typos that you find.*

Today, we will summarize the "Schur-Weyl toolbox" that we developed in lecture 6 to solve the spectrum estimation problem. We will then apply it to the task of compressing a quantum information source.

## 7.1 The Schur-Weyl toolbox

Let us recapitulate the machinary that we developed to solve the spectrum estimation problem. Just like any representation of $\mathrm{SU}(2)$, the Hilbert space of $n$ qubits can be decomposed in the form

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_j V_j \otimes \mathbb{C}^{m(n,j)}.$$

Last time, we discussed that the action of $\mathrm{SU}(2)$ could be extended first to $\mathrm{SL}(2)$ and then to arbitrary operators on $\mathbb{C}^2$: In eq. (6.3), we found that

$$X^{\otimes n} \cong \bigoplus_j T_X^{(n,j)} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}},$$

where

$$T_X^{(n,j)} = (\det X)^{n/2}\, T^{(j)}_{X/\sqrt{\det X}}$$

is a polynomial in the matrix elements of $X$ and hence makes sense for arbitrary $X$. You can verify this, e.g., by using the symmetric subspace model of the spin-$j$ representation. In particular, this formula applies to unitary matrices $U$. It follows that the operators $T_U^{(n,j)}$ define a representation of the unitary group $U(2)$, which we will denote by $V_{n,j}$. Here, $j$ tells us the spin of the representation when restricted to matrices in $\mathrm{SU}(2)$, and $n$ reminds us of the way that multiples $\alpha \mathbb{1}_{\mathbb{C}^2}$ of the identity matrix act by $\alpha^n$. Since every unitary can be written as $\alpha U$ with $\alpha \neq 0$ and $U \in \mathrm{SU}(2)$, this information specifies the representation completely. It is clear that the representations $V_{n,j}$ are irreducible, since they are even irreducible for the subgroup $\mathrm{SU}(2)$.

We can also consider $(\mathbb{C}^2)^{\otimes n}$ as a representation of the symmetric group $S_n$. Since $[R_\pi, U^{\otimes n}] = 0$, Schur's lemma (lemma 3.2) implies that

$$R_\pi \cong \bigoplus_j \mathbb{1}_{V_{n,j}} \otimes R_\pi^{(n,j)}$$

for some operators $R_\pi^{(n,j)}$ on $\mathbb{C}^{m(n,j)}$. So far, the Hilbert spaces $\mathbb{C}^{m(n,j)}$ were simply vectors spaces – but now we see that the operators $R_\pi^{(n,j)}$ turn them into representations of $S_n$. We will denote these

representations by $W_{n,j}$. The representations $W_{n,j}$ are irreducible and pairwise inequivalent. You will verify this and the following statements in problem 3.5.

Thus, we have the following decomposition of the Hilbert space of $n$ qubits:

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus V_{n,j} \otimes W_{n,j} \tag{7.1}$$

which holds as a representation of both $U(2)$ and $S_n$ (equivalently, of the product group $U(2) \times S_n$). The spaces $\{V_{n,j}\}$ and $\{W_{n,j}\}$ are pairwise inequivalent, irreducible representations of $U(2)$ and of $S_n$, respectively. Equation (7.1) shows that they are "paired up" perfectly in the $n$-qubit Hilbert space. This result is known as *Schur-Weyl duality*, and it has a number of important consequences.

For example, any operator that commutes with all $U^{\otimes n}$ is necessarily a linear combination of the operators $R_\pi$. Dually, any operator that commutes with all $R_\pi$ is necessarily a linear combination of operators of the form $X^{\otimes n}$ (even $U^{\otimes n}$). Mathematically, we say that the two representations span each other's *commutants*. Schur-Weyl duality also implies that the projectors

$$P_j \cong \bigoplus_{j'} \delta_{j,j'} \mathbb{1}$$

not only have both symmetries of the spectrum estimation problem (i.e., that they commute with both the $U^{\otimes n}$ and the $R_\pi$), but that they are in fact the most fine-grained projective measurement with this property.

Table 1 assembles all important facts and formulas about the representation theory of the $n$-qubit Hilbert space that we obtained past week (the "Schur-Weyl toolbox"). It contains one formula, eq. (7.5), which is proved just like eq. (6.8). We will use it to solve the quantum state tomography problem in section 7.3 below.

**Remark 7.1.** *So far, we have simply argued on abstract grounds that the Hilbert space of n qubits can be decomposed in the form* (7.1)*. Here, the notation* $\cong$ *means that there exists a unitary intertwiner from the left-hand side to the right-hand side. But if we want to implement, e.g., spectrum estimation in practice, we need to know what this unitary operator looks like. In other words, we need to find a unitary operator that implements the transformation from the product basis*

$$|x_1, \ldots, x_n\rangle = |x_1\rangle \otimes \ldots \otimes |x_n\rangle$$

*to the Schur-Weyl basis*

$$|j, m, k\rangle$$

*where* $j \in \{\ldots, \frac{n}{2} - 1, \frac{n}{2}\}$*,* $m \in \{-j, \ldots, j\}$*,* $k \in \{1, \ldots, m(n, j)\}$*. Note that the right-hand side is* not *a tensor product of three spaces, because the allowed values for* $m$ *and* $k$ *depend on* $j$*. However, we can embed it into a larger space where* $|j, m, k\rangle = |j\rangle \otimes |m\rangle \otimes |k\rangle$ *is a product basis vector. In lecture 10 we will learn how to implement this transformation – called the* quantum Schur transform *– by a quantum circuit.*

## Beyond qubits

How does the preceding generalize beyond qubits? This is best explained by making a simple coordinate change and instead of by $(n, j)$ parametrizing all representations by

$$\lambda = (\lambda_1, \lambda_2) = \left(\frac{n}{2} + j, \frac{n}{2} - j\right) \in \mathbb{Z}^2.$$

We can identify $\lambda$ with a so-called *Young diagram* with two rows, where we place $\lambda_1$ boxes in the first and $\lambda_2$ boxes in the second row. E.g.,

$$\lambda = (7,3) = \ydiagram{7,3}$$

We always demand that $\lambda_1 \geq \lambda_2$, corresponding to $j \geq 0$. Note that the total number of boxes is $\lambda_1 + \lambda_2 = n$, while $2j = \lambda_1 - \lambda_2$ is the difference of row lengths.

If we write $V_\lambda := V_{n,j}$ and $W_\lambda := W_{n,j}$, then the Schur-Weyl duality (7.1) becomes

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_\lambda V_\lambda \otimes W_\lambda, \tag{7.2}$$

where we sum over all Young diagrams with $n$ boxes and at most two rows.

**Remark 7.2.** *In example 3.1, we already met the irreducible representations of $S_3$ and labeled them by Young diagrams. The representations $W_{\ydiagram{3}}$ and $W_{\ydiagram{2,1}}$ that occur in $(\mathbb{C}^2)^{\otimes 3}$ are precisely the ones that we already met in example 3.1. You will verify this in problem 4.1.*

*On the other hand, because the antisymmetric subspace $\bigwedge^3 \mathbb{C}^2 = \{0\}$ is zero-dimensional, the sign representation $W_{\ydiagram{1,1,1}}$ does not appear at all.*

The notation $\lambda$ is quite suggestive. Indeed, let us define the *normalization* of a Young diagram $\lambda$ by $\bar{\lambda} = \lambda/n = (\lambda_1/n, \lambda_2/n)$, where $n = \lambda_1 + \lambda_2$. This is a probability distribution, and

$$\bar{\lambda}_1 = \frac{1}{2} + \frac{j}{n} = \hat{p}, \quad \bar{\lambda}_2 = \frac{1}{2} - \frac{j}{n} = 1 - \hat{p}.$$

Thus, spectrum estimation can be rephrased as follows: When we measure $\{P_\lambda\}$ on $\rho^{\otimes n}$ and the outcome is $\lambda$, then $\bar{\lambda}$ is a good estimate for the spectrum of $\rho$.

The key point now is the following: eq. (7.2) generalizes quite directly from qubits to arbitrary $d$. This is because the irreducible representations of $U(d)$ are labeled by Young diagrams with (at most) $d$ rows, while the irreps of $S_n$ are labeled by Young diagrams with $n$ boxes. See, e.g., Harrow (2005), Christandl (2006), Walter (2014) for further detail.

## 7.2 Quantum data compression

We will now discuss quantum data compression in more precise terms (Schumacher, 1995). We consider a *quantum information source* described by an ensemble $\{p_x, |\psi_x\rangle\}$ of qubit pure states. It emits sequences

$$|\psi(\vec{x})\rangle = |\psi_{x_1}\rangle \otimes \ldots \otimes |\psi_{x_n}\rangle \in (\mathbb{C}^2)^{\otimes n}$$

with probabilities

$$p(\vec{x}) = p_{x_1} \ldots p_{x_n}.$$

The task of *quantum data compression* is to design an compressor that encodes a sequence $|\psi(\vec{x})\rangle \in (\mathbb{C}^2)^{\otimes n}$ into some state of $Rn$ qubits and a corresponding decompressor – $R$ is called the *compression rate* at *block length* $n$. Unlike the state of a coin, we cannot in general hope to precisely recover the original state. Instead, the decompressor should produce a state $|\tilde{\psi}(\vec{x})\rangle$ that has high overlap with the original state (say, on average):

$$\sum_{\vec{x}} p(\vec{x})\, E\left[|\langle\psi(\vec{x})|\tilde{\psi}(\vec{x})\rangle|^2\right] \approx 1. \tag{7.6}$$

**Schur-Weyl duality:**

$$
\begin{aligned}
(\mathbb{C}^2)^{\otimes n} &\cong \bigoplus_{j=\ldots,\frac{n}{2}-1,\frac{n}{2}} V_{n,j} \otimes W_{n,j}, \\
X^{\otimes n} &\cong \bigoplus_j T_X^{(n,j)} \otimes \mathbb{1}_{W_{n,j}}, \quad \text{where} \quad T_X^{(n,j)} \coloneqq (\det X)^{n/2}\, T^{(j)}_{X/\sqrt{\det X}}, \\
R_\pi &\cong \bigoplus_j \mathbb{1}_{V_{n,j}} \otimes R_\pi^{(n,j)}.
\end{aligned}
$$

$V_{n,j}$ and $W_{n,j}$ are pairwise inequivalent, irreducible representations of $U(2)$ and $S_n$, respectively.

**Dimensions:**

$$
\begin{aligned}
\dim V_{n,j} &= 2j+1 \leq n+1, \\
\dim W_{n,j} &= m(n,j) \leq 2^{nh(\hat{p})}, \quad \text{where} \quad \hat{p} = \frac{1}{2} + \frac{j}{n}.
\end{aligned} \tag{7.3}
$$

**Estimates:**

$$
2^{-n\left[h(\hat{p})+\delta(\hat{p}\|p)\right]} \leq \operatorname{tr}\left[T_\rho^{(n,j)}\right] \leq (2j+1)2^{-n\left[h(\hat{p})+\delta(\hat{p}\|p)\right]} \quad \text{where } \rho \text{ has eigenvalues } \{p, 1-p\}, \tag{7.4}
$$

More generally, if $X \geq 0$ and $k > 0$:

$$
\operatorname{tr}\left[T_{X^k}^{(n,j)}\right] \leq (2j+1)2^{-nk\left[h(\hat{p})+\delta(\hat{p}\|x)\right]} (\operatorname{tr} X)^{kn}, \quad \text{where } \frac{X}{\operatorname{tr} X} \text{ has eigenvalues } \{x, 1-x\}. \tag{7.5}
$$

**Spectrum estimation:**

$$
\begin{aligned}
P_j &\cong \bigoplus_{j'} \delta_{j,j'} \mathbb{1}_{V_{n,j}} \otimes \mathbb{1}_{W_{n,j}}, \\
\rho^{\otimes n} &\cong \bigoplus_j T_\rho^{(n,j)} \otimes \mathbb{1}_{V_{n,j}} =: \bigoplus_j p_j\, \rho_{V_{n,j}} \otimes \tau_{W_{n,j}},
\end{aligned}
$$

and so

$$
\begin{aligned}
p_j = \operatorname{tr}\left[P_j \rho^{\otimes n}\right] &\leq (n+1)^2 2^{-n\delta(\hat{p}\|p)} \leq (n+1)^2 2^{-n\frac{2}{\ln 2}(\hat{p}-p)^2} \\
\operatorname{tr}\left[\widetilde{P}_n \rho^{\otimes n}\right] &\geq 1 - (n+1)^2 2^{-n\frac{2}{\ln 2}\varepsilon^2}
\end{aligned}
$$

where $\widetilde{P}_n \coloneqq \sum_{j:|\hat{p}-p|<\varepsilon} P_j$ is the projector onto the "$\varepsilon$-spectrum typical subspace" of $\rho$.

Table 1: The Schur-Weyl toolbox for i.i.d. quantum information theory (in the case of qubits).

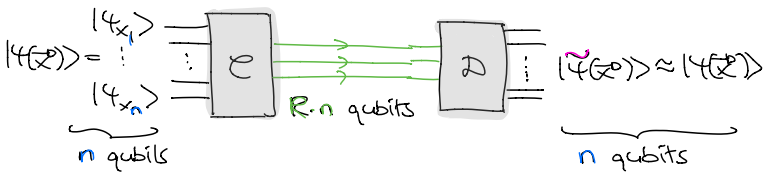



Figure 10: Illustration of the compression of a quantum information source.

The average value $E[\dots]$ refers to the fact that the decompressed state $|\tilde{\psi}(\vec{x})\rangle$ for a given $|\psi(\vec{x})\rangle$ is typically random. How should we go about solving this problem?

At the end of last lecture, we constructed, for every $p \in [\frac{1}{2}, 1]$ and $\varepsilon > 0$, projectors $\tilde{P}_n = \sum_{j:|p-\hat{p}|<\varepsilon} P_j$ onto a subspace $\tilde{\mathcal{H}}_n$ of $(\mathbb{C}^2)^{\otimes n}$ such that

$$\dim \tilde{\mathcal{H}}_n \leq (n+1)^2 2^{nh(p)+\delta(\varepsilon)},$$

and

$$\mathrm{tr}\left[\tilde{P}_n \rho^{\otimes n}\right] \to 1 \tag{7.7}$$

for all density operators $\rho$ with eigenvalues $\{p, 1-p\}$ (cf. table 1).

What is the density operator $\rho$ that we should care about? Every ensemble gives rise to a density operator $\rho = \sum_x p_x |\psi_x\rangle\langle\psi_x|$, describing the *average state* emitted by the qubit source (we discussed this in lecture 3).

**Remark.** *The states $|\psi_x\rangle$ emitted by the source do* not *have to be orthogonal. Thus, the eigenvalues $\{p, 1-p\}$ of $\rho$ used to construct $\tilde{P}_n$ are* not *in general the same as the probabilities $p_x$ of the ensemble. E.g., in problem 1.3 you computed that $\frac{1}{2}(|1\rangle\langle 1| + |-\rangle\langle -|)$ has eigenvalues $\frac{1}{2} \pm \frac{1}{2\sqrt{2}}$.*

This suggests the following two-step *quantum data compression protocol* that is completely analogous to the way by which we compressed sequences of coin flips in section 5.1:

- Alice measures the observable $\tilde{P}_n$ (i.e., she performs the projective measurement $\{\tilde{P}_n, \mathbb{1} - \tilde{P}_n\}$).
- If the outcome is 1, then the post-measurement state

$$\frac{\tilde{P}_n |\psi(\vec{x})\rangle}{\|\tilde{P}_n |\psi(\vec{x})\rangle\|} \in \tilde{\mathcal{H}}_n$$

  lives in the subspace $\tilde{\mathcal{H}}$ only. Thus, Alice can send this state over to Bob by transmitting roughly $n(h(p)+\delta)$ qubits.
- If the outcome is 0, she simply sends over some fixed state. (Alternatively, she might signal failure – as in our coin flip protocol.)

Bob then uses the sent-over state in $\tilde{\mathcal{H}} \subseteq (\mathbb{C}^2)^{\otimes n}$ as the decompressed state. For large $n$, this protocol achieves a quantum compression rate of roughly $R = h(p) + \delta$.

**Remark.** *As discussed in class, in order to be able to "send over" the post-measurement state we first need to identify the subspace $\tilde{\mathcal{H}}$ with $N \approx n(h(p)+\delta)$ many qubits. For example, Alice*

*could first apply a unitary $U$ that maps the subspace $\tilde{\mathcal{H}}$ into the subspace of states of the form $|\phi\rangle_{A_1\ldots A_N} \otimes |0\rangle_{A_{N+1}} \otimes \ldots \otimes |0\rangle_{A_n}$. Alice would then send over the first $N$ of her qubits. Upon receiving those, Bob would apply $U^\dagger$ to obtain the decompressed sate. Mathematically, this is not very interesting, but physically this is quite important because we usually do not get to choose our physical qubits!*

Let us analyze the average fidelity (7.6) achieved by our compression protocol. If the input state is $|\psi(\vec{x})\rangle$ then according to the Born rule the measurement of the observable $\tilde{P}_n$ yields outcome 1 with probability

$$q(\vec{x}) := \langle\psi(\vec{x})|\tilde{P}_n|\psi(\vec{x})\rangle\,.$$

As already mentioned above, the post-measurement state in this case is

$$\frac{\tilde{P}_n\,|\psi(\vec{x})\rangle}{\|\tilde{P}_n\,|\psi(\vec{x})\rangle\|} = \frac{\tilde{P}_n\,|\psi(\vec{x})\rangle}{\sqrt{q(\vec{x})}}$$

and so this is the state $|\tilde{\psi}(\vec{x})\rangle$ that Bob obtains at his end. Thus, eq. (7.6) can be bounded as follows:

$$\sum_{\vec{x}} p(\vec{x})\, E\left[|\langle\psi(\vec{x})|\tilde{\psi}(\vec{x})\rangle|^2\right] \geq \sum_{\vec{x}} p(\vec{x})\, q(\vec{x})\, |\langle\psi(\vec{x})|\frac{\tilde{P}_n\,|\psi(\vec{x})\rangle}{\sqrt{q(\vec{x})}}\rangle|^2 = \sum_{\vec{x}} p(\vec{x})\, |\langle\psi(\vec{x})|\tilde{P}_n|\psi(\vec{x})\rangle|^2$$

$$= \sum_{\vec{x}} p(\vec{x})\, q(\vec{x})^2 \geq \left(\sum_{\vec{x}} p(\vec{x})\, q(\vec{x})\right)^2$$

The first inequality is because we lower bound the overlap in the case that the outcome is 0; the second inequality is Jensen's inequality that we already used previously in class. But now note that

$$\sum_{\vec{x}} p(\vec{x})\, q(\vec{x}) = \sum_{\vec{x}} p(\vec{x})\ \mathrm{tr}\left[|\psi(\vec{x})\rangle\,\langle\psi(\vec{x})|\,\tilde{P}_n\right] = \mathrm{tr}\left[\left(\sum_{\vec{x}} p(\vec{x})\ |\psi(\vec{x})\rangle\,\langle\psi(\vec{x})|\right)\tilde{P}_n\right] = \mathrm{tr}\,\rho^{\otimes n}\tilde{P}_n \to 1$$

by eq. (7.7). Thus, our compression protocol will successfully compress a quantum information source with associated density operator $\rho$ at rate $h(p) + \delta$. We can make $\delta > 0$ arbitrarily small by choosing $\varepsilon > 0$ smaller and smaller (note, however, that this requires the block length $n$ to increase). This compression rate turns out to be optimal, as we will find in lecture 9.

This motivates us to define the *von Neumann entropy* of a density operator $\rho$ as

$$S(\rho) = -\,\mathrm{tr}\,\rho\log\rho.$$

For qubits, $S(\rho) = h(p)$, as you can verify by expanding the trace in the eigenbasis of $\rho$. Thus, the von Neumann entropy that you might already know from your quantum physics research has a well-defined *operational interpretation*: It is the optimal compression rate of *any* quantum information source with associated density operator $\rho$. This is in complete analogy to one of the many roles played by the Shannon entropy in classical information theory. Next time, we will discuss a number of other meanings of the von Neumann entropy related to entanglement.

**Remark.** *This emphasizes a fundamental idea in information theory: We often seek to find characterizations of entropic quantities as optimal rates for information processing tasks. In the asymptotic limit of $n \to \infty$, the von Neumann entropy plays a rather universal role. However, at finite block lengths $n < \infty$, there is not just one entropy but a whole zoo of entropic quantities that information theorists are interested in, each targeted at different tasks (Faist, 2013).*

An interesting fact about our compression protocol is that the projectors $\tilde{P}_n$ depended only on the eigenvalues $p$ and $1-p$, not on the eigenbasis of the density operator $\rho$. Thus the compression protocol designed above works for all qubit sources whose associated density operator has eigenvalues $\{p, 1-p\}$. On problem 3.3 you will show that by a very simple extension of this idea one obtains a truely *universal quantum compression protocol*: *It is targeted at a fixed compression rate $S_0$ and is able to compress an arbitrary qubit source whose density operator has entropy $S(\rho) < S_0$.* This universality is *not* automatic using the textbook approach to asymptotic quantum information theory, and it is one of the main advantages of the Schur-Weyl toolbox introduced in section 7.1.

## 7.3 Supplement: Quantum state tomography

Starting with our solution to the spectrum estimation problem, we can also solve the problem of estimating an unknown quantum state from many copies – a task that is also known as *quantum state tomography*. That is, given $\rho^{\otimes n}$, we would like to design a POVM measurement that yields an estimate $\hat{\rho} \approx \rho$ with high probability,

$$\rho^{\otimes n} \longrightarrow \hat{\rho} \approx \rho.$$

We follow the approach of Haah et al. (2015) (but see the original paper by Keyl (2006) and other exciting recent works by O'Donnell and Wright (2015, 2016)).

### The POVM measurement

The general idea is that we would like to design a POVM measurement $\{Q_{j,U}\}$ with *two* outcomes $j$ and $U$, such that the estimate is

$$\hat{\rho} = U \begin{pmatrix} \hat{p} & \\ & 1-\hat{p} \end{pmatrix} U^\dagger.$$

As before, $j$ is a discrete parameter that we will use for the eigenvalue estimate $\hat{p} = \frac{1}{2} + \frac{j}{n}$, while $U$ is a continuous parameter that rotates the diagonal matrix with eigenvalues $\{\hat{p}, 1-\hat{p}\}$ into the proper eigenbasis. In order for $\{Q_{j,U}\}$ to be a POVM, we need that $Q_{j,U} \geq 0$ as well as

$$\sum_j \int dU\, Q_{j,U} = \mathbb{1}, \tag{7.8}$$

where $\int dU$ denotes the *Haar measure* of the unitary group $U(2)$. This is the unique probability measure on $U(2)$ such that all expectation values are invariant under the substitution $U \mapsto VUW^\dagger$ for unitaries $V$, $W$. Moreover, we would like for the POVM $\{Q_{j,U}\{$ to be a refinement of $\{P_j\}$, so that the $j$ have the same meaning as before. That is, if we forget about the outcome $U$ then we would like to get the same statistics for $j$ as if we performed the measurement $\{P_j\}$. Mathematically, this means that we would like to demand that

$$\int dU\, Q_{j,U} = P_j \tag{7.9}$$

which clearly implies eq. (7.8). What does such a POVM look like?

We will make the ansatz

$$Q_{j,U} \propto P_j \hat{\rho}^{\otimes n} P_j = P_j U^{\otimes n} \begin{pmatrix} \hat{p} & \\ & 1-\hat{p} \end{pmatrix}^{\otimes n} U^{\dagger,\otimes n} P_j \cong T^{(n,j)}_{\hat{\rho}} \otimes \mathbb{1}_{W_{n,j}}.$$

To see that this is natural, we observe that, for $j = \frac{n}{2}$, $P_j$ is the projector $\Pi_n$ onto the symmetric subspace $\mathrm{Sym}^n(\mathbb{C}^2)$. Moreover, $\hat{p} = 1$, hence

$$\hat{\rho} = U|0\rangle\langle 0|U^\dagger =: |\hat{\psi}\rangle\langle\hat{\psi}|,$$

and so

$$Q_{n/2,U} \propto \Pi_n \hat{\rho}^{\otimes n} \Pi_n = |\hat{\psi}\rangle^{\otimes n} \langle\hat{\psi}|^{\otimes n}$$

is exactly the uniform POVM (2.8) that we used for pure state estimation in lecture 2. Thus, our POVM measurement $Q_{j,U}$ is a true generalization of what we did for pure states – that's already an encouraging sign. Moreover, note that $Q_{j,U}$ has permutation symmetry (i.e., $[R_\pi, Q_{j,U}] = 0$) and it is *covariant* with respect to the unitary group in the following sense: For all $V \in U(2)$,

$$\mathrm{tr}\left[\rho Q_{j,U}\right] = \mathrm{tr}\left[V\rho V^\dagger Q_{j,VU}\right],$$

where we note that estimate corresponding to the POVM element $Q_{j,VU}$ is $V\hat{\rho}V^\dagger$. We could summarize this as $\rho \mapsto V\rho V^\dagger \rightsquigarrow \hat{\rho} \mapsto V\hat{\rho}V^\dagger$.

We will now show that eq. (7.9) holds true by a suitable choice of normalization constant. For this, we first note that

$$\int dU\, Q_{j,U} \cong \underbrace{\int dU\, T^{(n,j)}_{\hat{\rho}}}_{\propto \mathbb{1}_{V_{n,j}}} \otimes \mathbb{1}_{W_{n,j}} \propto P_j$$

as a consequence of Schur's lemma. Indeed, the indicated operator is a self-intertwiner on the irreducible representation $V_{n,j}$, because

$$\begin{aligned} T^{(n,j)}_V \left(\int dU\, T^{(n,j)}_{\hat{\rho}}\right) T^{(n,j)}_{V^\dagger} &= T^{(n,j)}_V \left(\int dU\, T^{(n,j)}_U T^{(n,j)}_{\binom{\hat{p}}{1-\hat{p}}} T^{(n,j)}_{U^\dagger}\right) T^{(n,j)}_{V^\dagger} \\ = \int dU\, T^{(n,j)}_{VU} T^{(n,j)}_{\binom{\hat{p}}{1-\hat{p}}} T^{(n,j)}_{(VU)^\dagger} &= \int dU\, T^{(n,j)}_U T^{(n,j)}_{\binom{\hat{p}}{1-\hat{p}}} T^{(n,j)}_{U^\dagger} = \int dU\, T^{(n,j)}_{\hat{\rho}}; \end{aligned}$$

in the second to last step we used that the integral is invariant under the substitution $U \mapsto VU$. It is now easy to figure out the correct normalization constant – we merely need to compare traces. On the one hand, in view of the definition of $Q_{j,U}$, its trace that does not depend on $U$, and so

$$\mathrm{tr}\left[\int dU\, Q_{j,U}\right] = \mathrm{tr}\, Q_{j,U} = \mathrm{tr}\left[T^{(n,j)}_{\hat{\rho}}\right](\dim W_{n,j})$$

for any $U$ that we like. On the other hand,

$$\mathrm{tr}\, P_j = (\dim V_{n,j})(\dim W_{n,j}) = (2j+1)(\dim W_{n,j}).$$

Thus, the appropriately normalized POVM elements are

$$Q_{j,U} = \frac{2j+1}{\mathrm{tr}\left[T^{(n,j)}_{\hat{\rho}}\right]} P_j \hat{\rho}^{\otimes n} P_j.$$

### The fidelity between two quantum states

In section 4.3 we discussed the trace distance $T(\rho,\sigma)$ as a distance measure between quantum states (whether pure or mixed). Another very useful measure was the overlap, $|\langle\phi|\psi\rangle|$, which we only defined for pure states. The overlap also generalizes nicely to mixed states, but the expression is more complicated: It is the following quantity, known as the *fidelity*:

$$F(\rho,\sigma) = \operatorname{tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} = \operatorname{tr}\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}$$

As in problem 1.4, $\sqrt{M}$ denotes the square root of a positive semidefinite operator $M$, defined by taking the square root of all eigenvalues. The fidelity might seem like a strange definition – but actually it is precisely the maximal overlap that can be obtained between any two purifications. That is,

$$F(\rho,\sigma) = \max_{|\phi\rangle_{AB},|\psi\rangle_{AB}} |\langle\phi_{AB}|\psi_{AB}\rangle|$$

where we optimize over all pure states $|\phi\rangle_{AB}, |\psi\rangle_{AB}$ such that $\operatorname{tr}_B[|\phi\rangle\langle\phi|_{AB}] = \rho$, $\operatorname{tr}_B[|\psi\rangle\langle\psi|_{AB}] = \sigma$. In particular, if $\rho = |\phi\rangle\langle\phi|$ and $\sigma = |\psi\rangle\langle\psi|$ are themselves pure then the fidelity agrees with the overlap. (You can also check this explicitly from the definition, since in that case $\sqrt{\rho} = \rho$ and $\sqrt{\sigma} = \sigma$.) In general, the trace distance and fidelity are related by the *Fuchs-van de Graaf inequalities*:

$$1 - F(\rho,\sigma) \leq T(\rho,\sigma) \leq \sqrt{1 - F^2(\rho,\sigma)} \tag{7.10}$$

### Analysis of the measurement

Similarly as when analyzing the spectrum estimation measurement, our goal is to show that $\operatorname{tr}[Q_{j,U}\rho^{\otimes n}]$ is exponentially small unless $\rho \approx \hat{\rho}$. Thus, we want to bound For this, we will use the full strength of the Schur-Weyl toolbox. We start with

$$\begin{aligned}\operatorname{tr}\left[Q_{j,U}\rho^{\otimes n}\right] &= \frac{2j+1}{\operatorname{tr}\left[T_{\hat{\rho}}^{(n,j)}\right]}\operatorname{tr}\left[P_j\hat{\rho}^{\otimes n}P_j\rho^{\otimes n}\right] = \frac{2j+1}{\operatorname{tr}\left[T_{\hat{\rho}}^{(n,j)}\right]}\operatorname{tr}\left[T_{\hat{\rho}}^{(n,j)}T_{\rho}^{(n,j)}\otimes \mathbb{1}_{W_{n,j}}\right] \\ &= \frac{(2j+1)m(n,j)}{\operatorname{tr}\left[T_{\hat{\rho}}^{(n,j)}\right]}\operatorname{tr}\left[T_{\sqrt{\rho}\hat{\rho}\sqrt{\rho}}^{(n,j)}\right] = \frac{(2j+1)m(n,j)}{\operatorname{tr}\left[T_{\hat{\rho}}^{(n,j)}\right]}\operatorname{tr}\left[T_{\sqrt{\sqrt{\rho}\hat{\rho}\sqrt{\rho}}^{2}}^{(n,j)}\right].\end{aligned}$$

In the second to last step, we have used that $T_X^{(n,j)}T_Y^{(n,j)} = T_{XY}^{(n,j)}$ for arbitrary operators, as well as cyclicity of the trace. We now use the upper bound (7.3), the lower bound in (7.4) (observing that $\widehat{\rho}$ has eigenvalues $\{\hat{p}, 1-\hat{p}\}$), and the upper bound (7.5) (with $k=2$). The result is that

$$\operatorname{tr}\left[Q_{j,U}\rho^{\otimes n}\right] \leq \frac{(2j+1)2^{nh(\hat{p})}}{2^{-nh(\hat{p})}}(2j+1)2^{-2n(h(\hat{p})+\delta(\hat{p}\|x))}\left(\operatorname{tr}\sqrt{\sqrt{\rho}\hat{\rho}\sqrt{\rho}}\right)^{2n} \leq (n+1)^2F(\rho,\hat{\rho})^{2n},$$

where in the second step we used $\delta(\hat{p}\|x) \geq 0$ as well as $2j \leq n$. This is the desired upper bound. Indeed, it implies that

$$\begin{aligned}&\Pr(T(\hat{\rho},\rho) \geq \varepsilon) \leq \Pr(F(\hat{\rho},\rho)^2 \leq 1-\varepsilon^2) \leq \sum_j \int dU (n+1)^2(1-\varepsilon^2)^n \\ &\leq (n+1)^3 2^{n\log(1-\varepsilon^2)} \leq (n+1)^3 2^{-n\varepsilon^2}\end{aligned}$$

(The first inequality is a consequence of the upper bound in eq. (7.10). The last holds whenever $\varepsilon \leq \frac{1}{2}$ and is only for illustration.)

# Bibliography


Aram W Harrow. *Applications of coherent classical communication and the Schur transform to quantum information theory*. PhD thesis, 2005. arXiv:quant-ph/0512255.

Matthias Christandl. *The structure of bipartite quantum states-insights from group theory and cryptography*. PhD thesis, 2006. arXiv:quant-ph/0604183.

Michael Walter. *Multipartite quantum states and their marginals*. PhD thesis, 2014. arXiv:1410.6820.

Benjamin Schumacher. Quantum coding, *Physical Review A*, 51(4):2738, page 2738, 1995.

Philippe Faist. Welcome to the entropy zoo. 2013. URL `http://www.statslab.cam.ac.uk/biid2013/slides/EntropyZoo.pdf`. Beyond IID.

Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. 2015.

Michael Keyl. Quantum state estimation and large deviations, *Reviews in Mathematical Physics*, 18 (01):19–60, pages 19–60, 2006. arXiv:quant-ph/0412053.

Ryan O'Donnell and John Wright. Efficient quantum tomography. 2015. arXiv:1508.01907.

Ryan O'Donnell and John Wright. Efficient quantum tomography ii. 2016. arXiv:1612.00034.

**PHYSICS 491: Symmetry and Quantum Information** April 27, 2017

# Compression and entanglement, entanglement transformations

*Lecture 8* *Michael Walter, Stanford University*

*These lecture notes are not proof-read and are offered for your convenience only. They include additional detail and references to supplementary reading material. I would be grateful if you email me about any mistakes and typos that you find.*

Today we will discuss some entanglement theory of bipartite pure states (i.e., pure states $|\psi\rangle_{AB}$ with two subsystems). First, we will solve the problem of compressing subsystems of entangled states. Then we study transformations between pure states in order to compare them in their entanglement.

## 8.1 Compression and entanglement

Density operator do not only arise when describing statistical ensembles, but also when describing subsystems of entangled states. This suggests a second kind of quantum compression task (Schumacher, 1995): Given many copies of a bipartite pure state, $|\psi\rangle_{AB}^{\otimes n}$, we would like to send over the $B$-systems to Bob by first compressing the $B$-systems, sending over a minimal number of qubits, and decompressing at Bob's side (fig. 11). Thus, if $|\tilde{\psi}\rangle_{A^nB^n}$ is the state after compression and decompression, we would like that

$$|\tilde{\psi}\rangle_{A^nB^n} \approx |\psi\rangle_{AB}^{\otimes n}$$

(say, on average).

We can achieve this using the same protocol as before – but this time applied to the $B$-systems only. Let us accordingly write $\tilde{P}_{B^n}$ for the typical projector defined in terms of the eigenvalues $\{p, 1-p\}$ of $\rho_B := \mathrm{tr}_A[|\psi\rangle\langle\psi|_{AB}]$, and $\tilde{\mathcal{H}}_{B^n} \subseteq (\mathbb{C}^2)^{\otimes n}$. Then the protocol reads as follows:

- Measure the observable $\tilde{P}_{B^n}$.
- If the outcome is 1, then the post-measurement state lives in $(\mathbb{C}^2)^{\otimes n} \otimes \tilde{\mathcal{H}}_{B^n}$. We send over the $B$-systems using roughly $n(S(\rho) + \delta)$ qubits.
- If the outcome is 0, send over some arbitrary state (or simply fail).

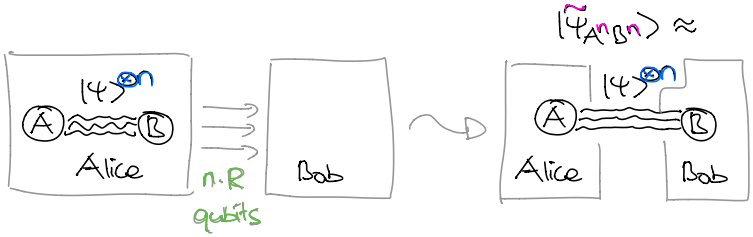



Figure 11: Alice wants to send half of her entangled states $|\psi\rangle_{AB}^{\otimes n}$ over to Bob at transmission rate $R$.

The probability that the measurement $\tilde{P}_{B^n}$ yields outcome 1 is given by

$$q := \langle \psi_{AB}^{\otimes n} | \mathbb{1}_{A^n} \otimes \tilde{P}_{B^n} | \psi_{AB}^{\otimes n} \rangle = \mathrm{tr}\left[ \rho_B^{\otimes n} \tilde{P}_{B^n} \right] \to 1.$$

In this case, the post-measurement state is

$$\frac{(\mathbb{1}_{A^n} \otimes \tilde{P}_{B^n}) |\psi_{AB}\rangle^{\otimes n}}{\sqrt{q}}$$

and its squared overlap with the original state is

$$\frac{1}{q} |\langle \psi_{AB}^{\otimes n} | \mathbb{1}_{A^n} \otimes \tilde{P}_{B^n} | \psi_{AB}^{\otimes n} \rangle|^2 = \frac{q^2}{q} = q \to 1.$$

It follows that the average overlap is at least

$$E |\langle \psi_{AB}^{\otimes n} | \tilde{\psi}_{A^n B^n} \rangle|^2 \geq q^2 \to 1.$$

Thus we have solved the problem of sending over half of an entangled state: Our compression protocol works at an asymptotic rate of $S(\rho) + \delta$ qubits. Again, it turns out that this rate is optimal – we will be able to prove this next time in lecture 9.

We thus obtain a second operational interpretation of the von Neumann entropy: When applied to the reduced density matrix $\rho_B$ of a bipartite pure state, it is the minimal rate of qubits required to send over the $B$-systems of many copies of the state from Alice to Bob. This is very intuitive and in line with our discussions in section 3.2 and problem 2.1: For pure states, the mixedness of the reduced density operators is a signature of entanglement. The more entanglement there is in $|\psi\rangle_{AB}$ the more qubits we need to send over to Bob in order to create this state between Alice and Bob. This gives a good justification why in the literature the expression

$$S_E(\psi_{AB}) = S(\rho_A) = S(\rho_B) \tag{8.1}$$

is often called the *entanglement entropy* of the bipartite pure state $|\psi\rangle_{AB}$.

**Example.** *If $|\psi\rangle_{AB} = |0\rangle_A \otimes |0\rangle_B$ then we do not need to send any quantum information – we can simply prepare the state $|0\rangle$ on Bob's end. If $|\psi\rangle_{AB} = \frac{1}{\sqrt{2}} (|00\rangle_{AB} + |11\rangle_{AB})$ is a maximally entangled state then we cannot compress the $B$-systems at all and need to send a rate of $S_E = 1$.*

The task that we just solved could be more aptly called "quantum state transfer", since we seek to transfer the state of the $B$-systems over to Bob while preserving all correlations with the purifying $A$-systems (sadly, this term is usually used with a different connotation). It is a special case of the more general problem of *quantum state merging*, where the receiver already possesses part of the state – we will have a peek at this next week.

**Remark.** *Again, note that our protocol only depended on the eigenvalues of $\rho_B$ (equivalently, of $\rho_A$). The same modification discussed in problem 3.3 allows us to build a universal protocol at fixed rate $S_0$ that works for all states whose entanglement entropy is bounded by $S_E < S_0$.*

**Remark.** *It is possible to show that the task of sending over half of a maximally entangled state at minimal qubit cost is a more difficult problem than the compression of quantum sources in the sense that whenever we have a protocol for the former we can use it to compress arbitrary quantum sources with associated density operator $\rho_B$.*

## 8.2 Entanglement transformations

Let us talk some more about entanglement. For pure states, $|\psi\rangle_{AB} \neq |\psi\rangle_A \otimes |\psi\rangle_B$ means that the state is entangled. But how can be compare and quantify different states in their entanglement? One approach is to assign to each state some arbitrary numbers that we believe reflect aspects of their entanglement properties – e.g., the entanglement entropy $S_E$ from eq. (8.1), the Rényi entropy from problem 2.1, or simply the collection of all eigenvalues of $\rho_A$ or $\rho_B$. Yet, this might seem somewhat ad hoc and so is not completely satisfactory.

A more operational approach would be to compare two states $|\phi\rangle_{AB}$ and $|\psi\rangle_{AB}$ by studying whether one can be transformed into the other: What family of operations should we consider in such a transformation? Since our goal is compare entanglement, we should only allow for operations that cannot create entanglement from unentangled states. We already briefly mentioned such a family when we discussed mixed-state entanglement in section 4.1: It is LOCC, short for *Local Operations and Classical Communication*. Here, we imagine that Alice and Bob each have their separate laboratory.

- Local operations refers to arbitrary quantum operations that can be done on Alice' and Bob's subsystems. We allow any combination of unitaries, adding auxiliary systems, performing partial traces, and measurements.
- Classical communication refers to Alice and Bob's ability to exchange measurement outcomes. Thus, Bob's local operations can depend on Alice's previous measurement outcomes, and vice versa.

Thus we are interested to study whether

$$|\psi\rangle_{AB} \overset{LOCC}{\longrightarrow} |\phi\rangle_{AB}\,.$$

If yes, then we could say that $|\psi\rangle_{AB}$ is at least as entangled as $|\phi\rangle_{AB}$ – indeed, the former is as useful as the latter for any nonlocal quantum information processing task, since we can always convert $|\psi\rangle_{AB}$ into $|\phi\rangle_{AB}$ when required.

**Remark.** *Note that the setup here is very different from quantum data compression – there, we wanted to minimize the amount of quantum communication sent. Here, we do* not *allow* any *quantum communication.*

**Example 8.1.** *Consider the* EPR pair *or* ebit $|\Phi_2^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$*, as well as its generalization, the* maximally entangled state *in d-dimensions*

$$|\Phi_d^+\rangle := \frac{1}{\sqrt{d}} \sum_i |ii\rangle\,.$$

*It is intuitive and also true that*

$$|\Phi_d^+\rangle \overset{LOCC}{\longrightarrow} |\Phi_{d'}^+\rangle$$

*if and only if $d \geq d'$. The "if" is only obvious if $d = 2^n$ and $d' = 2^{n'}$, since in this case the transformation can simply be achieved by tracing out $n - n'$ of the qubit. For the "only if", one can argue that the number of terms in the Schmidt decomposition, which is d for $|\Phi_d^+\rangle$, can never increase under LOCC. We will not prove this in class.*

*However, it might be instructive to see concretely how the conversion $|\Phi_3^+\rangle \to |\Phi_2^+\rangle$ can be achieved, since the general case can be proved completely analogously. The trick is to note that, while*

$$|\Phi_3^+\rangle = \frac{1}{\sqrt{3}} \left( |11\rangle + |22\rangle + |33\rangle \right),$$

*we can also write*

$$|\Phi_2^+\rangle = (\mathbb{1}_A \otimes U_B) \frac{1}{\sqrt{3}} \left( |\psi_1\rangle |1\rangle + |\psi_2\rangle |2\rangle + |\psi_3\rangle |3\rangle \right) \tag{8.2}$$

*where $U_B$ is some unitary on $B$. Here, the $|\psi_i\rangle \in \mathbb{C}^2$ are normalized but non-orthogonal states such that $\frac{1}{3} \sum_i |\psi_i\rangle \langle\psi_i| = \frac{1}{2} \sum_{i=1}^2 |i\rangle \langle i|$ (!). For example, you can use the three states constructed in example 2.1.*

*Alice and Bob can now apply the following LOCC protocol: First, Alice applies the isometry*

$$|i\rangle_A \mapsto |\phi_i\rangle_A \otimes \frac{1}{\sqrt{3}} \sum_j \omega^{ij} |j\rangle_{A'},$$

*where $\omega = e^{2\pi i/3}$ is a primitive third root of unity (as in problem 1.4, this can be realized by adding an auxiliary system and performing a unitary). The second system is necessary to ensure that this is indeed an isometry (recall that the $|\phi_i\rangle_A$ alone are not orthogonal). When applied to $|\Phi_3^+\rangle$, the resulting state is*

$$\frac{1}{3} \sum_{i,j} \omega^{ij} |\phi_i\rangle_A \otimes |j\rangle_{A'} \otimes |i\rangle_B .$$

*Alice now measures her auxiliary $A'$ system in the standard basis. The probability of each outcome is $1/3$. After discarding $A'$, the corresponding post-measurement state si*

$$\frac{1}{\sqrt{3}} \sum_i \omega^{ij} |\phi_i\rangle_A \otimes |i\rangle_B .$$

*This almost looks as desired – except for the phases. To get rid of them, Alice sends $j$ over to Bob, and Bob applies the diagonal unitary $|i\rangle_B \mapsto \omega^{-ij} |i\rangle_B$. We obtain*

$$\frac{1}{\sqrt{3}} \sum_i |\phi_i\rangle_A \otimes |i\rangle_B .$$

*At last, Bob applies to unitary $U_B$. Thus, Alice and Bob have obtained eq. (8.2) – done!*

The theory of exact interconversion is solved for bipartite pure states. However, there are many parameters – the entire spectrum of $\rho_A$ and $\rho_B$ matters (Nielsen, 1999, Nielsen and Vidal, 2001). It turns out that the asymptotic theory simplifies tremendouly, and we will discuss this now. The key idea is that instead of converting many copies of two arbitrary states into each other, we will study the conversion into (and from) a common resource or "currency" of entanglement. This common resource is the *maximally entangled state* or *ebit* $|\Phi_2^+\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle)$.

## 8.3 Entanglement concentration

The first problem that we want to study is the following: Given many copies of a state $|\psi\rangle_{AB}$, convert them by LOCC into as many ebits as possible:

$$|\psi\rangle_{AB}^{\otimes n} \overset{LOCC}{\longrightarrow} \approx |\Phi_2^+\rangle^{\otimes Rn}$$

Just as in the case of data compression, we are interested in the maximal rate $R$ that can be achieved with error going to zero for $n \to \infty$ (or rather its supremum). This is called the *distillable entanglement* $E_D(\psi)$ of the state $|\psi\rangle_{AB}$.

For example, $E_D(|\phi^+\rangle) = 1$ and, more generally, $E_D(|\Phi_d^+\rangle) = \log d$ (cf. example 8.1). Instead of proving directly, we will consider the general case right away.

We will approach this problem by first focusing on Alice' Hilbert space,

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_j V_{n,j}^A \otimes W_{n,j}^A,$$

where the superscripts indicate that we refer to Alice. If we write $\rho_A = \mathrm{tr}_B\left[|\psi\rangle\langle\psi|_{AB}\right]$, then

$$\rho_A^{\otimes n} \cong \bigoplus_j T_{\rho_A}^{(n,j)} \otimes \mathbb{1}_{W_{n,j}^A} = \bigoplus_j p_j \rho_{V_{n,j}^A} \otimes \tau_{W_{n,j}^A}.$$

On the right-hand side, we have written each direct summand as a probability ($p_j$) times a tensor product of density operators – this is possible since the direct summands are positive semidefinite. Note that $p_j$ is nothing but the probability of obtaining outcome $j$ when measuring $P_j$ on Alice's qubits, and recall that $\tau_{W_{n,j}^A} = \mathbb{1}_{W_{n,j}^A}/m(n,j)$ was our notation for a maximally mixed state. Now suppose that Alice does indeed perform the measurement $P_j$ on her qubits and receives outcome $j$. Then her post-measurement state is $\rho_{V_{n,j}^A} \otimes \tau_{W_{n,j}^A}$. What does the overall post-measurement state look like? Let us first guess a purification. We can purify $\rho_{V_{n,j}^A}$ to some arbitrary $|\tilde{\psi}\rangle_{V_{n,j}^A V_{n,j}^B}$, and $\tau_{W_{n,j}^A}$ to the maximally entangled state $|\Phi^+\rangle_{W_{n,j}^A W_{n,j}^B}$. Hence, a purification of her post-measurement states looks like

$$\begin{aligned}|\tilde{\psi}\rangle_{V_{n,j}^A V_{n,j}^B} \otimes |\Phi^+\rangle_{W_{n,j}^A W_{n,j}^B} &\in (V_{n,j}^A \otimes V_{n,j}^B) \otimes (W_{n,j}^A \otimes W_{n,j}^B) \\ &\cong (V_{n,j}^A \otimes W_{n,j}^A) \otimes (V_{n,j}^B \otimes W_{n,j}^B) \subseteq (\mathbb{C}^2)^{\otimes n} \otimes (\mathbb{C}^2)^{\otimes n}.\end{aligned}$$

We now use an important result that we have not met before: *Any two purifications of a quantum state are related by a unitary on the auxiliary Hilbert space.* In the present context, this means that the post-measurement state is precisely equal to

$$(\mathbb{1}_{A^n} \otimes U_{B^n})\left(|\tilde{\psi}\rangle_{V_{n,j}^A V_{n,j}^B} \otimes |\Phi^+\rangle_{W_{n,j}^A W_{n,j}^B}\right),$$

where $U_{B^n}$ is some unitary acting on Bob's Hilbert space. If Bob applies $U_{B^n}^\dagger$ and both parties discard their $V_{n,j}$-systems, they arrive at the maximally entangled state

$$|\Phi^+\rangle_{W_{n,j}^A W_{n,j}^B}.$$

But with high probability, $j$ will be such this is a maximally entangled state of dimension no smaller than $2^{n(S(\rho_A)-\delta)}$. According to example 8.1, we can convert this into $\lfloor n(S(\rho_A)-\delta)\rfloor$ ebits.

Thus we find that using the preceding entanglement concentration protocol, which is completely universal, we can distill entanglement at rates arbitrary close to the entanglement entropy $S_E(\psi) = S(\rho_A)$. In other words:

$$E_D(\psi) \geq S_E(\psi)$$

**Remark.** *Since $|\psi\rangle_{AB}^{\otimes n}$ is in the symmetric subspace $\mathrm{Sym}^n(\mathbb{C}^2 \otimes \mathbb{C}^2)$, we can identify the maximally entangled state much more precisely by using representation theory. This avoids the need to appeal to Uhlmann's theorem and makes the protocol quite a bit more concrete. You are welcome pursue this idea in problem 4.4.*

Is this rate optimal? Yes – we will show this next time using a "thermodynamics argument".

- First, we will study the reverse transformation (i.e., from perfects ebits to copies of $|\psi\rangle_{AB}$). We will show that the minimal rate of ebits required, known as the *entanglement cost* $E_C(\psi)$, is no more than $S_E(\psi)$.
- We can thus consider the "cyclic process" starting and ending at ebits:

$$\phi_+^{\otimes E_C(\psi)n} \to \psi^{\otimes n} \to \phi_+^{\otimes E_D(n)}$$

- Necessarily, $E_C(\psi) \geq E_D(\psi)$, because otherwise we could create ebits from nothing! Why is this not possible? See example 8.1 above for the exact case; the approximate case follows by tracking epsilons and deltas.
- By combing all results, we will find that $S_E(\psi) \geq E_C(\psi) \geq E_D(\psi) \geq S_E(\psi)$ so they are all equal:

$$S_E(\psi) = E_C(\psi) = E_D(\psi)$$

This is the main result of the bipartite entanglement of pure states, and it gives us two new operational interpretations of the von Neumann entropy which justify its use as an entanglemen measure for pure states: The von Neumann entropy measures the maximal rate at which ebits can distilled from many copies of a state $|\psi\rangle_{AB}$, as well as the minimal rate of ebits required to produce many copies of $|\psi\rangle_{AB}$ (up to arbitrarily high fidelity).

More generally, if we have two states $|\psi\rangle_{AB}$ and $|\phi\rangle_{AB}$ then we can convert the former into the latter by LOCC at optimal rate $S_E(\psi)/S_E(\phi)$ – this is a satisfyingly simple resolution of the question that we set out to solve.

## Discussion

Let us close with two remarks. First, the approach that we pursued above to study entanglement transformations was rooted in the idea of the ebit as a *resource*. This idea of setting up *resource theories* to compare different quantum states in their relative strength for certain tasks has been quite fruitful in quantum information theory, and there are many further examples (e.g., in quantum thermodynamics).

Second, you might wonder how the above story generalizes to mixed states $\rho_{AB}$. It turns out that in this case the entanglement theory is much more complicated. We already saw hints of this

in section 4.1 when we discussed that even deciding whether a given state $\rho_{AB}$ is separable is in general an NP-hard problem. In addition, while the same definitions can be made as above, there are many new phenomena. For example, in general we have that $E_C(\rho) > E_D(\rho)$, meaning that the conversion via ebits is in general asymptotically irreversible! In fact, there are entangled mixed states states such that $E_C(\rho) > 0$ while $E_D(\rho) = 0$. We call them *bound entangled states* – these states are entangled but no ebits can be distilled from them at a positive rate.

Relatedly (because every mixed state $\rho_{AB}$ can be purified to a tripartite pure states $|\psi\rangle_{ABC}$) the entanglement of pure states with more than two subsystems is similarly complicated.

## Bibliography


Benjamin Schumacher. Quantum coding, *Physical Review A*, 51(4):2738, page 2738, 1995.

Michael A Nielsen. Conditions for a class of entanglement transformations, *Physical Review Letters*, 83(2):436, page 436, 1999.

Michael A Nielsen and Guifré Vidal. Majorization and the interconversion of bipartite states., *Quantum Information & Computation*, 1(1):76–93, pages 76–93, 2001.

**PHYSICS 491: Symmetry and Quantum Information** May 2, 2017

# Entanglement dilution, quantum teleportation, resource inequalities

*Lecture 9* *Michael Walter, Stanford University*

*These lecture notes are not proof-read and are offered for your convenience only. They include additional detail and references to supplementary reading material. I would be grateful if you email me about any mistakes and typos that you find.*

Last time, we discussed a number of characterizations of the *entanglement entropy* $S_E(\psi) = S(\rho_A) = S(\rho_B)$ of bipartite pure state $|\psi\rangle_{AB}$:

(i) $S_E(\psi)$ describes the optimal quantum compression rate that can be achieved sending over the $B$-systems of large number of copies of $|\psi\rangle_{AB}$,

(ii) $S_E(\psi)$ is equal to both the distillable entanglement $E_D(\psi)$ and the entanglement cost $E_C(\psi)$, i.e., the rate of ebits that can be obtained from a large number of copies of $|\psi\rangle_{AB}$ and vice versa (with vanishing error as $n \to \infty$):

$$|\psi\rangle_{AB}^{\otimes n} \underset{LOCC}{\rightleftarrows} |\Phi_2^+\rangle^{\otimes nS_E(\psi)}$$

For (ii), we wanted to use the chain of inequalities

$$S_E(\psi) \geq E_C(\psi) \geq E_D(\psi) \geq S_E(\psi). \tag{9.1}$$

But we still need to prove the first inequality in eq. (9.1), i.e., that the entanglement cost is at most the entanglement entropy. Moreover, we had claimed without proof that the entanglement entropy is the optimal quantum compression rate in (i). Today, we will discuss both of these results.

## 9.1 Entanglement dilution

We first consider the task of *entanglement dilution*, where we try to construct many copies of a pure state $|\psi\rangle_{AB}$ from ebits at some rate $R$:

$$|\Phi_2^+\rangle^{\otimes Rn} \overset{LOCC}{\longrightarrow} |\psi\rangle_{AB}^{\otimes n}$$

Our idea is follows: Alice can always prepare the entangled state $|\psi\rangle_{AB}^{\otimes n}$ in her laboratory. According to (i), quantum data compression would allow her to transfer the $B$-systems to Bob at high fidelity by sending roughly $n(S_E(\psi) + \delta)$ qubits. However, sending qubits is disallowed in the current scenario. Can we instead use ebits and LOCC?

It turns out that this is indeed possible. The corresponding protocol is famously known as *quantum teleportation* (Bennett et al., 1993).

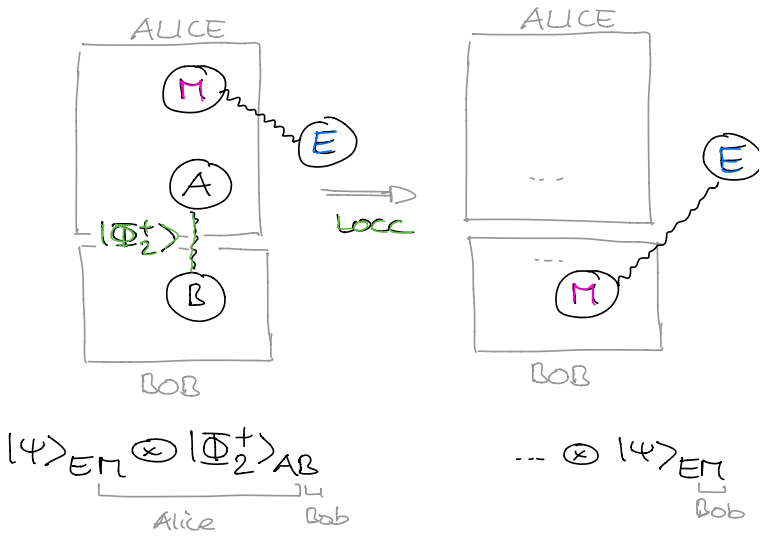



Figure 12: Illustration of the quantum teleportation task: Alice would like to send her $M$ qubit over to Bob, while preserving any entanglement with system $E$.

## 9.2 Quantum teleportation

In teleportation, Alice and Bob share an ebit $|\Phi_2^+\rangle_{AB}$ and the goal is for Alice to send an additional qubit $M$ (for "message") that is in her possession over to Bob. We will assume that the qubit $M$ is in a *completely unknown* state and that it might be entangled with some other system, denoted by $E$ (for "environment"). Just as in quantum data compression, we would like to preserve this entanglement. In mathematical terms, what we would like to achieve is the transformation

$$|\psi\rangle_{ME} \otimes |\Phi_2^+\rangle_{AB} \overset{LOCC}{\longrightarrow} |\psi\rangle_{ME},$$

where initially systems $AM$ are in Alice' possession and $B$ in Bob's possession and where we would like to end with $M$ in Bob's possession. See fig. 12 for an illustration.

The *no cloning* theorem suggests that we can only succeed with this task if Alice learns nothing about the state of $M$. On the other hand, it is clear that she has to apply *some* operation that couples her $A$ and $M$ systems in order to achieve the teleportation task. Since maximally entangled states are locally maximally mixed (problem 2.1), this suggests the following idea: Alice might measure $AM$ in a basis of maximally entangled states, such as

$$\begin{aligned}
|\phi_0\rangle &= \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right) = \left(\mathbb{1} \otimes \mathbb{1}\right)|\Phi_2^+\rangle, \\
|\phi_1\rangle &= \frac{1}{\sqrt{2}}\left(|00\rangle - |11\rangle\right) = \left(\mathbb{1} \otimes Z\right)|\Phi_2^+\rangle, \\
|\phi_2\rangle &= \frac{1}{\sqrt{2}}\left(|01\rangle + |10\rangle\right) = \left(\mathbb{1} \otimes X\right)|\Phi_2^+\rangle, \\
|\phi_3\rangle &= \frac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right) = \left(\mathbb{1} \otimes XZ\right)|\Phi_2^+\rangle,
\end{aligned} \tag{9.2}$$

which we may summarize by $|\phi_k\rangle = \left(\mathbb{1} \otimes U_k\right)|\Phi_2^+\rangle$. When she performs the projective measurement

$P_{AM,k} = |\phi_k\rangle\langle\phi_k|_{AM}$,

$$\begin{aligned}
\Pr(\text{outcome } k) &= \left(\langle\psi|_{ME} \otimes \langle\Phi_2^+|_{AB}\right)\left(P_{AM,k} \otimes \mathbb{1}_{EB}\right)\left(|\psi\rangle_{ME} \otimes |\Phi_2^+\rangle_{AB}\right) \\
&= \operatorname{tr}\left[P_{AM,k} \operatorname{tr}_{EB}\left[|\psi\rangle\langle\psi|_{ME} \otimes |\Phi_2^+\rangle\langle\Phi_2^+|_{AB}\right]\right] \\
&= \operatorname{tr}\left[P_{AM,k}\left(\operatorname{tr}_E\left[|\psi\rangle\langle\psi|_{ME}\right] \otimes \frac{\mathbb{1}_A}{2}\right)\right] \\
&= \frac{1}{2}\operatorname{tr}\left[|\phi_k\rangle\langle\phi_k|_{AM}\left(\operatorname{tr}_E\left[|\psi\rangle\langle\psi|_{ME}\right] \otimes \mathbb{1}_A\right)\right] \\
&= \frac{1}{2}\operatorname{tr}\left[\frac{\mathbb{1}_M}{2}\operatorname{tr}_E\left[|\psi\rangle\langle\psi|_{ME}\right]\right] \\
&= \frac{1}{4}\operatorname{tr}\left[|\psi\rangle\langle\psi|_{ME}\right] = \frac{1}{4}.
\end{aligned}$$

Thus her measurement outcome is completely random and uninformative, as desired. If the outcome is $k$, what is the corresponding post-measurement state on $ME$? It is given by

$$\begin{aligned}
&2\left(\langle\phi_k|_{AM} \otimes \mathbb{1}_{EB}\right)\left(|\psi\rangle_{ME} \otimes |\Phi_2^+\rangle_{AB}\right) \\
&= 2\left(\langle\phi_k|_{AM} \otimes \mathbb{1}_{EB}\right)\left(\mathbb{1}_{ME} \otimes |\Phi_2^+\rangle_{AB}\right)|\psi\rangle_{ME} \\
&= 2\left(\langle\Phi_2^+|_{AM} \otimes \mathbb{1}_{EB}\right)\left(\mathbb{1}_{ME} \otimes |\Phi_2^+\rangle_{AB}\right)\left(U_{M,k}^\dagger \otimes \mathbb{1}_E\right)|\psi\rangle_{ME} \\
&= 2\left(\mathbb{1}_E \otimes \underbrace{\left(\langle\Phi_2^+|_{AM} \otimes \mathbb{1}_B\right)\left(\mathbb{1}_M \otimes |\Phi_2^+\rangle_{AB}\right)}_{=?}\right)\left(U_{M,k}^\dagger \otimes \mathbb{1}_E\right)|\psi\rangle_{ME}.
\end{aligned}$$

Let's calculate the indicated term directly from its definition:

$$\begin{aligned}
&\left(\langle\Phi_2^+|_{AM} \otimes \mathbb{1}_B\right)\left(\mathbb{1}_M \otimes |\Phi_2^+\rangle_{AB}\right) \\
&= \frac{1}{2}\sum_{x,y}\left(\langle x|_A \otimes \langle x|_M \otimes \mathbb{1}_B\right)\left(|y\rangle_A \otimes \mathbb{1}_M \otimes |y\rangle_B\right) \\
&= \frac{1}{2}\sum_{x,y}\langle x|y\rangle|y\rangle_B\langle x|_M = \frac{1}{2}\sum_x |x\rangle_B\langle x|_M
\end{aligned}$$

Remarkably, this is nothing but the identity map from two qubit $M$ to $B$ (up to an overall factor 1/2)! As a direct consequence, we obtain that the post-measurement state is given by

$$\begin{aligned}
&2\left(\mathbb{1}_E \otimes \underbrace{\left(\langle\Phi_2^+|_{AM} \otimes \mathbb{1}_B\right)\left(\mathbb{1}_M \otimes |\Phi_2^+\rangle_{AB}\right)}_{=?}\right)\left(U_{M,k}^\dagger \otimes \mathbb{1}_E\right)|\psi\rangle_{ME} \\
&= \left(\mathbb{1}_E \otimes \sum_x |x\rangle_B\langle x|_M\right)\left(U_{M,k}^\dagger \otimes \mathbb{1}_E\right)|\psi\rangle_{ME} = \left(U_{B,k}^\dagger \otimes \mathbb{1}_E\right)|\psi\rangle_{BE},
\end{aligned}$$

where we write $|\psi\rangle_{BE}$ for the same state as $|\psi\rangle_{ME}$ but now living in the two-qubit Hilbert space corresponding to systems $BE$ rather than $ME$. If Alice sends over $k \in \{0,1,2,3\}$, which requires *two bits of classical communication*, then Bob can apply the unitary $U_{B,k}$ on his system. Thus, our two protagonists have produced the state $|\psi\rangle_{BE}$ (or $|\phi_k\rangle_{AM} \otimes |\psi\rangle_{BE}$, if we are interested in the state

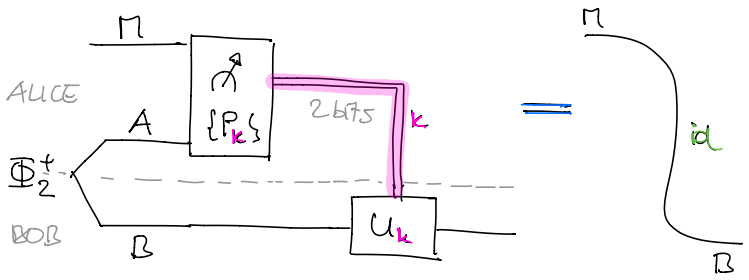



Figure 13: Quantum teleportation as a quantum circuit.

of all four quantum systems). This concludes the teleportation protocol – we have successfully sent over Alice' $M$ system to Bob while preserving all entanglement with $E$.

See fig. 13 for an illustration of the teleportation procedure in quantum circuit notation. The notation will be clear to you, but next time, in lecture 10, we will give a more systematic introduction to quantum circuits. Note that quantum teleportation is indeed an LOCC protocol – we only applied local operations and Alice needed to send over 2 bits of classical communication. We emphasize that no asymptotics was required and the teleportation procedure worked perfectly, without disturbing the sent-over state at all. Moreover, it is *composable* in the sense that we can send over a state of $N$ qubits by using $N$ ebits (and $2N$ bits of classical communication).

### From quantum compression to entanglement dilution

In particular, we can use this to convert any quantum data compression protocol into a entanglement dilution protocol at the same rate: Alice simply prepares the target state $|\psi\rangle_{AB}^{\otimes n}$ in her laboratory and then applies the data compression protocol, with quantum communication replaced by ebits and LOCC. In particular, this is true for an optimal quantum compression protocol. It follows that

$$S_E(\psi) \geq R_{\text{compr}}^{\text{opt}}(\psi) \geq E_C(\psi) \geq E_D(\psi) \geq S_E(\psi)$$

where we denote by $R_{\text{compr}}^{\text{opt}}(\psi)$ the optimal quantum compression rate for many copies of $|\psi\rangle_{AB}$. The first inequality holds because we know from lecture 7 that we can compress at rate $S_E(\psi)$; the second inequality holds by what we just discussed; and the remaining inequalities we had already justified last time. As a consequence,

$$S_E(\psi) = R_{\text{compr}}^{\text{opt}}(\psi) = E_C(\psi) = E_D(\psi).$$

We have thus proved *both* outstanding claims in points (i) and (ii) mentioned at the beginning of today's lecture.

### Entanglement swapping

Teleportation can also be used to establish entanglement between distant parties. For example, suppose that Alice and Bob are completely uncorrelated but that each of them shares an ebit with an intermediate party, Charlie, as displayed in fig. 14. Charlie and Bob can use their ebit $|\Phi_2^+\rangle_{C_2B}$ to teleport over Charlie's $C_1$ system to Bob. The result is a maximally entangled state $|\Phi_2^+\rangle_{AB}$

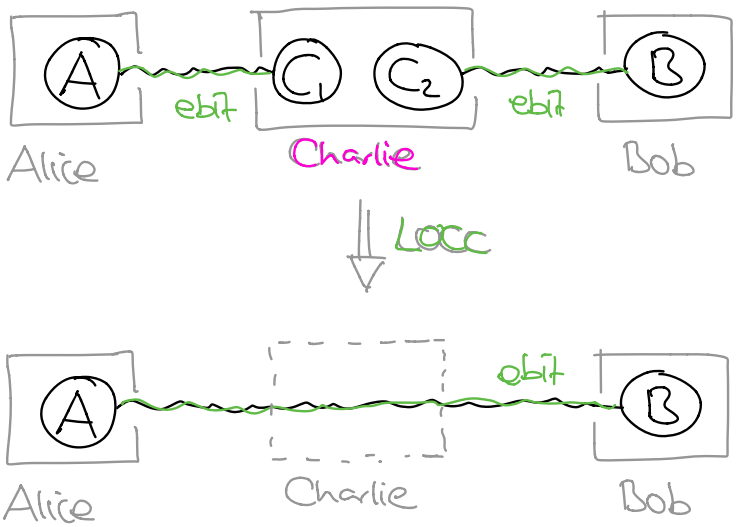



Figure 14: Entanglement swapping establishes entanglement by using quantum teleportation through intermediate parties (for simplicity, only a single intermediate party is displayed).

between Alice and Bob. (Here, we crucially used the fact that teleportation preserves the pre-existing entanglement between $C_1$ and $A$.)

The very same idea works if we have many intermediate parties. By successive teleportation, we can establish long-range entanglement between Alice and Bob. This protocol is known as *entanglement swapping*.

## 9.3 Resource inequalities

We have seen that it can be quite useful to compare different information processing resources with each other. In quantum information theory we like to use a formal notation for this. For example, we would write teleportation as a *resource inequality*

$$\text{ebit} + 2[c \to c] \geq [q \to q]. \tag{9.3}$$

This inequality means that an ebit and 2 bits of classical communication ($[c \to c]$) can be used to send one qubit of quantum communication ($[q \to q]$). Sometimes, ebits are also denoted by $[qq]$.

What other resource inequalities do we know? Clearly,

$$[q \to q] \geq \text{ebit},$$

since we can always prepare the ebit at Alice' side and send over half of it to Bob. However, $\text{ebit} \not\geq [q \to q]$, since entanglement alone cannot be used to communicate.

Another example is

$$[q \to q] \geq [c \to c],$$

since Alice can encode a classical bit $x$ into the state $|x\rangle$ of a qubit, send that qubit over, and have Bob measure $\{|x\rangle\langle x|\}$. However, $[q \to q] \not\geq 2[c \to c]$. This is a consequence of the *Holevo bound*, but we have not had time to discuss this in class.

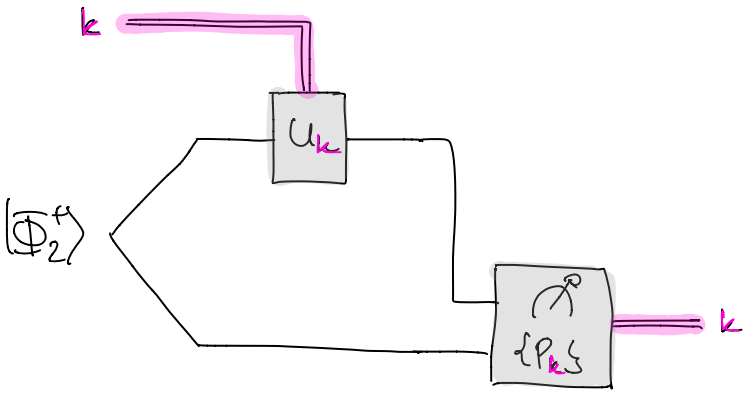



Figure 15: In superdense coding, Alice can communicate two classical bits to Bob by sending over a single qubit that is part of a shared ebit.

### Superdense coding

What is in fact possible, though, is to send over 2 classical bits by sending a qubit if we can use some entanglement:

$$\text{ebit} + [q \to q] \geq 2[c \to c]. \tag{9.4}$$

We can think of this as an analogue or "dual" of teleportation. However, it is *not* a converse, since both protocols use ebits as a resource. By combining eqs. (9.3) and (9.4), we find that

$$[q \to q] \equiv 2[c \to c] \pmod{\text{ebit}},$$

although this is not a very standard notation.

How can we achieve eq. (9.4)? The corresponding protocol is known as *superdense coding*, and it is in fact very simple: Suppose that Alice and Bob share an ebit $|\Phi_2^+\rangle_{AB}$. Alice first applies one out of the four unitaries $U_k$ to her qubit before sending it over to Bob. But now Bob has one of the four states $|\phi_k\rangle$ in his possession. Since they are orthogonal, he can simply perform the projective measurement $P_k = |\phi_k\rangle\langle\phi_k|$ to perfectly distinguish the four states and thereby recover $k$. In this way, Alice can send over an arbitrary message $k \in \{0, \ldots, 3\}$ to Bob, amounting to two bits of classical communication. See fig. 15 for an illustration.

### A glance at quantum channels

At this point, it would be natural to introduce *quantum channels* which are described mathematically by so-called completely positive, trace-preserving maps. They provide a unified framework for modelling general quantum information processing protocols. In this course, we only had time for a brief discussion at the end of today's lecture, but you are encouraged to have a look at, e.g., Wilde (2013).

# Bibliography

Charles H Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels, *Physical Review Letters*, 70(13):1895, page 1895, 1993.

Mark M Wilde. *Quantum information theory*. Cambridge University Press, 2013.

**PHYSICS 491: Symmetry and Quantum Information** May 4, 2017

# Quantum circuits, swap test, quantum Schur transform

*Lecture 10* *Michael Walter, Stanford University*

*These lecture notes are not proof-read and are offered for your convenience only. They include additional detail and references to supplementary reading material. I would be grateful if you email me about any mistakes and typos that you find.*

In the past two weeks, we used an important tool, the decomposition

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_j V_j \otimes \mathbb{C}^{m(n,j)} \tag{10.1}$$

of the $n$-qubit Hilbert space into irreducible representations of $\mathrm{SU}(2)$. We used the "Schur-Weyl toolbox" obtained in this way to solve the spectrum estimation problem, various data compression problems, and to study entanglement transformations (lectures 5, 6, 8 and 9). A fundamental role was played by the the projections $P_j$ onto the different sectors. But how would we realize these projections in practice?

Recall that the notation $\cong$ in eq. (10.1) refers to a unitary intertwiner

$$(\mathbb{C}^2)^{\otimes n} \to \bigoplus_j V_j \otimes \mathbb{C}^{m(n,j)}.$$

The $n$-qubit Hilbert space on the left-hand side has the product basis

$$|x_1, \dots, x_n\rangle = |x_1\rangle \otimes \dots \otimes |x_n\rangle,$$

while the right-hand side has a natural "Schur-Weyl basis" labeled by

$$|j, m, k\rangle$$

where $j \in \{\dots, \frac{n}{2} - 1, \frac{n}{2}\}$, $m \in \{-j, \dots, j\}$, $k \in \{1, \dots, m(n,j)\}$. Since the values of $m$ and $k$ are constrained by $j$, the right-hand side space is *not* a tensor product. However, we can safely think of it as a *subspace* of the tensor product space

$$\mathbb{C}^n \otimes \mathbb{C}^{n+1} \otimes \mathbb{C}^{2^n},$$

since (i) there are at most $n$ options for $j$, (ii) the dimension of $V_j$ is $\frac{2}{j} + 1 \leq n + 1$, and (iii) certainly $m(n,j) \leq 2^n$. Thus, we obtain an isometry

$$U_{\mathrm{Schur}} \colon (\mathbb{C}^2)^{\otimes n} \longrightarrow \mathbb{C}^n \otimes \mathbb{C}^{n+1} \otimes \mathbb{C}^{2^n} \tag{10.2}$$

This transformation is called the *quantum Schur transform* (fig. 16, (a)).

Why is this convenient? The isometry nicely separates the three pieces of information that we care about – the spin $j$ and the corresponding vectors in $V_j$ and in $\mathbb{C}^{m(n,j)}$ – into different subsystems. For example, we can now implement the spin measurement $\{P_j\}$ by first applying $U_{\mathrm{Schur}}$ and then measuring the first subsystem. In other words,

$$P_j = U_{\mathrm{Schur}}^\dagger \left(|j\rangle\langle j| \otimes \mathbb{1} \otimes \mathbb{1}\right) U_{\mathrm{Schur}}.$$

This is visualized in fig. 16, (b). The goal of today's lecture will be to design a *quantum circuit* for the quantum Schur transform.

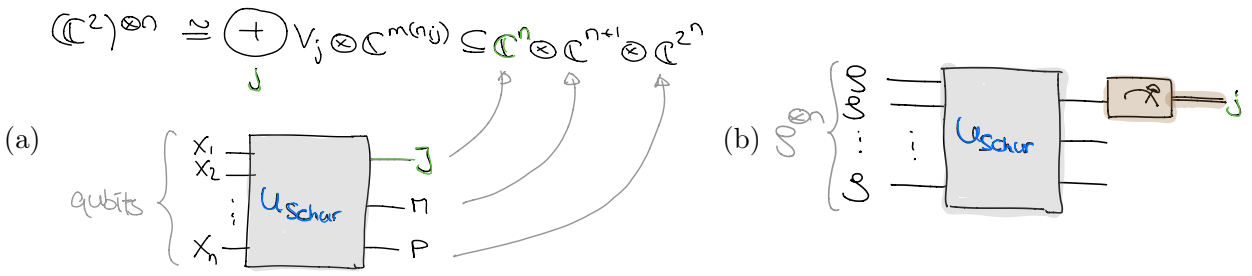



Figure 16: (a) The Schur transform (10.2). (b) We can implement the measurement $\{P_j\}$ by first applying the Schur transform and then measuring the $j$-system.

## 10.1 Quantum circuits

Just like we typically describe computer programs or algorithms in terms of simple elementary instructions, we are interested in constructing a unitary transformation $U$ of interest from "simple" building blocks. These building blocks are *quantum gates*, i.e., unitary operations that involve only a smaller number of qubits (or qu*d*its). We obtain a *quantum circuit* by connecting the output of some quantum gates by "wires" with the inputs of others. We will also allow *measurements* of individual qubits in the standard basis $\{|i\rangle\}$ as well as the *initialization* of qubits in basis states $|i\rangle$. For example, the circuit in fig. 17 first adds a qubit in state $|0\rangle$, then performs the unitary

$$(U_3 \otimes U_4)\,(\mathbb{1}_{\mathbb{C}^2} \otimes U_2 \otimes \mathbb{1}_{\mathbb{C}^2})\,(U_1 \otimes \mathbb{1}_{\mathbb{C}^2} \otimes \mathbb{1}_{\mathbb{C}^2})$$

and then measures one of the qubits. In the absence of measurements and initializations, a quantum circuit performs a unitary transformation from the input qubits to the output qubits. In the absence of measurements alone, the quantum circuit implements an isometry from the input qubits to the outputs qubits.

**Remark.** *The number of gates in a quantum circuit is known as the* (gate) complexity *of that circuit. Intuitively, the higher the complexity the longer it would take a quantum computer to run this circuit. This is because we expect that a quantum computer, in completely analogy to a classical computer, will be able to implement each gate and measurement in a small, fixed amount of time. Much of the field of* quantum computation *is concerned with finding quantum circuits and algorithms of minimal complexity – with a particular emphasis on finding quantum algorithms that outperform all known classical algorithms. For example, Peter Shor's famous factoring algorithm outperforms all known classical factoring algorithms. Just like quantum information theory, this is a very rich subject. In this course, we only have time for a glance, but I encourage you to look at Nielsen and Chuang (2002), Kitaev et al. (2002) for further detail if you are interested in this subject.*

To practice, let us consider some interesting gates. For any single-qubit unitary $U$, there is a corresponding *single-qubit gate*. For example, the Pauli $X$-operator $X = \left(\begin{smallmatrix} & 1 \\ 1 & \end{smallmatrix}\right)$ gives rise to the so-called $X$-gate or *NOT-gate*

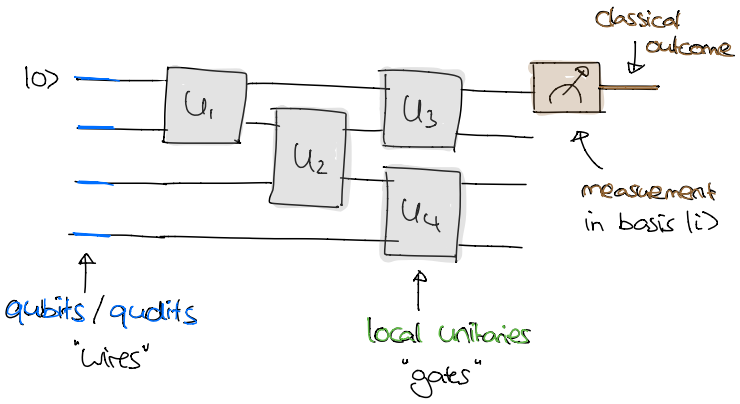
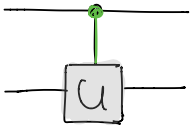



Figure 17: Illustration of a quantum circuit, composed of four unitary quantum gates and a single measurement. The first qubit is initialized in state $|0\rangle$ and the other three wires are inputs to the circuit.

which maps $X|0\rangle = |1\rangle$, $X|1\rangle = |0\rangle$. Another example is the so-called *Hadamard gate*


which maps $H|0\rangle = |+\rangle$, $H|1\rangle = |-\rangle$. Written as a unitary matrix, $H = \frac{1}{\sqrt{2}}\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$.

Single-qubit gates are not enough – for example, they do not allow us to create an entangled state starting from product states. A powerful class of gates can be obtained by performing a unitary transformation $U$ depending on the value of a *control qubit*. This is a standard but slightly misleading figure of speech, since we do not actually want to measure the value of the control qubit. To be more precise, we define the *controlled unitary gate*




by

$$\begin{aligned} \mathrm{C}U(|0\rangle \otimes |\psi\rangle) &= |0\rangle \otimes |\psi\rangle\,, \\ \mathrm{C}U(|1\rangle \otimes |\psi\rangle) &= |0\rangle \otimes (U|\psi\rangle) \end{aligned} \tag{10.3}$$

(and extend by linearity). It is easy to see that $\mathrm{C}U$ is indeed a unitary (indeed, $\mathrm{C}(U^\dagger)$ is its inverse).

**Remark 10.1.** *More generally, if $U_0, U_1$ are two unitaries then we can define a controlled unitary by $|x\rangle \mapsto U_x|x\rangle$. We will use this below when constructing a quantum circuit for the Clebsch-Gordan transformation.*

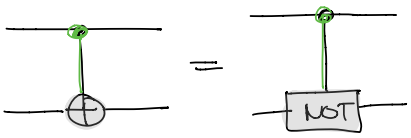
For example, if $U$ is the NOT-gate then the *controlled not (CNOT) gate* maps

$$\begin{aligned}\mathrm{CNOT}\,|0,0\rangle &= |0,0\rangle\,,\\ \mathrm{CNOT}\,|0,1\rangle &= |0,1\rangle\,,\\ \mathrm{CNOT}\,|1,0\rangle &= |1,1\rangle\,,\\ \mathrm{CNOT}\,|1,1\rangle &= |1,0\rangle\,,\end{aligned}$$

i.e.,

$$\mathrm{CNOT}\,|x,y\rangle = |x, x\oplus y\rangle\,,$$

where, as usual, $\oplus$ denotes addition modulo 2. This explains why the CNOT gate is often denoted by

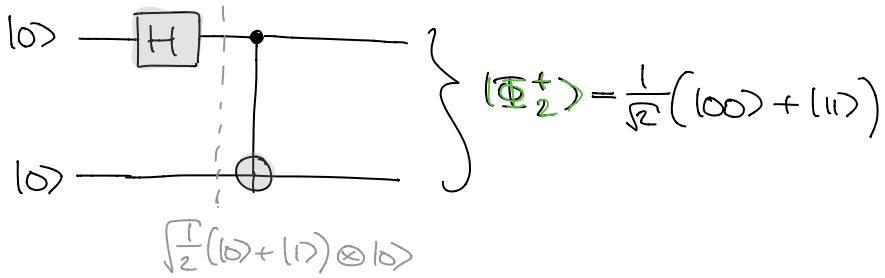



Using these ingredients, we can already build a number of interesting circuits.

**Remark.** *In fact, any $N$-qubit unitary can be to arbitrarily high fidelity approximated by quantum circuits composed only of CNOT-gates and single qubit gates. We say, that the CNOT gate together with the single qubit gates form a* universal gate set*. (In fact, CNOT together with a finite number of single qubit gates suffices.)*

### Entanglement and teleportation

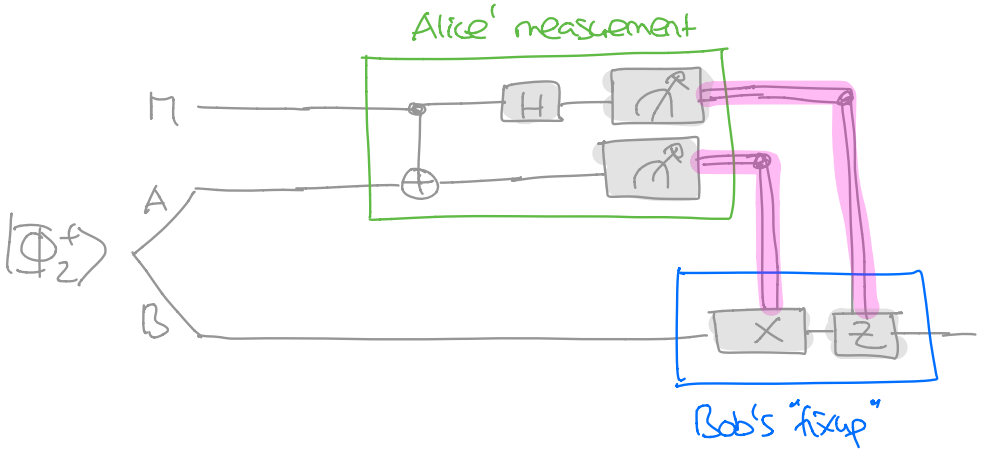
For example, consider the following circuit:




It is plain that this creates an ebit starting from the product state $|00\rangle$. More generally, for each product basis state $|xy\rangle$ the circuit produces one of the four maximally entangled basis vectors $|\phi_k\rangle$ from eq. (9.2) that we used in teleportation. Indeed, the circuit maps

$$|x,y\rangle \mapsto \frac{1}{\sqrt{2}}\left(|0\rangle + (-1)^x\,|1\rangle\right)\otimes|y\rangle = \frac{1}{\sqrt{2}}\left(|0,y\rangle + (-1)^x\,|1,y\rangle\right).$$

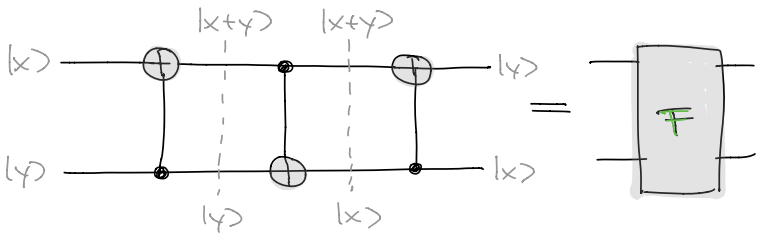
As a consequence, this allows us to write down a more detailed version of the teleportation circuit from last time (fig. 13):




The doubled wires (pink) denote the classical measurement outcomes (two bits $x$ and $y$, corresponding to the single integer $k \in \{0,1,2,3\}$ from last time). It is a fun exercise to verify that this circuit works as desired, i.e., that it implements an identity map from the input qubit $M$ to the output qubit $B$.

## 10.2 The swap test

We can implement the swap unitary $F\colon |xy\rangle \mapsto |yx\rangle$ by a quantum circuit composed of three CNOTs.




This is called the *swap gate*.

We can also write down a corresponding *controlled swap gate*, defined as in eq. (10.3) for $U = F$. Note that this is a *three* qubit gate. In problem 4.5, you will find a quantum circuit for the controlled swap gate that involves only single-qubit and two-qubit gates.

When we started studying the spectrum estimation problem in lecture 5, we first considered the case that we were given $n = 2$ two copies of our state as a "warmup" in example 5.3. The idea was that the two-qubit Hilbert space decomposes into the symmetric (triplet) and antisymmetric (singlet) subspaces,

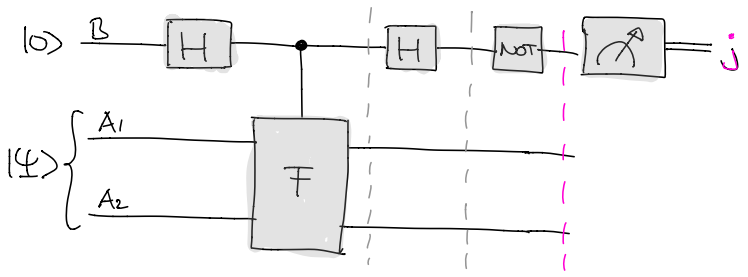
$$\mathbb{C}^2 \otimes \mathbb{C}^2 = \mathrm{Sym}^2(\mathbb{C}^2) \oplus \bigwedge\nolimits^2(\mathbb{C}^2),$$

which is of course a special case of eq. (10.1) since the triplet is a spin-1 irrep and the singlet a spin-0 irrep of $\mathrm{SU}(2)$. The swap operator $F$ acts by $+1$ on the triplet but by by $-1$ on the singlet, i.e.,

$$F = P_1 - P_0,$$

so measuring $F$ is completely equivalent to performing the projective measurement $\{P_0, P_1\}$.

How can we implement this measurement by a quantum circuit? Consider the following circuit, which uses the *controlled* swap gate discussed above:




(10.4)

Why does this circuit perform the desired measurement? Suppose that we initialize the $B$-wire in state $|0\rangle$ and the $A$-qubits in some arbitrary state $|\Psi\rangle$. The Hadamard gate sends $|0\rangle \mapsto |+\rangle$ and so the quantum state right after the controlled swap gate (first dashed line) is equal to

$$\frac{1}{\sqrt{2}} \left( |0\rangle_B \otimes |\Psi\rangle_A + |1\rangle_B \otimes F\,|\Psi\rangle_A \right)$$

After the second Hadamard gate (second dashed line), we obtain

$$\begin{aligned}
&\frac{1}{2} \left[ \left( |0\rangle_B + |1\rangle_B \right) \otimes |\Psi\rangle_A + \left( |0\rangle_B - |1\rangle_B \right) \otimes F\,|\Psi\rangle_A \right] \\
&= |0\rangle_B \otimes \frac{\mathbb{1} + F}{2}\,|\Psi\rangle_A + |1\rangle_B \otimes \frac{\mathbb{1} - F}{2}\,|\Psi\rangle_A \\
&= |0\rangle_B \otimes \Pi_2\,|\Psi\rangle_A + |1\rangle_B \otimes (\mathbb{1} - \Pi_2)\,|\Psi\rangle_A \\
&= |0\rangle_B \otimes P_1\,|\Psi\rangle_A + |1\rangle_B \otimes P_0\,|\Psi\rangle_A \,,
\end{aligned}$$

where $\Pi_2$ is the projector onto symmetric subspace, which for $n = 2$ qubits is nothing but the spin-1 projection $P_1$. The last NOT simply relabels $|0\rangle_B \leftrightarrow |1\rangle_B$, leading to

$$|1\rangle_B \otimes P_1\,|\Psi\rangle_A + |0\rangle_B \otimes P_0\,|\Psi\rangle_A \,.$$

In summary, the quantum circuit achieves the following task: It transforms an arbitrary input state $|\Psi\rangle_A$ into the following state right before the measurement of the $B$-qubit (last, pink dashed line)

$$|\Psi\rangle_A \mapsto \sum_{j=0,1} |j\rangle_B \otimes P_j\,|\Psi\rangle_A \,.$$

Hence

$$\Pr(\text{outcome } j) = \langle \Psi_A | P_j | \Psi_A \rangle \,,$$

and the post-measurement state on the $A$-qubits is proportional to $P_j\,|\Psi\rangle_A$. Thus, we have successfully implemented the measurement $\{P_0, P_1\}$. The quantum circuit (10.4) is known as the *swap test*.

## Applications

The swap test has many applications:

- If we choose $\rho^{\otimes 2}$ as input state for the A-qubits, then

$$\Pr(\text{outcome } j) = \operatorname{tr}\left[P_j \rho^{\otimes 2}\right],$$

i.e.,

$$\Pr(\text{outcome } 1) = \frac{1}{2}\left(1 + \operatorname{tr}\rho^2\right) = 1 - \Pr(\text{outcome } 0),$$

from which we can learn information about the spectrum of $\rho$. In particular, it allows us to estimate the *purity* $\operatorname{tr}\rho^2$ of the unknown quantum state (cf. example 5.3).

This was our original motivation for implementing the swap test.

- If we choose $|\psi\rangle_{A_1} \otimes |\phi\rangle_{A_2}$ as input state, then

$$\begin{aligned}\Pr(\text{outcome } 1) &= \frac{1}{2}\left(1 + \langle\psi_{A_1} \otimes \phi_{A_2}|F|\psi_{A_1} \otimes \phi_{A_2}\rangle\right) \\ &= \frac{1}{2}\left(1 + \langle\psi_{A_1} \otimes \phi_{A_2}|\phi_{A_1} \otimes \psi_{A_2}\rangle\right) = \frac{1}{2}\left(1 + |\langle\psi|\phi\rangle|^2\right),\end{aligned} \tag{10.5}$$

which allows us to estimate the overlap $|\langle\psi|\phi\rangle_{A_1}|$ between the pure states $|\psi\rangle$ and $|\phi\rangle$. Thus, the swap test can be used to test two unknown pure states for equality.

The swap test can be readily generalized to qu*d*its.

**Remark.** *There is a fun application of the swap test known as* quantum fingerprinting*, which we might discuss in class if there is enough time (Buhrman et al., 2001): The rough idea goes as follows: We can find $2^n$ many pure states $|\psi(\vec{x})\rangle \in \mathbb{C}^{cn}$, indexed by classical bit strings $\vec{x}$ of length $n$, with pairwise overlaps*

$$\langle\psi(\vec{x})|\psi(\vec{y})\rangle \leq \frac{1}{2}.$$

*Here $c > 0$ is some constant. Thus the quantum states live in a space of only order $\log n$ many qubits! (How can we justify the existence of such vectors? One way is to just choose them at random and estimate probabilities using a more refined version of our calculations for the symmetric subspace, see Harrow (2013) for more detail.) If we perform $k$ swap tests on $|\psi(\vec{x})\rangle^{\otimes k} \otimes |\psi(\vec{y})\rangle^{\otimes k}$ then we obtain*

$$\vec{x} \neq \vec{y} \quad \Rightarrow \quad \Pr(\textit{outcome 1 for all } k \textit{ swap tests}) = \left(\frac{3}{4}\right)^k \approx 0$$

*Thus the probability of outcome 1 is arbitrarily small, controlled only by the parameter $k$ (but not $n$). In this sense, we can use the states $|\psi(\vec{x})\rangle$ as short "fingerprints" for the classical bit strings $\vec{x}$. The latter are require $n$ bits to specify, while the fingerprints only need order $k \log n$ many qubits (this is not even optimal, but sufficient for our purposes).*

*Remarkably, while this allows us to test the fingerprints pairwise for equality with high certainty, it is* not *possible to determine the original bitstring $|\vec{x}\rangle$ from its fingerprint $|\psi(\vec{x})\rangle$ to good fidelity. This is ensured by the same Holevo bound mentioned last time in section 9.3, which ensures that we cannot communicate more than one classical bit by sending over a single qubit (in the absence of ebits).*

## 10.3 The quantum Schur transform

Now that we have acquired some familiarity with quantum circuitry, we will turn towards solving our actual goal for today – finding a quantum circuit for the Schur transform (10.2),

$$U_{\text{Schur}} : (\mathbb{C}^2)^{\otimes n} \cong \bigoplus_j V_j \otimes \mathbb{C}^{m(n,j)} \longrightarrow \mathbb{C}^n \otimes \mathbb{C}^{n+1} \otimes \mathbb{C}^{2^n}$$

(cf. fig. 16). We'll follow the exposition in Christandl (2010).

### The Clebsch-Gordan isometry

In lecture 6, we obtained the multiplicities $m(n,j)$ by successively applying the Clebsch-Gordan rule,

$$V_j \otimes V_{1/2} \cong \bigoplus_{j'=j-\frac{1}{2}}^{j+\frac{1}{2}} V_{j'}. \tag{10.6}$$

From your quantum mechanics class you know that the spin-$j$ representation $V_j$ has a basis $|j,m\rangle$ with $m = -j, \ldots, j$. The matrix elements of the basis transformation corresponding to (10.6) are known as the Clebsch-Gordan coefficients. They can packaged up in terms of unitary $2 \times 2$-matrices $U(j,m)$ such that

$$|j,m\rangle \otimes |\frac{1}{2}, s\rangle = \sum_{s'=-\frac{1}{2}}^{\frac{1}{2}} U(j,m)_{s,s'} \, |j+s', m+s\rangle \,. \tag{10.7}$$

for $s = \pm\frac{1}{2}$.

**Remark.** *Why is this the case, and how can these coefficients be computed? The defining property of the basis vectors $|j,m\rangle$ of $V_j$ is that*

$$\widetilde{Z} \, |j,m\rangle = 2m \, |j,m\rangle \,, \tag{10.8}$$

*where $\widetilde{Z}$ denotes the action of the "generator" $Z$ of* $\mathrm{SU}(2)$*, as discussed in remark 5.4. On the other hand, if we consider the action of the generator on the tensor product $V_j \otimes V_{1/2}$, then the generator $Z$ acts by*

$$\left(\widetilde{Z} \otimes \mathbb{1} + \mathbb{1} \otimes \widetilde{Z}\right) \left(|j,m\rangle \otimes |\frac{1}{2}, s\rangle\right) = 2(m+s) \left(|j,m\rangle \otimes |\frac{1}{2}, s\rangle\right).$$

*By comparing with eq.* (10.8)*, this means that $|j,m\rangle \otimes |\frac{1}{2}, s\rangle$ can indeed be written as a linear combination of $|j', m'\rangle$ with $m' = m + s$ – that is, in the form of eq.* (10.7)*.*

*How can the coefficients be determined? First, note that the only way of obtaining $m' = j + \frac{1}{2}$ is by choosing $m = j$ and $s = \frac{1}{2}$. Thus,*

$$|j + \frac{1}{2}, j + \frac{1}{2}\rangle = |j,j\rangle \otimes |\frac{1}{2}, \frac{1}{2}\rangle \,. \tag{10.9}$$

*Now you will remember from your quantum mechanics that the spin lowering operator $S_- = X - iY$ acts by*

$$\widetilde{S}_- \, |j,m\rangle = 2\sqrt{j(j+1) - m(m-1)} \, |j, m-1\rangle \,.$$

*By successively acting with $S_-$ on eq. (10.9) (i.e., by $\widetilde{S}_-$ on the left and by $\widetilde{S}_- \otimes \mathbb{1} + \mathbb{1} \otimes \widetilde{S}_-$ on the right), this allows us to obtain an expression of the form*

$$|j+\frac{1}{2}, m'\rangle = \#\,|j, m'-\frac{1}{2}\rangle \otimes |\frac{1}{2}, \frac{1}{2}\rangle + \#\,|j, m'+\frac{1}{2}\rangle \otimes |\frac{1}{2}, -\frac{1}{2}\rangle$$

*for some coefficients $\#$. Thus we have identified $V_{j+\frac{1}{2}}$ in $V_j \otimes V_{1/2}$. Next, we observe that*

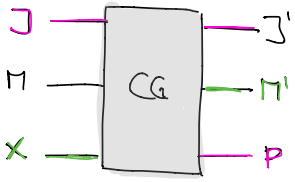
$$|j-\frac{1}{2}, j-\frac{1}{2}\rangle = \#\,|j, j-1\rangle \otimes |\frac{1}{2}, \frac{1}{2}\rangle + \#\,|j, j\rangle \otimes |\frac{1}{2}, -\frac{1}{2}\rangle \tag{10.10}$$

*is now uniquely determined by orthogonality to $|j+\frac{1}{2}, j-\frac{1}{2}\rangle$. We can now similarly obtain the coefficients in*

$$|j-\frac{1}{2}, m'\rangle = \#\,|j, m'-\frac{1}{2}\rangle \otimes |\frac{1}{2}, \frac{1}{2}\rangle + \#\,|j, m'+\frac{1}{2}\rangle \otimes |\frac{1}{2}, -\frac{1}{2}\rangle$$

*by successfully applying the action of the generator $S_-$ to eq. (10.10).*

We now define the Clebsch-Gordan isometry $U_{CG}$,



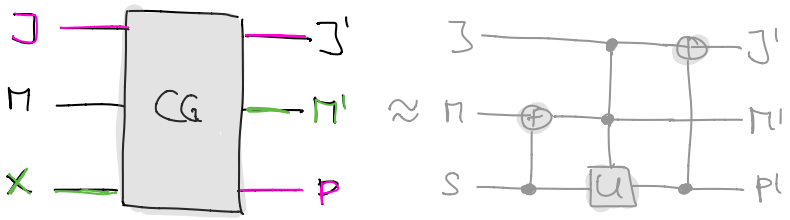

as the isometry that sends

$$\begin{aligned}|j, m, x\rangle \mapsto |j, m\rangle \otimes |\frac{1}{2}, s\rangle \mapsto U(j,m)_{s,\frac{1}{2}}\,|j+\frac{1}{2}, m+s\rangle \otimes |+\rangle \\ + U(j,m)_{s,-\frac{1}{2}}\,|j-\frac{1}{2}, m+s\rangle \otimes |-\rangle\,,\end{aligned}$$

where we first relabel the standard basis $|x\rangle$ of $\mathbb{C}^2$ to $|\frac{1}{2}, s\rangle$ of $V_{1/2}$, with $s := \frac{1}{2} - x \in \{\pm\frac{1}{2}\}$, and then apply the Clebssch-Gordan transformation. (To be precise, we should restrict the possible values of $j$ to some $j_{\max}$ to obtain a finite matrix.)

What is the meaning of the output $p$? In eq. (10.7), the left-hand side spin $j$ was fixed, but the spin $j$ is now part of the input. Since the same $j'$ can be obtained from two possible values of $j$, we use an additional output $p$ to remember the "direction" by which we arrived at $j'$ (that is, $j' = j + \frac{p}{2}$). Only then is $U_{CG}$ an isometry.

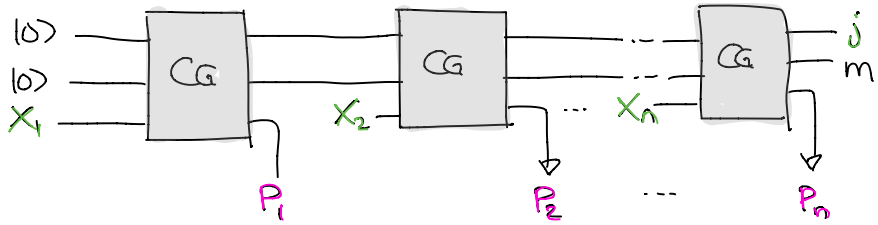
Schematically, the Clebsch-Gordan isometry $U_{CG}$ can be implemented by a quantum circuit of the following form




where the middle part uses the slightly more general notion of a controlled unitary described in remark 10.1, mapping $|j, m, s\rangle$ to $|j, m\rangle \otimes U(j,m)\,|s\rangle$.

### The quantum Schur transform

We now obtain the quantum Schur transform $U_{\text{Schur}}$ from eq. (10.2) by composing $n$ Clebsch-Gordan transformations:




We input the $n$ qubits into the wires $X_1, \ldots, X_n$ and the output consists of $J$, $M$, and $P = (P_1, \ldots, P_n)$. A moments thought shows that this indeed implements the desired transformation.

In particular, we can implement the spectrum estimation measurement $\{P_j\}$ by first applying the quantum Schur transform and then measuring the $J$-system in the standard basis (as in fig. 16, (b)).

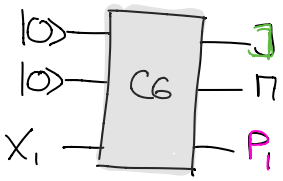
**Remark.** *We can expand*

$$U_{Schur}|\Psi\rangle = \sum_j \psi_{j,m,\vec{p}}\,|j\rangle_J \otimes |m\rangle_M \otimes |\vec{p}\rangle_P\,,$$

*where $\vec{p} \in \{\pm\}^n$. Then $\psi_{j,m,\vec{p}} \neq 0$ only if $\vec{p}$ is a sequence $|+-++-\ldots\rangle_P$ that corresponds to a path from $(0,0)$ to $(n,j)$ in fig. 9.*

At last, let us discuss some concrete examples to make sure that we fully understand what is going on:

**Example** (n=1)**.** *For a single qubit, the Schur transform is completely trivial:*

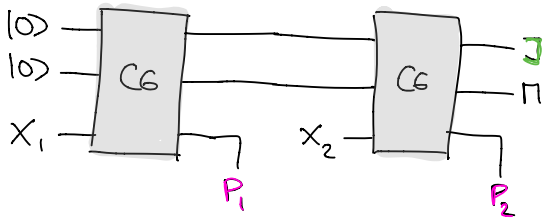



*It maps*

$$|0\rangle_X \mapsto |\frac{1}{2}\rangle_J \otimes |\frac{1}{2}\rangle_M \otimes |+\rangle_P$$
$$|1\rangle_X \mapsto |\frac{1}{2}\rangle_J \otimes |-\frac{1}{2}\rangle_M \otimes |+\rangle_P$$

*Note that the P-system is always in the $|+\rangle$ state, corresponding to the path $(0,0) \to (\frac{1}{2},1)$.*

**Example** (n=2)**.** *For two qubits, the Schur transform*



*maps*

$$|0,0\rangle_X \mapsto |1\rangle_J \otimes |1\rangle_M \otimes |{+}{+}\rangle_P$$
$$|1,1\rangle_X \mapsto |1\rangle_J \otimes |{-1}\rangle_M \otimes |{+}{+}\rangle_P$$

*(because those tensors are in the symmetric subspace, and $\tilde{Z}$ acts by $\pm 2$, respectively), while*

$$|0,1\rangle_X = \frac{1}{\sqrt{2}} \frac{|0,1\rangle + |1,0\rangle}{\sqrt{2}} + \frac{1}{\sqrt{2}} \frac{|0,1\rangle - |1,0\rangle}{\sqrt{2}} \mapsto \frac{1}{\sqrt{2}} |1\rangle_J \otimes |0\rangle_M \otimes |{+}{+}\rangle_P + \frac{1}{\sqrt{2}} |0\rangle_J \otimes |0\rangle_M \otimes |{+}{-}\rangle_P ,$$
$$|1,0\rangle_X = \frac{1}{\sqrt{2}} \underbrace{\frac{|0,1\rangle + |1,0\rangle}{\sqrt{2}}}_{\in \mathrm{Sym}^2(\mathbb{C}^2)} - \frac{1}{\sqrt{2}} \underbrace{\frac{|0,1\rangle - |1,0\rangle}{\sqrt{2}}}_{\in \bigwedge^2(\mathbb{C}^2)} \mapsto \frac{1}{\sqrt{2}} |1\rangle_J \otimes |0\rangle_M \otimes |{+}{+}\rangle_P - \frac{1}{\sqrt{2}} |0\rangle_J \otimes |0\rangle_M \otimes |{+}{-}\rangle_P .$$

**Exercise.** *Can you write down the Schur transform for $n = 3$?*

# Bibliography


Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.

Alexei Yu Kitaev, Alexander Shen, and Mikhail N Vyalyi. *Classical and quantum computation*, volume 47. American Mathematical Society Providence, 2002.

Harry Buhrman, Richard Cleve, John Watrous, and Ronald De Wolf. Quantum fingerprinting, *Physical Review Letters*, 87(16):167902, page 167902, 2001.

Aram W Harrow. The church of the symmetric subspace. 2013. arXiv:1308.6595.

Matthias Christandl. Symmetries in quantum information theory. 2010. URL `http://edu.itp.phys.ethz.ch/hs10/sqit/`.

**PHYSICS 491: Symmetry and Quantum Information** April 9, 2017

# Problem Set 1

*Michael Walter, Stanford University* *due April 18, 2017*

**Problem 1.1** (Classical and quantum strategies for the GHZ game)**.**
Three players and the referee play the GHZ game, following the same conventions as in class. In particular, the referee chooses each of the four questions $xyz$ with equal probability 1/4.

(a) Verify that the winning probability for a general quantum strategy, specified in terms of a state $|\psi\rangle_{ABC}$ and observables $A_x, B_y, C_z$, is given by

$$p_{\text{win,q}} = \frac{1}{2} + \frac{1}{8} \langle \psi_{ABC} | A_0 \otimes B_0 \otimes C_0 - A_1 \otimes B_1 \otimes C_0 - A_1 \otimes B_0 \otimes C_1 - A_0 \otimes B_1 \otimes C_1 | \psi_{ABC} \rangle \,. \quad (1.1)$$

(b) Suppose that Alice, Bob, and Charlie play the following randomized classical strategy: When they meet before the game is started, they flip a biased coin. Let $\pi$ denote the probability that the coin comes up heads. Depending on the outcome of the coin flip, which we denote by $\lambda \in \{\text{HEADS}, \text{TAILS}\}$, they use one of two possible deterministic strategies $a_\lambda(x), b_\lambda(y), c_\lambda(z)$ to play the game. Find a formula analogous to (1.1) for the winning probability $p_{\text{win,cl}}$ of their strategy.

(c) In class we argued that even randomized classical strategies cannot do better than $p_{\text{win,cl}} \leq 3/4$. Verify this explicitly using the formula you derived in (b).

(d) Any classical strategy can be realized by a quantum strategy. Show this explicitly for the randomized classical strategy described in (b) by constructing a quantum state $|\psi\rangle_{ABC}$ and observables $A_x, B_y, C_z$ such that $p_{\text{win,cl}} = p_{\text{win,q}}$.

**Problem 1.2** (Distinguishing quantum states)**.**
The *trace distance* between two quantum states $|\phi\rangle$ and $|\psi\rangle$ is defined by

$$T(\phi, \psi) = \max_{0 \leq Q \leq \mathbb{1}} \langle \phi | Q | \phi \rangle - \langle \psi | Q | \psi \rangle \,. \quad (1.2)$$

Here, $0 \leq Q \leq \mathbb{1}$ means that both $Q$ and $\mathbb{1} - Q$ are positive semidefinite operators.

(a) Imagine a quantum source that emits $|\phi\rangle$ or $|\psi\rangle$ with probability 1/2 each. Show that the optimal probability of identifying the true state by a POVM measurement is given by

$$\frac{1}{2} + \frac{1}{2} T(\phi, \psi).$$

Why can this probability never be smaller than 1/2?

(b) Conclude that only orthogonal states (i.e., $\langle \phi | \psi \rangle = 0$) can be distinguished perfectly.

(c) Show that the trace distance is a metric. That is, verify that $T(\phi, \psi) = 0$ if and only if $|\phi\rangle = e^{i\theta} |\psi\rangle$, that $T(\phi, \psi) = T(\psi, \phi)$, and prove the triangle inequality $T(\phi, \psi) \leq T(\phi, \chi) + T(\chi, \psi)$.

You will now derive an explicit formula for the trace distance. For this, consider the spectral decomposition $\Delta = \sum_i \lambda_i |e_i\rangle\langle e_i|$ of the Hermitian operator $\Delta = |\phi\rangle\langle\phi| - |\psi\rangle\langle\psi|$.

(d) Show that the operator $Q = \sum_{\lambda_i > 0} |e_i\rangle\langle e_i|$ achieves the maximum in (1.2), and deduce the following formulas for the trace distance:

$$T(\phi, \psi) = \sum_{\lambda_i > 0} \lambda_i = \frac{1}{2} \sum_i |\lambda_i|.$$

(e) Conclude that the optimal probability of distinguishing the two states in (a) remains unchanged if we restrict to projective measurements.

In class, we used another measure to compare quantum states, namely their overlap $|\langle\phi|\psi\rangle|$.

(f) Show that trace distance and overlap are related by the following formula:

$$T(\phi, \psi) = \sqrt{1 - |\langle\phi|\psi\rangle|^2}.$$

*Hint: Argue that it suffices to verify this formula for two pure states of a qubit, with one of them equal to $|0\rangle$, and use the formula derived in part (d).*

This exercise shows that states with overlap close to one are almost indinguishable by any measurement, justifying our intuition from class.

**Problem 1.3** (POVMs can outperform projective measurements; Nielsen & Chuang §2.2.6)**.**
Imagine a qubit source that emits either of the two states $|0\rangle$ and $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ with equal probability 1/2. Your task is to design a measurement scheme that allows to optimally distinguish these two cases. Unfortunately, the states $|0\rangle$ and $|+\rangle$ are not orthogonal, so you know that this cannot be done perfectly (e.g., from the previous problem).

Suppose now that your measurement scheme is *not allowed to ever give a wrong answer*! Instead, it is allowed to report one of *three* possible answers: that the true state is $|0\rangle$, that the true state is $|+\rangle$, or that the measurement outcome is inconclusive. We define the success probability of such a scheme as the probability that you identify the true state correctly.

(a) Show that for projective measurements the success probability is at most 1/4.

(b) Find a POVM measurement that achieves a success probability strictly larger than 1/4.

**Bonus Problem 1.4** (POVM measurements are physical)**.**
In this exercise, you will show that every POVM measurement can be realized by a projective measurement on a larger system. Thus, let $\{Q_x\}_{x\in\Omega}$ be an arbitrary POVM measurement on some Hilbert space $\mathcal{H}_A$. For simplicity, we will assume that the set of possible outcomes $\Omega$ is finite.

(a) Let $\mathcal{H}_B$ be a Hilbert space space with one basis vector $|x\rangle_B$ for each $x \in \Omega$, and fix some arbitrary $x_0 \in \Omega$. Show that the linear map

$$|\psi\rangle_A \otimes |x_0\rangle_B \mapsto \sum_x \sqrt{Q_x}\,|\psi\rangle_A \otimes |x\rangle_B \tag{1.3}$$

is an isometry (an isometry is a map that preserves inner products).[2]

[2]Every positive semidefinite operator such as $Q_x$ has a square root $\sqrt{Q_x}$, defined by taking the square root of each eigenvalue while keeping the same eigenspaces.

Any isometry from a subspace into a larger Hilbert space can be extended to a unitary operator on the larger space. Thus there exists a unitary $U_{AB}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$ that extends the isometry (1.3).

(b) Use $U_{AB}$ to design a projective measurement $\{P_{AB,x}\}$ on the joint system $\mathcal{H}_A \otimes \mathcal{H}_B$ such that

$$Q_x = (\mathbb{1}_A \otimes \langle x_0|_B)\, P_{AB,x}\, (\mathbb{1}_A \otimes |x_0\rangle_B)$$

for all outcomes $x \in \Omega$.

**PHYSICS 491: Symmetry and Quantum Information** April 16, 2017

# Problem Set 2

*Michael Walter, Stanford University* *due April 25, 2017*

**Problem 2.1** (Pure state entanglement)**.**
In this exercise you will study the entanglement of pure states $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$. In class, we discussed the Schmidt decomposition

$$|\psi\rangle_{AB} = \sum_{i=1}^{r} s_i \, |e_i\rangle_A \otimes |f_i\rangle_B$$

and its relation to the eigenvalues of the reduced density matrices. For simplicity we will assume that $\dim \mathcal{H}_A = \dim \mathcal{H}_B = d$.

(a) We say that $|\psi\rangle_{AB}$ is *maximally entangled* if $s_i = \frac{1}{\sqrt{d}}$ for all $i$. Show that $|\psi\rangle_{AB}$ is maximally entangled if and only if $\rho_A$ and $\rho_B$ are maximally mixed (i.e., proportional to $\mathbb{1}$).

(b) Show that $|\psi\rangle_{AB}$ is a product state if and only if $\rho_A$ and $\rho_B$ are pure states.

This suggests that the eigenvalues of the reduced density matrices $\rho_A$ and $\rho_B$ can be used to characterize the entanglement of $|\psi\rangle_{AB}$. As an example, consider the *Rényi-2 entropy*, defined by

$$S_2(A) = -\log \operatorname{tr} \rho_A^2.$$

(c) Find a formula for $S_2(A)$ in terms of the eigenvalues of the reduced density matrices.

(d) Show that $S_2(A) = 0$ for product states, $S_2(A) = \log d$ for maximally entangled states, and otherwise $0 < S_2(A) < \log d$.

You will now study the average entanglement of pure states in $\mathcal{H}_A \otimes \mathcal{H}_B$, drawn at random from the "uniform" probability distribution $d\psi_{AB}$ that you know from class.

(e) Let $F_A$ denote the swap operator on $\mathcal{H}_A^{\otimes 2}$ that sends $|a_1, a_2\rangle \mapsto |a_2, a_1\rangle$. Verify that

$$\operatorname{tr} \rho_A^2 = \operatorname{tr}\left[ (F_A \otimes \mathbb{1}_{BB}) \, |\psi\rangle_{AB}^{\otimes 2} \, \langle\psi|_{AB}^{\otimes 2} \right].$$

(f) Let $F_B$ denote the swap operator on $\mathcal{H}_B^{\otimes 2}$, defined in the same way as $F_A$. Show that

$$\int d\psi_{AB} \, |\psi\rangle_{AB}^{\otimes 2} \, \langle\psi|_{AB}^{\otimes 2} = \frac{1}{d^2(d^2+1)} \left( \mathbb{1}_{AA} \otimes \mathbb{1}_{BB} + F_A \otimes F_B \right).$$

*Hint: Remember the symmetric subspace.*

(g) Show that the average Rényi-2 entropy $S_2(A)$ of a random pure state is no smaller than $\log d - \log 2$.

*Hint: Jensen's inequality shows that* $\int d\psi \, \log f(|\psi\rangle) \leq \log \left( \int d\psi \, f(|\psi\rangle) \right)$.

**Problem 2.2** (Extensions of quantum states)**.**
In this exercise you will verify two important facts that we discussed in class:

(a) Show that any density operator admits a *purification*. That is, given a quantum state $\rho_A$ on some Hilbert space $\mathcal{H}_A$, construct a pure state $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$, where $\mathcal{H}_B$ is some auxiliary Hilbert space, such that

$$\rho_A = \mathrm{tr}_B \left[ |\psi\rangle \langle\psi|_{AB} \right].$$

*Hint: Consider the spectral decomposition of $\rho_A$.*

(b) Show that any extension of a pure state is a tensor product. That is, show that if $\rho_A$ is pure then any extension is of the form

$$\rho_{AB} = \rho_A \otimes \rho_B.$$

*Hint: You have already solved this problem in the case that $\rho_{AB}$ is pure.*

**Problem 2.3** (The symmetric subspace is irreducible)**.**
In this problem, you will show that the symmetric subspace is an irreducible representation of $\mathrm{SU}(d)$. We will start with $d = 2$. For any operator $M$ on $\mathbb{C}^2$, define a corresponding operator on $(\mathbb{C}^2)^{\otimes n}$ by

$$\widetilde{M} = M_1 + M_2 + \cdots + M_n.$$

Here we write $M_1 = M \otimes \mathbb{1} \otimes \ldots \otimes \mathbb{1}$, $M_2 = \mathbb{1} \otimes M \otimes \mathbb{1} \otimes \ldots \otimes \mathbb{1}$, etc. Now consider an arbitrary subspace $\mathcal{H} \subseteq \mathrm{Sym}^n(\mathbb{C}^2)$ that is invariant for $\mathrm{SU}(2)$.

(a) Show that $\widetilde{M}|\psi\rangle \in \mathcal{H}$ for any vector $|\psi\rangle \in \mathcal{H}$.

*Hint: If $H$ is Hermitian then $e^{iH}$ is unitary.*

In class, we observed that the symmetric subspace has natural occupation number basis. For $d = 2$, it is given by

$$\|t\rangle\!\rangle \propto |\underbrace{0, \ldots, 0}_{t}, \underbrace{1, \ldots, 1}_{n-t}\rangle + \text{permutations} \qquad (t = 0, \ldots, n).$$

(b) Find an operator $M$ such that $\widetilde{M}$ has the basis vectors $\|t\rangle\!\rangle$ as eigenvectors (with distinct eigenvalues). Conclude that $\mathcal{H}$ is spanned by a subset of the basis vectors $\|t\rangle\!\rangle$.

(c) Find operators $M_\pm$ such that $\widetilde{M_\pm}\|t\rangle\!\rangle \propto \|t \pm 1\rangle\!\rangle$. Conclude that $\mathcal{H}$ is either $\{0\}$ or all of $\mathrm{Sym}^n(\mathbb{C}^d)$.

Thus you have proved that $\mathrm{Sym}^n(\mathbb{C}^2)$ is indeed an irreducible representation of $\mathrm{SU}(2)$!

(d) Any irreducible representation of $\mathrm{SU}(2)$ can be labeled by its spin $j$. What is the spin of the symmetric subspace $\mathrm{Sym}^n(\mathbb{C}^2)$?

(e) *Optional:* Sketch how your proof can be generalized to show that $\mathrm{Sym}^n(\mathbb{C}^d)$ is an irreducible representation of $\mathrm{SU}(d)$.

**Bonus Problem 2.4** (Entanglement witnesses and convexity)**.**
An observable $X_{AB}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$ is called an *entanglement witness* for a quantum state $\rho_{AB}$ if

$$\mathrm{tr}[X_{AB}\,\rho_{AB}] < 0,$$

while

$$\mathrm{tr}[X_{AB}\,\sigma_{AB}] \geq 0 \tag{2.1}$$

for all separable states $\sigma_{AB}$.

(a) Construct an entanglement witness for the maximally entangled state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.

*Hint: Compute the overlap of $|\Phi^+\rangle$ with a pure product state $|\psi\rangle_A \otimes |\phi\rangle_B$. Why could this help?*

(b) Argue that for any entangled state $\rho_{AB}$ there exists an entanglement witness $X_{AB}$.

*Hint: You do not need to construct the entanglement witness explicitly.*

**Bonus Problem 2.5** (The extendibility hierarchy)**.**
In this problem, you will show that any quantum state that has an $n$-extension is close to a separable state if $n$ is large, as discussed in class.

(a) Imitate the proof of the quantum de Finetti theorem given in class to show that, for any pure state $|\Phi\rangle_{AB_1\ldots B_n} \in \mathcal{H}_A \otimes \mathrm{Sym}^n(\mathcal{H}_B)$,

$$\mathrm{tr}_{B_2\ldots B_n}[|\Phi\rangle\langle\Phi|] \approx \int d\psi\, p(\psi)\, |W_\psi\rangle\langle W_\psi|_A \otimes |\psi\rangle\langle\psi|_{B_1}$$

for large $n$. Here, the integral is over the set of pure states on $\mathcal{H}_B$, $p(\psi)$ is a probability density, and the $|W_\psi\rangle$ are pure states in $\mathcal{H}_A$.

Now suppose that $\rho_{AB}$ is an arbitrary quantum state that has an $n$-extension (i.e., that there exists some $\sigma_{AB_1\ldots B_n}$ such that $\sigma_{AB_k} = \rho_{AB}$ for all $k$).

(b) Show that $\rho_{AB}$ also has an $n$-extension $\rho_{AB_1\ldots B_n}$ that is permutation-invariant on the $B$-systems, i.e., $[\mathbb{1}_A \otimes R_\pi, \rho] = 0$ for all $\pi \in S_n$.

Any $n$-extension as in (b) admits a purification in $(\mathcal{H}_A \otimes \mathcal{H}_{A'}) \otimes \mathrm{Sym}^n(\mathcal{H}_B \otimes \mathcal{H}_{B'})$, where $\mathcal{H}_{A'} = \mathcal{H}_A$ and $\mathcal{H}_{B'} = \mathcal{H}_B$.

(c) Conclude that any $n$-extendible $\rho_{AB}$ is close to a separable state for large $n$.

*Hint: The trace distance does not increase when you take the partial trace.*

**PHYSICS 491: Symmetry and Quantum Information** April 23, 2017

# Problem Set 3

*Michael Walter, Stanford University* *due May 4, 2017*

**Problem 3.1** (The antisymmetric state)**.**
In class, we discussed the quantum de Finetti theorem for the symmetric subspace. It asserts that the reduced density matrices $\rho_{A_1\dots A_k}$ of a state on $\mathrm{Sym}^n(\mathbb{C}^d)$ are $\sqrt{kd/(n-k)}$ close in trace distance to a separable state (in fact, to a mixture of tensor power states).

The goal of this exercise is to show that a dependence on the dimension $d$ is unavoidable. To start, consider the *Slater determinant*

$$|S\rangle_{A_1\dots A_d} = |1\rangle \wedge \dots \wedge |d\rangle := \sqrt{\frac{1}{d!}} \sum_{\pi\in S_d} \mathrm{sign}(\pi)\, |\pi(1)\rangle \otimes \dots \otimes |\pi(d)\rangle \in (\mathbb{C}^d)^{\otimes d}.$$

We define the *antisymmetric state* on $\mathbb{C}^d \otimes \mathbb{C}^d$ by tracing out all but two subsystems,

$$\rho_{A_1A_2} = \mathrm{tr}_{A_3\dots A_d}\left[|S\rangle\langle S|\right].$$

(a) Show that $T(\rho_{A_1A_2}, \sigma_{A_1A_2}) \geq \frac{1}{2}$ for all separable states $\sigma_{A_1A_2}$.
*Hint: Consider the POVM element $Q = \Pi_2$ (i.e., the projector onto the symmetric subspace).*

Thus you have shown that the antisymmetric state is far from any separable state. However, note that $|S\rangle$ is *not* in the symmetric subspace.

(b) Show that $|S\rangle^{\otimes 2} \in \mathrm{Sym}^d(\mathbb{C}^d \otimes \mathbb{C}^d)$, while $\rho^{\otimes 2}$ is likewise far away from any separable state. Conclude that the quantum de Finetti theorem must have a dependence on the dimension $d$.

**Problem 3.2** (De Finetti and mean field theory)**.**
In this exercise you will explore the consequences of the quantum de Finetti theorem for mean field theory. Consider an operator $h$ on $\mathbb{C}^d \otimes \mathbb{C}^d$ and the corresponding *mean-field Hamiltonian*

$$H = \frac{1}{n-1} \sum_{i\neq j} h_{i,j}$$

on $(\mathbb{C}^d)^{\otimes n}$, where each term $h_{i,j}$ acts by the operator $h$ on subsystems $i$ and $j$ and by the identity operator on the remaining subsystems (e.g., $h_{1,2} = h \otimes \mathbb{1}^{\otimes(n-2)}$).

(a) Show that the eigenspaces of $H$ are invariant subspaces for the action of the symmetric group.

Now assume that the ground space is nondegenerate, and spanned by some $|E_0\rangle$. Then part (a) implies that $R_\pi |E_0\rangle = \chi(\pi)|E_0\rangle$ for some function $\chi$. This function necessarily satisfies $\chi(\pi\tau) = \chi(\pi)\chi(\tau)$.

(b) Show that $\chi(i \leftrightarrow j) = \chi(1 \leftrightarrow 2)$ for all $i \neq j$. Conclude that $|E_0\rangle$ is either a symmetric tensor or an antisymmetric tensor.
*Hint: First show that $\chi(\pi\tau\pi^{-1}) = \chi(\tau)$.*

If $n > d$, then there exist no nonzero antisymmetric tensors. Thus, in the thermodynamic limit of large $n$, the ground state $|E_0\rangle$ is in the symmetric subspace $\mathrm{Sym}^n(\mathbb{C}^d)$ and so the quantum de Finetti theorem is applicable.

(c) Show that, for large $n$, the energy density in the ground state can be well approximated by

$$\frac{E_0}{n} \approx \min_{|\psi\rangle} \langle \psi^{\otimes 2} | h | \psi^{\otimes 2} \rangle = \frac{1}{n} \min_{|\psi\rangle} \langle \psi^{\otimes n} | H | \psi^{\otimes n} \rangle .$$

This justifies the folklore that "in the mean field limit the ground state has the form $|\psi\rangle^{\otimes \infty}$".

**Problem 3.3** (Universal quantum data compression)**.**
In class, we discussed a quantum compression protocol that works for all qubit ensembles $\{p_x, |\psi_x\rangle\}$ for which the associated density operator $\rho = \sum_x p_x |\psi_x\rangle\langle\psi_x|$ has given eigenvalues $\{p, 1-p\}$.

Your task in this exercise is to design a *universal compression protocol* that works for all qubit ensembles with $S(\rho) < S_0$, where $S_0 > 0$ is a given target compression rate.

(a) Show that, for all $S_0 > 0$, there exist projectors $\tilde{P}_n$ on subspaces $\tilde{\mathcal{H}}_n$ of $(\mathbb{C}^2)^{\otimes n}$ such that:

(i) For all density operators $\rho$ with $S(\rho) < S_0$, $\mathrm{tr}\left[\tilde{P}_n \rho^{\otimes n}\right] \to 1$ as $n \to \infty$,

(ii) The dimension of $\tilde{\mathcal{H}}_n$ is at most $2^{n(S_0 + \delta(n))}$ for some function $\delta$ with $\delta(n) \to 0$ as $n \to \infty$.

*Hint: Use the spectrum estimation projectors $P_j$ in a clever way.*

(b) Use the projectors $\tilde{P}_n$ to construct a compression protocol with compression rate $S_0$ that works for all qubit ensembles with $S(\rho) < S_0$ (i.e., show that in the limit of large block length $n$, the average squared overlap between the original state and the decompressed state goes to one).

*Hint: Follow the same construction as in lecture 7.*

**Bonus Problem 3.4** (Bounds on entropies)**.**
In this exercise, you will prove two bounds that we used in class. Let $0 \leq p, q \leq 1$. The first bound concerns the binary entropy function $h(p) = -p \log p - (1-p)\log(1-p)$.

(a) Consider the function $\eta(x) = -x \log x$ and assume that $|p - q| \leq \frac{1}{2}$. Show that

$$|\eta(p) - \eta(q)| \leq \eta(|p-q|), \tag{3.1}$$

and deduce the following special case of *Fannes' inequality*:

$$|h(p) - h(q)| \leq 2\eta(|p-q|)$$

The second bound concerns the binary relative entropy $\delta(p\|q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}$.

(b) Derive the following special case of *Pinsker's inequality*:

$$\delta(p\|q) \geq \frac{2}{\ln 2}(p-q)^2.$$

*Hint: Remember that $\log x = \ln x / \ln 2$ is the logarithm to the base two.*

**Bonus Problem 3.5** (Schur-Weyl duality)**.**
In class, we discussed an important mathematical result known as Schur-Weyl duality. The goal of this exercise is to supply some last details and conclude its proof.

Recall that we decomposed the Hilbert space of $n$ qubits as a representation of $U(2)$. Using the same notation as in class,

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_j V_{n,j} \otimes \mathbb{C}^{m(n,j)},$$

such that, for all $X \in U(2)$,

$$X^{\otimes n} \cong \bigoplus_j T_X^{(n,j)} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}}, \tag{3.2}$$

and we discussed that this formula can be extended to arbitrary operators $X$ on $\mathbb{C}^2$.

(a) Show that the representation operators $R_\pi$ for $\pi \in S_n$ have the form

$$R_\pi \cong \bigoplus_j \mathbb{1}_{V_{n,j}} \otimes R_\pi^{(n,j)}. \tag{3.3}$$

Conclude that the operators $R_\pi^{(n,j)}$ turn the spaces $\mathbb{C}^{m(n,j)}$ into representations of $S_n$. We will denote these representations by $W_{n,j}$.

*Hint: Recall that* $[U^{\otimes n}, R_\pi] = 0$ *and use Schur's lemma.*

In view of eqs. (3.2) and (3.3), we observe that $[X^{\otimes n}, R_\pi] = 0$ for *arbitrary* operators $X$ on $\mathbb{C}^2$ .

(b) Show that, conversely, any operator that commutes with all $R_\pi$ can be written as a linear combination of operators of the form $X^{\otimes n}$.

*Hint: Compute* $\frac{d}{dt_1}\Big|_{t_1=0} \cdots \frac{d}{dt_n}\Big|_{t_n=0} (\sum_{i=1}^n t_i X_i)^{\otimes n}$. *Why does this help?*

(c) Conclude that the representations $W_{n,j}$ of $S_n$ are irreducible and pairwise inequivalent.

*Hint: Use Schur's lemma.*

You have thus proved the following result, known as *Schur-Weyl duality*: The decomposition

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_j V_{n,j} \otimes W_{n,j}$$

holds as a representation of both $U(2)$ and $S_n$. The spaces $V_{n,j}$ and $W_{n,j}$ are pairwise inequivalent, irreducible representations of $U(2)$ and of $S_n$, respectively. This has important consequences. E.g.:

(d) Show that any operator that commutes with all $U^{\otimes n}$ and $R_\pi$ is necessarily of the form $\sum_j z_j P_j$, with $z_j \in \mathbb{C}$. Conclude that $\{P_j\}$ is the most fine-grained projective measurement that has both symmetries of the spectrum estimation problem, as discussed in class.

*Hint: Use Schur's lemma.*

**PHYSICS 491: Symmetry and Quantum Information** May 2, 2017

# Problem Set 4

*Michael Walter, Stanford University* *optional*

**Problem 4.1** (Schur-Weyl duality)**.**
Your goal in this exercise is to concretely identify irreducible representations of $U(2)$ and $S_n$ in the $n$-qubit Hilbert space. Let $j$ be such that $\frac{n}{2} - j$ is a nonnegative integer.

(a) Show that the subspace

$$\mathcal{H}_{n,j} := \left\{ |\phi\rangle \otimes |\psi^-\rangle^{\otimes \frac{n}{2}-j},\ |\phi\rangle \in \mathrm{Sym}^{2j}(\mathbb{C}^2) \right\} \subseteq (\mathbb{C}^2)^{\otimes n}$$

is an irreducible $U(2)$-representation equivalent to $V_{n,j}$. Here, $|\psi^-\rangle = \frac{1}{\sqrt{2}}(|10\rangle - |01\rangle)$ is the singlet state. How can you obtain further $U(2)$-representations in $(\mathbb{C}^2)^{\otimes n}$ equivalent to $V_{n,j}$?

(b) Now construct an irreducible $S_n$-representation in $(\mathbb{C}^2)^{\otimes n}$ that is equivalent to $W_{n,j}$. How can you obtain further $S_n$-representations in $(\mathbb{C}^2)^{\otimes n}$ equivalent to $W_{n,j}$?

(c) Using part (b), confirm that the definition of $W_{\square\square\square}$ and $W_{\begin{smallmatrix}\square\square\\\square\end{smallmatrix}}$ via Schur-Weyl duality is equivalent to our original definition in lecture 3.

**Problem 4.2** (PPT criterion)**.**
In this exercise, you will study a simple, highly useful entanglement criterion. Given an operator $M_{AB}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$, we define its *partial transpose* as the operator $M_{AB}^{T_B}$ with matrix elements

$$\langle a, b|M_{AB}^{T_B}|a', b'\rangle = \langle a, b'|M_{AB}|a', b\rangle\,.$$

Note that this definition depends on the choice of basis for $\mathcal{H}_B$ (but not of the basis for $\mathcal{H}_A$).

(a) Show that $\mathrm{tr}\, M_{AB}^{T_B} = \mathrm{tr}\, M_{AB}$.

(b) Observe that if $M_{AB} = X_A \otimes Y_B$ then $M_{AB}^{T_B} = X_A \otimes Y_B^T$ and argue that this uniquely determines the partial transpose.

In particular, we can consider the partial transpose of a density operator $\rho_{AB}$.

(c) Show that if $\rho_{AB}$ is separable then $\rho_{AB}^{T_B} \geq 0$.

You thus obtain the so-called *PPT criterion*, short for positive partial transpose criterion: *If the partial transpose $\rho_{AB}^{T_B}$ is* not *positive semidefinite then $\rho_{AB}$ must be entangled.*

(d) Verify using the PPT criterion that the ebit $|\Psi_2^+\rangle$ is entangled.

(e) Consider the family of *isotropic two-qubit states*,

$$\rho_{AB}(p) := p\,\tau_{\mathrm{sym}} + (1-p)\tau_{\mathrm{anti}},$$

where $\tau_{\mathrm{sym}}$ denotes the maximally mixed state on the symmetric subspace of two qubits and $\tau_{\mathrm{anti}} = |\psi^-\rangle\langle\psi^-|$ the singlet state. For which values of $p \in [0,1]$ does the PPT criterion establish entanglement?

In general, the PPT criterion is only a sufficient, but not a necessary criterion for entanglement. If $\dim H_A \otimes H_B > 6$, then there exist entangled states with a positive semidefinite partial transpose.

**Problem 4.3** (Dual representations)**.**
This problem introduces the concept of a *dual representation*. To start, consider a representation $\mathcal{H}$ of some group $G$, with operators $\{R_g\}$. Let $\mathcal{H}^*$ denote the dual Hilbert space, whose elements are "bras" $\langle\phi|$, and define operators $R_g^*$ on $\mathcal{H}^*$ by $R_g^* \langle\phi| := \langle\phi| R_{g^{-1}}$.

(a) Verify that the operators $\{R_g^*\}$ turn $\mathcal{H}^*$ into a representation of $G$. This representation is called the *dual representation* of $\mathcal{H}$.

(b) Show that if $\mathcal{H}$ is irreducible then $\mathcal{H}^*$ is irreducible.

A representation $\mathcal{H}$ is called *self-dual* if $\mathcal{H}^* \cong \mathcal{H}$.

(c) Show that the irreducible representations of $\mathrm{SU}(2)$, and hence all its representations, are self-dual.

(d) Show that any representation of $S_3$ is self-dual.

It is true more generally that any representation of $S_n$ is self-dual.

**Problem 4.4** (Many copies of a bipartite pure state)**.**
In this exercise, we will revisit the universal entanglement concentration protocol discussed in lecture 8. Let $|\phi\rangle_{AB}$ be an arbitrary state of two qubits. Then $|\phi\rangle_{AB}^{\otimes n}$ is a vector in the Hilbert space

$$(\mathbb{C}^2)^{\otimes n} \otimes (\mathbb{C}^2)^{\otimes n} \cong \left(\bigoplus_j V_{n,j}^A \otimes W_{n,j}^A\right) \otimes \left(\bigoplus_{j'} V_{n,j'}^B \otimes W_{n,j'}^B\right) \cong \bigoplus_{j,j'} V_{n,j}^A \otimes V_{n,j'}^B \otimes W_{n,j}^A \otimes W_{n,j'}^B.$$

The superscripts $A$ refer to the Schur-Weyl decomposition of the $n$ $A$-systems, and likewise for $B$. Now consider the representation of $S_n$ on $W_{n,j}^A \otimes W_{n,j'}^B$ given by the operators $R_\pi^{(n,j)} \otimes R_\pi^{(n,j')}$. A vector in $W_{n,j}^A \otimes W_{n,j'}^B$ is called an *invariant vector* if it is left unchanged by all these operators.

(a) Show that if $j \neq j'$ then $W_{n,j}^A \otimes W_{n,j'}^B$ contains no nonzero invariant vector for $S_n$.

(b) Show that $W_{n,j}^A \otimes W_{n,j}^B$ contains a unique invariant vector (up to scalar multiples). Moreover, show that this vector is a maximally entangled state, which we denote by $|\Phi^+\rangle_{W_{n,j}^A W_{n,j}^B}$.

*Hint: Use problem 4.3 and Schur's lemma.*

(c) Conclude that $|\psi\rangle_{AB}^{\otimes n}$ can be written in the form

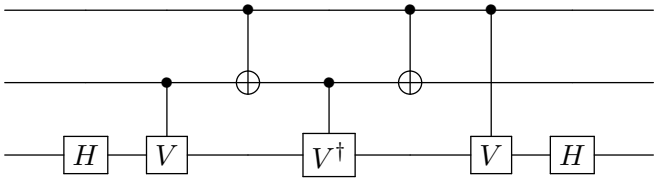
$$|\psi\rangle_{AB}^{\otimes n} \cong \sum_j \sqrt{p_j}\, |\Psi\rangle_{V_{n,j}^A V_{n,j}^B} \otimes |\Phi^+\rangle_{W_{n,j}^A W_{n,j}^B},$$

where $p_j = \mathrm{tr}[P_j \rho_A^{\otimes n}]$ and where the $|\Psi\rangle_{V_{n,j}^A V_{n,j}^B}$ are suitable pure states in $V_{n,j}^A \otimes V_{n,j}^B$.

(d) Use part (c) to analyze the universal entanglement concentration protocol discussed in class.

**Problem 4.5** (The controlled swap gate)**.**
In this exercise, you will decompose the controlled swap (CSWAP) gate into a quantum circuit that consists of single-qubit and two-qubit gates only.

(a) Compute the three-qubit unitary that corresponds to the following quantum circuit:




Here, $V = \begin{pmatrix} 1 & \\ & i \end{pmatrix}$ is a square root of the $Z$-gate.

The unitary from part (a) is known as the *Toffoli gate*.

(b) Show that the controlled swap gate can be implemented by a sequence of Toffoli gates.