Course FSMP, Fall'20: Interactions with Quantum Devices[1]

Thomas Vidick

September 4, 2025

[1] **(Thomas:** To the reader: feel free to make corrections where appropriate. If you would like me to notice a correction, use the macro "\typo{$\cdots$}". If you would like to insert a comment, use the macro "\com{$\cdots$}".)

Contents

Lecture 1

Introduction

1.1 Presentation of the course

Our goal in this course is to build towards a mostly self-contained presentation of two recent papers in quantum computing:

(a) *Classical verification of quantum computations*, by Mahadev [Mah18]. This paper addresses the question of *verification of quantum computation*: given *classical* data that is obtained from a quantum device that claims to have the ability to execute arbitrary quantum circuits (of polynomial size), how can a classical "verifier" ensure that the reported data indicates the correct outcome of the computation? For problems that have a natural classical certificate of correctness this is a simple task. For example, if the problem is to determine, given as input an integer n, if n has a prime factor larger than (say) $n^{1/4}$, then a positive answer can be certified by providing such a factor when it exists (and a negative answer can be certified by providing a complete prime decomposition of n). Such a factor, or more generally the prime decomposition of n, can be determined in quantum polynomial time using Shor's factoring algorithm; verifying it can be done in classical polynomial time.

However, not all problems that can be solved in quantum polynomial time are believed to lie in the class NP, i.e. not all quantum computations have outcomes that can be certified using an easily verifiable classical "witness". The well-known complexity-theoretic inclusion $\mathrm{BQP} \subseteq \mathrm{IP}$, that we discuss later in the course, implies that all problems in BQP have a classical randomized polynomial-time *interactive* verification procedure; however, in this procedure the prover may be asked to perform computations that are *harder* than BQP.

What Mahadev shows in her breakthrough result is that every polynomial time quantum computation can nevertheless be verified in classical randomized polynomial time by interacting with a quantum polynomial time device *as long as* one can ascertain (or believes) that the quantum device does not have the power to break a natural "post-quantum" cryptographic assumption (namely, the "Learning with Errors" (LWE) assumption).

(b) $\mathrm{MIP}^* = \mathrm{RE}$, by Ji, Natarajan, Vidick, Wright and Yuen [JNV$^+$20a]. The equality that gives this paper its title is an equality between complexity classes, i.e. computational problems that have a similar level of "difficulty" in a given model of computation (e.g. BQP and NP are complexity classes). Here, MIP^* designates all those computational problems that can be decided efficiently in classical randomized polynomial time (similar to the verification in the previous item) by asking (classical) questions to *two infinitely powerful quantum provers sharing entanglement*. Here the provers have unbounded

> computational capabilities and can hence solve any computational problem they like, including the one that the verifier is concerned with. However, it is assumed that the provers are not trusted: they will always try to convince the verifier that the answer to the problem in question is "yes" (e.g. "yes, this graph does have a valid 3-coloring"). The verifier has to employ certain tricks, or "tests", to detect any malicious behavior by the provers, so as never to make the wrong decision. The class RE denotes all problems for which there is an algorithm, running in any amount of time, that always eventually halts with the answer "yes" when this is the case (the algorithm does not need to halt in other cases). RE is a mind-boglingly large class of problems; it contains any decidable problem and even some undecidable ones such as the halting problem, which is complete for the class.
>
> The equality $\mathrm{MIP}^* = \mathrm{RE}$ is surprising in that it shows that access to untrusted quantum provers grants exceedingly large verification power to the polynomial-time verifier. In contrast, it is known that the same class without entanglement between the provers, denoted MIP, is much smaller, $\mathrm{MIP} = \mathrm{NEXP}$. Other motivations for the result tie it to questions in the foundations of quantum non-locality ("Tsirelson's problem") and the theory of operator algebras ("Connes' embedding problem") which we will discuss in due time.

There is a deep connection between the two papers mentioned above (as well as the many works that led to them—although the course is not meant to be a comprehensive survey, we will review the most relevant references in due course). Indeed, at their heart both works identify means by which a classical procedure is able to certify an appropriate "quantum computation workspace" within one (or two) quantum devices, using only a classical interaction with it. In this sense, and to borrow the title of one of the important earlier papers in this area [RUV13], both works provide techniques to tie a "classical leash" around a quantum system. In order to achieve this, both works ultimately have to tackle the same fundamental problem: what are classical signatures of quantum processes that can be leveraged to certify an entire computation? Very informally, we will see that in the case of (a) this signature is provided by the *uncertainty principle*: the fact that certain measurements in quantum mechanics are intrinsically incompatible; in the case of (b) it will be provided by *quantum non-locality*: the fact that entanglement allows distant parties to generate correlations that have no classical equivalent.

We have motivated our choice of topics by arguing that they tackle a fundamental problem, that of classically testing a quantum system. This is a problem of practical relevance (given an experimental quantum device, does it really do anything quantum?) as well as one that reaches deep to the power and limitations of the scientific method (for a discussion from an epistemological point of view, I recommend the great presentation by Aharonov and Vazirani [AV12]). It is also a problem that has turned out to stimulate many recent advances in quantum cryptography and complexity. For example, early works in delegated computation encouraged the development of quantum authentication codes, such as the Clifford code, that have found wide uses in cryptography; the study of quantum entangled-prover interactive proof systems has brought many discoveries in the foundations of quantum non-locality, such as dimension witnesses. The topics are connected through the common framework of interactive proof systems and share many techniques. Throughout the course we will make an effort to highlight the many open questions that stimulate research in this area and hope that our choice of results will provide a compelling entry point to it.

Having set the stage, let us discuss the structure of the course. We will start by tackling arguably the most fundamental question in this line of work: *what is a qubit*? How to define it mathematically, and how does one "certify" its existence? What does it mean to "test a qubit," and how can it be done? This question will occupy us for the coming four lectures. Exploring it will give us the opportunity to lay common foundations for the discussion of results (a) and (b) above. The remaining 6 lectures will be equally divided in two sets of 3 lectures each. The first set will examine the problem of delegation of polynomial-time quantum computations in general, and then focus on a presentation of Mahadev's result. The second set will discuss the theory of multi-prover interactive proof systems and build towards an understanding of the main ingredients in the result by Ji et al.

We will start the course slowly, so that the first few lectures are accessible to as broad a public as possible. The content will be mathematically precise and as self-contained as can be, so that it is possible to follow starting with only an elementary background in quantum computing at the level of the Nielsen & Chuang textbook [NC02]. Later lectures will be more involved and may require a somewhat more advanced knowledge of certain topics in cryptography or in complexity theory; I will aim to keep the presentation as self-contained as possible. My goal is that by the end of the course an assiduous participant may dive into either paper with a solid understanding of the main steps and key techniques in mind.

1.2 What is a qubit?

Let's start with the basics; as we will see shortly the simplest questions are not always the least interesting.

According to Wikipedia, "The qubit is the basic unit of information for quantum computers." The associated entry goes on to declare that "A qubit is a two-state (or two-level) quantum-mechanical system," the latter being defined as "a quantum system that can exist in any quantum superposition of two independent (physically distinguishable) quantum states." So a qubit is some kind of "system" whose state space has two fundamental states, and moreover such that the system can be in any "superposition" of these two states. Note that the definition makes an important distinction between the system (qubit) itself and its state. The latter is easier to define, and indeed it is through its state space that the "qubit" is generally introduced in quantum computing courses: the *state space of a qubit* is the set of unit vectors in the complex vector space $\mathbb{C}^2$; we denote this space as $\mathrm{S}(\mathbb{C}^2)$.[1] Thus the *state of a qubit* can be represented by a unit vector in $\mathbb{C}^2$; for example,

$$|0\rangle = \begin{pmatrix}1\\0\end{pmatrix} \qquad \text{and} \qquad |+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$$

are both valid states for a qubit.[2]

So we all know how to recognize a valid state for a qubit. But what about *the qubit itself*? We would like a precise definition that captures the intuition given in the Wikipedia entry: a qubit is a "two-state system" that can be in any "independent superposition" of its two states. In particular, we would like our definition to clearly distinguish this notion from that of a classical "probabilistic bit", which could also be considered a two-level system that can be in any "superposition" of its two states, $p\,\overline{0} + (1-p)\,\overline{1}$ for any $p \in [0,1]$. What distinguishes the real, "1-dimensional" degree of freedom $p \in [0,1]$ of the probabilistic bit from the complex, "2-dimensional" degree of freedom of the qubit?

As we will soon see, for our purposes it turns out to be most meaningful to attempt a definition in terms of the "Heisenberg representation" of quantum mechanics, that places *observable* quantities at a forefront. Informally, we will distinguish a quantum degree of freedom from a classical one by requiring that the quantum degree of freedom can be measured (observed) in two *mutually incompatible ways*. In order to make this precise we make a small detour to introduce the formalism associated with observable quantities in quantum mechanics.

[1] To be consistent with standard mathematical terminology we'd call this the complex projective space and denote it $P^2(\mathbb{C})$. We'll follow standard quantum computing conventions instead.

[2] For the entirety of this course (except the last lecture) we take the traditional perspective from quantum computer science: unless specified otherwise we consider that Hilbert spaces are always finite dimensional and computation is always performed on qubits in $\mathbb{C}^2$ or possibly higher-dimensional qudits in $\mathbb{C}^d$ for $d \geq 1$ a (finite) integer.

1.2.1 Observables

While the state space of a *qubit* is generally taken to be $\mathbb{C}^2$, more generally a quantum mechanical state lies in an arbitrary separable Hilbert space $\mathcal{H}$.[3] The term "observable" is used to denote any quantity that can in principle be obtained as the result of a measurement: for example, position, momentum, spin, energy (with respect to a certain Hamiltonian), are all observables. In full generality an observable is specified by a Hermitian operator O on $\mathcal{H}$.[4] The interpretation of this is that each eigenvalue of O represents a possible outcome under a measurement of the observable, and the associated eigenvectors denote states under which the observable deterministically yields that outcome. Thus if $O = \sum_i \lambda_i \Pi_i$ is a spectral decomposition of O, then any state $|\psi\rangle$ such that $\Pi_i|\psi\rangle = |\psi\rangle$ will deterministically yield the outcome λ_i when measured according to O. In particular, we see that the important part of an observable is its eigenprojections rather than the associated eigenvalues: the latter are real numbers that are associated to the different possible experimental outcomes. These numbers are generally associated to a physical meaning (such as position, momentum, etc.) but they can easily be changed by post-processing.

A collection of projection operators $\{\Pi_i\}$ that sum to identity is called a *Projector-Valued Measure* (PVM). When the Π_i are no longer required to be projections, it is called a *Positive Operator-Valued Measure* (POVM). This is the most general kind of measurement that is allowed in quantum mechanics. Given a POVM $\{\Pi_i\}$ the *Born rule* specifies that a measurement of an arbitrary state $|\psi\rangle$ under it will yield measurement outcome i with probability $\|\sqrt{\Pi_i}|\psi\rangle\|^2 = \langle\psi|\Pi_i|\psi\rangle$. Since the Π_i form a resolution of the identity ($\sum_i \Pi_i = \mathrm{Id}$) and $\||\psi\rangle\|^2 = 1$ we see that these probabilities always sum to 1, as they should. When the outcome i is obtained, the state evolves to a *post-measurement state* $|\psi'\rangle = \sqrt{\Pi_i}|\psi\rangle / \|\sqrt{\Pi_i}|\psi\rangle\|$. (Here $\sqrt{\Pi_i}$ is generally taken to be the positive square root of Π_i. However, other square roots can be used as well: since different square roots differ only by a unitary degree of freedom, choosing the one over the other is analogous to imposing an additional reversible evolution on the post-measurement state, which can be considered to be part of the measurement itself.)

In case the $\{\Pi_i\}$ are obtained from some observable O, we may use the associated eigenvalues to associate a real value to each measurement outcome i. In this case the *expectation* of the outcome of the measurement is

$$\sum_i \lambda_i \|\Pi_i|\psi\rangle\|^2 \,=\, \sum_i \lambda_i \langle\psi|\Pi_i|\psi\rangle \,=\, \langle\psi|O|\psi\rangle\,,$$

a quantity that is sometimes referred to as the "overlap" of $|\psi\rangle$ on O. We will use this formula often.

An observable such that $O^2 = \mathrm{Id}$ has at most two eigenvalues, which by convention we take to be -1 and $+1$. Such an observable is called a *binary* observable; it is the most frequent kind of observable that we will encounter.

1.2.2 First definition of a qubit

With a precise mathematical definition of an observable quantity we are ready to make precise our informal definition of a "qubit" as "a system that can be observed in two mutually incompatible ways".

[3] "Separable" means that $\mathcal{H}$ has a countable basis. In fact quantum states can also live in non-separable Hilbert spaces; we make the restriction for convenience. In fact for the purposes of these notes you may as well think of $\mathcal{H}$ as being finite-dimensional, i.e. $\mathcal{H}$ is isomorphic to the complex vector space $\mathbb{C}^d$, for some integer $d \geq 1$. Allowing infinite-dimensional spaces gives us a little more generality.

[4] For us "Hermitian" means that O is self-adjoint, $O = O^\dagger$, where $O^\dagger$ is the operator such that $\langle O^\dagger u|v\rangle = \langle u|Ov\rangle$ for all $|v\rangle$ in the domain of O. If $\mathcal{H}$ is finite-dimensional then a matrix representation of $O^\dagger$ is obtained by taking the conjugate-transpose of a matrix representation of O.

Definition 1.1 (Qubit, Take 1)**.** A *qubit* is a triple $(\mathcal{H}, X, Z)$ consisting of a separable Hilbert space $\mathcal{H}$ and a pair of Hermitian operators X, Z acting on a $\mathcal{H}$ such that $X^2 = Z^2 = \mathrm{Id}$ and $\{X, Z\} = XZ + ZX = 0$.

Let's see why Definition 1.1 captures the intuitive notion of "mutually incompatible" observables. Let the 'computational basis' be an eigenbasis of Z, and the 'Hadamard basis' an eigenbasis of X. Then we claim that the anticommutation relation $XZ + ZX = 0$ ensures that any vector in the former makes a 45° angle with any vector in the latter. To see this, let $|\psi\rangle$ be an eigenvector of X with associated eigenvalue $\varepsilon \in \{\pm 1\}$. Then $\langle\psi|XZ + ZX|\psi\rangle = 0$ immediately implies $2\varepsilon\langle\psi|Z|\psi\rangle = 0$. Given that Z only has -1 and $+1$ as eigenvalues, this relation implies that the projections of $|\psi\rangle$ on the two eigenspaces of Z have equal length; in other words, $|\psi\rangle$ lies exactly between the $+1$ and -1 eigenspaces of Z. (Yet another way of saying this is that all principal angles between an eigenspace of X and one of Z are $\frac{\pi}{4}$; we will make this precise soon.) In this sense any X and Z satisfying the conditions of the definition are "maximally incompatible": any definite state for the one is entirely undetermined (i.e. yields uniformly random outcomes when measured) under the other.

There is a problem with this definition: by allowing the underlying Hilbert space $\mathcal{H}$ to be arbitrary we seem to have all but dropped the earlier requirement that a qubit is a system whose state space is "two-level" and thus identifiable with the projective space $\mathrm{S}(\mathbb{C}^2)$. Luckily, the following lemma allows us to make the connection with this requirement.

Lemma 1.2. *Let $(\mathcal{H}, X, Z)$ be a qubit. Then there is a Hilbert space $\mathcal{H}'$ and an isomorphism $\mathcal{H} \simeq \mathbb{C}^2 \otimes \mathcal{H}'$ such that under the same isomorphism, $X \simeq \sigma_X \otimes \mathrm{Id}$ and $Z \simeq \sigma_Z \otimes \mathrm{Id}$.*[5] *Here, σ_X and σ_Z are the usual Pauli observables on $\mathbb{C}^2$: in matrix form,*

$$\sigma_X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \textit{and} \qquad \sigma_Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} .$$

Note that a consequence of the lemma is that qubits, as defined in Definition 1.1, only exist in spaces of even (or infinite) dimension! In particular, qubits don't exist in dimension 1; indeed, in dimension 1 all operators commute. This is satisfactory: intuitively, a situation in which all possible observables commute ought to be considered "classical" (for instance, because there is a complete set of simultaneous eigenvectors for all observables).

It will be essential for our later goals that Definition 1.1 does not *a priori* require $\mathcal{H}$ to be a two-dimensional space. Indeed, how would one test such a claim? One does not "see" the dimension of the state space; while it is possible to probe parts of it it can never be excluded that the state space is larger than what is accessible to the experimentalist's setup. In this sense Definition 1.1 has a nice "operational" flavor to it: it refers to *observables* of the system and their properties. Although much more work is needed before we are able to make any of these statements formal, we see the definition as a good step towards giving us the ability to "test" that a system "is a qubit". In addition, the definition clearly has meaningful consequences; in particular it implies that qubits do not have a "classical explanation", so that a "test for a qubit" can serve as a "test for quantumness", i.e. a test that distinguishes quantum from classical behavior.

The proof of the lemma makes use of an elementary but fundamental tool in the analysis of many quantum information protocols, the CS (for "Cosine-Sine") decomposition. This decomposition is also known as "Jordan's lemma," after Camille Jordan's *Traité des substitutions et des équations algébriques* from 1870 (see this EHESS PhD thesis [Bre06] for a masterful 730-page account of the history behind the

[5] The reader might wonder what happened to σ_Y... Don't we need it to define our qubit? Here we are taking the "operator algebraists"' perspective, which is that if the system supports X and Z observables then it also supports $Y = iXZ$. Because Y is determined by X and Z, we do not include it in the definition.

use of Jordan's name alongside this theorem). Given that we will use the lemma frequently we give it a self-contained treatment in the next section.

1.2.3 Jordan's lemma

The discussion of Jordan's lemma in this section is mostly borrowed from `https://cims.nyu.edu/~regev/teaching/quantum_fall_2005/ln/qma.pdf`. Let P, Q be two orthogonal projections on a separable Hilbert space $\mathcal{H}$, and consider their sum $R = P + Q$. Then R is Hermitian so it has an orthonormal set of eigenvectors that is also a basis for $\mathcal{H}$.[6] Let $|\varphi\rangle$ be any eigenvector of R with associated eigenvalue λ. Consider two cases for the vector $P|\varphi\rangle$. The first case is that $P|\varphi\rangle$ is parallel to $|\varphi\rangle$. In this case, since

$$Q|\varphi\rangle = R|\varphi\rangle - P|\varphi\rangle = \lambda|\varphi\rangle - P|\varphi\rangle \tag{1.1}$$

it follows that $Q|\varphi\rangle$ is also parallel to $|\varphi\rangle$, so $|\varphi\rangle$ is a common eigenvector of P and Q. The second case is that $P|\varphi\rangle$ is linearly independent from $|\varphi\rangle$. In this case the two-dimensional subspace $\mathcal{S}$ spanned by $|\varphi\rangle$ and $P|\varphi\rangle$ is stable under P, because being a projection P satisfies $P^2 = P$. Moreover, by (1.1) we see that $Q|\varphi\rangle$ lies in $\mathcal{S}$, and

$$QP|\varphi\rangle \;=\; Q(\lambda|\varphi\rangle - Q|\varphi\rangle) = (\lambda - 1)Q|\varphi\rangle$$

is also in that subspace. Using that P, Q are projections and that P is neither 0 nor the identity on $\mathcal{S}$ (otherwise we would have been in the first case) it follows that there is a basis of $\mathcal{S}$ such that in that basis,

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad Q = \begin{pmatrix} c^2 & cs \\ cs & s^2 \end{pmatrix}\,, \tag{1.2}$$

where $c = \cos\theta$ and $s = \sin\theta$ for some $\theta \in [0, \pi/2)$. (Other values of θ reduce to this case because the sign of cs can be flipped by negating the second basis vector for the chosen basis of $\mathcal{S}$.) Note finally that since $\mathcal{S}$ is stable by both P and Q it is stable by $R = P + Q$, so it has a basis made of eigenvectors of R, the vector $|\varphi\rangle$ that we started from and its orthogonal in R. Proceeding in this way inductively this lets us identify an eigenbasis of R such that its vectors are either isolated (stable by both P and Q) or in pairs (spanning a 2-dimensional subspace that is stable by both P and Q).

The following lemma summarizes the discussion so far.

Lemma 1.3. *Let P, Q be projections on a separable Hilbert space $\mathcal{H}$. Then there exists an orthogonal decomposition $\mathcal{H} = \oplus_i \mathcal{S}_i$ such that each $\mathcal{S}_i$ is a* 1- *or* 2*-dimensional subspace that is stable by P and Q. Furthermore, whenever $\mathcal{S}_i$ is* 2*-dimensional there is a basis for it in which P and Q take the form* (1.2)*, for some c_i and s_i that may depend on $\mathcal{S}_i$.*[7]

This very useful lemma informally says that, when only two projections are concerned, we can reduce the analysis to a 2-dimensional problem. Let's apply it to show Lemma 1.2.

Proof of Lemma 1.2. Let $\mathcal{H}$, X, Z be as in the statement of the lemma. Let $P = \frac{1}{2}(Z + \mathrm{Id})$ and $Q = \frac{1}{2}(X + \mathrm{Id})$. Then P, Q are projections on $\mathcal{H}$ so we can decompose them according to Lemma 1.3. Let $(|e_i\rangle, |f_i\rangle)$ be a basis for the i-th space $\mathcal{S}_i$ in which the matrices for P and Q have the form (1.2). Using $XZ + ZX = 0$ it follows that (i) there cannot be any 1-dimensional blocks, because these necessarily

[6] Here we are using that $\mathcal{H}$ is separable.

[7] For the case of 1-dimensional subspaces, since P and Q are projections they are each either identically 0 or identity in those subspaces.

commute, and (ii) in any two-dimensional block, the angle θ must equal $\pi/4$, as this is the only value in $[0, \pi/2)$ that leads to anti-commuting operators. Thus in each space $\mathcal{S}_i$, Z acts exactly as σ_Z and X as σ_X. Let $\mathcal{H}'$ have canonical basis $\{|i\rangle\}$, where i ranges over the block indices in the decomposition of P and Q. The required isomorphism is obtained by e.g. mapping $|e_i\rangle \in \mathcal{H}$ to $|0\rangle \otimes |i\rangle \in \mathbb{C}^2 \otimes \mathcal{H}'$ and $|f_i\rangle \in \mathcal{H}$ to $|1\rangle \otimes |i\rangle \in \mathbb{C}^2 \otimes \mathcal{H}'$. □

1.2.4 *n* qubits

Now that we have a working definition of a qubit, how about two, three, or even n qubits? What we mean when we say that a system "has n qubits" is that (i) it should have n copies of one qubit, so there should be $(X_1, Z_1), (X_2, Z_2), \dots, (X_n, Z_n)$ on $\mathcal{H}$ such that each pair satisfies the definition of a qubit, and moreover (ii) the qubits should be "independent": indeed, we wouldn't want something like $X_1 = X_2$ to happen. How do we prevent this? Intuitively "independence" of the qubits should be reflected in the fact that they can be observed "independently", in any order, such that if e.g. we " observe" qubits 1 and 2 and then discard the outcome associated with qubit 1, we should obtain an outcome that is identically distributed as if we had only observed qubit 2 in the first place. These considerations suggest the following definition.

Definition 1.4 (n qubits, Take 1)**.** A system of *n qubits* is a tuple $(\mathcal{H}, X_1, Z_1, \dots, X_n, Z_n)$ consisting of a separable Hilbert space $\mathcal{H}$ and n pairs of Hermitian operators (X_i, Z_i) for $i \in \{1, \dots, n\}$ acting on $\mathcal{H}$ such that

(i) For each $i \in \{1, \dots, n\}$, $(\mathcal{H}, X_i, Z_i)$ is a qubit;

(ii) For each $i \neq j \in \{1, \dots, n\}$, qubits i and j are independent:

$$[X_i, X_j] = [X_i, Z_j] = [Z_i, X_j] = [Z_i, Z_j] = 0 ,$$

where for arbitrary operators A, B on $\mathcal{H}$, $[A, B] = AB - BA$ denotes the algebra commutator.

The commutation condition (ii) indeed implies that measurements on different qubits can be performed "independently". For example, the expectation of a measurement of qubit 2 in the Hadamard basis *after* qubit 1 has been measured in the computational basis is given by (where we write $X_1 = X_1^0 - X_1^1$ for the spectral decomposition of the observable X_1)

$$\begin{aligned} \langle\psi|X_1^0 Z_2 X_1^0|\psi\rangle + \langle\psi|X_1^1 Z_2 X_1^1|\psi\rangle &= \langle\psi|X_1 Z_2 X_1|\psi\rangle \\ &= \langle\psi|Z_2 X_1^2|\psi\rangle \\ &= \langle\psi|Z_2|\psi\rangle , \end{aligned}$$

as desired. (Here, for the first equality we used that $\langle\psi|X_1^0 Z X_1^1|\psi\rangle = \langle\psi|X_1^1 Z X_1^0|\psi\rangle$, for the second we used the commutation condition $[X_1, Z_2] = 0$, and for the last we used $X_1^2 = \mathrm{Id}$.) This calculation can be done for any pair of qubits, or even any sequence of measurements of qubits, and it shows that item (ii) in Definition 1.4 indeed captures the idea that each of the n qubits can be measured independently.

Note however that this "independence" is not necessarily quite the same as there truly being n qubits. In particular, if our definition of a qubit only required the use of a single observable X, as a classical bit would, then taking $X_1 = \cdots = X_n$ would satisfy both (i) and (ii), since an operator always commutes with itself. Indeed, just as we're used to a qubit being defined through its state space $\mathrm{S}(\mathbb{C}^2)$, we're used to n qubits being defined through their state space $\mathrm{S}(\mathbb{C}^2) \otimes \cdots \otimes \mathrm{S}(\mathbb{C}^2)$. Where is the tensor product in Definition 1.4, isn't it missing? The following lemma shows that an n-fold tensor product is in fact implicit in the definition.

Lemma 1.5. *Let $(\mathcal{H}, X_1, Z_1, \ldots, X_n, Z_n)$ be a system of n qubits. Then there exists a Hilbert space $\mathcal{H}'$ and an isomorphism $\mathcal{H} \simeq (\mathbb{C}^2)^{\otimes n} \otimes \mathcal{H}'$ such that under the same isomorphism, for every $i \in \{1, \ldots, n\}$ and $W \in \{X, Z\}$, $W_i \simeq \sigma_{W,i} \otimes \mathrm{Id}_{\mathcal{H}'}$, where here $\sigma_{W,i}$ denotes the Pauli W operator acting on the i-th copy of $\mathbb{C}^2$.*

Proof. We show the lemma by induction on $n \geq 1$. The case $n = 1$ is provided by Lemma 1.2. Suppose the lemma shown for some $n \geq 1$, show it for $(n+1)$. Let $(\mathcal{H}, X_1, Z_1, \ldots, X_{n+1}, Z_{n+1})$ be a system of $(n+1)$ qubits. Since $(\mathcal{H}, X_1, Z_1, \ldots, X_n, Z_n)$ is a system n qubits we can apply the induction hypothesis to it. Let $\mathcal{H}'$ and π' be the promised space and isomorphism. The key step is provided by the following claim.

Claim 1.6. *Let W be an Hermitian operator on $\mathcal{H}$ such that $[W, X_i] = [W, Z_i] = 0$ for all $i \in \{1, \ldots, n\}$. Then there exists W' Hermitian acting on $\mathcal{H}'$ such that under π', $W \simeq \mathrm{Id}_{(\mathbb{C}^2)^{\otimes n}} \otimes W'$.*

The claim immediately gives us the induction step: by applying it to X_{n+1} and Z_{n+1} we find X'_{n+1} and Z'_{n+1} on $\mathcal{H}'$ such that $(\mathcal{H}', X'_{n+1}, Z'_{n+1})$ is a qubit. Applying Lemma 1.2 to this qubit and composing the isomorphism obtained with π' completes the induction step. Therefore, it only remains to prove the claim.

Proof of Claim 1.6. Clearly it suffices to prove the statement "under π", i.e. for the case where $X_i = \sigma_{X,i}$ and $Z_i = \sigma_{Z,i}$. We introduce the following notation: for $a, b \in \{0,1\}^n$,

$$\sigma_X(a) = \sigma_{X,1}^{a_1} \otimes \cdots \otimes \sigma_{X,n}^{a_n} \qquad \text{and} \qquad \sigma_Z(b) = \sigma_{Z,1}^{b_1} \otimes \cdots \otimes \sigma_{Z,n}^{b_n} .$$

Using that the four 1-qubit Pauli matrixes Id, σ_X, σ_Z and $\sigma_X \sigma_Z$ form a basis for the complex vector space of linear operators on $\mathbb{C}^2$, W has a decomposition

$$W = \sum_{a,b} \sigma_X(a)\sigma_Z(b) \otimes W_{a,b} ,$$

where $W_{a,b}$ are arbitrary operators on $\mathcal{H}'$ (they are not necessarily Hermitian). Let's write out the left and right products of W with $\sigma_X(c)\sigma_Z(d)$, for some $c, d \in \{0,1\}^n$:

$$\begin{aligned} \sigma_X(c)\sigma_Z(d)W &= \sum_{a,b} \sigma_X(c)\sigma_Z(d)\sigma_X(a)\sigma_Z(b) \otimes W_{a,b} \\ &= \sum_{a,b} (-1)^{a \cdot d} \sigma_X(a+c)\sigma_Z(b+d) \otimes W_{a,b} , \end{aligned} \tag{1.3}$$

where in the second line we used the anti-commutation relation $\sigma_Z(d)\sigma_X(a) = (-1)^{a \cdot d}\sigma_X(a)\sigma_Z(d)$, as well as the "additivity" relations $\sigma_X(a+c) = \sigma_X(a)\sigma_X(c)$, and similarly for σ_Z. Similarly,

$$\begin{aligned} W\sigma_X(c)\sigma_Z(d) &= \sum_{a,b} \sigma_X(a)\sigma_Z(b)\sigma_X(c)\sigma_Z(d) \otimes W_{a,b} \\ &= \sum_{a,b} (-1)^{b \cdot c} \sigma_X(a+c)\sigma_Z(b+d) \otimes W_{a,b} . \end{aligned} \tag{1.4}$$

Using that the $\sigma_X(a)\sigma_Z(b)$ are linearly independent we can identify terms in (1.3) and (1.4); it follows that for any a, b, c, d, $(-1)^{b \cdot c} W_{a,b} = (-1)^{a \cdot d} W_{a,b}$. For any (a, b) unless $a = b = 0$ we can find strings c, d such that the two terms in $W_{a,b}$ are given opposite signs. Thus $W_{a,b} = 0$ whenever $(a, b) \neq (0, 0)$, and $W = \mathrm{Id} \otimes W_{0,0}$. Since W is Hermitian, $W_{0,0}$ is also Hermitian, proving the claim. □

□

Remark 1.7. The statement of Lemma 1.5 can be reformulated in the language of group representation theory, and this reformulation will be useful later on. The "n qubit Weyl-Heisenberg group" is the $2 \cdot 4^n$-element group G_n that is generated by the n-qubit σ_X and σ_Z matrices; its elements are $(-1)^c \sigma_X(a)\sigma_Z(b)$ for $a, b \in \{0,1\}^n$ and $c \in \{0,1\}$. From any system of n qubits $(\mathcal{H}, X_1, Z_1, \ldots, X_n, Z_n)$ it is straightforward to specify a representation ϕ of G_n by setting $\phi((-1)^c \sigma_X(a)\sigma_Z(b)) = (-1)^c \prod_i X_i^{a_i} \prod_i Z_i^{b_i}$. The lemma can be adapted to show that any representation of G_n that in addition sends -1 to -1, as ϕ does, must be a direct sum of copies of the representation by Pauli matrices.

1.2.5 Approximate qubits

An important theme of this course will be that the objects that we observe and manipulate in our "protocols" or "experiments" generally cannot be assumed to be perfectly "clean" or "noise-free". In this respect, the following exercise is a simple test that we ought to make on our definition; furthermore, it is a good exercise to practice the use of the CS decomposition.

Exercise 1.1. Suppose that X and Z are binary observables on $\mathcal{H}$ such that $\|\{X, Z\}\| \leq \varepsilon$ for some $\varepsilon \geq 0$, where $\| \cdot \|$ denotes the operator norm (largest singular value). Show that there exists a qubit $(\mathcal{H}, X', Z')$ such that $\max\{\|X - X'\|, \|Z - Z'\|\} \leq \delta(\varepsilon)$. State the best dependence δ that you can get.

The exercise can be extended to consider n approximate qubits, but the proof is more delicate as some work is needed to keep the errors under control. The following is shown in [CRSV17].

Theorem 1.8. *Let $X_1, Z_1, \ldots, X_n, Z_n$ be binary observables on $\mathcal{H}$ and $\varepsilon \geq 0$ such that $\varepsilon/(1-\varepsilon)^2 \leq 1/(64n)$ and $\|\{X_i, Z_i\}\| \leq \varepsilon$ and $\|[S_i, T_j]\| \leq \varepsilon$ for all $i \neq j \in \{1, \ldots, n\}$ and $S, T \in \{X, Z\}$.*[8] *Then there exists binary observables $X'_1, Z'_1, \ldots, X'_n, Z'_n$ on $\mathcal{H}$ such that $\{X'_i, Z'_i\} = 0$, $[S'_i, T'_j] = 0$ and moreover $\|S'_j - S_j\| \leq 4n\varepsilon/(1-\varepsilon)^2 + \varepsilon$ for all $i \neq j \in \{1, \ldots, n\}$ and $S, T \in \{X, Z\}$.*

The theorem shows that "n approximate qubits" are close to "n exact qubits" according to our definitions. Note that there is a dependence of the error on n, but not on the dimension of $\mathcal{H}$. In [CRSV17] it is shown by an explicit example that a linear dependence on n is necessary.

1.2.6 An operational definition?

Earlier we qualified our definition of a qubit as being "operational". This term is generally used to refer to a definition that can be "observed", i.e. the definition should imply a natural "test" that can be performed experimentally and that "certifies" that a certain object satisfies the definition in some way or another. Definition 1.1, or even the approximate version of it suggested in the previous section, lacks severely in that matter: who gets to observe "operators"? How can the anti-commutator be "witnessed"? In the next lecture we will refine our definition so that we are able to answer these questions.

[8] We slightly abuse notation in writing e.g. S_i for X_i when $S = X$.

Lecture 2

Testing a qubit

Recall the definition of a qubit from the first lecture: a qubit is a triple $(\mathcal{H}, X, Z)$ of a Hilbert space $\mathcal{H}$ and a pair of binary observables X and Z on $\mathcal{H}$ such that $\{X, Z\} = 0$. Unfortunately, this definition is far from operational! The operator condition $\{X, Z\} = 0$ is not something that we can hope to test based on experimental data alone. A simple reason for this is that in general we may only hope to observe expectation values for observables W *evaluated on a certain state* $|\psi\rangle$. While we may be able to prepare quite a range of states $|\psi\rangle$ using our experimental system, there is no hope that we can prepare *all* possible states, which would be required for a full "tomography" of the observable.[1]

Today we tweak our definition to obtain a new one that we claim is truly "operational," and we start exploring means of justifying this claim by showing how the definition can be "tested".

2.1 Setup

Before we can make precise what we mean by "operational" we need to describe the framework in which we operate. This framework is inspired by an (idealized) perspective on how "real-life" experiments are made. An experiment can be formalized as an interactive process that involves two entities. One of the entities is the "experimentalist", whom we will refer to as the *verifier*. The other entity is the "quantum device" on which the experiment is being performed. We will personify that device and refer to it as the *prover*. (We motivate this terminology a little later.)

In an experiment the experimentalist generally has a model of how the device is expected to behave. This model can be used to predict the input-output behavior for the device, i.e. how it will react to various stimuli that the experimentalist might subject it to. For example, the device may be the combination of a laser, a sheet of paper with two slits on it, and a screen. This device takes inputs in $\{0, 1\}^2$ that model the experimentalist's choice of slits to open (0 for 'open' and 1 for 'closed'). The device's outputs are elements of, say, $\{R, G, B\}^{1000\times1000}$, i.e. a 1000×1000 pixel RGB image of the screen. The experimentalist's model makes a prediction for the device's output pic_{ab} on each possible input (a, b). In addition, if the experimentalist follows best practices in statistics they should decide *a priori* on a *scoring function* that determines, whenever the experiment is performed, a "success score" for the experimental outcome obtained. In our example we could use the normalized Hamming distance $10^{-6}d_H(pic_{ab}, res_{ab})$ where for $a, b \in \{0, 1\}$, pic_{ab} is the ideal outcome and res_{ab} the experimental outcome. The experimentalist would then repeatedly provide

[1] The problem is distinct from the "exponential scaling" of the Hilbert space: here, the issue is that we simply can't expect that the experimentalist has the ability to probe the system's entire Hilbert space; in particular, we cannot impose any dimension bound *a priori*.

inputs $(a^{(i)}, b^{(i)})$ to the device for $i = 1, \dots,$ chosen uniformly at random or with some smarter distribution (e.g. she could decide to never test the case $(1,1)$ corresponding to both slits closed), obtain a sequence of outputs $res^{(1)}, res^{(2)}, \dots$, and return an averaged score that quantifies agreement of the experiment with the theory.

Having set the stage with this rather loose description we make some important remarks:

1. Our notion of interactive experiment substantially restricts the means by which the experimentalist may interact with the device. The experimentalist is allowed to provide classical inputs and obtain classical outputs in return. While there may be some semantics associated with the inputs and outputs ("which slit is open", "an RGB image") when the experiment is performed there is no guarantee that these semantics correspond to any real-life phenomenon: inputs and outputs are strings of bits, nothing more. There is no a priori guarantee that the device has any number of "slits" that are being "opened" or "closed"; maybe the mysterious device contains a student equipped with a textbook on quantum mechanics that allows her to calculate a reasonable outcome for the experiment. We emphasize that "in real life" the experimentalist will typically make a number of explicit and implicit assumptions about the system that is being tested and how it is accessed; here we aim to minimize such assumptions to the extreme.

2. The insistence on classical inputs and outputs also means that we forbid the experimentalist from directly accessing the quantum state or measurements of the device. One of our basic goals is to devise tests that distinguish a classical device from a quantum one, and so we cannot assume any quantum access to it a priori. We will formalize this model of "black-box access" in more detail in the next section.

3. Nevertheless, we will assume throughout that quantum mechanics is a correct theory, i.e. the device can always be modeled using the framework of quantum mechanics: it has a quantum state (that may be entangled with the environment) that it evolves unitarily and measures according to the Born rule. What we will aim to test is e.g. that the device *does not* have a model in classical mechanics.

4. Assuming correctness of quantum mechanics will not suffice. Make sure that you can convince yourself of the following statement: "for any non-trivial experiment, i.e. such that for each possible input of the verifier there is at least one output that would be accepted, there is a classical device that is always accepted in the experiment." In other words, any meaningful experiment will need to place additional assumptions on the device: maybe the "valid" outcomes are hard to compute for a classical device, or maybe they are impossible to generate without entanglement or communication, etc. Are we contradicting our first item? It all depends on what the assumption is. We will aim for assumptions that require the least "faith" possible in the adequate execution of the experiment (i.e. the experimentalist's skills).

5. We ended the description of the double slit experiment by suggesting that the experimentalist may repeat the same experiment multiple times in order to collect statistics. In real life there is no guarantee that a device behaves identically from one experiment to the next; its behavior may naturally fluctuate with time, or it may have memory and adapt itself, etc. The assumption that the device can be accessed repeatedly without changing its behavior is called the "i.i.d. assumption", for "identically and independently distributed". We will make that assumption when it is convenient; more often than not it can be dropped at the cost of substantial technical work that we will not always have the opportunity to accomplish.

2.2 Interactive proofs

With this informal motivation for our notion of an "interactive experiment" in place we now give a more precise framework for modeling such experiments. For this we adapt the framework of *interactive proof systems* from cryptography and complexity theory. In this framework it is generally assumed that a trusted entity called the *verifier* interacts with an not-necessarily-trusted entity called the *prover*. The verifier is trying to verify some claim about the world (e.g. in complexity, that some input formula φ is satisfiable) or about the prover itself (e.g. in cryptography, that the prover has the right identifying information). Towards this the verifier may "interrogate" the prover in an interactive manner. At the end of the interaction the verifier makes a decision to accept or reject. Informally, the proof system will be called "sound" if whenever the verifier accepts, the claim is indeed true.

The formal definition of an interactive proof system makes use of the notion of "interactive Turing machine" to model the prover and verifier. Since this formalism will not be essential for us we refer the interested reader to [VW16, Chapter 4] for details. In complexity theory an interactive proof is always associated to a *language*, that is a collection of problems, usually specified by strings $x \in \{0,1\}^*$, such that some of the problems have an affirmative answer and some have a negative answer (e.g. the problems could be graphs, and the ones with affirmative answer those that have a proper 3-coloring). At the beginning of the interactive proof both prover and verifier are provided with a problem instance x, and the goal of the verifier is to leverage the prover's computational power to help her determine if x is a positive instance, all the while accounting for the fact that the prover may misbehave.

For our purposes we are led to slightly broaden the notion in an informal manner, so that we can not only associate interactive proof systems to formal languages but also to statements about the device itself, as is sometimes done in cryptographic applications of interactive proof systems. We will thus refer to an interactive proof system, or sometimes more simply a "test," for a *hypothesis* H as the specification of a verifier in an interactive protocol with the following properties: (In the protocol both verifier and prover may be provided with some auxiliary input, a classical x_V for the verifier and a quantum ρ_P for the prover.)

1. *Completeness:* This property means that whenever the hypothesis H (which may depend on x_V and/or ρ_P) is satisfied there is a way for the prover to be accepted in the protocol "with high probability." We will sometimes use a parameter $c \in [0,1]$ to designate the probability that a "honest prover" succeeds in the protocol.

2. *Soundness:* This property means that whenever the hypothesis H is not true no prover can succeed in the protocol with probability higher than a small quantity $s \in [0,1]$ termed the "soundness parameter".

We give a few examples. In the traditional setting of interactive proof systems the hypothesis H is that $x_V = \rho_P \in L$, where L is a fixed language, x_V is the verifier's input, and the prover's input ρ_P is assumed to equal x_V. For example if $L = 3COL$ then completeness states that whenever both V and P are provided with the valid description of a graph as input, and that this graph is 3-colorable, there must be a way for the prover to convince the verifier that this is so; soundness states that whenever x_V designates a graph that is not 3-colorable, irrespective of what ρ_P is there is no way for the prover to convince the verifier. (An interactive proof system that satisfies both conditions is one in which the verifier simply expects the prover to directly provide it with a proper coloring.)

As a second example, H could be the hypothesis that "P has the BB'84 state that is specified by x_V". In this case we expect that e.g. $x_V = (v,\theta)$ for $v,\theta \in \{0,1\}$ and $\rho_P = H^\theta|v\rangle$. Completeness states that if this is indeed the case then there should be a way for P to succeed; soundness states the converse. There is an easy quantum protocol for this hypothesis in which P is expected to provide its qubit to V, who verifies it by performing the appropriate measurement. But is there a classical protocol?

Finally, a less formal but more interesting for us example is that we could consider H to be the hypothesis that "P has a qubit.". In this case we do not make use of the auxiliary inputs; completeness states that for any prover that does have a qubit (i.e. P has access to observables X, Z on $\mathcal{H}$ such that $\{X, Z\} = 0$) then there should be a way for it to succeed in the protocol, whereas soundness states that conversely, any prover that succeeds in the protocol must "have a qubit."

2.3 An operational definition of a qubit

Given an interactive experiment of the sort described in the previous section, how do we model the actions of an arbitrary prover in the protocol? At each stage of the protocol the prover receives a question $x \in \mathcal{X}$ and is expected to provide an answer $a \in \mathcal{A}$. Here $\mathcal{X}$ and $\mathcal{A}$ are finite sets that are specified by the protocol. Although in general these sets may vary depending on the round in the protocol, for convenience we can assume that it is always the same set of questions and of answers that is used.

As discussed earlier we will in general make the assumption that the prover's actions can be modeled using quantum mechanics. Thus there must exist a Hilbert space $\mathcal{H}$ associated with the prover and a state $\rho \in \mathrm{D}(\mathcal{H})$ that the prover possesses at the start of the protocol.

Remark 2.1. Here we start using the density matrix representation for quantum states: we use the notation $\mathrm{D}(\mathcal{H})$ to represent the set of density matrices on $\mathcal{H}$, i.e. positive semidefinite matrices with trace 1. A density matrix is used to represent part of a quantum state $|\psi\rangle \in \mathcal{H} \otimes \mathcal{H}'$. Here $\mathcal{H}'$ designates the Hilbert space associated with the "environment", which is everything that is not in the possession of V or P. Since we do not want to rule out that the prover may share entanglement with the environment, we do not assume that their initial state is a pure state $|\psi\rangle$.

When P receives its question x it measures some observable $O_x = \sum_a \lambda_a \Pi^x_a$, where λ_a are arbitrary and Π^x_a projections that sum to identity, i.e. Π^x_a is a POVM. According to the Born rule, it obtains an answer distributed as $\Pr(a|x) = \mathrm{Tr}(\Pi^x_a \rho)$.[2] Finally the quantum state ρ of the prover gets updated as a function of the outcome obtained. This formalization is fully general; in particular it can be used to model classical deterministic strategies by setting $\Pi^x_a = 1_{f(x)=a}$ where f would be the function used by the prover to determine its answers. Similarly, randomized strategies can be represented by making use of a totally mixed state $\rho = \sum_r p_r |r\rangle\langle r|$, for some arbitrary distribution $\{p_r\}$, to capture the randomness.

When the interactive experiment is executed the only observable data that is accessible to the experimentalist is, at best, the probabilities $\Pr(a|x)$.[3] An important consequence of this is that we cannot hope to achieve a characterization of the prover's *observable itself*, but instead may only make assertions about the *action of the observable on the state*. That is, if O is an observable, $|\psi\rangle$ a state on which it acts, and U an arbitrary unitary,

$$\langle\psi|O|\psi\rangle = \langle U\psi|(UOU^\dagger)|U\psi\rangle \ .$$

Thus two models of the prover, using state $|\psi\rangle$ and observable O or using state $U|\psi\rangle$ and observable $UOU^\dagger$, lead exactly to the same observed data. Our earlier definition of a qubit, by ignoring the role played by the state and imposing constraints on the operators themselves, violates this. This leads us to update our first definition as follows.

[2] This is the generalization of the Born rule to density matrices. We recover the pure case by restricting to $\rho = |\psi\rangle\langle\psi|$, in which case using cyclicity of the trace, $\mathrm{Tr}(\Pi^x_a \rho) = \mathrm{Tr}(\Pi^x_a |\psi\rangle\langle\psi|) = \langle\psi|\Pi^x_a|\psi\rangle$.

[3] We write "at best" because the experimentalist does not get to see probabilities. Under the i.i.d. assumption it can sometimes estimate them to within an additive error. However, in the case where $\mathcal{A}$ is a large alphabet it may be that all probabilities are exponentially small. This will be the case in some of the experiments that we describe.

Definition 2.2 (Qubit, Take 2)**.** A *qubit* is a triple $(|\psi\rangle, X, Z)$ such that $|\psi\rangle \in \mathrm{S}(\mathcal{H})$, where $\mathcal{H}$ is a separable Hilbert space left implicit in the notation, and X and Z are Hermitian operators on $\mathcal{H}$ such that

$$\{X, Z\}|\psi\rangle = 0\;. \tag{2.1}$$

Note that the definition still makes the requirement that $X^2 = Z^2 = \mathrm{Id}$ as operators. This is because this requirement follows from the laws of quantum mechanics themselves; informally, it just means that each of X and Z has a spectral decomposition with two associated eigenprojections, i.e. they represent valid binary observables.

At this point there are two important questions we should be asking: (i) Is this definition meaningful? With the anti-commutator weakened as in (2.1), does the definition still capture our intuitive notion of a qubit? (ii) We weakened the definition in an arbitrary-looking way by inserting a dependence on the state vector $|\psi\rangle$. Can we justify this, i.e. are we now able to develop protocols that test the definition?

In the remainder of the lecture we provide partial answers to these two questions. To answer the first, we show the following.

Lemma 2.3. *Let $(|\psi\rangle, X, Z)$ be a qubit on $\mathcal{H}$. Then there exists a Hilbert space $\mathcal{H}'$ and an isometry $V : \mathcal{H} \to \mathbb{C}^2 \otimes \mathcal{H}'$ such that*

$$VX|\psi\rangle = (\sigma_X \otimes \mathrm{Id})V|\psi\rangle \qquad \textit{and} \qquad VZ|\psi\rangle = (\sigma_Z \otimes \mathrm{Id})V|\psi\rangle\;. \tag{2.2}$$

The following diagram illustrates the situation guaranteed by the lemma:

$$\begin{tikzcd} \mathcal{H} \arrow[r, "V"] \arrow[d, "{X, Z}"'] & \mathbb{C}^2 \otimes \mathcal{H}' \arrow[d, "{\sigma_X\otimes\mathrm{Id},\, \sigma_Z\otimes\mathrm{Id}}"] \\ \mathcal{H} \arrow[r, "V"] & \mathbb{C}^2 \otimes \mathcal{H}' \end{tikzcd} \tag{2.3}$$

Note that the lemma no longer says that X is *equal* to $\sigma_X \otimes \mathrm{Id}$ (under the isomorphism π), but only that *it has the same action on the state*, up to the isometry V. In particular, it is now possible for $\mathcal{H}$ to have odd dimension. This is necessary: for example, we can set

$$|\psi\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\;, \quad X = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\;, \quad Z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and still satisfy Definition 2.2. Here, the third dimension has been added to the operators but since none of $|\psi\rangle$, $X|\psi\rangle$ or $Z|\psi\rangle$ has support on it it is "inaccessible" to any experiment that involves only this state and operators. However, it is good to verify that the definition is non-trivial, and in particular requires $\dim(\mathcal{H}) \geq 2$. Indeed, suppose that $X|\psi\rangle$ and $Z|\psi\rangle$ are colinear. Then by (2.2) it follows that $(\sigma_X \otimes \mathrm{Id})V|\psi\rangle$ and $(\sigma_Z \otimes \mathrm{Id})V|\psi\rangle$ are colinear. As we saw in the previous lecture, due to $\{\sigma_X, \sigma_Z\} = 0$ this is impossible.

Proof. The proof is very similar to the proof of Lemma 1.2. Using Jordan's lemma we find a decomposition $\mathcal{H} = \oplus_i \mathcal{S}_i$ such that for each i, $\mathcal{S}_i$ is stable by both X and Z and moreover either $\mathcal{S}_i$ is 1-dimensional or $\mathcal{S}_i$ is 2-dimensional and in a well-chosen basis, $Z = \sigma_Z$ and $X = \begin{pmatrix} c_i & s_i \\ s_i & -c_i \end{pmatrix}$ for some $c_i = \cos 2\theta_i$, $\theta_i \in [0, \pi/2)$. For the one-dimensional blocks the anti-commutator equals (2). For a two-dimensional block we compute $\{X, Z\}^2_{|\mathcal{S}_i} = 4c_i^2\,\mathrm{Id}$. Decompose $|\psi\rangle = \sum_i \alpha_i |\psi_i\rangle$ with $|\psi_i\rangle \in \mathcal{S}_i$. Then we immediately see that if $\mathcal{S}_i$ is a 1-dimensional block, or a 2-dimensional block such that $c_i \neq 0$, then $\alpha_i = 0$. This proves the lemma. □

Note that the proof of the lemma shows something slightly stronger than is captured by the statement of the lemma: informally, that for any of the subspaces $\mathcal{S}_i$ on which $|\psi\rangle$ "has nonzero mass", it must be that $\{X, Z\}_{|\mathcal{S}_i} = 0$, as operators. But we can't conclude anything about blocks where $|\psi\rangle$ "has no mass".

The proof that we gave easily extends to the approximate case.

Exercise 2.1. Say that $(|\psi\rangle, X, Z)$ is an ε-approximate qubit if $\|\{X, Z\}|\psi\rangle\| \leq \varepsilon$. Show that there is an isometry $V : \mathcal{H} \to \mathbb{C}^2 \otimes \mathcal{H}'$ such that for $W \in \{X, Z\}$,

$$\big\|\big(W - V^\dagger(\sigma_W \otimes \mathrm{Id})V\big)|\psi\rangle\big\|^2 \leq O(\varepsilon) .$$

[Hint: Use $2(1 - \sin\theta) \leq \sqrt{4\cos^2\theta}$ *for* $\theta \in [0, \pi)$*.]*

We end the section with a semi-informal definition of "self-testing" that connects the notion of interactive experiment that we discussed earlier with the definition of qubit that we arrived at. For convenience we state the definition for the setting of an experiment that involves a single round of interaction: a question x is selected by the experimentalist, and an answer a is provided by the device. The "observable data" of such an experiment is completely captured in the family of distributions $\{p(\cdot|x)\}_{x\in\mathcal{X}}$ over $\mathcal{A}$, and so the starting point for the definition is that data only.

Definition 2.4. We say that the family of conditional distributions $\{p(\cdot|x)\}_{x\in\mathcal{X}}$ *self-tests a qubit* if for any state $|\psi\rangle \in \mathrm{S}(\mathcal{H})$ and family of POVM $\{P^x_a\}_{a\in\mathcal{A}}$ for $x \in \mathcal{X}$ such that $p(a|x) = \langle\psi|P^x_a|\psi\rangle$ for all a, x there is an isometry $V : \mathcal{H} \to \mathbb{C}^2 \otimes \mathcal{H}'$ and $x_0, z_0 \in \mathcal{X}$ such that the measurements P^{x_0} and P^{x_1} have only two possible outcomes $0, 1$ and moreover

$$V(P^{x_0}_0 - P^{x_0}_1)|\psi\rangle = (\sigma_X \otimes \mathrm{Id})V|\psi\rangle \quad \text{and} \quad V(P^{z_0}_0 - P^{z_0}_1)|\psi\rangle = (\sigma_Z \otimes \mathrm{Id})V|\psi\rangle . \qquad (2.4)$$

As you can see the definition is a little uncomfortable to state; not only does the notation quickly get pretty heavy but one also has to be quite careful to make a meaningful statement for the applications that one has in mind.

The use of the isometry in the definition may come as a surprise, because it allows us to "artificially" extend the space in which the operators live. This is necessary because as discussed below Definition 2.2 in general the dimension of $\mathcal{H}$ may not be even, whereas any space in which we can write something like "$\sigma_X \otimes \mathrm{Id}$" must have even dimension. For the time being you can think of V as an artifact that may create additional dimensions in which $V|\psi\rangle$ has no "mass" at all, but are still needed to give the desired form to the operators. As discussed below Lemma 2.3, even with the isometry the conclusion of the lemma is not trivial since it at least implies that $\dim \mathcal{H} \geq 2$.

Unfortunately, it is not hard to see that the definition is not "achievable" in the sense that without further assumptions, no family of distributions $\{p(\cdot|x)\}_{x\in\mathcal{X}}$ self-tests a qubit in the sense of the definition. This is simply because in general one cannot avoid that, say, $|\psi\rangle = 1 \in \mathbb{C}$ and $P^x_a = p(a|x)$ for all a and x, which is a valid POVM.[4] As such one should only treat this definition as "indicative" and we use it for inspiration only. In the future we will generally establish special-purpose statements that are more precise depending on the situation we're in.

[4] It is also possible to get a trivial realization using projective measurements by taking $|\psi\rangle$ to be sufficiently many EPR pairs, or a more general entangled state, so as to instantiate the randomness required to implement the distribution.

2.4 A first test for a qubit

We proceed to give a first answer to our second question, "is the definition testable?" Our answer today will not be completely satisfactory, but it's a start. Most important is that it will allow us to practice the notions introduced so far and put in place techniques that will be useful later on.

In order to analyze the protocol that we give in Section 2.4.2 we will need some elementary notions about density matrices and entanglement. The reader already familiar with these notions may skip the next section, which contains a very brief introduction; as usual we refer to [NC02] for a much more leisurely, and comprehensive, discussion.

2.4.1 Entanglement and density matrices

A pure state, as we know, is a unit vector in a Hilbert space $\mathcal{H}$. A pure bipartite state is $|\psi\rangle \in \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$, where the "bipartition" of $\mathcal{H} = \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$ is often implicit from context. We use subscripts A or B to denote subsystems, sometimes also called "registers". Any pure bipartite state has a *Schmidt decomposition*

$$|\psi\rangle_{\mathsf{AB}} = \sum_i \sqrt{\lambda_i} |u_i\rangle_\mathsf{A} |v_i\rangle_\mathsf{B} \tag{2.5}$$

where the λ_i are non-negative reals that sum to $\||\psi\rangle\|^2 = 1$ and $\{|u_i\rangle\}$ and $\{|v_i\rangle\}$ are orthonormal bases of $\mathcal{H}_\mathsf{A}$ and $\mathcal{H}_\mathsf{B}$ respectively. Here in the notation we sometimes, but not always, include a subscript A or B (or both) on a "ket" to indicate which subsystem the state lies in. The coefficients λ_i in (2.5) are called *Schmidt coefficients* and are uniquely defined. The $\{|u_i\rangle\}$ and $\{|v_i\rangle\}$ are called Schmidt vectors. The reduced state of $|\psi\rangle_{\mathsf{AB}}$ on $\mathcal{H}_A$ is described by a density matrix $\rho_A = \sum_i \lambda_i |u_i\rangle\langle u_i|$, that one can interpret as a distribution over pure states $|u_i\rangle$. More generally, if ρ_{AB} is a density matrix on $\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$ we use the notation $\rho_\mathsf{A} = \mathrm{Tr}_\mathsf{B}(\rho_{\mathsf{AB}})$ to denote its reduced density on $\mathcal{H}_\mathsf{A}$, and $\rho_\mathsf{B} = \mathrm{Tr}_\mathsf{A}(\rho_{\mathsf{AB}})$ for $\mathcal{H}_\mathsf{B}$. These reduced densities can be computed by extending the definition given for pure states by linearity, or in any other of a number of equivalent ways.

For us, the EPR pair is a specific bipartite state $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. It has the interesting property that for any orthonormal basis $|u_0\rangle, |u_1\rangle$ of $\mathbb{C}^2$,

$$|\phi^+\rangle = \frac{1}{\sqrt{2}}(|u_0\rangle\overline{|u_0\rangle} + |u_1\rangle\overline{|u_1\rangle}) \,. \tag{2.6}$$

This is because more generally for a linear operator A on $\mathbb{C}^2$, $(A \otimes \mathrm{Id})|\phi^+\rangle = (\mathrm{Id} \otimes A^T)|\phi^+\rangle$. In particular, we see that if a measurement of the first qubit is made in the basis $\{|u_0\rangle\langle u_0|, |u_1\rangle\langle u_1|\}$ and the outcome $b \in \{0,1\}$ is obtained then the state of the second qubit reflects this fact, becoming $\overline{|u_b\rangle\langle u_b|}$.

2.4.2 The protocol

Consider the following protocol between a "verifier" V and a "prover" P. Although ultimately our goal is to have protocols that involve a purely classical verifier, in this first protocol V has some quantum capabilities; its goal is to use this to ascertain that P has similar capabilities. In particular for this protocol we assume that V has a quantum communication channel to P.

1. V selects two bits $v, \theta \in \{0,1\}$ uniformly at random. She prepares a single-qubit state $|v^\theta\rangle = H^\theta |v\rangle$, where $H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ is the Hadamard matrix, and sends it to P.

2. V waits for a few seconds.

3. V sends θ to P.

4. P returns a value $v' \in \{0,1\}$.

5. V declares that P has succeeded if and only if $v' = v$.

We claim that any prover that succeeds with probability 1 in this protocol "has a qubit". Before showing this, let's discuss a few points.

- *What do you mean, the prover has a qubit? Of course it has a qubit–the verifier sent it to him!* Aha, but remember the discussion surrounding our definition of a qubit! What the prover gets is the *state* of a qubit. A good model for a "classical" prover would be one that quickly measures ("decoheres") any state it receives in the computational basis, recovering the classical information only. Certainly, such a prover would not count as having a "qubit", because any measurement they are able to make is in the computational basis, and in particular commutes. And indeed, it is easy to verify that such prover only succeeds with probability at most $3/4$ in the test. This is why step 2., the few seconds' pause, is inserted in the protocol. As we will see from the proof, we will be able to show that the prover still "has a qubit" at step 3, when it receives the value θ from v.

- *Didn't we say that the verifier is classical? How come they can prepare qubits?* That's a good point. As our analysis will show it is possible to show that the same protocol remains valid if we remove the assumption that the verifier prepares the claimed state. That is, we can assume that an arbitrary entity prepares an arbitrary $(1+N)$-qubit state and sends one qubit to V and the others to P. In that case the only thing that we need to assume is that the verifier has the ability to measure σ_X and σ_Z. So, using that the verifier has a qubit, they can check that the prover also has a qubit. It's not so trivial!

- *How can you check that the prover succeeds with probability* 1*?* Of course, we can't. Assuming that the prover behaves in an i.i.d. fashion, repeating the protocol $K \sim (1/\varepsilon)\log(1/\delta)$ times and observing K successes would let us conclude, with confidence $1-\delta$, that the prover's "intrinsic" probability of succeeding is at least $1-\varepsilon$.[5]

Lemma 2.5. *Suppose that a prover P succeeds with probability* 1 *in the protocol. Then P has a qubit.*

To connect the statement of the lemma to Definition 2.4 we could also try to say that in this protocol, the family of distributions $\{p(v'|\theta,v) = 1_{v'=v}\}_{v,\theta\in\{0,1\}}$ "self-tests" a qubit. As we had predicted however, the protocol does not neatly fit the definition for multiple reasons: it is not a 1-round protocol, there is quantum communication, and the verifier maintains private information (the value v).

Proof. Before we show anything let's first model precisely what goes on in this protocol; this modeling step is often the most important one in the analysis of a protocol. For the proof it is more convenient (and also more general) to consider the "purified" variant hinted at above. First, note the following equivalent description of the protocol:

1. The verifier prepares a two-qubit EPR pair $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. She keeps the first qubit to herself and and sends the second to the prover.

[5] In other words, we'd show that any prover whose probability of succeeding is $< 1-\varepsilon$ only has a chance at most δ to succeed in K repetitions.

2. The prover applies an arbitrary quantum map to their qubit, yielding the shared state $|\tilde{\phi}\rangle = (\mathrm{Id} \otimes W)|\phi^+\rangle$, where W is the prover's operation. In general, the prover's map can be any isometry[6] as the prover may append ancilla qubits if it so desires.[7]

3. The verifier flips a coin $\theta \in \{0,1\}$ and measures her qubit in the standard basis (i.e. the eigenbasis of σ_Z) if $\theta = 0$ and the Hadamard basis (the eigenbasis of σ_X) if $\theta = 1$, obtaining an outcome $v \in \{0,1\}$.

4.-5. Same.

Using the observation (2.6) (and the discussion that follows it) we see that Step 4 has the effect of projecting the prover's share of the joint state to $W|v^\theta\rangle$, which is effectively the state that it would be in had we proceeded according to the original description of the protocol. So, the two descriptions are equivalent.

This "purified" description of the protocol has one major advantage, which is that it allows us to "delay" Alice's choice of θ and v until step 3; as we will see this is very helpful. Yet the version that we wrote down is still not so easy to analyze, mainly due to the fact that the isometry W may be completely arbitrary. As it turns out it is more convenient to analyze another variant in which the prover is given *more* power, so that showing this variant secure will immediately imply the same for the original one. In the new variant we replace the first two steps by imagining that both verifier and prover are handed out a share of an arbitrary initial state $|\psi\rangle_{\mathsf{AB}} \in \mathbb{C}^2_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$. Here the verifier gets the first subsystem, that is assumed to be of dimension 2, and the prover gets the second subsystem, whose dimension is arbitrary. This is more general because the state of the verifier's qubit is no longer characterized (except for its dimension, that we fix to 2). However, we will show that even in this variant in order to succeed the prover must "have a qubit".

Remark 2.6. It will generally be convenient to assume that any measurement that the prover makes can be modeled by a projective measurement. Abstractly, this can be guaranteed by Naimark's theorem. We will not review the theorem here, but if you are not familiar with it it is a good idea to make sure that you understand its formulation. In particular, any use of Naimark's may require extending the Hilbert space by adding ancilla qubits to $|\psi\rangle$. This operation is an isometry that one should not forget to include in the conclusion one is making—it is another reason for including the isometry V from Definition 2.4.

Continuing our modeling effort, at step 4 of the protocol the prover has in its hands (i) the qubit it received from the verifier, that we model as the second half of some $|\psi\rangle_{\mathsf{AB}} \in \mathbb{C}^2_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$ (where the extension to a larger space $\mathcal{H}_{\mathsf{B}}$ may have occurred as a result of some map that the prover applied during the course of step 2 in the protocol), and (ii) the value $\theta \in \{0,1\}$ it has received at step 3. Given this information, it is expected to return a value $v' \in \{0,1\}$. In full generality we can model this by saying that for each $\theta \in \{0,1\}$ the prover has a measurement $\{P^\theta_0, P^\theta_1\}$ that it performs on its share of $|\psi\rangle_{\mathsf{AB}}$ in order to obtain v'. Using the remark we may further assume that this measurement is projective, and so we can associate a binary observable $P^\theta = P^\theta_0 - P^\theta_1$, for $\theta \in \{0,1\}$, to it. While this requires to enlarge the prover's space to apply Naimark's theorem, since here we already allow the space to be arbitrary there is no loss in generality with assuming at the outset that $\{P^\theta_0, P^\theta_1\}$ is projective.

With all this modeling in place we are ready to write a formal expression for the prover's success prob-

[6] An isometry is a linear length-preserving map into a larger space. Formally, $W : \mathbb{C}^2 \mapsto \mathcal{H}$ such that $W^\dagger W = \mathrm{Id}$. For example, $W|u\rangle = |u\rangle|0\rangle$ is an isometry, which simply appends a qubit in state $|0\rangle$ to its input.

[7] We can use the ancilla to model a classical prover as well; here, W would simply copy the qubit to an environment register that would become inacessible to the prover. This effectively decoheres the qubit that remains in the prover's possession.

ability in the test. By definition it is

$$\begin{aligned}\Pr(v = v') = \frac{1}{2}\big(\langle\psi|(|0\rangle\langle 0| \otimes P_0^0)|\psi\rangle + \langle\psi|(|1\rangle\langle 1| \otimes P_1^0)|\psi\rangle\big) \\ + \frac{1}{2}\big(\langle\psi|(|+\rangle\langle +| \otimes P_0^1)|\psi\rangle + \langle\psi|(|-\rangle\langle -| \otimes P_1^1)|\psi\rangle\big)\ .\end{aligned} \tag{2.7}$$

Here the factors $\frac{1}{2}$ represent the probabilities that the verifier chooses $\theta = 0$ (measurement in the standard basis) and $\theta = 1$ (measurement in the Hadamard basis) respectively, and inside each bracket each of the two terms represents the probability that the prover and verifier obtain the same measurement outcome $v = v' = 0$ for the first term and $v = v' = 1$ for the second. Using the identities

$$|0\rangle\langle 0| = \frac{1}{2}\big(\mathrm{Id} + \sigma_Z\big)\ , \quad |1\rangle\langle 1| = \frac{1}{2}\big(\mathrm{Id} - \sigma_Z\big) \quad \text{and} \quad |+\rangle\langle +| = \frac{1}{2}\big(\mathrm{Id} + \sigma_X\big)\ , \quad |-\rangle\langle -| = \frac{1}{2}\big(\mathrm{Id} - \sigma_X\big)$$

as well as the symmetric ones

$$P_0^0 = \frac{1}{2}\big(\mathrm{Id} + P^0\big)\ , \quad P_1^0 = \frac{1}{2}\big(\mathrm{Id} - P^0\big) \quad \text{and} \quad P_0^1 = \frac{1}{2}\big(\mathrm{Id} + P^1\big)\ , \quad P_1^1 = \frac{1}{2}\big(\mathrm{Id} - P^1\big)$$

together with some simple manipulations we can rewrite the expression (2.7) as

$$\Pr(v = v') = \frac{1}{2} + \frac{1}{4}\big(\langle\psi|\sigma_Z \otimes P^0|\psi\rangle + \langle\psi|\sigma_X \otimes P^1|\psi\rangle\big)\ . \tag{2.8}$$

This equality is the central equality in the proof, so it is worth looking at it closely. The expression quantifies some form of “correlation” between the verifier’s observables, σ_Z and σ_X, and the prover’s, P^0 and P^1. Each of the numbers inside the brackets on the right-hand side is a real number in $[-1, 1]$ that is the expectation value of the product of their outcomes, when interpreted as values in ± 1. For the success probability to equal 1 the outcomes must always match. Note, however, that the verifier is making two *incompatible* measurements on their share of the state. The following claim shows that in this situation the prover’s observables must also be incompatible, i.e. anti-commute.

Claim 2.7. *Let $|\psi\rangle \in \mathbb{C}^2 \otimes \mathcal{H}$ be an arbitrary state and X, Z arbitrary observables on $\mathcal{H}$ such that*

$$\langle\psi|\sigma_X \otimes X|\psi\rangle = \langle\psi|\sigma_Z \otimes Z|\psi\rangle = 1\ . \tag{2.9}$$

Then $(\mathrm{Id} \otimes \{X, Z\})|\psi\rangle = 0$.

Intuitively, if X and Z were *not* incompatible then, since X can be used to predict the outcome of σ_X and Z that of σ_Z, by simultaneously measuring the *compatible* observables X and Z we would be able to simultaneously predict the outcomes of a measurement in the *incompatible* observables σ_X and σ_Z, a contradiction. Let’s see the proof.

Proof. Using that all operators have norm at most 1 and that $\||\psi\rangle\| = 1$ the equality (2.9) implies that

$$\mathrm{Id} \otimes X|\psi\rangle = \sigma_X \otimes \mathrm{Id}\,|\psi\rangle \quad \text{and} \quad \mathrm{Id} \otimes Z|\psi\rangle = \sigma_Z \otimes \mathrm{Id}\,|\psi\rangle\ .$$

Using these identities,

$$\begin{aligned} XZ|\psi\rangle &= \sigma_X\sigma_Z|\psi\rangle \\ &= -\sigma_Z\sigma_X|\psi\rangle \\ &= -ZX|\tilde{\psi}\rangle\ , \end{aligned}$$

as required. □

Exercise 2.2. The proof can be adapted to show a bit more than we extracted from it. By using Lemma 2.3 show that under the same assumptions as in the claim there must exist an isometry $V : \mathcal{H} \to \mathbb{C}^2 \otimes \mathcal{H}'$ on $\mathcal{H}$ under which $(\mathrm{Id}_{\mathbb{C}^2} \otimes V)|\psi\rangle = |\phi^+\rangle \otimes |\psi'\rangle$, where $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle$ is an EPR pair and $|\psi'\rangle$ an arbitrary state on $\mathcal{H}'$. That is, even if we do not assume a priori that the two parties share an EPR pair, they must do so in order to win with probability 1.

Applying Claim 2.7 to $Z = P^0$ and $X = P^1$ and using (2.8) to verify that (2.9) is satisfied we obtain that $(|\psi\rangle, P^0, P^1)$ is a qubit according to Definition 2.2. This concludes the proof. □

2.5 Scaling it up: a test for quantum memory

The main drawback of the test presented in the previous section is that it requires one qubit on the verifier side to test one qubit on the prover side. In general we are interested in tests where the verification effort is much smaller than the effort that is certified of the quantum device, or "prover".

The test we presented has a natural "scaled up" variant, in which the verifier sequentially prepares single-qubit states $|v_i^{\theta_i}\rangle$ for $i = 1, \dots, n$ and sends them to the prover. Once all qubits have been sent, the verifier announces $\theta_1, \dots, \theta_n$ and expects $v'_1, \dots, v'_n \in \{0, 1\}$ such that $v'_i = v_i$ for all i. In this test the verifier can accomplish her share of the work using a single-qubit computer only, since she can prepare and send the n qubits one at a time and use classical memory to store the entire strings $v, \theta \in \{0, 1\}^n$. A similar analysis as the one presented in the previous section would then demonstrate that the prover "has n qubits", where the notion of having n qubits is the state-dependent version of the n-qubit definition we saw in the previous lecture, Definition 1.4. Moreover, an n-qubit variant of Lemma 2.3 also holds, so that we'd effectively have shown that the prover does require a quantum memory of dimension 2^n in order to successfully accomplish its task.

Unfortunately, the method that we introduced so far does not extend well to success probabilities smaller than 1. In general it is unrealistic to make the assumption that the prover succeeds perfectly in the protocol, as this cannot be verified. A more reasonable assumption is that the prover succeeds with probability $1 - \varepsilon$, for some constant $\varepsilon > 0$ that can be made smaller and smaller with higher and higher confidence by repeating the protocol, but can never be driven down to exactly zero. As far as I can tell applying the method from the previous section to this case only yields a good bound on the quantum dimension of the prover when $\varepsilon = O(1/n)$, which then requires $\Omega(n)$ executions of the protocol to certify. (See [CRSV18, Theorem 2.1] for a quantitative statement of the kind that would apply here.)

Instead in this section we propose a different method to analyze the scaled up protocol, that uses information-theoretic technique to quantify the intuition from previous section that the quantumness of the prover arises from its need to make predictions for incompatible observables. This method based on information theory has the advantage that it generally yields much better quantitative results. The main drawback is that it allows us to certify less—here, we will be able to certify the prover's quantum dimension but not the observables that it makes use of.

Before stating the protocol precisely and giving the analysis we introduce a variant of Heisenberg's uncertainty principle "for qubits" that we will make use of in the proof.

2.5.1 Uncertainty relations

Our notion that anti-commuting observables are "incompatible" can be quantified through the uncertainty principle. Here is an elementary formulation that applies to our context, due to Maassen and Uffink. To state it we recall the definition of the Shannon entropy,

$$H(\{p_i\}) \;=\; -\sum_i p_i \log p_i \;,$$

for any distribution $\{p_i\}$. Note that here we use a variant using base 2 logarithms, which is the standard used for the extension to density matrices, that we give a little later.

Theorem 2.8. *Let R and S be observables on $\mathcal{H}$. Let $c = \max|\langle\psi|\phi\rangle|^2$ where $|\psi\rangle$ (resp. $|\phi\rangle$) ranges over all eigenvectors of R (resp. S). Let $|\psi\rangle$ be an arbitrary state on $\mathcal{H}$ and let R and S be random variables distributed as the outcome of a measurement of R and S on $|\psi\rangle$, respectively. Then*

$$H(R) + H(S) \;\geq\; \log_2\frac{1}{c}\;. \tag{2.10}$$

In the case when R and S are binary observables then c is precisely the squared cosine of the smallest principal angle between an eigenspace of R and an eigenspace of S. If R and S have an eigenvector in common then $c = 1$ and the right-hand side in (2.10) vanishes, as one would expect since taking $|\psi\rangle$ to be that eigenvector yields zero entropy on the left-hand side. If R and S anti-commute, all the principal angles are $\pi/4$ and so $c = \frac{1}{2}$. In this case, the uncertainty principle states that among the two binary variables R and S there is at least one bit of entropy. This is a quantitative version of an observation that we made in the first lecture, which was that any state that is determined for one observable must be “fully random” with respect to the other: in that case we get $H(R) = 0$ and $H(S) = 1$ (or vice-versa). Theorem 2.8 shows that there is always a quantitative trade-off between these two extremes.

For our purposes we need an extension of this relation to the case of quantum memory. To motivate it, interpret Theorem 2.8 as a statement about the difficulty of a *prediction task*:

1. The “adversary” prepares an arbitrary pure state $|\psi\rangle$ and sends it to the “challenger”.

2. The challenger selects a uniformly random $\theta \in \{0,1\}$ and measures $|\psi\rangle$ using the observable R (case $\theta = 0$) or S (case $\theta = 1$), obtaining an outcome r or s respectively. It sends θ to the prover.

3. The prover returns a guess r' or s', depending on θ.

4. The adversary succeeds if its guess is correct.

Intuitively Theorem 2.8 states that there is no way for the adversary to succeed in this game, because however it prepares $|\psi\rangle$ at least one of r or s will have randomness in it, making it impossible to predict without additional information. (There is a precise quantitative connection between optimal success probability in this game and the left-hand side in (2.10), but making this precise would take us a little too far in the discussion of entropies.)

Now suppose for simplicity that $R = \sigma_X$ and $S = \sigma_Z$. Then the game studied in the previous section suggests that there is a way for the adversary to succeed perfectly in this game, if they are allowed to have quantum memory: in this case, they would prepare an EPR pair $|\phi^+\rangle_{\mathsf{AB}}$ and give only the first qubit to the challenger. To match the challenger’s outcome, they would simply make the same measurement on the qubit that they had kept to themselves. In this case their prediction will be correct with probability 1!

This example suggests that Theorem 2.8 should be modified to account for this more complex scenario. Berta et al., based on work by Christandl, Winter, and others, introduced the following refinement. To state the refinement we recall the definition of the conditional von Neumann entropy: for a bipartite state ρ_{AB},

$$H(A|B)_\rho = H(\rho_{\mathsf{AB}}) - H(\rho_{\mathsf{B}})\;, \tag{2.11}$$

where ρ_{B} is the reduced density matrix of ρ_{AB} on $\mathcal{H}_B$ and for a density matrix σ, $H(\sigma)$ is its von Neumann entropy

$$H(\sigma) \;=\; -\mathrm{Tr}(\sigma \ln \sigma) \;=\; -\sum_i \lambda_i \log \lambda_i \;,$$

where the λ_i are the nonzero eigenvalues of σ.

Note that this quantity can be negative! However, it can never be more negative than the "quantum dimension" of B. To make this notion precise we introduce the notion of a "classical-quantum", or CQ state. A CQ state is simply a bipartite state such that the first system is classical. More precisely, a CQ state $\rho_{\mathsf{AB}} \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ has a decomposition

$$\rho_{\mathsf{AB}} \;=\; \sum_x |x\rangle\langle x|_{\mathsf{A}} \otimes (\rho_x)_{\mathsf{B}} \;,$$

where here x ranges over the standard basis of $\mathcal{H}_{\mathsf{A}}$. The first system is "classical" in the sense that a measurement of it in the standard basis does not change ρ. Now, for an arbitrary state $\rho \in \mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$ suppose that we can decompose the B part into a "classical" part C and a "quantum" part Q:

$$\rho_{\mathsf{B}} = \sum_c p_c \, |c\rangle\langle c| \otimes \rho_c \in \mathcal{H}_{\mathsf{C}} \otimes \mathcal{H}_Q \;,$$

where $\{p_c\}$ is an arbitrary distribution. Then $\rho_{AB} = \sum_c p_c |c\rangle\langle c| \otimes \rho'_c$, with $\rho'_c \in \mathrm{D}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{Q}})$ such that $\mathrm{Tr}_{\mathsf{A}}(\rho'_c) = \rho_c$. Then,

$$\begin{aligned}
H(A|B)_\rho &= H(\rho_{\mathsf{AB}}) - H(\rho_{\mathsf{B}}) \\
&= \Big(H(\{p_c\}) + \sum_c p_c H(\rho'_c)\Big) - \Big(H(\{p_c\} + \sum_c p_c H(\rho_c)\Big) \\
&\geq \min_c \big(H(\rho'_c) - H(\rho_c)\big) \\
&= \min_c H(A|Q)_{\rho'_c} \\
&\geq -\log \dim \mathcal{H}_{\mathsf{Q}} \;. \qquad (2.12)
\end{aligned}$$

Here, the first line is by definition, the second is the chain rule, the third and fourth are clear, and the last is again by definition.

Theorem 2.9. *Let R and S be observables on $\mathcal{H}_{\mathsf{A}}$ and $c = \max |\langle\psi|\phi\rangle|^2$ where $|\psi\rangle$ (resp. $|\phi\rangle$) ranges over all eigenvectors of R (resp. S). Let ρ_{AB} be an arbitrary density matrix on $\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$. Let $R(\rho), S(\rho) \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ denote the post-measurement states after a measurement of A using the observables R and S respectively.*[8] *Then*

$$H(A|B)_{R(\rho)} + H(A|B)_{S(\rho)} \;\geq\; \log_2 \frac{1}{c} + H(A|B) \;. \qquad (2.13)$$

If B is empty and $|\psi\rangle_{\mathsf{A}}$ is a pure state then $H(A|B) = 0$ and we recover the previous relation. If B is empty and ρ_{A} is a mixed state than $H(A|B) > 0$ and we obtain a strengthening of Theorem 2.8. If $\rho_{AB} = |\phi^+\rangle\langle\phi^+|_{\mathsf{AB}}$ is an EPR pair then we can compute $H(A|B)_{\phi^+} = -1$, so the inequality does not imply any non-trivial bound on the left-hand side, as we expect it to based on the discussion above.

[8] Equivalently, this is like decohering A in the eigenbasis of R or S respectively.

2.5.2 A test for large quantum memory

The information-theoretic tools introduced in the previous section allow us to introduce a neat "scaling up" of the single-qubit test from Section 2.4. Towards this we consider the following variant of the protocol from Section 2.4.2, which was introduced in [CR20]. Important changes are highlighted in blue.

1. V selects a bit $\theta \in \{0,1\}$ and a string $v \in \{0,1\}^n$ uniformly at random. For $i = 1,\ldots,n$ she successively prepares the single-qubit states $|v_i^\theta\rangle = H^\theta|v_i\rangle$ and sends them to P.

2. V waits for a few seconds.

3. V sends θ to P.

4. P returns a string $v' \in \{0,1\}^n$.

5. V declares that P has succeeded if and only if $v' = v$.

Note that compared to the naïve repetition of the single-qubit protocol, here we elected to use the same basis for all qubits. This brings a very minor saving in communication, and in parameters, that comes for free from the tools that we use to analyze the protocol.

Lemma 2.10. *Suppose that P succeeds with probability 1 in this protocol. Then P has quantum memory of dimension 2^n.*

Proof. As in the proof of Lemma 2.5 it is convenient to consider a purified version of the protocol in which the verifier prepares n EPR pairs and measures her n halves in the basis θ at step 3. Moreover, we can give more power to the adversary by considering that the joint state between V and P at the end of step 2 is an arbitrary $\rho \in (\mathbb{C}^2)^{\otimes n} \otimes \mathcal{H}$ where the n copies of $\mathbb{C}^2$ correspond to the verifier's qubits, and the remaining space $\mathcal{H}$ is in the hands of the prover. Let R (resp. S) be the observable associated with a measurement of the verifier's n qubits in the computational (resp. Hadamard) basis.

First we show that the fact that the prover succeeds with probability 1 in the protocol implies that necessarily the left-hand side in (2.13) equals 0. This is due to the *data processing inequality*, which states that the conditional entropy can only *increase* as a result of any quantum information processing performed on the system that is being conditioned on. (This is intuitive: in an information-theoretic sense post-processing can only increase uncertainty, by discarding information, and not reduce it.) Starting from e.g.

$$\rho' := R(\rho) = \sum_{v\in\{0,1\}^n} \Pr(v)\,|v\rangle\langle v| \otimes \rho'_v\,,$$

where v denotes the verifier's outcome, $\Pr(v)$ denotes the probability that the verifier obtains the outcome v (conditioned on having chosen to perform a measurement in the standard basis) and $\rho'_v \in \mathrm{D}(\mathcal{H})$ the prover's system conditioned on the verifier having obtained v, the prover's measurement that returns v' yields

$$\rho'' := \sum_{v,v'} \Pr(v)\ \Pr(v'|v)|v\rangle\langle v| \otimes |v'\rangle\langle v'| = \sum_{v} \Pr(v)\,1|v\rangle\langle v| \otimes |v\rangle\langle v|\,,$$

since $v = v'$ with probability 1. On the state ρ'', $H(A|B)_{\rho''} = 0$ since both systems are classical and perfectly correlated. Since ρ'' is obtained by a measurement on register B in ρ', it follows that $H(A|B)_{\rho'} \leq 0$.[9] Of course a similar argument applies to S, showing that the left-hand side in (2.13) is non-positive (and in fact, equal to 0).

[9] This inequality suffices for our purposes, but using that A is classical it is possible to conclude that $H(A|B)_{\rho'} = 0$.

Using that the maximum overlap c between an eigenvector of R and an eigenvector of S is $c = 1/2^n$ and applying Theorem 2.9 we deduce that necessarily

$$H(A|B)_\rho \leq -\log_2 \frac{1}{c} = -n$$

We conclude by (2.12): the dimension of Bob's quantum memory must be at least 2^n. □

An advantage of using information theory is that it is generally a very robust technique, i.e. with all the machinery that we put in place it is not hard to extend the proof of Lemma 2.10 to cover the case where the prover's success probability is not necessarily 1 and may even be quite small. It is also possible to analyze a "fault-tolerant" variant of the protocol in which the verifier accepts the outcome v' reported by P as long as it matches v in a fraction at least $(1-\alpha)$ of positions, where $\alpha \in (0, 1/2]$ is an arbitrary constant. The following is shown in [CR20].

Theorem 2.11 (Theorem 2.1 in [CR20])**.** *If P succeeds with probability p in the protocol, then its Hilbert space must have dimension d such that*

$$\log_2(d) \geq \big((1-H(\alpha))2p-1\big)n - 2H(p)\,,$$

where H is the binary entropy function.

This result is quite strong, as even for small constant values of α and values of p that are sufficiently close to 1 we get a bound on the number of qubits of memory that Bob needs to keep that scales linearly with n. Aside from the "dimension test" presented here these kinds of bounds have found numerous applications in cryptography such as to proving security in the bounded storage model, where it is assumed that the adversary has a limited amount of storage available (for the reader familiar with cryptography but who hasn't seen this before, we probably already said enough to start suggesting a protocol for some variant of oblivious transfer...).

While techniques based on information theory have quantitative advantages, they will generally not suffice for our purposes. In particular, note the difference between Lemma 2.5 and Lemma 2.10: while the latter guarantees "one qubit" the former certifies "dimension 2^n": the former is quantitatively stronger, but qualitatively weaker as it does not give us access to information about the prover's observables. From now on we will mostly abandon the use of information theory, as it is too coarse grained for our purposes.

Lecture 3

Testing a qubit under spatial assumptions

(**Comment:** To insert a comment, use the macro “\com{·}”)

In this lecture we introduce a new assumption in addition to our overarching assumption that all parties in a protocol can be modeled using quantum mechanics; as argued in the previous lecture additional assumptions are necessary to develop a test for a qubit with classical verifier. In physical terms our new assumption consists in requiring that the device that is being tested is made of two parts that are “physically isolated,” in the sense that no communication can take place between the two parts for the duration of the protocol. For the case where the protocol consists of a single round of interaction (a question from the verifier and an answer from the prover) one can imagine enforcing this assumption by e.g. placing the provers and verifier on a line $P_1 - V - P_2$ and ensuring that the round-trip interaction between the verifier and either prover takes place sufficiently fast that the verifier is confident, based on relativistic considerations (information does not travel faster than light), that no information can be exchanged between the provers between the times when they receive their question and have to send their answer. Mathematically, the assumption is reflected by modeling the device’s Hilbert space $\mathcal{H}$ as $\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$ and writing that each sub-device’s observables act on its Hilbert space only. Following tradition we will use the symbols A and B to denote “registers” (a word loosely used to refer to the physical substrate modeled by the mathematical Hilbert space) associated with each device and, oftentimes personify the devices as “provers” or “players” with the lovely names of “Alice’ and “Bob” respectively.

As we will see in this lecture as well as in the last third of the course this assumption of “localization” allows the verifier to gain much leverage over the device. Some intuition for this may be gained from thinking about a situation where a detective (the verifier) interrogates two suspects (the provers). Clearly the detective has more leverage over the suspects if she interrogates them in isolation and cross-examines their answers. Be warned however that this intuition only goes so far, because it only explains why interactive proofs with two provers may be more powerful than single-prover interactive proofs; it does not give insight into why specifically quantum aspects of the provers may manifest themselves in this framework. The fact that quantum mechanics allows a broader set of behavior for the provers than classical mechanics does is evidenced in the EPR paradox [EPR35], whose authors puzzle over the “non-local” nature of quantum mechanics. A precise framework for describing this non-locality was set in place by Bell [Bel64] who identified simple “inequalities” that separate classical from quantum behavior in bipartite scenario. Here we take the modern tack on Bell’s inequalities and introduce them directly through the framework of nonlocal games.

3.1 Nonlocal games

A non-local game is a cooperative game of imperfect information between a referee and two players. The referee is a trusted party that executes the game by sending a question to each player, collecting answers from them, and deciding whether the players' answers satisfy a winning criterion. The rules of the game (the distribution on questions used by the referee, the possible answers, the winning criterion) are public and known to the players, who cooperate in order to maximize their chances of winning. The only source of uncertainty is that each player is only revealed their question, but not the other's; this point is what makes the difference between a single-player and a multi-player game.

Remark 3.1. To make the connection with interactive proof systems of the kind that we described in the previous lecture, somewhat informally a *multi-prover interactive proof system* for a language L is specified by a collection of non-local games $\{G_x\}_{x\in\{0,1\}^*}$, one for each possibe input x. These games should have the property that if $x \in L$ then there is a strategy for the players that succeeds with high probability (this is the completeness property) and if $x \notin L$ then no strategy will make them win with high probability (the soundness property).[1] In the notes we freely interchange between the terminology of non-local games, referees and players and that of interactive proofs, verifier and prover depending on context.

One may rightfully wonder what is the benefit of associating games, or interactive proof systems, to computational problems. One element that we can point out is that a game itself is a computational problem—is the maximum winning probability high (larger than some c) or low (smaller than some s)? By providing a different, "dynamic" perspective on e.g. a 3SAT formula the framework of games has historically been instrumental in proving results in hardness of approximation for constraint satisfaction problems. In a completely different direction, they are a natural setting for cryptography where they were introduced in the context of zero-knowledge proofs.

For the sake of concreteness let us see an example. Consider the language L that is the collection of all strings x such that x represents a satisfiable 3SAT formula φ.[2] For example, φ could be "$y_1 \vee y_2 \vee \overline{y_3}$ AND $\overline{y_2} \vee y_3 \vee y_4$," which is obviously satisfiable. Since it is in general believed to be hard to determine satisfiability of such a formula, let's make the provers work and design an interactive proof systems for the hypothesis "φ is satisfiable".[3]

Here is a first candidate, which involves a single prover:

1. The verifier sends φ to the prover.[4]

2. The prover returns a $\{0,1\}$-valued assignment $(y_1, y_2, \ldots)$ to all variables in φ.

3. The verifier accepts if and only if the assignment satisfies φ.

This proof system has completeness 1 (if there is a solution, a prover that sends it will be accepted with probability 1) and soundness 0 (if there is no solution, no prover has any chance of being accepted). Unfortunately, while the verifier is more efficient than solving the formula herself (by e.g. trying out all possible

[1] For a formal connection between interactive proofs and games one would also have to insist that the games be "uniformly generated" from x, and that the verifier in each game is described by a circuit of size $\mathrm{poly}(|x|)$.

[2] In the following we use the notation φ and x interchangeably: we think of x as a string of bits and φ as a formula, but we assume fixed an efficiently computable bijection between the two.

[3] We emphasize that the goal of the proof system is not to find a more efficient method to solve the formula itself, as someone —the prover— still has to do the work. The goal rather is to provide a different framework in which to think about the complexity of the computational problem "decide if φ is satisfiable".

[4] In the theory of interactive proof systems it is always assumed that the prover has access to the instance that is being decided, so this step is not necessary.

solutions) she still has to read a lot of information in order to make her decision.[5] In keeping with our goal of making verifiers more efficient, let's see a more succinct proof system with two provers. Let φ consist of equations $E_1, \ldots, E_m$ such that E_j has the form $y_{j_1}^{c_{j_1}} \vee y_{j_2}^{c_{j_2}} \vee y_{j_3}^{c_{j_3}}$ with $c_j \in \{0,1\}$ and for a variable y, $y^0 = y$ and $y^1 = (1-y)$.

1. The verifier selects $j \in \{1, \ldots, m\}$ uniformly at random and $k \in \{1,2,3\}$ uniformly at random. She sends j to the first prover and j_k to the second, where j_k is the index of the k-th variable on which clause E_j acts in some canonical ordering. (Importantly, this ordering hides k, i.e. the prover only knows that its variable appears in *some* clause, but not which clause or which position the variable appears in it.)

2. The first prover returns a triple $(a_1, a_2, a_3) \in \{\pm 1\}$. The second prover returns a value $b \in \{\pm 1\}$.

3. The verifier accepts if and only if *(consistency check:)* $a_k = b$ and *(equation check:)* (a_1, a_2, a_3) satisfy clause E_j.

We make the following claim regarding completeness and soundness of this proof system:

Claim 3.2. *The two-prover proof system described above has completeness* 1 *and soundness at most* $1 - \frac{1}{3m}$, *where* m *is the number of clauses in the input formula.*

Note that while our proof system brought us gains in terms of communication, the soundness has degraded quite substantially, from 0 to $1 - \frac{1}{3m}$. It is possible to obtain improved variants of this proof system that have roughly similar communication complexity but much better soundness, say $\frac{1}{100}$ or even less. However, this requires much more work and is essentially the content of the *PCP theorem*, to which we will return in the last part of the course.

Proof. The completeness is easy to verify. For soundness, consider an arbitrary strategy for the two provers that succeeds with some probability p. In order to analyze this strategy we first need to accomplish the usual modeling step: how do we represent a two-prover strategy? The most "naïve" way to do so is to use a representation of each prover as a function from questions to answers and declare that the provers' joint strategy is the combination (direct product) of these functions: the first prover, Alice, employs a function $f_A : \{1, \ldots, m\} \to \{0,1\}^3$ and the second prover, Bob, a function $f_B : \{1, \ldots, n\} \to \{0,1\}$; their joint strategy is simply the function $f = (f_A, f_B)$ that goes from pairs of questions (x, y) in the protocol to pairs of answers (a, b). If one gives a little more thought to the question then it is not at all obvious that this is the right answer. Nevertheless, let's postpone any further thinking for now and finish the proof of the claim using this model for the provers.

Fix a strategy (f_A, f_B) of this form for the provers. We distinguish two cases. Either the strategies "match", meaning that for any clause E_j it holds that

$$f_A(j) = (f_B(j_1), f_B(j_2), f_B(j_3)) , \tag{3.1}$$

where y_{j_1}, y_{j_2}, y_{j_3} are the three variables involved in E_j. In this case we interpret the list of values $f_B(1), \ldots, f_B(n)$ as an assignment to the n variables of φ. Since by assumption φ is not satisfiable there must exist a j such that $(f_B(j_1), f_B(j_2), f_B(j_3))$ do not satisfy clause E_j. By (3.1), $f_A(j)$ does not satisfy E_j

[5] It is possible to argue that for a proof system of this form it is necessary for the prover to send a total number of bits that scales linearly with the length of an NP (i.e. non-interactive) proof for the same statement, see e.g. [GVW01].

either. Hence whenever the verifier sends a question of the form (j,k) for $k \in \{j_1, j_2, j_3\}$ the provers fail in the equation check.

In the second case, the strategies do not match, i.e. there is a pair $(j,k) \in \{1,\ldots,m\} \times \{1,2,3\}$ such that the k-th entry of $(f_A(j))$ does not match $f_B(j_k)$. In this case the provers fail in the consistency check when the question (j,k) is sent.

In all cases there is at least one question on which the provers must fail one of the verifier's checks. Since there are $3m$ possible questions in total and the verifier's distribution on them is uniform this completes the proof of the claim. □

3.2 Non-local strategies

In the proof of Claim 3.2 we were faced with the problem of modeling precisely how the assumption that the provers do not communicate affects the class of strategies that they may employ. While we dodged the question there, let's turn to it more seriously now. First of all, note that the object we are trying to represent is a family of bipartite conditional probability distributions $\{p(\cdot,\cdot|x,y)\}_{x,y \in \mathcal{X}\times\mathcal{Y}}$ over $\mathcal{A} \times \mathcal{B}$, where $\mathcal{X},\mathcal{Y}$ and $\mathcal{A},\mathcal{B}$ are finite sets of questions and answers respectively associated with each player. The question then is, *what families of bipartite conditional distributions can be generated by non-communicating provers?* (equivalently, players, devices, etc.)

3.2.1 Classical and non-signaling correlations

Let's examine two extremes. The first extreme is to require that each prover performs an entirely local computation. In this case the first prover's answer a_1 to their question x_1 is determined by a function $f_1 : \mathcal{X} \to \mathcal{A}$, and similarly for the second prover. This is the answer that we adopted in the proof of Claim 3.2. More generally, being familiar with randomized computation we could allow each prover to make use of a randomized computation, in which case their respective input-output behavior can be modeled by a family of conditional distributions $\{p_{\mathsf{A}}(\cdot|x)\}_{x\in\mathcal{X}}$ on $\mathcal{A}$, and similarly for the second prover. The joint distributions of answers that they provide to the verifier would then be required to factorize as

$$\forall (x,y) \in \mathcal{X}\times\mathcal{Y}\,,\quad \forall (a,b) \in \mathcal{A}\times\mathcal{B}\,,\qquad p(a,b|x,y) \;=\; p_{\mathsf{A}}(a|x)p_{\mathsf{B}}(b|y)\;. \tag{3.2}$$

Since we allowed randomness it may also be natural to allow the randomness to be shared, i.e. allow the more general class of distributions that can be represented as

$$\forall (x,y) \in \mathcal{X}\times\mathcal{Y}\,,\quad \forall (a,b) \in \mathcal{A}\times\mathcal{B}\,,\qquad p(a,b|x,y) \;=\; \int_\lambda p_{\mathsf{A}}(a|x,\lambda)p_{\mathsf{B}}(b|y,\lambda)d\lambda\;, \tag{3.3}$$

where λ ranges over any measurable set and for each λ, $\{p_{\mathsf{A}}(\cdot|x,\lambda)\}_{x\in\mathcal{X}}$ is a family of conditional distributions on $\mathcal{A}$, and similarly for the other prover. It is not hard to see that the proof of Claim 3.2 generalizes to this case: briefly, this is because for a strategy of the form (3.3) to succeed with probability p in the protocol it is necessary that the product strategy obtained by fixing λ succeeds with probability p for at least some choice of λ.

The second extreme is to allow the most general family of bipartite conditional distributions that does not "imply communication". A natural formalization of the latter requirement, usually referred to as the "non-signaling assumption" on p, is that for every a,x and y,y',

$$\sum_b p(a,b|x,y) \;=\; \sum_b p(a,b|x,y')\;. \tag{3.4}$$

In words, the answer a given by the first prover should have a marginal distribution that is independent of the question y given to the second prover. Of course, a symmetric condition should hold with the provers' roles exchanged.

At first it may seem that these two extreme classes "ought to" be the same. Are there distributions that satisfy (3.4) but are not of the form (3.3)? The answer is yes. Here is a simple example: let $\mathcal{X} = \mathcal{Y} = \mathcal{A} = \mathcal{B} = \{0,1\}$. For $(x,y) \neq (1,1)$ let $p(\cdot,\cdot|x,y)$ be uniform over $\{(0,0),(1,1)\}$. For $(x,y) = (1,1)$ let $p(\cdot,\cdot|x,y)$ be uniform over $\{(0,1),(1,0)\}$. It is easy to see that this distribution cannot be expressed in factorized form, or even as a convex combination of factorized forms as in (3.3). (Showing this is a good exercise which we leave to the reader.[6]) However, the distribution clearly satisfies (3.4) since all marginals are uniform. We will see another example in Section 3.3.1.

Having observed that there are at least two possible models for the "non-communicating provers," which one is it most appropriate? Conventionally we call the first model "classical" because it can be realized physically using local computation only, together with possibly a source of shared randomness. The second model is called "non-signaling" and is considered non-physical even though it does not strictly violate the no-communication assumption, because we do not have a credible physical theory in which arbitrary distributions in that model can be generated at locations that are space-time isolated (in other words, there is no physical theory that allows us to describe an experiment which would be able to generate any kind of correlation that is in principle allowed by special relativity; there are other constraints that relativity itself does not provide a means to model). Interestingly, the kind of correlations that can be generated by *quantum* provers lies strictly in-between the two extremes. Let's explore those correlations next.

3.2.2 Quantum (tensor product) correlations

The most natural way to measure spatial isolation in non-relativistic quantum mechanics is to associate a distinct Hilbert space with each device (or prover), $\mathcal{H}_\mathsf{A}$ for Alice and $\mathcal{H}_\mathsf{B}$ for Bob, such that the joint Hilbert space is $\mathcal{H} = \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$. Upon receiving a question x the first prover performs a POVM $\{A_a^x\}_{a\in\mathcal{A}}$ on $\mathcal{H}_\mathsf{A}$ to obtain an outcome a that it sends back to the verifier; similarly, the second prover performs a POVM $\{B_b^y\}_{b\in\mathcal{B}}$ to obtain its answer b. The class of correlations that can be generated in this model is all families of bipartite conditional distributions that take the form

$$p(a,b|x,y) \;=\; \langle\psi|A_a^x \otimes B_b^y|\psi\rangle\;, \tag{3.5}$$

where A_a^x and B_b^y are as above and $|\psi\rangle \in \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$ is an arbitrary state. We will succinctly write $(|\psi\rangle, A, B)$ to represent a triple of a bipartite state and families of POVM measurements on each subsystem as above, and refer to such a triple as a *strategy* for a given two-player game.

Definition 3.3. Given a two-player one-round[7] game G with question sets $\mathcal{X}$ and $\mathcal{Y}$ and answer sets $\mathcal{A}$ and $\mathcal{B}$, a *strategy* for G is a triple $(|\psi\rangle, A, B)$ where $|\psi\rangle \in \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$ is a quantum state on the tensor product of finite-dimensional Hilbert spaces $\mathcal{H}_\mathsf{A}$ and $\mathcal{H}_\mathsf{B}$ and $A = \{A^x\}$ and $B = \{B^y\}$ are collections of POVM $\{A_a^x\}$ and $\{B_b^y\}$ on $\mathcal{H}_\mathsf{A}$ and $\mathcal{H}_\mathsf{B}$ respectively, for every $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

As a sanity check we can verify that (3.5) allows us to recover (3.3). To see this, set $\mathcal{H}_\mathsf{A}$ and $\mathcal{H}_\mathsf{B}$ to be separable Hilbert spaces with a basis indexed by all possible values of λ,[8] $|\psi\rangle = \sum_{\lambda\in\Omega}\sqrt{p_\lambda}|\lambda\rangle_\mathsf{A}|\lambda\rangle_\mathsf{B}$,

[6] Hint: "fix the randomness" and consider the values that each prover returns to each of their two possible questions. Show that these four values together cannot all lie exclusively among the allowed pairs for all four possible pairs of questions.

[7] A non-local game, just as an interactive proof system, can in principle involve multiple rounds of interaction. Here we always always restrict ourselves to the single-round case, that is simpler to analyze and captures everything that we need.

[8] Assume for simplicity that the probability space is countable.

$A^x_a = \sum_\lambda p_{\mathsf{A}}(a|x,\lambda)|\lambda\rangle\langle\lambda|_{\mathsf{A}}$, and similarly for B^y_b. One immediately verifies that these define valid POVM and that they lead to the desired correlation (3.3). Moreover, the POVM condition $\sum_b B^y_b = \mathrm{Id}$ for all y implies that the correlations (3.5) always satisfy the no-signaling condition (3.4). However, the model is strictly more general than the classical model (3.3), as we show next by identifying a non-local game in which the use of quantum correlations leads to a strictly higher winning probability than classical correlations could. (It is also possible to show that it is not as general as the non-signalling model, as the example of a non-signaling correlation given above cannot be realized in quantum mechanics.)

3.3 Binary Linear System Games

For this section we borrow some material from the lecture notes by Richard Cleve available at this url.

Binary linear system games, or BLS games for short, are a class of two-player one-round games introduced in [CM14] and inspired by Mermin's proofs of Bell's theorem [Mer90, Mer93]. These games capture the flavor of the "clause-vs-variable" game considered in the previous section, except that the underlying formula involves parity constraints of the form $y_{j_1} \oplus \cdots \oplus y_{j_\ell} = c_j$ as opposed to the disjunctions we had for the case of 3SAT.

Definition 3.4. A BLS game is specified by integers $m, n \geq 1$, a matrix $E \in \{0,1\}^{m\times n}$ and a vector $c \in \{\pm 1\}^m$. (This information is available to both the referee and the players in the game.) The game proceeds as follows:

1. The referee samples $j \in \{1, \dots, m\}$ uniformly at random and sends j to the first player. Let ℓ be the number of nonzero entries in the j-th row of E. The referee samples $k \in \{1, \dots, \ell\}$ uniformly at random and sends the index of the k-th nonzero entry of the j-th row of E to the second player.
2. The referee expects answers $(a_1, \dots, a_\ell) \in \{\pm 1\}^\ell$ from the first player and $b \in \{\pm 1\}$ from the second.[9]
3. The referee declares that the players win if and only if both the following conditions hold: *(consistency check:)* $a_k = b$ and *(equation check:)* $\prod_i a_i = c_j$.

The class of BLS games has many interesting properties. In particular, there is a direct correspondence between the existence of perfect strategies in different models and certain kinds of 'solutions' to the system of equations implied by E and c. (Precisely, for $j \in \{1, \dots, m\}$ the j-th row of E and c can be interpreted as a constraint $y_1^{E_{j,1}} \cdots y_n^{E_{j,n}} = c_j$ on n variables $y_1, \dots, y_n \in \{\pm 1\}$.) For the case of classical strategies, following the proof of Claim 3.2 we easily see that the game has a perfect strategy if and only if the system of equations has a solution over $\{\pm 1\}$, which in this case can be determined by Gaussian elimination. For quantum strategies in the model introduced above (i.e. the "tensor product model" (3.5)) there is a correspondence between perfect strategies and "operator solutions" to the system of equations. This correspondence will allow us to make use of a specific BLS game called the "Magic Square game" in order to develop our first test of a qubit that can be executed by an entirely classical verifier. We introduce the Magic Square game in the next section.

Remark 3.5. The correspondence between strategies and (operator) solutions goes further than the classical and tensor product models. In particular one can say interesting things about quantum strategies in an extended model called the "commuting-operator model", but we don't discuss this here. See for example [CLS17] and follow-up works.

[9] For later convenience we adopt a multiplicative $\{\pm 1\}$ convention for the variables, instead of the more usual $\{0,1\}$ convention.

3.3.1 An example: the Magic Square game

The Magic Square game is the following BLS game with 6 constraints on 9 variables. The constraints are best visualized by picturing the variables arranged in the entries of a 3 × 3 square, as follows:

$$\begin{array}{ccccc} y_1 & y_2 & y_3 & & +1 \\ y_4 & y_5 & y_6 & & +1 \\ y_7 & y_8 & y_9 & & +1 \\ & & & & \\ +1 & +1 & -1 & & \end{array}$$

As indicated on the picture the 6 constraints are that the product of all variables in any given row should equal $+1$ and that the product of all variables in any column should equal $+1$ *except* for the last column, where it should equal -1.

This system of equations does not have a solution (make sure you can show this!), and so the associated BLS game, as described in Definition 3.4, does not have a perfect classical strategy: it is not hard to see that the maximum success probability that classical players can achieve is $\frac{17}{18}$, matching the bound of Claim 3.2.

A remarkable fact is that there is a perfect quantum strategy for this game ("perfect" means that the strategy succeeds with probability 1 in the game). This is remarkable because, as we just saw, the underlying system of equations *does not have a solution*! Yet quantum players are able to *always* give answers that are accepted by the referee. For this to be possible these answers necessarily have to be generated "on the fly", freshly every time a question is asked: if this were not the case then the same proof as that of Claim 3.2 would apply. Quantum provers are able to win with certainty, yet there is no way to extract a satisfying assignment from them. What feature of the system of equations makes this possible? Can quantum provers win *any* BLS game with probability 1, irrespective of any truth value of the underlying system of equations?

To gain insight into this question let us describe an explicit quantum strategy for the players that succeeds with probability 1. The key observation is that even though as we saw the system of equations associated with the magic square does not have a solution with values in $\{\pm 1\}$, it has an *operator solution*

$$\begin{array}{ccc} I\otimes\sigma_Z & \sigma_Z\otimes I & \sigma_Z\otimes\sigma_Z \\ \sigma_X\otimes I & I\otimes\sigma_X & \sigma_X\otimes\sigma_X \\ \sigma_X\otimes\sigma_Z & \sigma_Z\otimes\sigma_X & \sigma_Y\otimes\sigma_Y \end{array} \tag{3.6}$$

where $\sigma_Y = i\sigma_X\sigma_Z$. Observe that in each row or column the three observables always commute; moreover, the product of the three observables in each row or column is always $+I$ except for the last column, where it is $-I$. This is what we mean by "operator solution".

Definition 3.6. An operator solution to a BLS (E,c) is a collection of binary observables $Y_1,\ldots,Y_n$ on the same Hilbert space $\mathcal{H}$ such that for each equation (specified by a row of E) $y_{j_1}\cdots y_{j_\ell} = c_j$ the observables $Y_{j_1},\ldots,Y_{j_\ell}$ commute and their product equals c_j Id.

It is not too hard to show that for any BLS, an operator solution immediately translates into a perfect quantum strategy for it.

Lemma 3.7. *Suppose given an operator solution $Y_1,\ldots,Y_n$ to a BLS (E,c) such that each Y_j is a binary observable on a finite-dimensional Hilbert space $\mathcal{H}$. Then the following strategy succeeds with probability* 1 *in the BLS game:*

- *The players share the* maximally entangled *state*

$$|\psi\rangle_{AB} = \frac{1}{\sqrt{d}}\sum_i |i\rangle_A \otimes |i\rangle_B \in \mathcal{H}_A \otimes \mathcal{H}_B\,, \tag{3.7}$$

  where d is the dimension of $\mathcal{H}$, each of $\mathcal{H}_A$ and $\mathcal{H}_B$ is a copy of $\mathcal{H}$, and $\{|i\rangle\}$ an orthonormal basis for it.[10]

- *On question j, Alice sequentially measures the observables $Y_{j_1}, Y_{j_2}, \dots, Y_{j_\ell}$ on her share of $|\psi\rangle$, where $j_1, \dots, j_\ell$ are the indices of the nonzero entries of the j-th row of E. She obtains outcomes $a_1, \dots, a_\ell$ that she returns as her answer.*

- *On question $k \in \{1\dots,n\}$ Bob measures the observable Y_k^T on his share of $|\psi\rangle$. He obtains an outcome $b \in \{\pm 1\}$ that he returns as his answer.*

Proof. First we note that the strategy described in the lemma is valid: since by definition of an operator solution the observables $Y_{j_1}, Y_{j_2}, \dots, Y_{j_\ell}$ always commute it is possible for Alice to measure them simultaneously.

The following relation holds the key to the proof: for any operators A on $\mathcal{H}_A$ and B on $\mathcal{H}_B$ it holds that

$$\langle\psi|A \otimes B|\psi\rangle \;=\; \frac{1}{d}\mathrm{Tr}(AB^T)\,, \tag{3.8}$$

where $|\psi\rangle$ is as in (3.7). This relation follows easily from the relation $(\mathrm{Id} \otimes B)|\psi\rangle \;=\; (B^T \otimes \mathrm{Id})|\psi\rangle$ that we saw in the previous lecture and the fact that the reduced density matrix of $|\psi\rangle$ on either subsystem is the totally mixed stated $d^{-1}\,\mathrm{Id}$. Using this relation it is a matter of direct calculation to verify that the prover's answers always satisfy the verifier's checks in the game. In more detail,

- For the consistency check, we note that the probability that the two players return consistent answers on question (j,k) is

$$\frac{1}{2} + \frac{1}{2}\langle\psi|Y_{jk} \otimes Y_{jk}^T|\psi\rangle \;=\; \frac{1}{2} + \frac{1}{2}\frac{1}{d}\mathrm{Tr}\big(Y_{jk}^2\big) \;=\; 1\,,$$

  where the first equality follows from (3.8) and the second holds since Y_{jk} is a binary observable so $Y_{jk}^2 = \mathrm{Id}$.

- For the equation check, we note that the probability that Alice's answers satisfy the check for the j-th equation is

$$\frac{1}{2} + \frac{c_j}{2}\langle\psi|Y_{j_1}\cdots Y_{j_\ell} \otimes \mathrm{Id}\,|\psi\rangle \;=\; \frac{1}{2} + \frac{c_j}{2}\langle\psi|c_j\,\mathrm{Id} \otimes \mathrm{Id}\,|\psi\rangle \;=\; 1\,,$$

  where the first equality holds since $Y_{j_1}\cdots Y_{j_\ell} = c_j\,\mathrm{Id}$ by definition of an operator solution.

□

Remark 3.8. The reader will have noticed that in Lemma 3.7 we carefully added the assumption that the operator solution is finite-dimensional, and indeed this seems necessary for the state $|\psi\rangle$ to be well-defined. It is possible to show that infinite-dimensional operator solutions to a BLS correspond to *commuting-operator* strategies for the associated game, and conversely; this correspondence is established in [CLS17]. Commuting-operator strategies are a strict superset of tensor-product strategies

[10] The maximally entangled state is a natural generalization of the EPR pair which can be defined on any tensor product of (finite-dimensional) isomorphic Hilbert spaces.

Combining Lemma 3.7 with the operator solution to the magic square given by (3.6) we obtain a perfect strategy for the magic square game that uses two qubits per player, and two EPR pairs shared between them. Since we saw that the magic square does not have a perfect strategy this strategy gives us another example of a non-signaling correlation that is not classical.

3.3.2 Characterization of optimal strategies

The following converse to Lemma 3.7 is shown in [CM14].

Lemma 3.9. *Suppose given a BLS (E, c) and a strategy $(|\psi\rangle, A, B)$ for the associated game that succeeds with probability* 1*. Then the BLS has a finite-dimensional operator solution.*

Proof. We give the proof for the special case of the Magic Square game, as the general case is similar. We start with the modeling step: a strategy $(|\psi\rangle, A, B)$ for the magic square game is given by a bipartite state $|\psi\rangle \in \mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$ for finite-dimensional $\mathcal{H}_{\mathsf{A}}$ and $\mathcal{H}_{\mathsf{B}}$ as well as the following measurements. For the first player (Alice), for each row or column x there is a 9-outcome projective measurement $\{A^x_a : a \in \{\pm 1\}^3\}$ on $\mathcal{H}_{\mathsf{A}}$. For the second player (Bob), for each variable (square) y there is an observable B_y on $\mathcal{H}_{\mathsf{B}}$. Note that here we assumed that the measurements made by each player are projective, which is without loss of generality by applying Naimark's theorem and enlarging the spaces $\mathcal{H}_{\mathsf{A}}$ and $\mathcal{H}_{\mathsf{B}}$ if necessary.

To each of Alice's questions we can associate three observables that correspond to the three bits of her answer. For example, for question $j = 1$ (first row) we can define

$$A_1 = \sum_{a_1,a_2,a_3 \in \{\pm 1\}} a_1\, A^1_{a_1 a_2 a_3}\;, \quad A_2 = \sum_{a_1,a_2,a_3 \in \{\pm 1\}} a_2\, A^1_{a_1 a_2 a_3}\;, \quad A_3 = \sum_{a_1,a_2,a_3 \in \{\pm 1\}} a_3\, A^1_{a_1 a_2 a_3}\;.$$

We can similarly proceed to define $A_4, \ldots, A_9$ from the rows and $A'_1, \ldots, A'_9$ from the columns. Next we show that success with probability 1 in the consistency checks implies that

$$\forall y \in \{1, \ldots, 9\}\;, \qquad A_y = A'_y = B_y^T\;. \tag{3.9}$$

Take for example the consistency check on question $(1, 2)$ (first row to Alice, second entry to Bob). It is easy to show that success in that check implies that

$$\langle \psi | A_2 \otimes B_2^T | \psi \rangle = 1\;. \tag{3.10}$$

We use the following claim.

Claim 3.10. *Suppose that $|\psi\rangle$ is a bipartite state A, B observables such that $\langle \psi | A \otimes B | \psi \rangle = 1$. Let $|\psi\rangle = \sum_t \lambda_t |u_t\rangle |v_t\rangle$ be the Schmidt decomposition of $|\psi\rangle$, with $\lambda_t > 0$ for all t and $\{|u_t\rangle\}$ and $\{|v_t\rangle\}$ orthonormal families. Let $S_A = Span\{|u_t\rangle\} \subseteq \mathcal{H}_A$ and $S_B = Span\{|v_t\rangle\} \subseteq \mathcal{H}_B$. Then S_A is stable by A and S_B is stable by B. Moreover, letting A_S denote the matrix of the restriction of A to S_A in the basis $\{|u_t\rangle\}$ and similarly for B, it holds that $A_S = B_S^T$.*

Proof sketch. Let $K = \sum_t \lambda_t |u_t\rangle\langle v_t|$. Then the equality $\langle \psi | A \otimes B | \psi \rangle = 1$ is equivalent to $AKB^T = K$. Identifying left and right eigenspaces we see that A and B must each preserve the eigenspaces of K associated with any given eigenvalue. Thus $AKB^T = K$ decomposes in block form $\oplus_\lambda A_\lambda B_\lambda^T = \mathrm{Id}_\lambda$, where for each block we indicated with a subscript λ the restriction of each operator to the eigenspace of K associated with eigenvalue λ. This shows the claim. □

Using Claim 3.10 and the implications of the form (3.10) for the consistency checks, (3.9) follows, where the operators and the transpose should be understood to be written with respect to the Schmidt bases of $|\psi\rangle$. To conclude we claim that $B_1^T, \dots, B_9^T$ (precisely, their restriction to the support of $|\psi\rangle$ on $\mathcal{H}_\mathsf{B}$) are an operator solution to the Magic Square. Commutation in each row or column follows from (3.9) and the definition of the A_y (which by definition commute by rows) and A'_y (by columns). The constraints follow from the fact that e.g. for the first row, $\langle\psi|A_1A_2A_3 \otimes \mathrm{Id}\,|\psi\rangle = +1$, which using Claim 3.10 implies that $A_1A_2A_3 = \mathrm{Id}$ and hence $B_1^TB_2^TB_3^T = \mathrm{Id}$. (Of course we could remove the transpose signs and still have a valid solution.) □

3.4 A nonlocal test for a qubit

We now have everything that we need in order to give our first classical-verifier test for a qubit (in fact, as we will see, for two qubits!). To motivate this, observe that the proof of Lemma 3.9 says a bit more than is stated in the lemma itself: not only did we show that the Magic Square has an operator solution, we also exhibited such a solution directly from the second player's observables in the game. Let's show the following simple fact.

Claim 3.11. *Suppose given an operator solution $Y_1, \dots, Y_9$ to the magic square. Then Y_2 and Y_4 anti-commute.*

Proof. We first rewrite the product Y_2Y_4 by rows to obtain

$$\begin{aligned} Y_2Y_4 &= Y_1Y_3 \cdot Y_6Y_5 \\ &= Y_1 \cdot Y_9 \cdot Y_5 , \end{aligned}$$

where the second line is by the last column constraint. Next we write the product Y_4Y_2 by columns:

$$\begin{aligned} Y_4Y_2 &= Y_1Y_7 \cdot Y_8Y_5 \\ &= Y_1 \cdot (-Y_9) \cdot Y_5 , \end{aligned}$$

where the second line is by the last row constraint. Combining both equations it follows that $Y_2Y_4 = -Y_4Y_2$, as claimed. □

The following lemma is immediate from the proof of Lemma 3.9 and Claim 3.11. We state the lemma using the language of "self-testing" from the previous lecture.

Lemma 3.12. *Suppose that two non-communicating quantum devices A and B generate correlations*

$$p(a,b|x,y) \;=\; \langle\psi|A_a^x \otimes B_b^y|\psi\rangle$$

that perfectly satisfy the referee's tests in the Magic Square game. Let S_B denote the support of the reduced density ρ_B of $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ on $\mathcal{H}_B$. Then the observables $B_1, \dots, B_9$ stabilize S_B, and their restriction to that space form an operator solution to the Magic Square. In particular, the device's joint state $|\psi\rangle_{AB}$ together with observables B_2 and B_4 of device B associated with inputs $y = 2$ and $y = 4$ respectively form a qubit $(|\psi\rangle, B_2, B_4)$.

Proof. The first part of the lemma follows from the proof of Lemma 3.9. By Claim 3.11 the observables associated to $y = 2$ and $y = 4$ anti-commute. □

The preceding lemma shows that two of device B's observables, B_2 and B_4, must anti-commute. As we saw in Lemma 1.2[11] this means that up to an isomorphism on the device's space these observables must take the form $B_2 \simeq \sigma_Z \otimes \mathrm{Id}$ and $B_4 \simeq \sigma_Z \otimes \mathrm{Id}$, which is exactly the form that they take in the solution given in (3.6). What about the other observables? In other words, are the constraints that underlie the Magic Square game *rigid*?

Lemma 3.13. *Under the same assumptions as Lemma 3.12 there is a unitary U on the space $\mathcal{H}_B$ associated with device B such that the observables $U^\dagger B_k U$ for $k \in \{1,\ldots,9\}$ take the form described in* (3.6). *in particular, the dimension of (the span of the support of $|\psi\rangle$ on) $\mathcal{H}_B$ is a multiple of* 4.

Proof. The main ingredient in this proof is the qubit lemma, Lemma 1.2, together with Claim 1.6 which allows us to argue that observables that commute with σ_Z and σ_X on a copy of $\mathbb{C}^2$ must act as identity on $\mathbb{C}^2$.

First note that the proof of Lemma 3.12 immediately extends to show that any two observables not in the same row or column anti-commute. Furthermore, by definition the condition that the 9 observables $B_1,\ldots,B_9$ form an operator solution to the Magic Square implies that all observables in the same row or column must commute. Using this condition and the characterization of B_2 and B_4 given in Lemma 3.12 it follows from Claim 1.6 that $B_1 \simeq \mathrm{Id} \otimes B_1'$ and $B_5 \simeq \mathrm{Id} \otimes B_5'$, for some observable B_1' and B_5' on $\mathcal{H}'$ that anti-commute. Using Lemma 1.2 again it follows that there is an isometry U' on $\mathcal{H}'$ such that as operators on $\mathcal{H}'$, $B_1' \simeq \sigma_Z \otimes \mathrm{Id}$ and $B_5' \simeq \sigma_X \otimes \mathrm{Id}$, with the identity acting on some new ancilla space $\mathcal{H}''$ such that $\mathcal{H}' \simeq \mathbb{C}^2 \otimes \mathcal{H}''$. Combining U and U' together, we have shown that there is an isomorphism $U'U$ under which

$$\begin{array}{ll} B_1 \simeq \mathrm{Id} \otimes \sigma_Z \otimes \mathrm{Id} & B_2 \simeq \sigma_Z \otimes \mathrm{Id} \otimes \mathrm{Id} \\ B_4 \simeq \sigma_X \otimes \mathrm{Id} \otimes \mathrm{Id} & B_5 \simeq \mathrm{Id} \otimes \sigma_X \otimes \mathrm{Id} \end{array}.$$

The remaining entries of the table are immediately filled in from the row and column constraints, which uniquely determine them. □

As a last step we show that we can also characterize the entangled state used by any strategy. Interestingly, this characterization comes as a consequence of the characterization of the observables, which we obtained without talking much about the state. This is based on the following general lemma, that we will often make use of.

Lemma 3.14. *Let $|\psi\rangle_{ABE} \in (\mathbb{C}^2)_A^{\otimes n} \otimes (\mathbb{C}^2)_B^{\otimes n} \otimes \mathcal{H}_E$ be such that for every $i \in \{1,\ldots,n\}$ it holds that*

$$(\sigma_{X,i})_A \otimes (\sigma_{X,i})_B |\psi\rangle_{ABE} = (\sigma_{Z,i})_A \otimes (\sigma_{Z,i})_B |\psi\rangle_{ABE} = |\psi\rangle_{ABE},$$

where the Pauli operators act on the i-th copy of $\mathbb{C}^2$ in register A and B respectively. Then $|\psi\rangle_{ABE} = |\phi^+\rangle_{AB}^{\otimes n} \otimes |aux\rangle$, for some state $|aux\rangle$ on $\mathcal{H}$.

Proof. Note that $\sigma_X \otimes \sigma_X$ and $\sigma_Z \otimes \sigma_Z$ commute, hence are simultaneously diagonalizable. The proof immediately follows from the observation that the only simultaneous eigenvalue-1 eigenstate of $\sigma_X \otimes \sigma_X$ and $\sigma_Z \otimes \sigma_Z$ is the EPR pair $|\phi^+\rangle$. □

Exercise 3.1. Show that the conclusion of Lemma 3.14 holds under the following weaker assumption: $|\psi\rangle_{\mathsf{ABE}} \in (\mathbb{C}^2)_{\mathsf{A}}^{\otimes n} \otimes \mathcal{H}_{\mathsf{B}}^{\otimes n} \otimes \mathcal{H}_{\mathsf{E}}$ with $\mathcal{H}_{\mathsf{B}}$ arbitrary, and for every $i \in \{1,\ldots,n\}$,

$$(\sigma_{X,i})_{\mathsf{A}} \otimes (X_i)_{\mathsf{B}} |\psi\rangle_{\mathsf{ABE}} = (\sigma_{Z,i})_{\mathsf{A}} \otimes (Z_i)_{\mathsf{B}} |\psi\rangle_{\mathsf{ABE}} = |\psi\rangle_{\mathsf{ABE}},$$

[11] Here we can apply the "state-independent" version of the qubit lemma because Lemma 3.12 states that the observables themselves, or rather their restriction to the support of $|\psi\rangle$, satisfy the operator constraints.

with X_i and Z_i arbitrary binary observables on $\mathcal{H}_\mathsf{B}$ (in particular, we are not assuming any a priori qubit structure on $\mathcal{H}_\mathsf{B}$). *[Hint: Remember Claim 2.7]*

3.4.1 Consequences

The characterization of perfect strategies given in Lemma 3.13 together with Lemma 3.14 have some nice consequences. First of all, they imply that the Magic Square game tests not one, but two qubits: any perfect strategy must have a 4-qubit entangled state, two qubits per player, and Bob's observables specify two qubits, e.g. B_2 and B_4 for the first and B_1 and B_5 for the second. We even have access to more: for example, we know that when Bob is asked question 9 the observable he applies is $\sigma_Y \otimes \sigma_Y$. Although we clearly have some distance to go, these are first steps towards testing that Bob implements a certain computation; for now, we are able to test that he applies specific observables.

Another consequence of the characterization has to do with the problem of randomness certification. At this point we know that, in any perfect strategy, whenever Bob is asked question 2 he measures the first qubit of an EPR pair in the standard basis. This has the following implications:

1. The answer reported by Bob on question 2 (and, in fact, on *any* question) is a uniformly random bit. In particular, no deterministic strategy can succeed in the game! We knew this already, because deterministic strategies are classical. As such, any game for which quantum strategies can succeed with strictly higher probability than classical strategies can serve as a "test for randomness".

2. More importantly, the randomness that is generated by Bob at each execution of the game is "fresh" and "private". What we mean by this is that Bob's random bit is (1) independent of any information at the verifier's side, including Bob's question, and (2) uncorrelated to the environment. Indeed, since Bob's bit is the result of a measurement of half an EPR pair, the only party that can obtain correlated information is Alice, who holds the other half of the EPR pair. By the rigidity theorem this EPR pair *must* be in control of Alice: she needs it for them to succeed in the game. Therefore the verifier has the guarantee that the bit she obtains (1) cannot have been "planted" *a priori* in the devices, and (2) cannot be learned, even partially, by any third party distinct from A and B, even if the party could a priori have kept entanglement with the devices—this is because, using the notation of Lemma 3.14, the third party would only at best have access to the entirety of system E, which is uncorrelated with AB.

These observations are important for cryptography, where the use of high-quality randomness that is uncorrelated from any possible eavesdropper or adversary is an essential resource. Indeed, the observations we just made form the basis for the so-called "device-independent" analysis of quantum cryptography protocols.

Remark 3.15. We presented the fact that the Magic Square game tests two qubits, instead of one, as a "feature". But what if one only cares about a single qubit, is there a simple test for this? There is such a test, but it is not an BLS game: it is the CHSH game. The proof that this game tests a qubit was recognized early on, see e.g. [SW87] or [MYS12] for a more modern treatment. Unfortunately the game does not have "quantum completeness 1", in the sense that the optimal quantum strategy for it achieves a success probability that is greater than the optimal classical, but less than 1 (precisely, it is $\cos^2 \pi/8 \approx 0.85$). This makes it less convenient to use as a building block in larger protocols, and so here we will stick with the Magic Square game that is the simplest value-1 game which self-tests at least one qubit that we know of.

An important drawback of our analysis so far is that it is limited to the case of perfect strategies, i.e. strategies that succeed with probability 1 in the game. In practice one may only reasonably assume, after multiple executions of the game, that a given strategy succeeds with some probability that is close to one,

$1-\varepsilon$ for some $\varepsilon \geq 0$ that can be made small but not 0. In the next section we discuss how the results can be extended to that case.

3.4.2 The approximate case

The following theorem gives the flavor of an approximate version of the lemmas from the previous section. It is taken from [CS17], where more general statements are shown for any BLS that satisfies appropriate conditions. (A similar result specialized to the case of the Magic Square game is shown in [WBMS16].)

Theorem 3.16. *Suppose that a strategy $(|\psi\rangle, A, B)$ succeeds with probability $1-\varepsilon$ in the Magic Square game, for some $\varepsilon \geq 0$. Then there are isometries $V_D : \mathcal{H}_D \to \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathcal{H}_{D'}$ for $D \in \{A, B\}$ such that*

$$\big\| V_A \otimes V_B |\psi\rangle_{AB} - |\phi^+\rangle \otimes |\phi^+\rangle \otimes |aux\rangle \big\|^2 \;=\; O\big(\sqrt{\varepsilon}\big)\;,$$

for some state $|aux\rangle$ on $\mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$, and

$$\big\| \mathrm{Id}_A \otimes \big(V_B B_2 - (\sigma_Z \otimes \mathrm{Id} \otimes \mathrm{Id}) V_B\big) |\psi\rangle_{AB} \big\|^2 = O(\varepsilon)\;,$$
$$\big\| \mathrm{Id}_A \otimes \big(V_B B_4 - (\sigma_X \otimes \mathrm{Id} \otimes \mathrm{Id}) V_B\big) |\psi\rangle_{AB} \big\|^2 = O(\varepsilon)\;,$$

and similar relations hold for the remaining seven observables on Bob's side.

Note that the theorem only characterizes the player's observables "up to isometry", as opposed to "up to isomorphism" as we were able to in the perfect case (Lemma 3.13). As discussed in the previous lecture (Section 2.3) this is unavoidable in general.

Later we will see a general method to derive statements such as Theorem 3.16 based on the use of approximate group representation theory. For now, we let it serve as a good illustration of the kind of statements we aim to prove in this course. It is worth reflecting on the strength of what we have achieved: using only classical data and a single physical assumption (our model for spatial isolation based on the use of tensor products) we have arrived at a very simple test that can be used to fully characterize the quantum state of a 16-dimensional system (4 qubits) as well as elementary operations performed on it. This conclusion is much stronger than the "standard" conclusion that motivates the study of Bell inequalities in the first place: that they require entanglement.[12] There is no equivalent to this in classical theory!

[12] Note that the fact that the isometries V_{A} and V_{B} are "local", each acting only on one half of the total Hilbert space, is important because it means that they couldn't have artificially create the entanglement present in $|\phi^+\rangle \otimes |\phi^+\rangle$: that entanglement must "exist" even independently of the application of the isometry, which only serves to "package" it in the neat form of two EPR pairs. Of course, the state $|aux\rangle$ may or may not contain entanglement itself.

Lecture 4

Testing a qubit under computational assumptions

As we discussed in the first lecture, a simple interactive "test of quantumness" under computational assumptions consists in asking the device to factor a large integer n; under the assumption that factoring is hard for classical computers this test adequately distinguishes classical from quantum devices.

The main limitation of this test that is generally pointed out is that in order for a device to successfully demonstrate its "quantumness" it needs to have the capability to implement a large, fault-tolerant quantum computation. In contrast, the test based on spatial isolation that we saw in the previous lecture can be executed with a pair of two-qubit devices.[1]

A second limitation that is relevant for us is that the test does not seem to provide a means to certify a qubit. Modeling the prover in a "factoring test" would give us a family of POVM $\{\Pi^n_{\{p_i\}}\}$, indexed by integers n and whose outcomes are lists of primes $\{p_i\}$. In fact, for each n there should be a single POVM element —the one with the correct outcome— such that $\Pi^n_{\{p_i\}}|\psi\rangle = |\psi\rangle$ with $|\psi\rangle$ the initial state of the device; in somewhat informal notation that POVM element is supposed to be obtained as $U^\dagger|\{p_i\}\rangle\langle\{p_i\}|U$ with U a circuit implementation of Shor's algorithm. But to identify a qubit we know that we need *two* observables acting on the same space as well as some indication that these observables ought to be "incompatible" (anti-commutation). It is not at all clear how to identify such observables here.

A key insight from this lecture is that in order to obtain a computational test for a qubit we will need to assume that a certain problem is hard not for classical computers but also for quantum computers. This is because our model requires us to identify two observables X and Z in the device such that the device has the ability to perform either X or Z but not both simultaneously. In Lecture 2 this incompatibility arose from the necessity for X and Z to yield predictions that matched those of σ_X and σ_Z. In the third lecture, it arose from some form of information-theoretic impossibility—in a loose sense, had X and Z (recall that in the notation from the previous lecture these were identified as B_4 and B_2) been compatible, then the Magic Square would have had a classical solution—which it doesn't.

In this lecture the impossibility of measuring X and Z will be based on considerations of computational difficulty. We will show that, if the quantum device was able to measure X and Z jointly then it would break a computational problem that is assumed to be hard *even for quantum computers* — we will formulate this later as a form of "computational uncertainty principle". Since by definition the device can measure X

[1] The test does require the ability to distribute entanglement across a large distance—if one uses relativity to certify the assumption of spatial isolation—and to perform fairly high-fidelity measurements on it. This is not at all easy, but it can be done today; implementations of Shor's algorithm that outperform the best classical factoring algorithms are not expected within the next decade.

and Z separately, if they commuted then it could also measure them jointly. Therefore, the computational assumption gives rise to an *information-theoretic* consequence on the observables X and Z: they must form a qubit.

4.1 Simon's algorithm

As a warm-up we start by reminding ourselves of a typical example of a kind of task for which the manipulation of quantum information provides a computational advantage.

4.1.1 The algorithm

The input to an instance of Simon's problem is a function $f : \{0,1\}^n \to \{0,1\}^n$ that has the property that f is 2-to-1 (every value in the range has exactly two preimages) and moreover there is a string $s \in \{0,1\}^n$ such that for every $x, y \in \{0,1\}^n$, $f(x) = f(y)$ if and only if $y = x$ or $y = x + s$, where addition is performed coordinate-wise and modulo 2. The goal is to recover the string s. It is not hard to see that in the worst case any classical algorithm requires at least $\Omega(2^{n/2})$ evaluations of f to determine s. This is because on the one hand for any deterministic algorithm that makes a smaller number of evaluations there is a function f such that all values returned by f are distinct, so no information about s is gained; similarly one can show that for any randomized algorithm if f is chosen at random then it is unlikely that the algorithm will gain any information about s in $\ll 2^{n/2}$ evaluations. On the other hand, by making roughly $\Omega(2^{n/2})$ evaluations at random points then by the birthday paradox one will likely obtain $x \neq y$ such that $f(x) = f(y)$, which immediately reveals $s = x + y$.

Simon showed that there is a quantum algorithm that can solve this problem using only $O(n)$ evaluations, provided that the function f can be evaluated "in superposition". The algorithm first evaluates f on a uniform superposition of inputs, as follows:

$$\begin{aligned} |0^n\rangle|0^n\rangle &\mapsto \frac{1}{\sqrt{2^n}}\sum_x |x\rangle|0^n\rangle \\ &\mapsto \frac{1}{\sqrt{2^n}}\sum_x |x\rangle|f(x)\rangle \ . \end{aligned}$$

It then measures the last register in the computational basis, yielding some $y = f(x_0) = f(x_1)$ where x_0 and $x_1 = x_0 + s$ are the two preimages of y under f. The re-normalized post-measurement state is

$$\frac{1}{\sqrt{2}}\big(|x_0\rangle + |x_1\rangle\big)|y\rangle \ . \tag{4.1}$$

Measuring the first register in the Hadamard basis yields a uniformly random $d \in \{0,1\}^n$ such that $d \cdot s = 0$. Repeating the entire procedure $O(n)$ times yields $(n-1)$ linearly independent such d's, which suffices to recover s with high probability.

4.1.2 Instantiating the black box

The main limitation of Simon's problem is that it only provides a *black-box* separation: the quantum advantage holds under the assumption that the classical or quantum algorithms are allowed to evaluate the function f, but they are not given an explicit description of it. Showing that the separation still holds for an explicit choice of the function f is much harder, because it is difficult to rule out some smart behavior for the classical algorithm that would take advantage of specific code for f; indeed, showing such a separation would be a major breakthrough in quantum algorithms.

This difficulty shouldn't prevent us from toying with the question: Can we identify natural candidates? For example one could take $f(x) = Ax$ for $A \in \mathbb{F}_2^{n\times n}$ a matrix of rank exactly $(n-1)$. In that case the kernel of A is spanned by a single vector $s \in \mathbb{F}_2^n$, and f is exactly 2-to-1: $f(x_0) = f(x_1)$ if and only if $A(x_0 - x_1) = 0$, i.e. $x_0 - x_1$ is either 0 or s. Unfortunately this f is not a good candidate, because there happens to be an efficient classical algorithm that directly solves Simon's problem for it: Gaussian elimination.[2] The example shows that at a minimum we need a function f that is 2-to-1 but such that finding any colliding pair of inputs (x_0, x_1) with $f(x_0) = f(x_1)$ is computationally difficult. In the next section we introduce some background from cryptography that will allow us to make this requirement precise.

4.2 Computational assumptions

So far everything that we have done in the course is "information-theoretic", in the sense that we have not had to fix a model of computation nor discuss efficiency of the various operations that we had our verifiers and provers implement (the only distinction we made is if an operation is classical or if it requires a quantum component). In this lecture we start making assumptions of a computational nature, such as "this problem cannot be solved by this class of adversaries". In this section we briefly review the formalism for making such assumptions precise and apply it to a specific scenario of interest for the lecture.

4.2.1 PPT and QPT procedures

The first thing that we need to make precise is our computational model. Since the protocols we consider involve interaction between a verifier and prover(s) we focus on modeling such devices as machines that perform a computation. Loosely speaking, each device operates in a number of rounds where at each round the device performs a computation that takes it from a certain internal state as well as an input (a message received from another device) to a new internal state and an output (a message that it returns). We will model each such computation as a circuit. A circuit is a sequence of elementary operations called "gates" that operate either on a classical state (in which case the gates can be things like an AND, an OR, a NOT, etc.) or a quantum state (in which case the gates can be things like a 1-qubit Hadamard, a σ_X or σ_Z, a 2-qubit controlled NOT, etc.).[3] To recap, for us a verifier or a prover is specified by a sequence of classical or quantum circuits. We will always assume that the circuits explicitly specify which spaces they are meant to operate on (e.g. verifier's space, message from verifier to prover, etc.).

Next we discuss what it means for a verifier (or prover) to be "efficient". To make this precise we need to talk about *families* of verifiers. We will imagine that there is an underlying size parameter $n \in \mathbb{N}$ (for example, n could be the size of a 3SAT formula that the verifier aims to check, or the number of qubits that she aims to certify) and that the verifier (or prover) is specified by a classical Turing machine M that on input 1^n returns an explicit classical description of a sequence of circuits that can be used to implement the verifier (or prover) for problems of size n. We will say that the verifier (or prover) is *probabilistic polynomial time* (PPT) (resp. *quantum polynomial time* (QPT)) if this Turing machine runs in time polynomial in its input (i.e. polynomial in n; this is why we always assume that n is passed in unary to M) and returns a

[2] In lecture 7 we will see that a "noisy" version of f provides a partial workaround.

[3] To be fully precise we would need to fix a finite gate set for classical circuits and another for quantum circuits. What gate set is used will not matter for us; the only important point is that there exists finite universal gate sets and that all such gate sets are roughly equivalent in terms of how many gates are required to decompose any larger unitary.

family of classical (resp. quantum) circuits. Note that the assumption that the Turing machine is polynomial time immediately implies that the circuits it returns act on polynomially many bits (resp. qubits) and have a polynomial number of classical (resp. quantum) gates.

In a cryptographic context we will generally allow M to take a second input 1^λ for $\lambda \in \mathbb{N}$ called the *security parameter*. While the input size n is governed by the size of the problem, the security parameter can be chosen at will; the larger it is the more "secure" the protocol is supposed to be (for example, the smaller the probability that the verifier makes an incorrect decision or the higher the quality of the certified qubits).

4.2.2 Claw-free functions

Similarly to circuits in the previous section, when we talk about computational *difficulty* of a certain problem we always need to refer to *families* of objects. This is because e.g. for any given function f there is nothing "hard" about the task of recovering specific information about f: if f is fixed everything about it is fixed as well; in particular in the case when f has the structure required for Simon's problem there is a simple algorithm that identifies s, and this is the algorithm that writes s down starting from any initial state.

For this reason we will always consider families of functions $\{f_{pk} : \{0,1\}^{m(\lambda)} \to \{0,1\}^{m(\lambda)}\}_{pk \in \{0,1\}^{k(\lambda)}}$ where the index $\lambda \in \mathbb{N}$ is called *security parameter* and k and m are polynomially bounded functions of λ; the idea is that for each λ there is a collection of functions, indexed by strings of length $k(\lambda)$ and with the same domain and range, such that the larger the λ the more "complex" the functions are. For example, we could take $k(\lambda) = \lambda^2$, $m(\lambda) = \lambda$, and f_{pk} to be multiplication by the matrix $A \in \{0,1\}^{\lambda \times \lambda}$ obtained by "reshaping" the λ^2-bit string pk into a $\lambda \times \lambda$ square.

Let's give our first definition of a cryptographic property that applies to a family of functions.

Definition 4.1 (Claw-free function family)**.** A family $\mathcal{F} = \{f_{pk} : \{0,1\}^{m(\lambda)} \to \{0,1\}^{m(\lambda)}\}_{pk \in \{0,1\}^{k(\lambda)}}$ is *claw-free* against classical (resp. quantum) adversaries if the following conditions hold:

- f_{pk} can be efficiently evaluated: there is a PPT procedure that given pk and x as inputs returns $f_{pk}(x)$.
- For every $\lambda \in \mathbb{N}$ and $pk \in \{0,1\}^{k(\lambda)}$, f_{pk} is 2-to-1.
- For every PPT (resp. QPT) procedure $\mathcal{A}$ the following holds: (the procedure $\mathcal{A}$ is often personified as the "adversary" trying to demonstrate that the function family is *not* claw-free) there exists a negligible[4] function $\mu : \mathbb{N} \to \mathbb{N}$ such that for every λ,

$$\Pr_{pk \leftarrow_R \{0,1\}^{k(\lambda)}} \big((x_0, x_1) \leftarrow \mathcal{A}(1^\lambda, pk) :\ x_0 \neq x_1,\ f_{pk}(x_0) = f_{pk}(x_1) \big) \leq \mu(\lambda) .$$

In words, the third condition states that there is no polynomial-time algorithm that given a uniformly chosen index pk for a function from the family is able to return two distinct inputs for the function that constitute a claw.[5]

Remark 4.2. In the definition we require the function family to be parametrized by arbitrary strings pk. In general this requirement can be relaxed; in fact there could even be a single function for every λ. In cryptographic constructions the function family generally comes equipped with a PPT *key generation procedure* GEN that takes 1^λ as input and returns pk.

[4] A function $\mu : \mathbb{N} \to \mathbb{R}$ is called negligible if for every polynomial p, $p(\lambda)\mu(\lambda) \to_{\lambda \to \infty} 0$.

[5] A triple (x_0, x_1, y) such that $x_0 \neq x_1$ and $f(x_0) = f(x_1) = y$ is called a *claw*. To see why, picture the arrows $x_0 \to y$ and $x_1 \to y$ drawn with x_0, x_1 on top of each other on the left and y on the right.

An example of a claw-free family of functions against PPT adversaries can be constructed as follows. (This construction appears in [GMR85], where it is used to construct a digital signature scheme.) Let $N = pq$ be a product of two primes $p \equiv 3 \bmod 8$ and $q \equiv 7 \bmod 8$. This choice ensures that -1 and 2 are not squares $\bmod N$; moreover, if Q_N denotes the set of quadratic residues (i.e. squares) modulo N then $f_0(x) = x^2 \bmod N$ and $f_1(x) = 4x^2 \bmod N$ are both permutations of Q_N. (This fact requires proof but it is a simple exercise in arithmetic.) However, suppose given a claw (x_0, x_1) such that $x_0, x_1 \in Q_N$ and $f_0(x_0) = f_1(x_1)$. Then $x_0^2 = 4x_1^2$ but $x_0 \neq \pm 2x_1 \bmod N$ because $\pm 2x_1 \notin Q_N$. Thus computing the GCD of N with $x_0 \pm 2x_1$ recovers a nontrivial factor.

While this family of functions is claw-free with respect to PPT adversaries, it is clearly not claw-free against QPT adversaries, that can use Shor's algorithm to factor efficiently. We will construct such a function family in the next lecture; for the time being we assume its existence.

4.2.3 Hardcore bits

Following the initial steps of Simon's algorithm as described in Section 4.1.1 when instantiated with any 2-to-1 function enables a quantum device to generate strings $d \in \{0,1\}^m$ such that $d \cdot (x_0 + x_1) = 0$, where (x_0, x_1) are preimages of some $y \in \mathbb{F}_2^m$ by f. Intuitively one might expect that this represents a computational advantage, because d provides an equation in $x_0 + x_1$, which is some information about both preimages together. For example in the case where $x_0 + x_1 = s$, where s is some fixed secret independent of y, we saw that provided the equation d can be assumed to be uniformly distributed among all valid equations in s then running the procedure $O(m)$ times gives sufficiently many equations to recover s.

Unfortunately, as discussed in Section 4.1.2 for the only function that we could think of that has this property it is in fact easy to recover s, even for a classical computer. This suggests that in general the assumption that f satisfies the structure required for Simon's algorithm might be too strong to obtain an explicit candidate. Moreover, recall that at the start of the lecture we pointed out that our goal is not directly to find a task for which there is a quantum computational advantage, but instead we are trying to identify *two* tasks that the quantum device can perform *separately* but not *simultaneously*—if someone, such as a classical device, was able to execute both tasks simultaneously then it would break the computational assumption. What could those two tasks be here?

Starting from the state (4.1) it is natural to measure in the Hadamard basis, obtaining as before an equation d such that $d \cdot (x_0 + x_1) = 0$, but also in the computational basis, obtaining either x_0 or x_1. Given that "honest" measurements in the computational and Hadamard basis are incompatible, these are natural candidates for our "qubit." However, we also saw that if $x_0 + x_1 = s$ for some fixed secret s then the Hadamard measurements alone allow us to recover s. So a quantum procedure could recover s "on the side" and then, knowing f explicitly, succeed in any reasonable "test" by using classical operations alone—this would make it very hard for us to identify the "qubit" that the device should have used to recover s (this is similar to the example of Shor's algorithm given at the start of the lecture). But what if the structure of f is a little more complicated, so that e.g. $x_0 + x_1 = g(x_0, s)$ for some function g? In that case a single equation in $g(x_0, s)$ might not be so useful; even many such equations for varying x_0 could be useless since without knowledge of x_0 itself one cannot determine what the equation is about. However, if one was able to obtain x_0 simultaneously with the equation then one would obtain a sequence of (possibly non-linear, depending on g) constraints on s. These considerations motivate the following computational assumption:

Assumption 1 (Adaptive hardcore bit)**.** *There is a claw-free family of functions $\mathcal{F} = \{f_{pk}\}$ such that for* *any QPT adversary $\mathcal{A}$ there is a negligible function μ such that*

$$\Big|\frac{1}{2} - \Pr_{pk \leftarrow_R \{0,1\}^{k(\lambda)}} \Big((x,d) \leftarrow \mathcal{A}(1^\lambda, pk),\ \{x_0, x_1\} \leftarrow f_{pk}^{-1}(f_{pk}(x)) :\ (d \neq 0^m) \wedge (d \cdot (x_0 + x_1) = 0)\Big)\Big| \leq \mu(\lambda)\ . \tag{4.2}$$

In words, the assumption is that no *quantum* polynomial-time algorithm can *simultaneously* return an element x in the domain of f and an equation d such that, letting $\{x_0, x_1\}$ be the two preimages of $f_{pk}(x)$ under f_{pk} it holds that $d \neq 0^m$ and $d \cdot (x_0 + x_1) = 0$. Note that although we required $\{f_{pk}\}$ to be claw-free, this requirement is stronger, since any algorithm that can find a claw (x_0, x_1) can be used to break (4.2).

Remark 4.3. Assumption 1 is called *adaptive hardcore bit* for the following reason. Given a function f a hardcore bit for f is a 1-bit function h such that given $f(x)$ (but not x) it is hard to predict $h(x)$. Here, the hardcore bit that underlies the assumption is the function $h(x) = d \cdot (x_0 + x_1)$ for any $d \neq 0^m$: the Goldreich-Levin theorem implies that if f is indeed claw-free then it is hard to predict $h(x)$ for a *random* d. The “adaptive” qualifier refers to the fact that in (4.2) we allow the adversary $\mathcal{A}$ itself to select the equation d without requiring that this equation is uniformly distributed (we will see why this is needed in the next section); in particular $\mathcal{A}$ may return always the same d, and this invalidates the classic Golreich-Levin argument. This makes the property harder to satisfy, because more power is given to the adversary. (Note in particular that we had to explicitly require $d \neq 0^m$, as otherwise there is an easy adversary that always succeeds.)

4.3 A computational test for a qubit

We conclude the lecture by giving a “proof of concept” that it is possible to test a qubit based on computational assumptions. Our presentation is a simplified version of the protocol and analysis from [BCM$^+$18]. For a related approach, see [CCKW19].

The main assumption that we make is that there is a function family $\{f_{pk}\}$ that satisfies the hardcore bit assumption, Assumption 1. In fact we will need a little bit more. Let’s summarize the requirements as follows:

(F.1) There is a 2-to-1 claw-free function family $\mathcal{F} = \{f_{pk}\}$ equipped with an efficient key generation procedure $\textsc{Gen}(1^\lambda)$ such that for each key pk the function f_{pk} can be evaluated efficiently.

(F.2) The function family $\mathcal{F}$ satisfies the adaptive hardcore bit assumption, Assumption 1.

(F.3) $\mathcal{F}$ is equipped with a trapdoor: in addition to pk, $\textsc{Gen}(1^\lambda)$ returns a trapdoor td such that given pk, td and any y in the range of f_{pk} it is possible to efficiently recover the two preimages x_0 and x_1 of y.

(F.4) For any pk and any y in the range of f_{pk} the two preimages of y are labelled ‘x_0’ and ‘x_1’ using some canonical efficient procedure. That is, given a key pk and an x in the domain of f_{pk} it is possible to efficiently determine if x is the ’x_0’ or the ’x_1’ preimage of $y = f(x)$. Let $b : \{0,1\}^m \to \{0,1\}$ be this labeling procedure; b may depend on pk.

Let us fix a function family $\mathcal{F}$ satisfying the assumptions (**F.1**) to (**F.4**). We give a protocol based on $\mathcal{F}$. The protocol describes the interaction between a classical polynomial-time verifier and a (possibly quantum) polynomial-time prover. Here, the input to both parties is the security parameter λ; when we refer to PPT or QPT we mean with respect to λ. The protocol is described in Figure 4.1. For future reference we refer to it as “protocol $\mathfrak{Q}$.”

Let $\mathcal{F}$ be a function family and $\lambda \in \mathbb{N}$ a security parameter.

1. The verifier generates $(pk, td) \leftarrow \textsc{Gen}(1^\lambda)$. It sends pk to the prover.
2. The prover returns $y \in \{0,1\}^m$, where $m = m(\lambda)$.
3. The verifier selects a uniformly random challenge $c \leftarrow_R \{0,1\}$ and sends c to the prover.
4. (a) *(pre-image test:)* In case $c = 0$ the prover is expected to return an $x \in \{0,1\}^m$. The verifier accepts if and only if $f_{pk}(x) = y$.

   (b) *(equation test:)* In case $c = 1$ the prover is expected to return a $d \in \{0,1\}^m$. The verifier uses td to determine the two preimages (x_0, x_1) of y by f_{pk}. She accepts if and only if $d \cdot (x_0 + x_1) = 0$.

Figure 4.1: Protocol $\mathfrak{Q}$, the computational test for a qubit. The protocol is parametrized by a function family $\mathcal{F}$ satisfying assumptions (**F.1**) to (**F.4**).

Theorem 4.4. *Let $\mathcal{F}$ satisfy the assumptions (**F.1**) to (**F.4**). Then the following hold for protocol $\mathfrak{Q}$.*

- *(Completeness:) There is a QPT prover P which succeeds with probability 1 in the protocol.*
- *(Soundness:) Suppose that a QPT prover P succeeds with probability 1 in the protocol. Then P has a (near-perfect) qubit.*

As we are now accustomed to, we note the slightly informal nature of the theorem and make a few comments:

- First of all, we make explicit the computational assumption: combining Assumption 1 with the requirement that the prover P is QPT effectively means that we are assuming that P does not "have the ability" to violate (4.2). Slightly more formally, in the proof we will show that if P *does not* "have a qubit" then it can be used to construct an adversary $\mathcal{A}$ that violates (4.2). Note also that in the soundness case it should be assumed that P is in fact a family of $\{P_\lambda\}$, one for each possible choice of λ, that can be uniformly generated from λ (i.e. there is a classical Turing machine that takes 1^λ as input and returns a description of a family of circuits that can be used to implement P_λ).
- Second, we ought to be a little more precise as to how P's qubit is specified. The two observables X and Z that define it will be derived from the two measurements that P makes based on the challenges $c = 0$ or $c = 1$. Since these measurements in general have outcomes in $\{0,1\}^n$ some post-processing will be required. Interestingly, the post-processing for the X observable will not be efficient, in the sense that it will require knowledge of td. So, our proof will show that there exists two anti-commuting observables on the Hilbert space of P that can be defined from P's operations *and some classical post-processing*. Since the post-processing is classical we can nevertheless claim in good faith that the "qubit" is located on the prover's space, as we are not injecting any external "quantumness" in it.
- Third, we make the usual comment regarding the assumption that the prover succeeds with probability 1: this assumption is, of course, unrealistic. As in other protocols that we have seen so far the assumption can be lifted at the cost of some amount of work. We will discuss this in more detail when we build on the present protocol to construct a more complex protocol for verifying an entire quantum computation in the next few lectures.

- Finally, an explanation is in order regarding the "(near-perfect)" qualifier. This is an unavoidable consequence of the fact that the protocol relies on a computational assumption. Indeed, consider the following possible behavior for the prover. The prover first devotes a small amount of time to trying their luck at breaking the underlying computational assumption (in our case, the prover could randomly generate candidate trapdoors td' and check if they allow it to invert the function f_{pk}). If the prover succeeds then it can pass in the protocol without manipulating any quantum state, using the fake td' to find a claw that allows it to answer both types of challenges. If it does not succeed then it behaves honestly in the protocol. Such a prover succeeds with probability 1, but the measurement operators associated with its answers have a part that is "classical" and from which we have no hope of extracting a qubit.

Proof of Theorem 4.4. The completeness part of the theorem is clear. In the first phase the prover proceeds exactly as in Simon's algorithm to obtain the state (4.1). In the second phase, it measures the preimage register in the standard basis in case $c = 0$ and in the Hadamard basis in case $c = 1$, returning the n-bit outcome obtained as its answer. This prover is always accepted with probability 1 in the protocol.

To show the soundness part of the theorem we start with the usual (and, here, crucial) modeling step.

Step 1: Modeling Since we will not need to model the prover's actions in the first phase of the protocol in detail we directly give a name to the state of the prover at the end of step 2; let it be $|\psi\rangle \in \mathcal{H}_{\mathsf{P}}$. This state depends on pk as well as on y; for clarity we suppress this dependence from the notation. Moreover, in general $|\psi\rangle$ may be a mixed state, and we represent it as a pure state for convenience only; in general one could assume that we included a register E to denote an "environment" that holds a purification $|\psi\rangle_{\mathsf{PE}}$ of a general $\rho \in \mathrm{D}(\mathcal{H}_{\mathsf{P}})$.

At the second stage of the protocol the prover is given a challenge $c \in \{0, 1\}$ and tasked with responding with an n-bit string, x or d depending on the challenge. In general, x is obtained by performing a POVM $\{\Pi_x\}$ on the prover's entire space, and similarly d is the outcome of a POVM $\{M_d\}$.[6] We make the following observations that allow us to simplify the presentation of these POVM:

- Without loss of generality both Π and M are projective measurements. This is because we can enlarge P and add sufficiently many ancilla qubits initialized to $|0\rangle$ so as to apply Naimark's theorem.
- Without loss of generality, assume that the prover has access to an m-qubit register X initialized to $|0^m\rangle$.
- Without loss of generality, assume that Π is obtained by first applying a unitary transformation U_0 on $\mathcal{H}_{\mathsf{X}} \otimes \mathcal{H}_{\mathsf{P}}$ followed by a standard basis measurement of $\mathcal{H}_{\mathsf{X}} \simeq (\mathbb{C}^2)^{\otimes m}$. Any projective measurement can be put in this form by letting U_0 be any unitary extension of the map
$$|0\rangle|\psi\rangle \in \mathcal{H}_{\mathsf{X}} \otimes \mathcal{H}_{\mathsf{P}} \mapsto \sum_x |x\rangle\sqrt{\Pi_x}|\psi\rangle ,$$
as this map is easily verified to be an isometry on $|0\rangle_{\mathsf{X}} \otimes \mathcal{H}_{\mathsf{P}}$.
- Similarly, without loss of generality assume that M is obtained by first applying a unitary transformation U_1 on $\mathcal{H}_{\mathsf{X}} \otimes \mathcal{H}_{\mathsf{P}}$ followed by a Hadamard basis measurement of $\mathcal{H}_{\mathsf{X}}$.
- Without loss of generality assume that $U_0 = \mathrm{Id}$. This is because we can always redefine the prover's state at the end of step 2 to be $|\psi'\rangle = U_0|\psi\rangle$, in which case $U_0' = \mathrm{Id}$ and $U_1' = U_1 U_0^\dagger$. Since $U_0 = \mathrm{Id}$, we simply use U to denote U_1.

[6] We do not need to explicitly mark any dependence of Π or M on pk and y, because without loss of generality the prover has kept a classical copy of these strings in its quantum state $|\psi\rangle$, which can be used as a classical control by both Π and M.

We now introduce observables Z and X on $\mathcal{H}_{\mathsf{X}}$ associated with the prover. For Z, we define it to be

$$Z = \sum_{x\in\{0,1\}^m} (-1)^{b(x)}|x\rangle\langle x| \,, \tag{4.3}$$

where $b : \{0,1\}^n \to \{0,1\}$ is the function from assumption (**F.4**). Z is efficiently computable since b is. For X, we define it to be

$$X = \sum_{d\in\{0,1\}^m} (-1)^{d\cdot(x_0+x_1)}\, U^\dagger\big(H_{\mathsf{X}}^{\otimes m}\otimes \mathrm{Id}_{\mathsf{P}}\big)^\dagger\big(|d\rangle\langle d|_{\mathsf{X}}\otimes \mathrm{Id}_{\mathsf{P}}\big)\big(H_{\mathsf{X}}^{\otimes m}\otimes \mathrm{Id}_{\mathsf{P}}\big)U \,, \tag{4.4}$$

where x_0 and x_1 are the two preimages under f_{pk} of the string y returned by P at step 2. (There is an observable X for each possible string y, but we suppress this dependence from the notation for clarity.) Note that X is *not* efficient, because we are not assuming that determining $x_0 + x_1$ from y is efficient in general. However, X can be computed in a straightforward manner by applying the prover's efficient measurement $\{M_d\}$ followed by (non-efficient) classical post-processing. (We insist on this point to clarify that our definition is not injecting "quantumness" artificially.) Informally, X can be thought of as the observable that determines if the equation d returned by the prover on challenge $c = 1$ is correct or not. In particular, later we will use that for a prover that always succeeds to a challenge $c = 1$ we have $X|\psi\rangle = |\psi\rangle$, i.e. $|\psi\rangle$ is a $+1$ eigenstate of X.

Step 2: Establishing a qubit The goal for the remainder of the proof is to show that $(|\psi\rangle, Z, X)$ form a qubit, i.e. that the two observables X and Z anticommute on $|\psi\rangle$. Informally, this is because if X and Z were jointly measurable then they could be used to simultaneously obtain a preimage of y and a valid equation d in $x_0 + x_1$, thereby violating (**F.2**). We proceed with the details. The heart of the proof is the following claim.

Claim 4.5. *For any $b \in \{0,1\}$,*

$$\big|\langle\psi|Z_b X Z_b|\psi\rangle\big| = \mathrm{negl}(\lambda)\,,$$

where $Z_b = (\mathrm{Id} + (-1)^b Z)/2$, $\mathrm{negl}(\lambda)$ denotes some negligible function of λ, and the expression on the left should be understood on average over $pk \leftarrow \textsc{Gen}(1^\lambda)$ and the distribution of y as returned by P in the protocol.

Proof. We do the proof for the case $b = 0$, the other case being similar. Suppose for contradiction that there is a polynomial $q : \mathbb{N} \to \mathbb{R}_+$ such that

$$\big|\langle\psi|Z_0 X Z_0|\psi\rangle\big| \geq \frac{1}{q(\lambda)} \tag{4.5}$$

for infinitely many values of λ. We use this assumption to construct an adversary in (4.2). The adversary proceeds as follows. Given as input 1^λ and pk the adversary first executes the first phase of the prover, obtaining an outcome y and a state $|\psi\rangle$. Then, the adversary measures the m qubits in register X in the computational basis to obtain a value $x \in \{0,1\}^m$. If $b(x) = 0$ then the adversary applies the unitary V and measures register X (again) in the Hadamard basis to obtain $d \in \{0,1\}^n$. The adversary returns the pair (x, d). If $b(x) = 1$ then the adversary chooses $d \in \{0,1\}^m$ uniformly at random and returns d. Since the prover P and b are both efficient, $\mathcal{A}$ is efficient.

Note that this adversary does something "unusual" in the sense that it sequentially applies two operators that the prover would never have applied simultaneously in the protocol. It is to make sense of this sequential application that we made the structural simplifications at the start of the proof. Let's analyze the success probability of $\mathcal{A}$ by using (4.5). There are two cases. Suppose first that the adversary obtains an x such that $b(x) = 0$. Then since P is assumed to succeed with probability 1 in case $c = 0$, we know that necessarily $x = x_0$, and moreover prior to the measurement the support of $|\psi\rangle$ on X contained only the two values $|x_0\rangle$ and $|x_1\rangle$ (as otherwise there would be a chance that the prover returns an invalid preimage to the challenge $c = 0$). Thus by definition of Z in (4.3) the post-measurement state is $Z_0|\psi\rangle$ (suitably re-normalized). The probability that $\mathcal{A}$ obtains $b(x) = 0$ and then a correct equation is then, by definition of X in (4.3) and $X_0 = \frac{1}{2}(\mathrm{Id} + X)$, precisely

$$\langle\psi|Z_0X_0Z_0|\psi\rangle \;=\; \frac{1}{2}\big(\langle\psi|Z_0|\psi\rangle + \langle\psi|Z_0XZ_0|\psi\rangle\big)\;.$$

For the second case assume that $\mathcal{A}$ obtains an x such that $b(x) = 1$. In this case it returns a uniformly random equation; since $x_0 + x_1 \neq 0$ this has probability exactly $\frac{1}{2}$ of being correct. Overall, the adversary's success probability is

$$\frac{1}{2}\langle\psi|Z_1|\psi\rangle + \frac{1}{2}\big(\langle\psi|Z_0|\psi\rangle + \langle\psi|Z_0XZ_0|\psi\rangle\big) \;=\; \frac{1}{2} + \frac{1}{2}\langle\psi|Z_0XZ_0|\psi\rangle\;,$$

where the equality uses $Z_0 + Z_1 = \mathrm{Id}$ to combine the first two terms. Using (4.5), this violates (4.2). □

To conclude the proof of the theorem we need the following simple calculation.

Claim 4.6. *Let X, Z be any two binary observables on $\mathcal{H}$. Then*

$$\frac{1}{4}\{X,Z\}^2 \;=\; XZ_0XZ_0 + Z_1XZ_1X\;.$$

Proof. This can be verified by direct calculation. Using that X and Z are Hermitian and square to identity we get by expanding the square

$$\{X,Z\}^2 = 2 + XZXZ + ZXZX\;. \tag{4.6}$$

Expanding $Z = Z_0 - Z_1$,

$$ZXZ = Z_0XZ_0 + Z_1XZ_1 - Z_0XZ_1 - Z_1XZ_0\;.$$

Moreover, using $Z_0 + Z_1 = \mathrm{Id}$ it follows that

$$Z_0XZ_0 + Z_1XZ_1 + Z_0XZ_1 + Z_1XZ_0 = X$$

Putting the two equations together, $ZXZ = 2\big(Z_0XZ_0 + Z_1XZ_1\big) - X$. Plugging back into (4.6) and using $X^2 = \mathrm{Id}$ proves the claim. □

Combining Claim 4.6 and Claim 4.5 we make the following calculation:

$$\begin{aligned}\frac{1}{4}\big\|\{X,Z\}|\psi\rangle\big\|^2 &= \langle\psi|XZ_0XZ_0|\psi\rangle + \langle\psi|Z_1XZ_1X|\psi\rangle\\ &= \langle\psi|Z_0XZ_0|\psi\rangle + \langle\psi|Z_1XZ_1|\psi\rangle\\ &= \mathrm{negl}(\lambda)\;,\end{aligned}$$

where to obtain the second line we used that $X|\psi\rangle = |\psi\rangle$ since the prover is assumed to succeed with probability 1 in the protocol (and hence always return a correct equation). This shows that $(|\psi\rangle, Z, X)$ is a near-perfect qubit, completing the proof. □

Lecture 5

Delegating Quantum Computations

So far we have been entirely focused on the problem of certifying a "qubit", i.e. that a certain device to which the experimentalist, or verifier, has access to and is willing to make simple assumptions about (the device has two spatially isolated components/the device is computationally bounded) is, at some point in its execution, making a pair of anti-commuting measurements.

Our goal in the next three lectures is to go beyond the certification of a single qubit, to the verification that the device implements an entire quantum computation of the verifier's choice. In this lecture we define this problem of *delegating quantum computations* to an untrusted party and give an overview of existing approaches. Towards the end of the lecture we describe a protocol due to Fitzsimons and Morimae [MF16] that involves quantum communication from the prover to the verifier. In subsequent lectures we combine that protocol with the computational qubit test from the previous lecture and a few additional ideas to obtain a purely classical protocol due to Mahadev [Mah18].

5.1 Problem statement

5.1.1 Quantum circuits and the class BQP

For us, a quantum circuit $\mathcal{C}$ is specified by an integer n and an ordered sequence of elements of the form (G, i, j) where $G \in \{H, CNOT, T\}$ and $i, j \in \{1, \ldots, n\}$. Letting

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\,, \qquad CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}\,, \qquad T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}\,, \tag{5.1}$$

the circuit $\mathcal{C} = ((G_1, i_1, j_1), \ldots, (G_T, i_k, j_k))$ represents the unitary U on $(\mathbb{C}^2)^{\otimes n}$ obtained as $U = U_T \cdots U_1$ where for all $t \in \{1, \ldots, T\}$, U_t acts as the unitary associated with G_t in (5.1) on qubits i_t and j_t and as identity on the other qubits. (In case $G_t \in \{H, T\}$ it is required that $i_t = j_t$.) The Solovay-Kitaev theorem shows that any n-qubit unitary can be arbitrary well-approximated, in operator norm, by the unitary derived from a circuit; however, the size of the circuit may (in fact, must) in general grow exponentially fast with n. Those unitaries that can be represented by small circuits are called "efficient".

[0] Some of the material for this lecture is taken from an overview of Mahadev's result written for a mathematical audience and published in the Bulletin of the AMS [Vid20]. Some of it is reproduced from lecture notes prepared for a winter school at UCSD: `http://cseweb.ucsd.edu/~slovett/workshops/quantum-computation-2018/`.

Given a quantum circuit $\mathcal{C}$ acting on n qubits and $x \in \{0,1\}^m$ for some $m \leq n$ we say that "$\mathcal{C}$ accepts input x with probability p" if the probability of obtaining the outcome 1 after a measurement in the computational basis of the first qubit of the n-qubit state obtained by applying the unitary $\mathcal{C}$ to the input state $|x\rangle|0^{n-m}\rangle$ is p.

Definition 5.1. We say that a promise language $L = (L_{yes}, L_{no})$ is in BQP if there exists a family of polynomial-time generated quantum circuits $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ such that for all integer n and $x \in \{0,1\}^n$,[1]

- (Completeness:) If $x \in L_{yes}$ then $\mathcal{C}_n$ accepts x with probability at least $\frac{2}{3}$;
- (Soundness:) If $x \in L_{no}$ then $\mathcal{C}_n$ accepts x with probability at most $\frac{1}{3}$.

Note the requirement that the family $\{\mathcal{C}_n\}$ is polynomial-time generated. This means that there exists a classical Turing Machine that on input 1^n runs in time $\text{poly}(n)$ and returns a description of $\mathcal{C}_n$ as a sequence of gates taken from a fixed universal set—here we use (5.1), but the specific choice will not matter for us.

The definition of BQP sets arbitrary values $2/3$ and $1/3$ for the completeness and soundness parameters. Error amplification works just as for the case of BPP, by repeating the circuit sequentially. This requires intermediate measurements, but it is not hard to show that these can be postponed till the end of the computation by the use of ancilla qubits and CNOT gates. As a result, any choice of a, b such that $a - b > \text{poly}^{-1}(n)$ gives the same definition: for any such a, b, and for any fixed polynomial q, $\text{BQP}(a,b) = \text{BQP} = \text{BQP}(1 - 2^{-q}, 2^{-q})$.

Exercise 5.1. Show that BQP is included in PP, the class of languages for which there exists a probabilistic Turing machine that accepts YES inputs with probability $> 1/2$, and rejects NO inputs with probability $> 1/2$. (Hint: first show inclusion in PSPACE by giving space-efficient implementations of basic linear algebra operations. Inclusion in PP follows from similar arguments, but is a bit more delicate.)

The class PP lies outside of the polynomial hierarchy. The most commonly-held belief is that the intersection of BQP and PH is non-trivial: it is neither BPP, nor PH itself. Recently Raz and Tal [RT19] showed that an oracle problem introduced by Aaronson [Aar10] is in BQP but not in PH.

Recall the notion of interactive proof system that we introduced informally in Section 2.2. We end this section by defining a family of complexity classes associated with interactive proof systems.

Definition 5.2 (Adapted from [AG17])**.** Given complexity classes $\mathcal{P}$ and $\mathcal{Q}$, $\text{IP}[\mathcal{P},\mathcal{Q}]$ is the class of (promise) languages L such that there is a polynomial-time Turing machine M that on input 1^n returns the description of classical circuits for the verifier V_n in an interactive protocol with a prover P such that

- (Completeness:) There is a family of provers $\{P_n\}_{n\in\mathbb{N}}$ that lie in the class $\mathcal{P}$ such that for all $x \in L_{yes}$ the interaction of $V_{|x|}$ and $P_{|x|}$ on common input x accepts with probability at least $\frac{2}{3}$.
- (Soundness:) For any family of provers $\{P_n\}_{n\in\mathbb{N}}$ that lie in the class $\mathcal{Q}$, for all $x \in L_{no}$ the interaction of $V_{|x|}$ and $P_{|x|}$ on common input x accepts with probability at most $\frac{1}{3}$.

When the classes $\mathcal{P}$ and $\mathcal{Q}$ coincide we simply write $\text{IP}[\mathcal{P}]$ for $\text{IP}[\mathcal{P},\mathcal{P}]$. We use the standard notation $\text{IP} = \text{IP}[\text{BPP}, \text{ALL}]$ with ALL the class of all languages (i.e. soundness is proves without any restriction on the prover).

The definition is slightly informal, because for some classes $\mathcal{P}$ it may not be clear what it means for the prover to lie in $\mathcal{P}$. For us the meaning will always be clear from context, as $\mathcal{P}$ and $\mathcal{Q}$ will always be either BPP, BQP or ALL.

[1] Note that in general, $\mathcal{C}_n$ may act on $\text{poly}(n)$ qubits, the first n of which are by convention destined to receive the input x and the first of which also serves as output qubit.

5.1.2 Delegating quantum computations

The fact that BQP is not (believed to be) in NP implies that in general we do not expect there to exist classically verifiable proofs for the correctness of an arbitrary quantum computation. This poses a challenge: as we see quantum computers emerging, how will we test their predictions? This is a practical problem — will anyone trust the "quantum cloud" — but also a philosophical one — is quantum mechanics a testable theory? (For more on this, see [AV13].)

Not all is lost. What we *do* know is that BQP is included in PSPACE, the class of languages that can be decided using polynomial space (and arbitrary time); in fact Exercise 5.1 asked you to show a stronger statement. And even though it is not a trivial result, it is known that PSPACE $=$ IP. So all languages in BQP have *classical* interactive proofs, with an efficient classical verifier! Unfortunately there is a major caveat to this observation. The proof that PSPACE is in IP is based on the classical SUM-CHECK protocol, which in general requires the server to execute PSPACE-complete computations (essentially, the server has to compute exponentially large sums in order to determine answers that will satisfy the client). (For an exposition of the proof we refer to the book [AB09].)

So, even though a protocol exists, it is unknown if there is such a protocol in which a honest server is only required to have the power of BQP. Today this is a major open question:

Open Question 5.3. Is $\mathrm{BQP} \subseteq \mathrm{IP}[\mathrm{BQP}, \mathrm{ALL}]$? In words, do all languages in BQP have single-server interactive proofs in which the client has the power of BPP and for which completeness holds with a BQP server and soundness holds against any server?

In [ACGK17] some partial impossibility results are shown for the case of blind delegation protocols. If the inequality $\mathrm{BQP} \subseteq \mathrm{IP}[\mathrm{BQP}, \mathrm{ALL}]$ does hold it is likely to require protocols with polynomially many rounds of interaction, because constant-round protocols lie in AM and BQP is not believed to be in AM. Such protocols are known where completeness holds for provers that require more power than BQP but not necessarily the entire power of PSPACE; see e.g. [AG17]. If, however, one allows slightly more power to the verifier then there are scenario in which the question is known to have a positive answer:

1. The client has access to a limited quantum computer, such as the ability to prepare single qubits in arbitrary states and send them to the server, or receive single qubits from the server and make simple measurements on them;

2. The client is allowed to interact with multiple quantum servers sharing entanglement.

The question as formulated above asks for *verifiable* delegation: given a quantum circuit (deciding some BQP language L), is there a protocol that allows a classical client to extract the outcome of the circuit from a BQP server, in a way that any cheating server, attempting to convince the client of the wrong outcome, will be detected? A second desirable property of a delegation protocol is *blindness*: while the client would like to learn the valid outcome of her circuit, she might not want to disclose the particular circuit or input she is interested in to the server. This is a distinct property from verifiability; in particular, one may ask for blindness in the "honest-but-curious" model, where verifiability is trivial. The following definition introduces these properties slightly more formally.

Definition 5.4 (Delegated computation)**.** In the task of delegated computation, a client (sometimes called the *verifier*) has an input $(x, \mathcal{C})$, where x is a classical string and $\mathcal{C}$ the classical description of a quantum circuit. The client has a multiple-round interaction with a quantum server (sometimes also called *server*). At the end of the interaction, Alice either returns a classical output y, or she aborts. A protocol for delegated computation is called:

- *Correct* if whenever both the client and the server follow the protocol, with high probability Alice accepts (she does not abort) and $y = \mathcal{C}(x)$. (This property is sometimes called *completeness*.)
- *Verifiable* if for any server deviating from the protocol, the client either aborts or returns $y = \mathcal{C}(x)$. (This property is analogous to what we have been calling *soundness*.)
- *Blind* if for any server deviating from the protocol, at the end of the protocol the server has no information at all about the client's input $(x, \mathcal{C})$.

The definition remains rather informal. For example, how should we formalize the "information" that the server has at the end of the computation? This can be rather delicate, especially once one starts taking into account a small chance ε of deviation from the perfect properties. A precise definition satisfying all the desired properties (universal composability in particular) would take us too far. Such a definition was given using the framework of *abstract cryptography* in [DFPR14].

The informal definition will be sufficient for our purposes. Note that in spite of being rather similar neither of the properties of verifiability or blindness is known to directly implies the other. In practice verifiability often follows from blindness by arguing, using "traps", that if a protocol is already blind then the server's trustworthiness can be tested by making it run "dummy" computations for which Alice already knows the output, without the server being able to distinguish whether it is asked to do a real or dummy computation. We will see an example of this technique later on.

Open Question 5.5. Is there a general transformation from any protocol satisfying blindness, to a protocol satisfying both verifiability and blindness? See [Mor18] for how to achieve this by making use of post-hoc verification (cf. Section 5.2), and [KMW17] for another approach.

Remark 5.6. The problem of delegating computation is interesting even for classical computation. In this case the client herself could directly execute the classical circuit $\mathcal{C}$. But it makes sense to be even more demanding, and seek protocols where the client is super-efficient: the best we could hope for is a client that runs in time *linear* in the input length, and independent of the size of the circuit. In addition, we would like the overhead for the server to be as small as possible, so that the honest behavior requires a server effort of the same order as the size of the circuit, $|\mathcal{C}|$. This kind of interactive proofs are called *doubly efficient* interactive proofs [GKR08]. The paper [RRR16] shows how to achieve such proofs with client runtime that is linear in the input length, polynomial in the space required by $\mathcal{C}$, and polylogarithmic in $|\mathcal{C}|$. If one is willing to make computational assumptions (essentially, subexponential LWE) then even more efficient delegation is possible [KRR14], with client runtime that is linear in the input size and poly-logarithmic in $|\mathcal{C}|$.

These results usually do not put emphasis on the requirement of blindness: they focus on verifiability alone. One reason for this is that blindness is "trivially solved" by employing homomorphic encryption [Gen09]. This, however, requires computational assumptions, and induces significant computational overhead.

5.1.3 Approaches to delegating quantum computation

There are three main types of approaches: *prepare-and-send* (the client has ability to prepare single-qubit states and send them to the server), *receive-and-measure* (the client has the ability to receive single-qubit states from the server and measure them), and *two-server* (the client interacts classically with two spatially isolated servers). A great recent survey describing these approaches in detail is [GKK19]. Here we focus on a specific protocol of receive-and-measure type, that we will later build on to obtain a protocol with a classical verifier, under computational assumptions on the prover.

5.2 The Fitzsimons-Morimae protocol

We describe the receive-and-measure protocol from [MF16], as it will form the basis for the Mahadev protocol.

5.2.1 The circuit-to-Hamiltonian reduction

The Cook-Levin theorem showing NP-completeness of the 3SAT problem is based on what could be called a "circuit-to-formula" reduction: given a classical circuit, the computation performed by the circuit on some input is represented as a "tableau" such that the property of being a valid tableau can be encoded in a formula whose variables represent the state of any given wire in the circuit and whose constraints enforce correct propagation of the gates of the circuit.

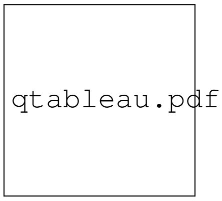

Figure 5.1: Two different ways to create a tableau from a quantum circuit. The state $|\psi_{tableau}\rangle$ is the tensor product of the state of the circuit at each time step. The state $|\psi_{hist}\rangle$ is their superposition, indexed by a clock register that goes from $|0\rangle$ to $|T\rangle$.

For quantum circuits the idea of a tableau of the computation is less straightforward. The most direct analogue is to consider the juxtaposition of the quantum state of a T-gate circuit at each step of the computation, i.e. the tensor product $|\psi_0\rangle \otimes \cdots \otimes |\psi_T\rangle$ of the states $|\psi_i\rangle$ obtained by executing the circuit from scratch and stopping after i gates have been applied. While this is a well-defined $n(T+1)$-qubit quantum state (see Figure 5.1) the property of being a valid "quantum tableau" cannot be enforced using *local* constraints! The reason is subtle, and has to do with the possible presence of entanglement at intermediate steps of the computation. Indeed, there are quantum states that are very different, in the sense that they are perfectly distinguishable by some *global* observable, yet cannot be distinguished at all by any *local* observable, that would act on at most, say, half the qubits. An example is given by the two n-qubit "cat" (named after the homonymous animal) states

$$|\psi_\pm\rangle = \frac{1}{\sqrt{2}}\big(|0\cdots 0\rangle \pm |1\cdots 1\rangle\big) .$$

The two states $|\psi_+\rangle$ and $|\psi_-\rangle$ are easily seen to be orthogonal, so that they can be perfectly distinguished by a measurement. But it is an exercise to verify that for any observable that acts on at most $(n-1)$ of the n qubits, both states give exactly the same expectation value. (Informally, this is because any measurement on a strict subset of the qubits of the state necessarily destroys the coherence; the only relevant information, the $\pm$ sign, is encoded "globally" and cannot be accessed locally.) Note that this is a uniquely quantum phenomenon: if two classical strings of bits have each of their bits equal, one pair at a time, then the strings are "globally" identical. Not so for quantum states.

So naïve tableaus will not do. In the late 1990s Alexei Kitaev introduced a very powerful idea that provides a solution. Kitaev's idea is to replace the juxtaposition of snapshot states by their *superposition* (see Figure 5.1). A special ancilla system, called the "clock", is introduced to index different elements of the superposition. Thus, instead of defining a tableau as $|\psi_0\rangle \cdots |\psi_T\rangle$, Kitaev considers the state

$$|\psi_{hist}\rangle \;=\; \frac{1}{\sqrt{T+1}} \sum_{t=0}^{T} |t\rangle|\psi_t\rangle \;. \tag{5.2}$$

Note that this takes less qubits to store, but this is not the important point. Kitaev showed that, assuming the clock register is encoded in unary, it is possible to check the correct propagation of every step of the circuit directly on this superposition by only applying local observables: there is a set of observables H_{in} that checks that $|\psi_0\rangle$ has the right format; a set of observables H_{prop} that checks propagation of the circuit, and an observable H_{out} that checks that the output qubit of the circuit is in the right state. (In addition, there is a term H_{clock} that checks that the clock register is well-formed, i.e. contains the representation of an integer in unary. This can be done locally by penalizing configurations of the form "$\cdots 10 \cdots$".) The key point that makes this possible is that, while equality of quantum states cannot be decided locally when the states are juxtaposed, it becomes possible when they are given in superposition. As an exercise, we can verify that a measurement of the first qubit of the state

$$|\psi_{SWAP}\rangle \;=\; \frac{1}{\sqrt{2}}\big(|0\rangle|\psi_0\rangle + |1\rangle|\psi_1\rangle\big)$$

in the Hadamard basis $\{|+\rangle, |-\rangle\}$ returns the first outcome with probability exactly $\frac{1}{2}(1 + |\langle\psi_0|\psi_1\rangle|^2)$. With more work, replacing the use of gadgets in the classical Cook-Levin reduction by techniques from perturbation theory, it is possible to write the resulting observables as a linear combination of local terms that all take a particularly simple form. The result is the following theorem from [CM16].

Theorem 5.7. *For any integer $n \geq 1$ there are $n' = \mathrm{poly}(n)$, $a = a(n)$ and $\delta \geq 1/\mathrm{poly}(n)$ such that the following holds. Given a T-gate quantum circuit $\mathcal{C} = ((G_1, i_1, j_1), \ldots, (G_T, i_T, j_T))$ acting on n qubits, such that $T = \mathrm{poly}(n)$, and an input x for the circuit, there exist efficiently computable real weights $\{J_{ij},\, i, j \in \{1, \ldots, n'\}\}$ such that $|J_{ij}| \leq 1$ for all i, j and moreover if*

$$H_{\mathcal{C}} \;=\; -\sum_{i,j} \frac{J_{ij}}{2} \left(\sigma_{X,i}\sigma_{X,j} + \sigma_{Z,i}\sigma_{Z,j}\right) \,, \tag{5.3}$$

where $\sigma_{X,i}$ and $\sigma_{Z,j}$ denote single-qubit Pauli X and Z operators acting on the i-th and j-th qubit respectively, then:

- *(Completeness) If the circuit $\mathcal{C}$ accepts its input x with probability at least $2/3$, then the smallest eigenvalue of $H_{\mathcal{C}}$ is at most a;*
- *(Soundness) If the circuit $\mathcal{C}$ accepts its input x with probability at most $1/3$, then the smallest eigenvalue of $H_{\mathcal{C}}$ is at least $a + \delta$.*

Remark 5.8. It is possible to modify Theorem 5.7 so that the completeness and soundness statements specify that "if there exists a state $|\phi\rangle$ such that $\mathcal{C}$ accepts on input $(x, |\phi\rangle)$ with probability at least $2/3$..." and "if there does not exist a state $|\phi\rangle$ such that $\mathcal{C}$ accepts on input $(x, |\phi\rangle)$ with probability greater than $1/3$..." respectively. Thus, Theorem 5.7 can be adapted to show that the problem of estimating the minimal energy of a Hamiltonian of the form (5.3) is a QMA-complete problem.

Theorem 5.7 provides us with a roadmap for the verification of quantum circuits: it is sufficient to verify the *existence* of a quantum state that yields certain statistics, when some of its qubits are measured in the computational (σ_Z observable) or Hadamard (σ_X observable) basis. The reason this can be considered progress is that we no longer need to check the time evolution of a quantum state under a quantum cicuit; it is sufficient to collect measurement statistics and estimate the "energy" $\langle\psi|H|\psi\rangle$. In particular, the theorem readily leads to a verification protocol in a model where the prover has a full quantum computer, and the verifier only has a limited quantum device — namely, a one-qubit memory, together with the ability to measure the qubit using either the σ_X or σ_Z observables.

5.2.2 The protocol

Such a verification protocol was introduced by Fitzsimons and Morimae and refined in a paper with Hadjušek. The protocol is summarized in Figure 5.2. In the protocol, the prover is required to prepare a smallest eigenstate of the Hamiltonian $H_{\mathcal{C}}$ given in (5.3). While it may not be immediately obvious at the level of our description, it is possible to prepare such a "history state" (5.2) by executing a quantum circuit that is only mildly more complex than the original circuit $\mathcal{C}$.

Let $\mathcal{C}$ be a quantum circuit provided as input, and $H_{\mathcal{C}}$ the n'-qubit Hamiltonian obtained from $\mathcal{C}$ as in (5.3).

1. The verifier initializes a counter γ to 0. She executes the following interaction with the prover independently $N = \frac{C}{\delta^2}\binom{n'}{2}^2 \ln(1/\varepsilon)$ times, where C is a large enough universal constant:
   (a) The prover creates an eigenstate $|\psi\rangle$ of H with smallest eigenvalue.
   (b) The prover sends the qubits of $|\psi\rangle$ one by one to the verifier.
   (c) The verifier selects a measurement $W \in \{X, Z\}$ uniformly at random, and measures each qubit in the associated basis upon reception. Let $b_{W,i} \in \{-1, 1\}$ be the outcome for the i-th qubit.
   (d) The verifier selects $i \neq j \in \{1, \ldots, n'\}$ uniformly at random. She updates her counter $\gamma \leftarrow \gamma - J_{ij} b_{W,i} b_{W,j}$.
2. If $\frac{\gamma}{N}\binom{n'}{2} \leq a + \delta/2$ the verifier accepts the interaction. Otherwise, she rejects.

Figure 5.2: The Fitzsimons-Hadjuček-Morimae verification protocol, parametrized by a quantum circuit $\mathcal{C}$ and an accuracy parameter $\varepsilon > 0$.

We note that in the protocol, the verifier measures the qubits in a randomly chosen basis, and then selects a single pair (i, j) such that $J_{ij} \neq 0$ uniformly at random to update her counter. One could imagine small optimizations where e.g. a maximum matching of such pairs is measured at each step. Such optimizations only bring marginal improvements in efficiency of the protocol; moreover they complicate the extension to a classical verifier that we will see later. For this reason, we prefer to keep the simplest expression possible for the protocol.

Theorem 5.9. *Let $\mathcal{C}$ be a quantum circuit and $H_{\mathcal{C}}$ the Hamiltonian associated to it as in* (5.3)*. Let x be an input to the circuit $\mathcal{C}$ and $\varepsilon > 0$ a parameter for the protocol. Then the following hold:*

- *(Completeness:) If $\mathcal{C}$ accepts x with probability at least $2/3$, then there is a QPT prover that is accepted with probability at least $1 - \varepsilon$*

- *(Soundness:) If $\mathcal{C}$ accepts x with probability at most $1/3$, then any prover is accepted with probability at most ε.*

Note that in the theorem, the soundness statement does not place any computational assumption on the prover.

Proof. The key calculation that underlies the proof is the following.

Claim 5.10. *Let ρ be the density matrix that represents the mixture over the N n'-qubit states sent by the prover in the protocol (in general these states may be entangled). Then the expectation of γ/N is exactly*

$$\mathrm{E}\Big[\frac{\gamma}{N}\Big] \;=\; -\frac{1}{\binom{n'}{2}}\sum_{i\neq j}\frac{J_{ij}}{2}\,\mathrm{Tr}\big((\sigma_X^i\sigma_X^j+\sigma_Z^i\sigma_Z^j)\rho\big) \;=\; \frac{1}{\binom{n'}{2}}\,\mathrm{Tr}\big(H\rho\big)\;. \tag{5.4}$$

Moreover, for N chosen as in the protocol for a large enough choice of the constant C it holds that

$$\Pr\Big(\Big|\frac{\gamma}{N}\binom{n'}{2}-\mathrm{Tr}\big(H\rho\big)\Big|>\frac{\delta}{2}\Big)\;\leq\;\varepsilon\;. \tag{5.5}$$

Proof. For $t\in\{1,\ldots,n\}$ let G_t denote the product of the two outcomes $b_{W,i}$ and $b_{W,j}$ obtained by the verifier at step (c) of the protocol, where W, i and j are as sampled at step (d). Then the random variables $G_t\in\{-1,1\}$ are i.i.d. such that for each t, $\mathrm{E}[G_t]=\mathrm{Tr}(\sigma_W^i\sigma_W^j\rho)$, with W, i and j are the values sampled in step t. Since $\gamma=-\sum_t J_{ij}G_t$, averaging over those choices gives (5.4). Using $|J_{ij}|\leq 1$, by Hoeffding's inequality for any $s>0$

$$\Pr\big(\big|\gamma-\mathrm{E}[\gamma]\big|>s\big)\;\leq\;e^{-\frac{2s^2}{4N}}\;.$$

By choosing N sufficiently large with respect to $\binom{n'}{2}^2\delta^{-2}\ln(1/\varepsilon)$ we get (5.5). □

Based on Claim 5.10 the proof of Theorem 5.9 follows rather directly. For the completeness, we take $\rho=|\psi\rangle\langle\psi|$ such that $\langle\psi|H|\psi\rangle\leq a$, whose existence is guaranteed by the completeness case of Theorem 5.7. As noted above, this ρ can be prepared efficiently by a QPT prover. Using (5.5) it follows that this prover is accepted with probability at least $1-\varepsilon$. For the soundness, ρ is arbitrary. Using the soundness case of Theorem 5.7 it must be that $\mathrm{Tr}(H\rho)\geq a+\delta$, so that the conclusion follows again from (5.5). □

Even though the verifier's "quantumness" in this protocol is limited — she only needs to hold one qubit at a time — this capability is crucial for the analysis, as it is used to guarantee the "existence" of the state that is being measured: it allows us to meaningfully talk about "the state ρ whose first qubit is the first qubit received by the verifier; whose second qubit is the second qubit received by the verifier; etc.". These qubits are distinct, because the verifier has seen and then discarded them (it would be a different matter if they were returned to the prover). In particular, the fact that a one-qubit computer can be trivially simulated on a classical piece of paper is immaterial to the argument.

With a classical verifier things become substantially more delicate. How can we verify the existence of an n-qubit state with certain properties, while having only access to classical data about the state, data that, for all we know a priori, could have been generated by a simple — classical — laptop? To achieve this we need to find a way for the verifier to establish that the prover holds an n-qubit state, without ever having the ability to directly probe even a single qubit of that state. In the previous lecture we saw a means to achieve this for a single qubit based on the computational hardness of certain functions called "claw-free". In the next lecture we extend that method to introduce a protocol by which the prover can certify the existence of any single-qubit state that is a low-energy eigenstate of a single-qubit Hamiltonian. In the following lecture we combine this extension with the Fizsimons-Morimae protocol to obtain a protocol for delegating quantum computations with a classical client.

Lecture 6

Verifying a single qubit-Hamiltonian

In the previous lecture we introduced the circuit-to-Hamiltonian construction, that given a quantum circuit $\mathcal{C}$ and an input x to it returns a Hamiltonian $H_{\mathcal{C}}$ of the form (5.3) such that the completeness and soundness properties stated in Theorem 5.7 hold. This construction allowed us to reduce the problem of delegating a quantum computation to the problem of deciding if a certain publicly known, explicitly specified exponential-size Hermitian matrix $H_{\mathcal{C}}$ has an eigenvalue below a certain threshold a, or all its eigenvalues are above $b + \delta$ for a δ that is at least inverse polynomial in the number of qubits n on which $H_{\mathcal{C}}$ acts.[1] We then introduced the Fitzsimons-Morimae protocol (Figure 5.2) that is a protocol with one-way communication for verifying this fact.

Our goal in the next two lectures is to combine the Fitzsimons-Morimae protocol with the computational test for a qubit from lecture 4, Section 4.3 to obtain a classical protocol with similar guarantees to the Fitzsimons-Morimae protocol. For this we will develop a test that allows one to verify that a prover "has" a quantum state $|\psi\rangle$ with certain properties (e.g. it satisfies $\langle\psi|H|\psi\rangle \leq a + \delta/2$, i.e. certifies that the outcome of the computation is '1'. Note that even though in principle it is sufficient for the verifier to be convinced that such a $|\psi\rangle$ exists to make the right decision, we will see from the proofs that we can go a little further and give a precise meaning to the notion that the prover 'has' $|\psi\rangle$. This, however, will not be as strong as the claim that the prover 'has n qubits in state $|\psi\rangle$' in the sense that we gave to the phrase 'has n qubits', i.e. we will not quite exhibit $2n$ Pauli operators X_i, Z_i that satisfy all the required relations.

Remark 6.1. In passing to the Hamiltonian model of computation we relaxed our main goal, from obtaining a value $b \in \{0,1\}$ that is distributed as a measurement of the output qubit of the quantum circuit $\mathcal{C}$ in the standard basis to obtaining a value that is 1 whenever this measurement returns 1 with probability larger than $\frac{2}{3}$, and 0 whenever it is less than $\frac{1}{3}$. In particular, we make no requirement for circuits that are "undecided", e.g. return a random bit as output. This is typical to applications in complexity where it is assumed that circuits of interest make a clear-cut decision, 0 or 1; this is the setting discussed in Section 5.1. By tweaking the definition of $H_{\mathcal{C}}$ it is in fact possible to guarantee that any state $|\psi\rangle$ such that $\langle\psi|H_{\mathcal{C}}|\psi\rangle \leq a + \delta/2$ is such that a measurement of the first qubit of $|\psi\rangle$ in the standard basis yields an outcome whose distribution is within total variation distance, say, $\frac{1}{100}$ from a measurement of the output qubit of $\mathcal{C}$. Using this observation the protocol given at the end of this lecture can be adapted to return outcomes that are distributed close to the circuit output distribution, even in cases where the output is not assumed to be biased one way or the other. For simplicity we leave this extension as an exercise to the reader.

[1] In the previous lecture this number of qubits was called n', with n the number of qubits of the circuit $\mathcal{C}$. For the next two lectures, $\mathcal{C}$ disappears and so we re-use n to measure the size of $H_{\mathcal{C}}$.

6.1 A test for a specific single-qubit Hamiltonian

We start with an "easy" case: we show how the computational test for a qubit from lecture 4, protocol $\mathfrak{Q}$, can be cast as a verification protocol for the claim that the Hamiltonian $H = -\sigma_Z$ has a "low" eigenvalue, equal to -1. We go a little further by showing how such an eigenstate can be "extracted" from any successful prover in the protocol.

6.1.1 An explicit isometry

Our main result on the computational qubit test, Theorem 4.4, states that any successful prover in the protocol must "have a qubit". The proof achieves slightly more than that, as it explicitly states what the observables Z in (4.3) and X in (4.4) that define the qubit $(|\psi\rangle, Z, X)$ are. As we saw in Lemma 2.3 in lecture 2 the qubit implies the existence of an isometry $V : \mathcal{H} \to \mathbb{C}^2 \otimes \mathcal{H}'$ under which $Z \simeq \sigma_Z$, $X \simeq \sigma_X$, and $|\psi\rangle \simeq |\psi'\rangle \in \mathbb{C}^2 \otimes \mathcal{H}'$, giving us an identification of the "abstract" qubit $(|\psi\rangle, Z, X)$ with a "concrete" qubit, i.e. the space $\mathbb{C}^2$ and its algebra of operators, of which σ_Z and σ_X form a linear basis.

With our present goal of "extracting" a specific quantum state (a low-energy eigenstate for the Hamiltonian H_C) it is worthwhile making V a little more explicit. Indeed, an important point that we did not emphasize so far is that this identification is not "canonical". If you remember the proof of Lemma 2.3, it involves an application of Jordan's lemma to identify a block structure such that in each block, Z and X act like σ_Z and σ_X respectively. These blocks were obtained by diagonalizing the operator $(X + Z)$. In the case where X and Z anti-commute this operator has only two eigenvalues, $\pm\sqrt{2}$, and the associated eigenspaces are highly degenerate. Any choice of a basis for one of the eigenspaces can be used to specify an isometry V (a basis for the other eigenspace is determined by the first). (That there would be such a degeneracy is easily seen by observing that composing V by any unitary on $\mathcal{H}'$ still gives a valid isometry with the same properties.)

It is, in fact, possible to define a canonical choice for the isometry. This choice has the advantage that it is explicit and from a computational viewpoint leads to a circuit for V that can be constructed from circuits for X and Z. The idea behind the definition is to use the operators X and Z to "teleport" the abstract qubit $(|\psi\rangle, Z, X)$ into a "concrete" qubit $(|\varphi\rangle, \sigma_Z, \sigma_X)$ by means of an EPR pair. This is done in the following proposition.

Proposition 6.2. *Let $(|\psi\rangle, Z, X)$ be a qubit on $\mathcal{H}$. Let $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle) \in \mathbb{C}^2 \otimes \mathbb{C}^2$ be the state of an EPR pair. Let $\mathcal{H}' = \mathbb{C}^2_{\mathsf{A}} \otimes \mathcal{H}$ and $V : \mathcal{H} \to \mathbb{C}^2_{\mathsf{Q}} \otimes \mathcal{H}'$ defined by*

$$\forall|\varphi\rangle \in \mathcal{H}\,, \quad V|\varphi\rangle = \frac{1}{2}\big(\mathrm{Id} \otimes \mathrm{Id}_{\mathsf{A}} \otimes \mathrm{Id}_{\mathsf{Q}} + X \otimes \sigma_X \otimes \mathrm{Id}_{\mathsf{Q}} + Z \otimes \sigma_Z \otimes \mathrm{Id}_{\mathsf{Q}} + XZ \otimes \sigma_X\sigma_Z \otimes \mathrm{Id}_{\mathsf{Q}}\big)|\varphi\rangle|\phi^+\rangle_{\mathsf{AQ}}\,, \tag{6.1}$$

where the systems in the range of V are re-ordered so that the first factor $\mathbb{C}^2$ is associated with the second qubit of $|\phi^+\rangle_{\mathsf{AQ}}$ in (6.1)*, and $\mathcal{H}'$ consists of the state of the first qubit of $|\phi^+\rangle$, i.e. register A, as well as the part of the state in $\mathcal{H}$. Then V is an isometry and for all $W \in \{X, Z\}$,*

$$VW|\psi\rangle = (\sigma_W \otimes \mathrm{Id})V|\psi\rangle\,. \tag{6.2}$$

Proof. The proof is by direct calculation. First we verify that V is indeed an isometry. This is simply because the four states $\{(\sigma_X(a)\sigma_Z(b) \otimes \mathrm{Id})|\phi^+\rangle,\ a, b \in \{0,1\}\}$ are orthonormal[2] and X and Z are observables, so

[2] For $a, b \in \{0,1\}$ we use the notation $\sigma_X(a)$ for σ_X^a and similarly $\sigma_Z(b)$ for σ_Z^b. The motivation for this notation will be seen later when we consider n-qubit Pauli operators.

that for normalized $|\varphi\rangle$ each of the four terms on the right-hand side of (6.1) has norm exactly 1. Note that this does not require any other condition on X, Z than that they are observables (in fact, unitarity suffices). In particular, they do not need to anti-commute. Next we verify (6.2). Take $W = X$. Then

$$\begin{aligned}
VX|\psi\rangle &= \frac{1}{2}\big(X \otimes \mathrm{Id} \otimes \mathrm{Id} + \mathrm{Id} \otimes \sigma_X \otimes \mathrm{Id} - XZ \otimes \sigma_Z \otimes \mathrm{Id} - Z \otimes \sigma_X\sigma_Z \otimes \mathrm{Id}\big)|\psi\rangle|\phi^+\rangle \\
&= \frac{1}{2}\big(X \otimes \mathrm{Id} \otimes \mathrm{Id} + \mathrm{Id} \otimes \sigma_X \otimes \mathrm{Id} - XZ \otimes \sigma_Z \otimes \mathrm{Id} - Z \otimes \sigma_X\sigma_Z \otimes \mathrm{Id}\big)|\psi\rangle(\sigma_X \otimes \sigma_X)|\phi^+\rangle \\
&= \frac{1}{2}\big(X \otimes \sigma_X \otimes \sigma_X + \mathrm{Id} \otimes \mathrm{Id} \otimes \sigma_X + XZ \otimes \sigma_X\sigma_Z \otimes \sigma_X + Z \otimes \sigma_Z \otimes \sigma_X\big)|\psi\rangle|\phi^+\rangle \\
&= (\sigma_X \otimes \mathrm{Id})V|\psi\rangle\,,
\end{aligned}$$

where for the first line we used that X and Z anti-commute on $|\psi\rangle$, for the second that $\sigma_X \otimes \sigma_X|\phi^+\rangle = |\phi^+\rangle$, for the third that σ_X and σ_Z anti-commute, and for the last we re-ordered terms. □

6.1.2 Extraction of prover's qubit

In the proof of Theorem 4.4 we defined a specific X and Z from the prover's actions and showed that they anti-commute. Moreover, we showed that for a prover that always succeeds in the equation test (case $c = 1$) it must be the case that the state $|\psi\rangle$ of the prover is a $+1$ eigenstate of X, $X|\psi\rangle = |\psi\rangle$. For the definition of $|\psi\rangle \in \mathcal{H}_\mathsf{X} \otimes \mathcal{H}_\mathsf{P}$, recall that we had assumed that the prover directly measures the qubits in register X on challenge $c = 0$, and applies an arbitrary unitary U before measuring in the Hadamard basis on challenge $c = 1$. This means that after the isometry V, the prover's state $|\psi\rangle$ is mapped to a $+1$ eigenstate of σ_X, i.e. the state $|+\rangle$. The following corollary summarizes this discussion.

Corollary 6.3. *Suppose that a prover succeeds with probability* 1 *in the protocol. Then the isometry V defined from the observables Z in* (4.3) *and X in* (4.4) *sends $|\psi\rangle$ to $|+\rangle_\mathsf{Q}|aux\rangle$, for some state $|aux\rangle$ on $\mathcal{H}'$.*

In the context of this lecture we interpret Corollary 6.3 as our first test for a quantum computation in the Hamiltonian-based model: in this test, the verifier is effectively checking that the prover has prepared a $+1$ eigenstate of the Hamiltonian σ_X, or in other words a ground state of $H = -\sigma_X$. While we know that this eigenstate always exists, the analysis of the protocol shows that in some sense the prover has prepared the state. This additional observation allows us to make stronger conclusions from the protocol. For example, just as a measurement of $|+\rangle$ in the computational basis yields an unbiased random bit, we are able to deduce that the value of $b(x)$ with x the prover's answer on challenge $c = 0$ is an unbiased random bit. This ties in to the discussion in Section 3.4.1, where we observed that high success probability in the magic square game could be used to certify the generation of a random bit, and makes the protocol potentially useful for cryptographic applications where the generation of certified unbiased randomness serves as a resource.

In the development of our test we were greatly aided by the fact that we *know* what is the ground state of $H = -\sigma_X$, and in particular we know that a Hadamard basis measurement of it yields the outcome 0 (for '$+$') with probability 1. This "knowledge" was indirectly encapsulated in the test performed for the case $c = 1$, the analysis of which led us to conclude that $X|\psi\rangle = +|\psi\rangle$. But what if we didn't? What if H is a general Hamiltonian of the form (5.3), for which we can't a priori predict measurement outcomes?

6.2 Extracting a qubit: general case

In Section 6.1.1 we made the important observation that the map V in (6.1) is a well-defined isometry for any choice of the two observables X and Z. In particular, this map allows us to make a meaningful definition of a *space* for a qubit, and a *state* for that qubit, associated with *any* prover in protocol $\mathfrak{Q}$, the computational test for a qubit described in Figure 4.1. This definition does not guarantee that the prover "has a qubit," because it does not say anything about how the prover's observables operate on it. However, it still allows us to define a *candidate* for a single-qubit state on which σ_X and σ_Z measurements can *in principle* be made. The next claim evaluates how outcomes of these measurements when performed on the extracted qubit are a distributed as a function of the observables X and Z on the prover's state $|\psi\rangle$.

Claim 6.4. *Let $|\psi\rangle \in \mathcal{H}$ and X, Z observables on $\mathcal{H}$ be arbitrary. Let V be defined as in* (6.1). *Then the following hold:*

$$\langle\psi|V^\dagger\big(\sigma_Z \otimes \mathrm{Id}\big)V|\psi\rangle = \langle\psi|Z|\psi\rangle\;, \tag{6.3}$$

$$\langle\psi|V^\dagger\big(\sigma_X \otimes \mathrm{Id}\big)V|\psi\rangle = \frac{1}{2}\big(\langle\psi|X|\psi\rangle - \langle\psi|ZXZ|\psi\rangle\big)\;, \tag{6.4}$$

where the σ_Z and σ_X operators act on the first tensor factor $\mathbb{C}^2$ in the range of V and the identities act on $\mathcal{H}'$.

Recalling the diagram 2.3 introduced to illustrate Lemma 2.3, Claim 6.4 can similarly be illustrated as follows, where E_b denotes the average over b:

$$\begin{tikzcd}
\mathcal{H} \arrow[r, "V"] \arrow[d, swap, "\substack{Z \\[2em] \mathrm{E}_{b\in\{0,1\}}(-1)^b\, Z^b X Z^b}"] & \mathbb{C}^2 \otimes \mathcal{H}' \arrow[d, "\substack{\sigma_Z\otimes\mathrm{Id} \\[2em] \sigma_X\otimes\mathrm{Id}}"] \\
\mathcal{H} \arrow[r, "V"] & \mathbb{C}^2 \otimes \mathcal{H}'
\end{tikzcd} \tag{6.5}$$

We emphasize that this diagram is purely illustrative and should be understood exactly in the sense of (6.3) and (6.4); i.e. it does not imply a relation on the operators but only on the expectation values on the state $|\psi\rangle$. Informally, when considering expectation values only the isometry has the effect of applying a Z-twirl to the Hadamard basis observable X.

Proof. Let $W \in \{X, Z\}$. Expanding from the definition of V,

$$\begin{aligned}
\langle\psi|V^\dagger\big(\sigma_W \otimes \mathrm{Id}\big)V|\psi\rangle &= \frac{1}{4}\sum_{P,Q\in\{I,X,Z,XZ\}} \langle\psi|P^\dagger Q|\psi\rangle \cdot \langle\phi^+|\sigma_P^\dagger\sigma_Q \otimes \sigma_W|\psi\rangle \\
&= \frac{1}{4}\sum_{P,Q:\,\sigma_P^\dagger\sigma_Q=\sigma_W} \langle\psi|P^\dagger Q|\psi\rangle\;,
\end{aligned}$$

where for the second line we used that $\langle\phi^+|\sigma_W \otimes \sigma_{W'}|\phi^+\rangle = \delta_{W,W'}$ with δ the Kronecker symbol. In case $W = Z$ the pairs P, Q that appear in the last summation above are $(X, I), (I, X), (XZ, X)$ and (X, XZ). Using $X^2 = \mathrm{Id}$ we obtain (6.3). In case $W = X$ then the summation is over $(Z, I), (I, Z), (XZ, Z)$ and (Z, XZ) and has a minus sign for the last two terms due to $\sigma_X\sigma_Z = -\sigma_Z\sigma_X$. Thus we get (6.4) as well. □

Observe that if X and Z anti-commute then Claim 6.4 gives us the result that we expect: in this case $(|\psi\rangle, Z, X)$ is a qubit so Proposition 6.2 applies and the isometry "intertwines" measurements X and Z on $|\psi\rangle$ with σ_X and σ_Z respectively on the first factor of $V|\psi\rangle$. At the other extreme, if X and Z commute then (6.4) indicates that a measurement in the Hadamard basis of the extracted qubit returns an unbiased random bit. This is expected of a "classical" state, which always leads to uniformly random results in the Hadamard basis. The lemma in some sense interpolates between these results. Importantly, it allows us to associate a qubit with the state of an arbitrary prover in the protocol, that is such that the distribution of measurements on the extracted qubit can be related to quantities that involve the prover's state and observables in the protocol. For convenience we make this into a definition.

Definition 6.5 (Extracted qubit)**.** Let P be a prover in protocol $\mathfrak{Q}$. Let $|\psi\rangle$ be the state of P after having sent y in the first round of interaction. Let V be defined as in (6.1). Then we call the reduced density of $V|\psi\rangle$ on the first factor $\mathbb{C}^2$, associated with register Q, the *extracted qubit* and denote it by ρ_{Q}.

Lemma 6.6. *Let P be a prover that succeeds with probability 1 in the pre-image test of protocol $\mathfrak{Q}$ and such that the string d returned in the equation test is $d = 0^m$ with probability that is negligibly small in λ. (No other assumption is made on the equation test.) Let ρ be the extracted qubit, as defined in Definition 6.5. Then the following hold:*

- *(Z-measurement:) The outcome of measuring ρ in the computational basis is identically distributed to the bit $(-1)^{b(x)}$ computed from the prover's answer x in case $c = 0$.*
- *(X-measurement:) Under assumption (**F.2**), the outcome of measuring ρ in the Hadamard basis is* computationally indistinguishable *from the bit $(-1)^{d\cdot(x_0+x_1)}$ where d is obtained from the prover in case $c = 1$.*

Remark 6.7 (Computational distinguishability). The statement of the lemma refers to two distributions being computationally indistinguishable. Informally, this means that no computationally efficient procedure can distinguish a sample taken from one distribution from a sample taken from the other. Formally, families of distributions $D = \{D_\lambda\}$ and $D' = \{D'_\lambda\}$ on universes $\{\mathcal{X}_\lambda\}$ are said to be computationally indistinguishable if for any PPT (or QPT for computational indistinguishability against quantum adversaries) procedure $\mathcal{A}$ there is a negligible function μ such that for every λ,

$$\Big| \Pr_{x \leftarrow D_\lambda} \big(\mathcal{A}(1^\lambda, x) = 1\big) - \Pr_{x' \leftarrow D'_\lambda} \big(\mathcal{A}(1^\lambda, x') = 1\big) \Big| \leq \mu(\lambda) .$$

Here, when we refer to computational indistinguishability we will always mean against QPT adversaries. Note that for distributions on a family of universes $\{\mathcal{X}_\lambda\}$ such that $|\mathcal{X}_\lambda|$ grows at most polynomially with λ the notion of computational indistinguishability is equivalent to statistical indistinguishability, i.e. the total variation distance between D_λ and D'_λ goes to 0 as fast as some negligible function. (Showing this formally is a good exercise to practice with the definitions.)

Proof. The first item follows immediately from (6.3) in Claim 6.4 and the definition of Z in (4.3), which guarantees that the bit $(-1)^{b(x)}$ obtained from the prover in case $c = 0$ has expectation precisely $\langle\psi|Z|\psi\rangle$.

To show the second item we assume for contradiction that the two distributions are computationally distinguishable. Since the distributions are over a single bit, as recalled in Remark 6.7 this is equivalent to statistical distinguishability: there must exist a polynomial $q(\lambda)$ such that for infinitely many values of λ,

$$\big|\langle\psi|X|\psi\rangle + \langle\psi|ZXZ|\psi\rangle\big| > \frac{1}{q(\lambda)} , \tag{6.6}$$

where recall that the expression on the left should be understood on average over the generation of pk by the verifier and the message y sent by the prover in the first round of interaction. We derive a contradiction with (**F.2**) by constructing an adversary in (4.2). Given λ and pk as input, $\mathcal{A}$ prepares the state $|\psi\rangle$. $\mathcal{A}$

then measures register X in the standard basis to obtain an outcome x. Using the assumption that the prover succeeds with probability 1 in the pre-image test, $f_{pk}(x) = y$ and the (unnormalized) post-measurement state is exactly $Z_{b(x)}|\psi\rangle$, where as usual $Z_b = (\mathrm{Id} + (-1)^b Z)/2$. Finally, the adversary applies the prover's unitary U and measures in the Hadamard basis to obtain a string d. It returns the pair (x, d). The expected value of $(-1)^{d\cdot(x_0+x_1)}$ under this procedure is

$$\langle\psi|Z_0 X Z_0|\psi\rangle + \langle\psi|Z_1 X Z_1|\psi\rangle \;=\; \frac{1}{2}\big(\langle\psi|X|\psi\rangle + \langle\psi|ZXZ|\psi\rangle\big)\;,$$

which can be seen by expanding $Z_b = (\mathrm{Id} + (-1)^b Z)/2$ for $b \in \{0,1\}$ and canceling cross-terms. Using (6.6), $\mathcal{A}$ violates (4.2).[3] □

6.3 A single-qubit verification protocol

In the previous section we showed how to identify a "qubit" such that for any prover in the protocol, as long as the prover succeeds in the preimage test then it is possible for the verifier to infer from the prover's answers a bit whose distribution is statistically indistinguishable from outcomes of σ_Z or σ_X measurements on a well-defined quantum state. In order to turn this into a verification protocol for a single-qubit Hamiltonian, we are missing the completeness statement: while in Section 4.3 we saw how a prover could behave in such a way that the extracted qubit is a $|+\rangle$ state, we do not yet know if it is possible to use the protocol for the verification of other single-qubit states. In order for this to work out, we make the following assumption that replaces assumption (**F.4**):

(**F.4'**) For any pk and any y in the range of f_{pk} the two preimages of y take the form (b, x_b) where $b \in \{0,1\}$ and $x_b \in \{0,1\}^{m(\lambda)-1}$. In particular, the function $b : \{0,1\}^m \to \{0,1\}$ returns the first bit of its input.

This assumption is mainly for convenience and holds for most constructions of claw-free functions, including the one that we sketch in the next lecture. Given a 2-to-1 function family that satisfies (**F.4'**) the following lemma shows how a prover can behave in the protocol so that the extracted qubit defined in the previous section is a state $|\varphi\rangle$ of its choice.

Lemma 6.8. *Let $|\varphi\rangle \in \mathbb{C}^2$ be any state. Then there is a way for the prover to behave in protocol $\mathfrak{Q}$ such that the prover is accepted with probability* 1 *in the preimage test and moreover the extracted qubit satisfies $\rho_{\mathsf{Q}} = |\varphi\rangle\langle\varphi|$.*

Proof. Let $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$ for $\alpha, \beta \in \mathbb{C}$ such that $|\alpha|^2 + |\beta|^2 = 1$. The prover performs the following steps:

1. Prepare the initial state

$$|\psi^{(0)}\rangle_{\mathsf{BXY}} \;=\; |\varphi\rangle_{\mathsf{B}} \otimes \Big(\frac{1}{2^{m-1}} \sum_{x\in\{0,1\}^{m-1}} |x\rangle_{\mathsf{X}}\Big)|0\rangle_{\mathsf{Y}}\;,$$

where B is a one-qubit register, X an $(m-1)$ and Y an m-qubit register, for $m = m(\lambda)$.

[3] The end of the proof glosses over a detail: one needs to guarantee that the equation d returned by $\mathcal{A}$ is not 0. While the lemma assumes that this is the case when the equation is measured directly on $|\psi\rangle$, here $\mathcal{A}$ measures after $|\psi\rangle$ has already been measured using the observable Z. To show that the assumption that $d \neq 0^m$ with probability negligibly close to 1 still holds one needs to use the "collapsing" property of f_{pk}, that we will introduce in the next lecture.

2. Upon receipt of the function index pk, coherently evaluate f_{pk} on the input in registers BX, writing the output in register Y to obtain the state
$$|\psi^{(1)}\rangle_{\mathsf{BXY}} \,=\, \frac{\alpha}{2^{m-1}} \sum_{x\in\{0,1\}^{m-1}} |0\rangle_{\mathsf{B}}|x\rangle_{\mathsf{X}}|f(0x)\rangle_{\mathsf{Y}} + \frac{\beta}{2^{m-1}} \sum_{x\in\{0,1\}^{m-1}} |1\rangle_{\mathsf{B}}|x\rangle_{\mathsf{X}}|f(1x)\rangle_{\mathsf{Y}}\,.$$
3. Measure the last register to obtain a y. Let $(0, x_0)$ and $(1, x_1)$ be the two preimages of y under f_{pk}. Then the re-normalized post-measurement state is
$$|\psi^{(2)}\rangle_{\mathsf{BXY}} \,=\, \big(\alpha|0\rangle_{\mathsf{B}}|x_0\rangle_{\mathsf{X}} + \beta|1\rangle_{\mathsf{B}}|x_1\rangle_{\mathsf{X}}\big)|y\rangle_{\mathsf{Y}}\,.$$
4. Upon receipt of challenge c, perform as the honest prover in protocol $\mathfrak{Q}$: if $c = 0$ measure registers BX in the standard basis and return the outcome $x = (b, x_b)$; if $c = 1$ measure in the Hadamard basis and return the outcome d.

This prover always returns a valid preimage in the case of a challenge $c = 0$, so it is accepted with probability 1. Observe that the operator Z associated to this prover is equal to a σ_Z on register B. Regarding the operator X, a simple calculation reveals that the action of X restricted to the span of $|0, x_0\rangle_{\mathsf{BX}}$ and $|1, x_1\rangle_{\mathsf{BX}}$ consists in exchanging these two basis states. Using the explicit form of the isometry V given in (6.1) one can verify that
$$V|\psi^{(2)}\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle_{\mathsf{B}}|x_0\rangle_{\mathsf{X}}|0\rangle_{\mathsf{A}} + |1\rangle_{\mathsf{B}}|x_1\rangle_{\mathsf{X}}|1\rangle_{\mathsf{A}}\big) \otimes \big(\alpha|0\rangle_{\mathsf{Q}} + \beta|1\rangle_{\mathsf{Q}}\big) \otimes |y\rangle_{\mathsf{Y}}\,,$$
where AQ are the two registers introduced to hold the EPR pair $|\phi^+\rangle_{\mathsf{AQ}}$ used in the definition of V. Register Q contains the extracted qubit. □

Let $\mathcal{F}$ be a 2-to-1 trapdoor claw-free function family and $\lambda \in \mathbb{N}$ a security parameter. Let $\varepsilon, \delta > 0$ be accuracy parameters. Let $\gamma = 0$ and $N = \frac{C}{\delta^2}\ln(1/\varepsilon)$ for some large constant C. The verifier and prover repeat the following interaction N times.

1. The verifier generates $(pk, td) \leftarrow \textsc{Gen}(1^\lambda)$. It sends pk to the prover.
2. The prover returns $y \in \{0,1\}^m$, where $m = m(\lambda)$.
3. The verifier selects a uniformly random challenge $c \leftarrow_R \{0,1\}$ and sends c to the prover.
4. (a) *(Computational basis, $c = 0$:)* In case $c = 0$ the prover is expected to return an $x \in \{0,1\}^m$. If $f_{pk}(x) \neq y$ then the verifier immediately aborts. The verifier sets $a \leftarrow (-1)^{b(x)}$ and $\gamma \leftarrow \gamma - J_Z a$.
   (b) *(Hadamard basis, $c = 1$:)* In case $c = 1$ the prover is expected to return a $d \in \{0,1\}^m$. The verifier uses td to determine the two preimages (x_0, x_1) of y by f_{pk}. She sets $b \leftarrow (-1)^{d\cdot(x_0+x_1)}$ and $\gamma \leftarrow \gamma - J_X b$.

If the verifier has not aborted at any of the steps $c = 0$, she returns the real number $o = \frac{1}{N}\gamma$.

Figure 6.1: Verification protocol for a single-qubit Hamiltonian $H = -\frac{J_X}{2}\sigma_X - \frac{J_Z}{2}\sigma_Z$.

The following proposition summarizes what we have achieved so far, a verification protocol for single-qubit Hamiltonians and a completely classical verifier. (Of course a simpler protocol would be to have the verifier classically do the computation themselves! The point is that this protocol is not too hard to extend to n qubits, as we will see in the next lecture.) The protocol, which combines protocol $\mathfrak{Q}$ with the Fitzsimons-Morimae verification protocol, is summarized in Figure 6.1.

Proposition 6.9. *Let $H = -\frac{J_X}{2}\sigma_X - \frac{J_Z}{2}\sigma_Z$ be a single-qubit Hamiltonian and $\delta, \varepsilon > 0$ accuracy parameters. Then the verification protocol from Figure 6.1 has the following properties:*

1. *(Completeness:) For any single-qubit state $|\varphi\rangle$, there is a QPT prover that is accepted with probability 1 in the protocol and such that the value o returned by the verifier at the end of the protocol satisfies $\mathrm{E}[o] = \langle\varphi|H|\varphi\rangle$.*

2. *(Soundness:) For any QPT prover that is accepted with probability 1 in the protocol, there is a single-qubit state ρ such that the value o returned by the verifier at the end of the protocol satisfies $\mathrm{E}[o] = \mathrm{Tr}(H\rho)$.*

Moreover, with the value of N specified in the protocol in both cases it holds that $\Pr(|o - \mathrm{Tr}(H\rho)| > \delta) \leq \varepsilon$.

Remark 6.10. The assumption that the prover succeeds with probability negligibly close to 1 in the protocol can be relaxed to a constant sufficiently close to 1, where the distance to 1 will affect the distance $|\mathrm{E}[o] - \mathrm{Tr}(H\rho)|$. First we observe that a success probability negligibly close to 1 is sufficient; this can be verified by going through the argument again, and nothing needs to be changed. Second, it is possible to show that any prover with success probability $1-\kappa$ for some $\kappa \geq 0$ can be transformed to a prover with success probability negligibly close to 1, affecting the distribution of o proportionately to κ. Intuitively, the new prover will test if the value y that the old prover would have returned will lead to success on challenge $c = 0$, in case that the test is actually executed by the verifier. This can be done efficiently by the prover by evaluating the pre-image condition on register X. If this test fails then the new prover simply re-executes the old prover from scratch, until it is certain to achieve success. With probability negligibly close to 1 this iterative procedure will stop in a polynomial number of steps, and using the "pretty-good lemma" it is possible to show that the prover's distribution of outcomes is affected by some $O(\sqrt{\kappa})$ in statistical distance. We omit the details.

Proof. The completeness statement follows from Lemma 6.8. For soundness we use Lemma 6.6. This shows that the expectation of the bit a in step 4.(a) satisfies $\mathrm{E}[a] = \mathrm{Tr}(\sigma_Z\rho)$ where ρ is the extracted qubit. Similarly, the bit b in step 4.(b) satisfies $\mathrm{E}[b] = \mathrm{Tr}(\sigma_Z\rho)$. Since by definition

$$\mathrm{E}[o] = -\frac{J_Z}{2}\mathrm{E}[a] - \frac{J_X}{2}\mathrm{E}[b]$$

the proposition follows. □

Lecture 7

Verification for n qubit Hamiltonians in $XX - ZZ$ form

Moving beyond the case of a single qubit, our goal in this lecture is to generalize the results from Section 6.2 to a procedure for extracting n qubits from a prover, together with a statement that allows us to relate measurements in the standard or Hadamard basis of the extracted qubits to measurements performed by the prover in the protocol. Once this has been put in place we will not be far from a delegation protocol along the lines of the Fitzsimons-Morimae protocol from Section 5.2, but, crucially, with classical verifier and communication.

7.1 Setup

To set the stage we first give a straightforward generalization of the single-qubit verification protocol from Figure 6.1 to the case of n qubits. Recall from Section 5.2.2 that for our purposes it suffices to consider Hamiltonians that take the form (5.3). This form allows us to restrict our attention to a collection of measurement outcomes on an n-qubit state such that all qubits are measured in the same basis, computational or Hadamard. In particular we do not need to consider "mixed" measurements, with some qubits measured in the standard basis and other qubits in the Hadamard basis, because (5.3) does not have mixed terms such as $\sigma_{X,i}\sigma_{Z,j}$. (See Remark 7.1 regarding extensions to the mixed case.) It is then natural to request that the honest prover behaves exactly as in the single-qubit verification protocol, except that each action should be repeated independently for each of the n qubits: in the first phase the verifier sends the information for n functions $f_{pk_1}, \dots, f_{pk_n}$, the prover executes the encoding procedure from the proof of Lemma 6.8 independently for each of the n qubits of its claimed low-energy eigenstate $|\varphi\rangle$ of H_C and reports the n images $y_1, \dots, y_n$ obtained; in the second phase the verifier sends a single-bit challenge $c \in \{0, 1\}$ and the prover measures all its qubits in the computational or Hadamard basis and returns the outcomes $x_1, \dots, x_n$ or $d_1, \dots, d_n$ respectively. Note that the only part that is not repeated is the challenge, which is identical for each of the n concurrent repetitions. The reason that we can restrict ourselves to such challenges is due to the form of H_C from (5.3) and, as we will see, greatly simplifies the analysis. The complete protocol is given in Figure 7.1.

Before proceeding to the analysis of the protocol we examine the question, "Where are the qubits?" For the single-qubit verification protocol our initial intuition came from the qubit computational test from lecture 4, for which we were able to argue that the prover indeed has a qubit $(|\psi\rangle, Z, X)$. For the verification protocol seen in the last lecture we saw that in order to allow verification of other states than the $|+\rangle$ state

Let $\mathcal{F}$ be a 2-to-1 trapdoor claw-free function family and $\lambda \in \mathbb{N}$ a security parameter. Let $\varepsilon, \delta > 0$ be accuracy parameters. Let $\gamma = 0$ and $N = \frac{C}{\delta^2}\binom{n}{2}^2 \ln(1/\varepsilon)$. The verifier and prover repeat the following interaction N times.

1. The verifier selects a pair $i \neq j \in \{1, \ldots, n\}$ and $W \in \{X, Z\}$ uniformly at random.
2. For $\ell = 1, \ldots, n$ the verifier generates $(pk_\ell, td_\ell) \leftarrow \textsc{Gen}(1^\lambda)$. It sends $(pk_1, \ldots, pk_n)$ to the prover.
3. The prover returns $y_1, \ldots, y_n \in \{0,1\}^m$, where $m = m(\lambda)$.
4. The verifier selects a uniformly random challenge $c \leftarrow_R \{0,1\}$ and sends c to the prover.
5. (a) *(Computational basis, $c = 0$:)* In case $c = 0$ the prover is expected to return $x_1, \ldots, x_n \in \{0,1\}^m$. If $f_{pk_\ell}(x_\ell) \neq y_\ell$ for any ℓ then the verifier immediately aborts. The verifier sets
$$\gamma \leftarrow \gamma - J_{ij}\,(-1)^{b(x_i)}(-1)^{b(x_j)}\ .$$
   (b) *(Hadamard basis, $c = 1$:)* In case $c = 1$ the prover is expected to return $d_1, \ldots, d_n \in \{0,1\}^m$. The verifier uses td_i and td_j to determine the preimages $(x_{i,0}, x_{i,1})$ of y_i by f_{pk_i} and $(x_{j,0}, x_{j,1})$ of y_j by f_{pk_j} respectively. She sets
$$\gamma \leftarrow \gamma - J_{ij}\,(-1)^{d_i\cdot(x_{i,0}+x_{i,1})}(-1)^{d_j\cdot(x_{j,0}+x_{j,1})}\ .$$

If the verifier has not aborted at any of the steps $c = 0$, she returns the real number $o = \frac{1}{N}\binom{n}{2}\gamma$.

Figure 7.1: Verification protocol $\mathfrak{V}_n$ for an n-qubit Hamiltonian $H_{\mathcal{C}} = -\sum_{i,j} \frac{J_{ij}}{2}(\sigma_{X,i}\sigma_{X,j} + \sigma_{Z,i}\sigma_{Z,j})$.

we had to remove some of the tests done by the verifier (specifically, the equation check) and that due to this we were no longer able to guarantee a qubit in the sense of Definition 2.2. Nevertheless we were able to get around this by defining an abstract *extracted qubit* that did not directly correspond to the prover's observables but was still such that measurement outcomes on the extracted qubit could be shown to have a distribution that is negligibly close to outcomes obtained from the prover in the actual protocol (Lemma 6.6).

For the case of demonstrating n qubits a priori one would have to show that the prover has a state $|\psi\rangle$ and two families of observables $\{X(a) : a \in \{0,1\}^n\}$ and $\{Z(b) : b \in \{0,1\}^n\}$ that satisfy the Pauli commutation and anti-commutation relations when they act on $|\psi\rangle$. Indeed, a straightforward generalization of Lemma 2.3 then guarantees the existence of a suitable isometry with the space of n actual qubits. Showing this is challenging; luckily, for our purposes it is also not necessary.[1] Indeed, just as in the single-qubit case it is worth emphasizing that in the context of verification we do not need to guarantee that the prover has a certain quantum state, nor that it is able to perform certain measurements on it. The only real requirement is that a state $|\varphi\rangle$ *exists* such that $\langle\varphi|H_C|\varphi\rangle \leq a$. Thus, as we did in the analysis of the single-qubit verification protocol we will first introduce a abstract *extracted n qubit* defined from the prover's state and actions in the protocol but that also include additional ingredients that make it at first unclear how they relate to the prover itself. The definition of the extracted qubits is given in Section 7.2. Once this has been defined we will perform the second, crucial step, which is to relate the distribution of measurement outcomes on the extracted qubits to quantities that are directly observable in the protocol. This is done in Section 7.3. Finally in Section 7.4 we put everything together and show the completeness and soundness properties of the verification protocol given in Figure 7.1. In addition in Section 7.5 we will sketch a construction of a function family based on the Learning With Errors (LWE) problem that (approximately) satisfies all required assumptions and can thus be used to instantiate the protocol.

7.2 The n extracted qubits

7.2.1 Modeling the prover

We start by introducing notation that allows us to model an arbitrary prover in the protocol. Similarly to how we modeled the prover for the analysis of the computational qubit test in Section 4.3, a prover in the n-qubit verification protocol from Figure 7.1 can be represented using the following objects:

1. A state $|\psi\rangle$, that may depend on $pk_1, \ldots, pk_n$ and $y_1, \ldots, y_n$, such that $|\psi\rangle \in \mathcal{H}_{\mathsf{X}_1} \otimes \cdots \mathcal{H}_{\mathsf{X}_n} \otimes \mathcal{H}_{\mathsf{P}}$ with each space $\mathcal{H}_{\mathsf{X}_i}$ isomorphic to $(\mathbb{C}^2)^{\otimes m}$. The state $|\psi\rangle$ represents the state of the prover and the message registers at the end of step 3 in the protocol.

2. For the case $c = 0$, the prover directly measures all the X registers in the standard basis to obtain $x_1, \ldots, x_n$ that it returns to the verifier. For a string $a \in \{0,1\}^n$ we let

$$Z(a) = \sum_{x_1,\ldots,x_n} (-1)^{a_1 \cdot b_1(x_1)} \cdots (-1)^{a_n \cdot b_n(x_n)} |x_1\rangle\langle x_1| \otimes \cdots |x_n\rangle\langle x_n| \,, \tag{7.1}$$

where the functions b_i are not necessarily all equal since they may depend on pk_i. This is analogous to (4.3).

[1] In the last three lectures of the course we will see how in the context of spatial assumptions it is known how to test n qubits in this sense; for computational assumptions we do not yet know how to do it.

3. For the case $c = 1$, the prover applies an arbitrary unitary U followed by a measurement of the qubits in X in the Hadamard basis to obtain $d_1, \dots, d_n$. For a string $b \in \{0,1\}^n$ we let

$$X(b) = \sum_{d_1,\dots,d_n} (-1)^{b_1(d_1\cdot(x_{1,0}+x_{1,1}))} \cdots (-1)^{b_n(d_n\cdot(x_{n,0}+x_{n,1}))}$$
$$\cdot\, U^\dagger \big(H_{\mathsf{X}}^{\otimes nm} \otimes \mathrm{Id}_{\mathsf{P}}\big)^\dagger \big(|d_1,\dots,d_n\rangle\langle d_1,\dots,d_n|_{\mathsf{X}} \otimes \mathrm{Id}_{\mathsf{P}}\big)\big(H_{\mathsf{X}}^{\otimes nm} \otimes \mathrm{Id}_{\mathsf{P}}\big) U\,. \quad (7.2)$$

Remark 7.1. Note that the fact that the protocol only has two different challenges, $c = 0$ and $c = 1$, allows us to have a simple description for all $Z(a)$ and all $X(b)$ observables that involves only one "adversarial" unitary U. If we had to design a protocol that allows more general Hamiltonians with mixed terms of the form $\sigma_{X,i}\sigma_{Z,j}$ we would need to consider more challenges, and this would require a more complex analysis. This is done in [Mah18].

7.2.2 The isometry V

Next we define the n-qubit isometry V, and the extracted qubits.

Claim 7.2. *Let $|\psi\rangle \in \mathcal{H}$ and for every $a, b \in \{0,1\}^n$, $X(a)$ and $Z(b)$ observables on $\mathcal{H}$ such that all $X(a)$ (resp. all $Z(b)$) mutually commute and moreover $X(a)X(a') = X(a + a')$ for any $a, a' \in \{0,1\}^n$. Let $V : \mathcal{H} \to \mathcal{H}_{\mathsf{Q}} \otimes \mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}'$ where each of $\mathcal{H}_{\mathsf{Q}}$ and $\mathcal{H}_{\mathsf{A}}$ is $(\mathbb{C}^2)^{\otimes n}$ and $\mathcal{H}' \simeq \mathcal{H}$ be defined for all $|\varphi\rangle \in \mathcal{H}$ as*

$$V|\varphi\rangle = \Big(\frac{1}{2^n}\sum_{a,b} \mathrm{Id} \otimes \sigma_X(a)\sigma_Z(b) \otimes X(a)Z(b)\Big)|\phi^+\rangle^{\otimes n}|\varphi\rangle\,, \quad (7.3)$$

where each EPR pair $|\phi^+\rangle$ has one qubit in register Q and the other in register A and the σ_X and σ_Z operators act on register A. Then V is an isometry.

The proof of the claim is immediate and only uses that the family of states

$$\big\{(\sigma_X(a)\sigma_Z(b) \otimes \mathrm{Id})|\phi^+\rangle^{\otimes n} : a, b \in \{0,1\}^n\big\}$$

is orthonormal. Similarly to Definition 6.5 we can now define the n extracted qubits.

Definition 7.3 (Extracted qubits)**.** Let P be a prover in the verification protocol $\mathfrak{V}_n$ described in Figure 7.1. Let $|\psi\rangle$ be the state of P after having sent $y_1, \dots, y_n$ at step 3 of the t-th iteration, for some $t \in \{1, \dots, N\}$. Let V be defined in (7.3). Then we call the reduced density of $V|\psi\rangle$ on register Q the *extracted qubits* (implicitly, at iteration t) and denote them by $\rho_{\mathsf{Q}_1\cdots\mathsf{Q}_n}$.

7.3 Measurements on the extracted qubits

We start with the following analogue to Claim 6.4, which gives an explicit formula for the distribution of measurements in the standard or Hadamard basis on the n extracted qubits as a function of the prover's state and observables.

Claim 7.4. *The following hold for any prover, with ρ the n extracted qubits at any iteration (Definition 7.3):*

$$\forall b \in \{0,1\}^n\,, \qquad \mathrm{Tr}\big(\sigma_Z(b)\rho\big) = \langle\psi|Z(b)|\psi\rangle\,, \quad (7.4)$$

$$\forall a \in \{0,1\}^n\,, \qquad Tr\big(\sigma_X(a)\rho\big) = \frac{1}{2^n}\sum_b (-1)^{a\cdot b}\langle\psi|Z(b)X(a)Z(b)|\psi\rangle\,. \quad (7.5)$$

The claim can be illustrated using the following generalization of (6.5)

$$\begin{tikzcd}
\mathcal{H} \arrow[r, "V"] \arrow[d, "\substack{Z(b) \\ \\ \mathrm{E}_{b\in\{0,1\}}(-1)^{b\cdot a}\, Z(b)X(a)Z(b)}"'] & \mathbb{C}^2 \otimes \mathcal{H}' \arrow[d, "\substack{\sigma_Z(b)\otimes\mathrm{Id} \\ \\ \sigma_X(a)\otimes\mathrm{Id}}"] \\
\mathcal{H} \arrow[r, "V"] & \mathbb{C}^2 \otimes \mathcal{H}'
\end{tikzcd} \tag{7.6}$$

Proof. Eq. (7.4) is immediate using that $X(a)$ are observables and $\langle\phi^+|^{\otimes n}\sigma_X(a')\sigma_Z(b')\otimes\sigma_Z(b)|\phi^+\rangle^{\otimes n}$ is zero unless $a'=0$ and $b=b'$. Eq. (7.5) is shown similarly by direct calculation, using $X(a')X(a'')=X(a'+a'')$ and $\sigma_Z(b)\sigma_X(a)\sigma_Z(b)=(-1)^{a\cdot b}\sigma_X(a)$. □

The next lemma is the key lemma. It argues that for computationally bounded provers, the quantity on the right-hand side of (7.5) is close to the simpler quantity $\langle\psi|X(a)|\psi\rangle$, that in particular can be inferred in the protocol from the prover's outcomes y_i and d_i (for those i such that $a_i=1$). Before we can state the lemma we need to introduce one last assumption on the function family $\mathcal{F}$. Intuitively, this assumption is a natural quantum analogue of the classical property of collision resistance, but is stronger than it.

(**F.5**) Consider the following abstract game between an arbitrary "adversary" (think prover) and a trusted (quantum) "challenger" (think verifier). First, the adversary is provided a label pk (generated at random by the challenger) and required to prepare an arbitrary state of the form $|\phi\rangle=\sum_x \alpha_x|x\rangle$, where x ranges over the domain of f_{pk}. (In general the adversary may keep an additional register entangled with this state. For ease of notation we do not consider such entanglement in this description.) The adversary hands the state $|\phi\rangle$ over to the challenger, who evaluates f_{pk} in superposition on $|\phi\rangle$ and measures the image register, obtaining a y in the range of f_{pk} and the (suitably re-normalized) post-measurement state $|\phi'\rangle=\sum_{x:f_{pk}(x)=y}\alpha_x|x\rangle$. The challenger then returns to the adversary the string c together with *either* the state $|\phi'\rangle$ *or* the probabilistic mixture $\sum_{x:f(x)=c}|\alpha_x|^2|x\rangle\langle x|$ obtained by measuring the same state $|\phi'\rangle$ in the computational basis (and throwing away the outcome). The adversary wins if it correctly guesses which is the case. Assumption (**F.5**) on the function family $\mathcal{F}$ states that for any QPT adversary $\mathcal{A}$ there is a negligible function μ such that for any λ, $\mathcal{A}$ succeeds in this game with probability that deviates from $\frac{1}{2}$ by at most $\mu(\lambda)$.

Remark 7.5. Assumption (**F.5**) is referred to as the "collapsing" property for the function family $\mathcal{F}$. This property was introduced by Unruh as a strengthening of the classical property of collision resistance required for his work on the security of commitment protocols that are computationally binding against quantum adversaries [Unr16]. The reason that this assumption implies collision resistance is that, if the function were not collision resistant, the adversary could identify a colliding pair (x_0,x_1) and submit $|\phi\rangle=\frac{1}{\sqrt{2}}(|x_0\rangle+|x_1\rangle)$ to the challenger. It could then measure the challenger's response in a basis containing the two states $\frac{1}{\sqrt{2}}(|x_0\rangle\pm|x_1\rangle)$ and guess that, in case the "$-$" outcome is obtained, the challenger must have measured; in the other case, the adversary guesses at random.

Note that assumption (**F.2**) *also* trivially implies collision resistance, since the ability to identify a claw allows one to generate arbitrary equations in it. It is possible to show that both (**F.2**) and (**F.5**) are strictly stronger than collision resistance. It is likely that the two assumptions are incomparable, but I have not tried to show this explicitly.

We can now state and prove the key lemma.

Lemma 7.6. *Let P be a prover that succeeds with probability 1 in protocol $\mathfrak{V}_n$. Let ρ be the n extracted qubits, as defined in Definition 7.3 (for any iteration). Then the following hold for any $i \neq j \in \{1,\dots,n\}$:*

- *(Z-measurement:) The outcome of measuring qubits i and j of ρ in the computational basis is identically distributed to the bits $b(x_i)$ and $b(x_j)$ obtained from the prover in case $c = 0$.*
- *(X-measurement:) Under assumptions (**F.2**) and (**F.5**) the outcome of measuring qubits i and j of ρ in the Hadamard basis is* computationally indistinguishable *from the pair of bits $d_i \cdot (x_{i,0} + x_{i,1})$ and $d_j \cdot (x_{j,0} + x_{j,1})$ where d_i and d_j are obtained from the prover in case $c = 1$.*

As already noted in the previous lecture, for distributions on two bits the notions of computational and statistical indistinguishability are essentially equivalent. The lemma generalizes to the joint distribution of any number of bits, and in this case it is only the weaker computational indistiguishability that is obtained. For simplicity we restrict ourselves to proving the lemma for the setting of two bits only.

Proof. For the case of a measurement in the computational basis the lemma follows directly from (7.4) in Claim 7.4 and the definition of the extracted qubits. For the case of a measurement in the Hadamard basis we proceed in two steps.

In the first step we show that for any $b \in \{0,1\}^n$ the states $|\psi\rangle$ and $Z(b)|\psi\rangle$ are computationally indistinguishable. We show this by performing a reduction to an adversary that breaks assumption (**F.5**). Fix any $i \in \{1,\dots n\}$ and suppose for contradiction that there exists an efficient observable R such that

$$\big|\langle\psi|R|\psi\rangle - \langle\psi|Z(e_i)RZ(e_i)|\psi\rangle\big| > \frac{1}{q(\lambda)}\,,$$

for some polynomial q and where the left-hand side should be understood on expectation over the creation of a state $|\psi\rangle$ according to the first three steps of protocol $\mathfrak{V}_n$. Let i be a position in which $b_i \neq 0$. Our goal is to reach a contradiction with (**F.5**). Towards this we construct an adversary $\mathcal{A}$ to the collapsing game that underlies assumption (**F.5**). Upon input pk, $\mathcal{A}$ creates the state $|\psi\rangle$ and returns to the challenger only the m-qubit register X_i. Note that the first part of the challenger's actions in the game does not change $|\psi\rangle$, since the prover has already collapsed it to a pair of preimages. The two cases correspond to the challenger returning either the mixed state $\sum_{b\in\{0,1\}} Z_{b,i}|\psi\rangle\langle\psi|Z_{b,i}$ or $|\psi\rangle\langle\psi|$, where $Z_{b,i} = (\mathrm{Id} + (-1)^b Z(e_i))/2$. The adversary $\mathcal{A}$ measures R and returns the outcome. The advantage of $\mathcal{A}$ in distinguishing the two cases is

$$\Big|\langle\psi|R|\psi\rangle - \sum_b \langle\psi|Z_{b,i}RZ_{b,i}|\psi\rangle\Big| = \frac{1}{2}\big|\langle\psi|R|\psi\rangle - \langle\psi|Z(e_i)RZ(e_i)|\psi\rangle\big|\,,$$

where the equality follows by definition of $Z_{b,i}$. Since by (**F.5**) this advantage should be negligible, we deduce that for every $i \in \{1,\dots,n\}$ and every efficient observable R it must be that

$$\langle\psi|R|\psi\rangle \approx \langle\psi|Z(e_i)RZ(e_i)|\psi\rangle\,. \tag{7.7}$$

Since for any b the observable $Z(b)RZ(b)$ itself is efficient, applying (7.7) n times (with different choices of R) we deduce that for any b and efficient R,

$$\begin{aligned}\langle\psi|R|\psi\rangle &\approx \langle\psi|Z(b_1e_1)RZ(b_1e_1)|\psi\rangle\\ &\approx \langle\psi|Z(b_1e_1+b_2e_2)RZ(b_1e_1+b_2e_2)|\psi\rangle\\ &\approx \cdots\\ &\approx \langle\psi|Z(b)RZ(b)|\psi\rangle\,.\end{aligned}$$

We now extend the preceding reasoning to show that for any a of the form $a = a_i e_i + a_j e_j$ with e_i, e_j the canonical basis vectors and $a_i, a_j \in \{0,1\}$,

$$\Big|\frac{1}{2^n}\sum_b (-1)^{a\cdot b}\langle\psi|Z(b)X(a)Z(b)|\psi\rangle - \frac{1}{4}\sum_{b_i,b_j\in\{0,1\}}(-1)^{a_ib_i+a_jb_j}\langle\psi|Z(b_ie_i+b_je_j)X(a)Z(b_ie_i+b_je_j)|\psi\rangle\Big| \leq \mu(\lambda)\,, \tag{7.8}$$

for some negligible function ν. Supposing this were not the case, by the triangle inequality and an averaging argument there must exist a b such that

$$\big|\langle\psi|Z(b)X(a)Z(b)|\psi\rangle - \langle\psi|Z(b_ie_i+b_je_j)X(a)Z(b_ie_i+b_je_j)|\psi\rangle\big| > \frac{1}{q(\lambda)}\,,$$

for some polynomial q. This leads to a contradiction with (**F.5**) using the same reasoning as before, because from the point of view of the statement of (**F.5**) for qubits not in positions i and j, the observable $X(a)$ is efficient, as its computation only requires trapdoors td_i and td_j.

To obtain the second part of the claim it remains to handle the $Z(e_i)$ and $Z(e_j)$ operators. For the positions i and j the associated trapdoor information *is* used in the computation of $X(a)$, so the preceding reasoning cannot be applied. Instead, we proceed similarly to the proof of Lemma 6.6, by reduction to the adaptive hardcore bit property, assumption (**F.2**). Note that if $a_i = 0$ or $a_j = 0$ then our task is exactly the task handled in Lemma 6.6. So assume $a_i = a_j = 1$. We perform a reduction to Lemma 6.6 via a simple hybrid argument. Suppose for the sake of contradiction that

$$\Big|\langle\psi|X(a)|\psi\rangle - \frac{1}{4}\sum_{b_i,b_j\in\{0,1\}}(-1)^{a_ib_i+a_jb_j}\langle\psi|Z(b_ie_i+b_je_j)X(a)Z(b_ie_i+b_je_j)|\psi\rangle\Big| > \frac{1}{q(\lambda)}\,.$$

Then by the triangle inequality and averaging it must be that either

$$\Big|\langle\psi|X(a)|\psi\rangle - \frac{1}{2}\sum_{b_i\in\{0,1\}}(-1)^{a_ib_i}\langle\psi|Z(b_ie_i)X(a)Z(b_ie_i)|\psi\rangle\Big| > \frac{1}{q(\lambda)}\,,$$

or

$$\Big|\langle\psi|Z(b_ie_i)X(a)Z(b_ie_i)|\psi\rangle - \frac{1}{2}\sum_{b_j\in\{0,1\}}(-1)^{a_jb_j}\langle\psi|Z(b_ie_i+e_jb_j)X(a)Z(b_ie_i+e_jb_j)|\psi\rangle\Big| > \frac{1}{q(\lambda)}\,.$$

The first case is ruled out directly by Lemma 6.6. The second case is ruled out by the same lemma, simply considering a prover that creates the state $Z(b_j)|\psi\rangle$ instead of $|\psi\rangle$ at the 3rd step (i.e. the step where $|\psi\rangle$ is defined). □

7.4 An n-qubit verification protocol

The following theorem is the main result of the past four lectures. It generalizes Proposition 6.9 to the case of an n-qubit Hamiltonian.

Theorem 7.7. *Let $\mathcal{F}$ be a function family satisfying (**F.1**), (**F.2**), (**F.3**), (**F.4'**) and (**F.5**). Let H_C be an n-qubit Hamiltonian of the form* (5.3) *and $\delta, \varepsilon > 0$ accuracy parameters. Then the verification protocol from Figure 7.1 has the following properties:*

1. *(Completeness:) For any n-qubit state $|\varphi\rangle$, there is a QPT prover that is accepted with probability* 1 *in the protocol and such that the value o returned by the verifier at the end of the protocol satisfies* $\mathrm{E}[o] = \langle\varphi|H|\varphi\rangle$.

2. *(Soundness:) For any QPT prover that is accepted with probability* 1 *in the protocol, there is an n-qubit state ρ such that the value o returned by the verifier at the end of the protocol satisfies* $\Pr(|o - \mathrm{Tr}(H\rho)| > \delta) \leq \varepsilon$.

Remark 7.8. The protocol in Figure 7.1, as the one in Figure 6.1, involves N repetitions of an elementary 4-message procedure. It is possible to parallelize the protocol to a single repetition in which the prover is asked to perform measurements on all N qubits of a ground state of $H_{\mathcal{C}}$. This however requires more work, because in the parallelized protocol the verifier needs to request "mixed" measurements from the prover; see Remark 7.1.

Remark 7.9. We pause to insist on how amazing Theorem 7.7 is. Due to Kitaev's circuit-to-Hamiltonian construction (Section 5.2.1) it is known that, under the widely believed assumption that QMA $\neq$ QCMA (where QCMA is the class of languages that admit classical proofs verifiable by QPT verifiers), there exist families of Hamiltonians of the form $H_{\mathcal{C}}$ such that any sufficiently low-energy eigenstate of $H_{\mathcal{C}}$ cannot have a simple classical description; in particular, there is no small quantum circuit to prepare such eigenstates, they must have high entanglement, etc. Yet Theorem 7.7 states that through an efficient classical interaction with a device that has the ability to prepare such states it is possible to *efficiently* verify their *existence*. There are two ways in which one might aim to strengthen that statement. First, in the spirit of "proofs of knowledge" we might aim to show that the prover *has* such a state, and not only that it *exists*. Showing this requires a formalization of the notion of the prover "having" a certain quantum state, but it can be done without any modification to the protocol itself; see [VZ20]. Second, in the spirit of our "test for a qubit" we might aim to show that the prover *has n qubits*. This we do not know how to show in the computational setting: it is an open question. (See the full notes at http://users.cms.caltech.edu/ vidick/teaching/fsmp/fsmp.pdf for how to achieve this broader goal under a different assumption, that of spatial assumption between two provers.)

Remark 7.10. The assumption that the prover succeeds with probability 1 that is made in the soundness statement is not difficult to relax; see Remark 6.10.

Proof. The completeness statement is entirely analogous to the same statement for Proposition 6.9. In slightly more detail, at each of the N iterations the honest prover prepares a fresh copy of the state $|\varphi\rangle$ and then applies the procedure described in the proof of Lemma 6.8 independently to each of the n qubits of $|\varphi\rangle$, using the key pk_i for the i-th qubit and obtaining an outcome y_i. For each qubit the post-measurement state is in an m-qubit register $\mathcal{X}_i$ that the prover measures in the standard basis in case of challenge $c = 0$, and Hadamard basis in case $c = 1$. It can then be verified by direct calculation that in case $c = 0$ for any pair $i \neq j$ the parity $(-1)^{b(x_i)+b(x_j)}$ is distributed as a measurement of $\sigma_Z(e_i + e_j)$ on $|\varphi\rangle$, and similarly in case $c = 1$ for any pair $i \neq j$ the parity $(-1)^{d_i\cdot(x_{i,0}+x_{i,1})+d_j\cdot(x_{j,0}+x_{j,1})}$ is distributed as a measurement of $\sigma_X(e_i + e_j)$ on $|\varphi\rangle$.

For soundness we use Lemma 7.6. The lemma shows that for any iteration $t = 1, \ldots, N$ in the protocol we can define a state ρ_t such that averaging over the verifier's choice of qubits i and j it holds that, whenever $c = 0$ then

$$\mathrm{E}\big[J_{ij}\,(-1)^{b(x_i)}(-1)^{b(x_j)}\big] \;=\; J_{ij}\,\mathrm{Tr}\big(\sigma_{Z,i}\sigma_{Z,j}\rho_t\big)\;.$$

and whenever $c = 1$ then

$$\mathrm{E}\big[J_{ij}\,(-1)^{d_i\cdot(x_{i,0}+x_{i,1})}(-1)^{d_j\cdot(x_{j,0}+x_{j,1})}\big] \;\approx\; J_{ij}\,\mathrm{Tr}\big(\sigma_{X,i}\sigma_{X,j}\rho_t\big)\;,$$

where the approximation is up to some negligible quantity in λ. Averaging these two quantities we see that on average over all the rounds,

$$\begin{aligned}
\mathrm{E}[o] &\approx \frac{1}{N}\sum_{t=1}^{N}\sum_{i\neq j}\Big(-\frac{1}{2}J_{ij}\operatorname{Tr}\big(\sigma_{Z,i}\sigma_{Z,j}\rho_t\big)-\frac{1}{2}J_{ij}\operatorname{Tr}\big(\sigma_{X,i}\sigma_{X,j}\rho_t\big)\Big)\\
&=\frac{1}{N}\sum_{t=1}^{N}\operatorname{Tr}\big(H_{\mathcal{C}}\rho_t\big)\\
&=\operatorname{Tr}\big(H_{\mathcal{C}}\rho\big)\ ,
\end{aligned}$$

where we defined $\rho=\frac{1}{N}\sum_t\rho_t$. The more quantitative statement given in the soundness part of the theorem follows directly by using a martingale concentration argument, provided the constant C in the definition of N is chosen large enough. □

7.5 Construction of a claw-free function family $\mathcal{F}$

The presentation of this section is adapted from [Vid20].

In Section 4.3 we have identified four assumptions (we added a fifth one in Section 7.3) on a family of functions $\{f_{pk(\lambda)}:\{0,1\}^{m(\lambda)}\to\{0,1\}^{m(\lambda)}\}_{\lambda\in\mathbb{N}}$, such that the five assumptions together are sufficient for the resulting delegated computation protocol to be sound. Can the five assumptions be simultaneously satisfied? Strictly speaking, we do not know the answer. In this section we sketch a construction that *nearly* satisfies the assumptions. The construction appears in [BCM$^+$18], and a mild modification of it is used in Mahadev's protocol. Even though the desired assumptions will not all be strictly satisfied by the construction,[2] it is possible to verify that the protocol itself remains sound.

7.5.1 The LWE problem

Our starting point is the *Learning with Errors* problem, introduced by Regev [Reg09]. The hardness of this problem has become a widely used computational assumption in cryptography, for at least three reasons. The first is that it is very versatile, allowing the implementation of advanced primitives such as fully homomorphic encryption [Gen09, BV14], attribute-based encryption [GVW15], program obfuscation [WZ17, GKW17], traitor tracing [GKW18], and many others. The second is that the assumption can be reduced to the hardness of *worst-case* computational problems on lattices: an efficient procedure that breaks the LWE assumption *on average* can be used to solve the closest vector problem in (almost) any lattice. The third reason, that is most relevant to the use of the LWE assumption made here, is that in contrast to the RSA assumption on the hardness of factoring or the discrete logarithm problem so far it is believed that the LWE problem may be hard for quantum computers, so that cryptographic schemes based on it remain (to the best of published knowledge) secure against quantum attacks.

The LWE assumption comes in multiple flavors, all roughly equivalent. Here we formulate the *decisional LWE* assumption on the difficulty of distinguishing samples from two distributions. To state the problem, fix a size parameter $n\geq 1$, an integer modulus $q\geq 2$, a number of equations $m\geq n\log q$, and an

[2] In particular, we construct functions from $\mathbb{Z}_q^m$ to Z_q^m for some q that is required to be large and may not necessarily be chosen even. The definition of assumption (**F.2**) considers equations modulo 2, and this is naturally tailored to the capabilities of a quantum prover, for whom it is possible to generate such equations by measuring in the Hadamard basis. The family of functions constructed in this section can be shown to possess the hardcore bit property over $\mathbb{Z}_q$, but proving it over $\mathbb{Z}_2$ requires more work.

error distribution χ over $\mathbb{Z}_q$.[3] Given χ, write χ^m for the distribution over $\mathbb{Z}_q^m$ that is obtained by sampling each entry of a vector independently according to χ. The decisional LWE assumption is the following.

> *(Decisional LWE, informal)* Let A be a uniformly random matrix in $\mathbb{Z}_q^{m\times n}$, s a uniformly random vector in $\{0,1\}^n$, e a random vector in $\mathbb{Z}_q^m$ drawn from χ^m, and r a uniformly random vector in $\mathbb{Z}_q^m$. Then no classical or quantum probabilistic polynomial-time procedure can distinguish $(A, As+e)$ from (A,r).

Note that the distribution of $(A, As+e)$ and the distribution of (A,r) are in general very far from each other: provided m is sufficiently larger than n a random vector r will not lie in the column span of A, nor even be close to it. What the (decisional) LWE assumption asserts is that, even though in principle these distributions are far from each other, it is computationally difficult, given a sample from the one or the other, to tell which is the case. Note that without the error vector e the task would be easy: given (A,y), solve for $As = y$ and check whether the solution has coefficients in $\{0,1\}$. The LWE assumption is that the inclusion of e makes the task substantially more arduous. In particular, it is well-known that Gaussian elimination is very sensitive to errors, which rules out the most natural approach.

The definition we gave is informal because we have not specified how the parameters n,m and q should be chosen as a function of the security parameter λ, and we have not specified the distribution χ. In general one can make the decisional LWE assumption for any choice of these parameters—but for some choices the assumption will be invalidated by existing algorithms. We comment on come choices of parameters that are made in cryptography. The integer n should generally be thought of as commensurate with the security parameter λ, i.e. $n = \Theta(\lambda)$. The modulus q should be at least polynomial in n, but can be as large as exponential; this will be the case in our construction. The error distribution χ can be chosen in multiple ways. A common choice is to set χ a discretized centered Gaussian distribution with variance αq, for some small parameter α (typically chosen as an inverse polynomial function of n); this is generally denoted $D_{\mathbb{Z}_q,\alpha q}$. For more details on LWE and its applications, we refer to the survey [P$^+$16].

7.5.2 Construction

To specify the function family $\mathcal{F}$ we first describe how public and private parameters for the function are chosen. Let λ be the security parameter (i.e. the number 2^λ is thought of as an estimate of the time required to break assumptions such as (**F.2**)).

First, integers n,m and a modulus q are chosen such that $n = \Omega(\lambda)$, $q \geq 2$ is a prime, and $m = \Omega(n\log q)$. Then, a matrix $A \in \mathbb{Z}_q^{m\times n}$ is sampled at random, together with a "trapdoor" in the form of a matrix $R \in \mathbb{Z}_q^{\ell\times m}$, where $n \leq \ell \leq m$ is a parameter. The sampling procedure has the property that the distribution of A is statistically close to uniform, and R is such that $G = RA \in \mathbb{Z}_q^{\ell\times n}$ is a "nice" matrix, in the sense that given $b = Gs+e$, for any $s \in \mathbb{Z}_q^n$ and e small enough, it is computationally easy to recover s.[4] That such a sampling procedure would exist and be efficiently implementable is non-trivial, and relies on the underlying lattice structure given by the columns of A; see [MP12]. Finally, a uniformly random $s \in \{0,1\}^n$, and a random $e \in \mathbb{Z}_q^m$ distributed according to $D_{\mathbb{Z}_q,\alpha q}$ with α of order $1/(\sqrt{mn\log q})$,[5]

[3] The use of the parameters n,m and q is local to this section. In particular, the m that specifies the domain and range of the function f_{pk} is not identical to the m here; see below.

[4] One can think of G as a matrix whose rows are almost orthonormal, so that Gaussian elimination on G induces only small propagation of the errors.

[5] The precise choice of α is delicate, and the parameters given here should only be treated as indicative; we refer to [BCM$^+$18, Section 8] for the right setting of parameters.

are sampled. The public information is $pk = (A, z = As + e)$. The trapdoor information is the pair $td = (R, s)$. Note that pk is not uniformly distributed, but pairs (pk, td) can be sampled in randomized polynomial time in λ.

Next we discuss how the function $f = f_{pk}$ can be evaluated, given the public parameters $pk = (A, z)$. We define two functions f_0, f_1 that should be understood as $f(0\|\cdot)$ and $f(1\|\cdot)$ respectively. Each function goes from $\mathbb{Z}_2^{wn}$ to $\mathbb{Z}_2^{wm}$ for $w = \lceil \log q \rceil$. For $b \in \{0,1\}$ the function f_b takes as input an $x \in \mathbb{Z}_q^n$ (that can be seen as an element of $\mathbb{Z}_2^{wn}$ through its binary representation) and returns $Ax + e' + bz$, which is an element of $\mathbb{Z}_q^m \subseteq \mathbb{Z}_2^{wm}$. Here, e' is a vector sampled at random from a distribution $D_{\mathbb{Z}_q,\alpha' q}$ such that α' is "much larger" than α. The inclusion of e' makes f a "randomized" function, which is the main way in which the construction differs from the requirements expressed in Section 4.3. A formal way around this is to think of f_b as the function that returns not $Ax + e' + bz$, but the *distribution* of $Ax + e' + bz$, when $e' \sim D_{\mathbb{Z}_q,\alpha' q}$ and all other variables are fixed. In practice, the evaluation of f on a quantum computer (as required of the honest prover in the verification protocol) involves preparing a weighted superposition over all error vectors, and computing the function in superposition.

We would, of course, rather do away with this complication. Why is the error vector necessary? It is there to satisfy the important requirement that the functions f_0 and f_1 are injective with overlapping ranges, so that f itself is 2-to-1. Injectivity follows from the existence of the trapdoor for A and an appropriate setting of the standard deviation of the error distribution, which guarantee that (given the trapdoor) x can be recovered from $Ax + e' + bz$ (with high probability over the choice of e'). To make the function ranges overlap, we need the distribution of $Ax + e'$ to be statistically close to the distribution of $Ax' + e' + z = A(x' + s) + (e' + e)$. The first distribution considers an arbitrary vector in the column span of A, shifted by e; the second considers the same, except that the shift is by $(e' + e)$. For the two distributions to (almost) match, we need the distribution of e' to (almost) match the distribution of $e + e'$. This is possible as long as the standard deviation $\sigma' = \alpha' q$ is substantially larger than the standard deviation $\sigma = \alpha q$; provided this holds it is an exercise to compute the statistical distance between the two Gaussian and verify that it can be made very close to 1.

With this important caveat in place, we have specified the function f and verified property (**F.1**). Property (**F.3**) follows from the existence of the secret information $td = (R, s)$. Given a $b \in \{0,1\}$ and an element $y = Ax + e' + bz = A(x + bs) + (e' + be)$ in the range of f_b it is possible to use the trapdoor matrix R to recover $x + bs$ and subtract bs to deduce the preimage x of z under f_b. Property (**F.4**) holds trivially from the construction. Note that the function f has domain and range that are different. In particular, here the domain is larger than the range, and in case q is not a power of 2 f is only defined on a subset of its natural domain $\mathbb{Z}_2^{wn}$. These points are not very important and can be ignored at the level of our discussion.

Showing the hardcore bit property (**F.2**) and the collapsing condition (**F.5**) require more work, and we refer to [BCM$^+$18] for a detailed exposition.[6] Similar "hardcore bit" properties to (**F.2**) have been shown for many LWE-based cryptographic schemes (see e.g. [AGV09]). Usually the property states that "for any vector $d \in \mathbb{Z}_q^n \backslash \{0\}$, the value $d \cdot s \in \mathbb{Z}_q$ is indistinguishable from uniform, even given a sample $(A, As + e)$". Our property (**F.2**) is subtly stronger, in that the adversary may choose the vector d itself, possibly as a function of the sample $(A, As + e)$. An additional difficulty stems from the specific equation that the adversary is asked to return. In the definition of Assumption (**F.2**) this is a d such that $d \cdot (x_0 + x_1) = 0$, where x_0, x_1 are the *binary representation* of the two preimages in $\mathbb{Z}_q^n$ of the prover's first message string $y \in \mathbb{Z}_q^m$. (The use of the binary representation comes from the requirements on the honest prover, that is asked to perform a measurement in the Hadamard basis, yielding a binary string of outcomes.) So here

[6] The collapsing condition is not shown in [BCM$^+$18]. It is implicitly shown in [Mah18], where it can be seen to follow from property 2 in Definition 4.4 of an extended trapdoor claw-free family. (The connection is made explicit in [GV19].)

$x_0 = (0, r_0)$ and $x_1 = (1, r_1)$ such that r_0, r_1 are binary representations for two elements $x'_0, x'_1 \in \mathbb{Z}_q^n$ such that $x'_1 = x'_0 - s$ over $\mathbb{Z}_q$. Since the binary representation is not linear the equation obtained is not directly a linear equation in the secret s. Completing the argument showing that a procedure that returns the information asked for in Assumption (**F.2**), i.e. the pair $(x = (b, r_b), d)$, can be turned into a procedure that breaks the decisional LWE assumption, requires a little more work; this is where we need to assume that the secret vector s is a binary vector.

Lecture 8

Multiprover interactive proof systems

In lecture 5 we considered the class $\mathrm{IP}[\mathcal{P}]$ of languages that can be decided by a BPP verifier interacting with a prover in class $\mathcal{P}$. We showed how to construct verification protocols for all of BQP in this model: informally, the result of the previous three lectures is that $\mathrm{BQP} \in \mathrm{IP}[\mathrm{BQP}]$ under the Learning with Errors assumption.[1] What if we do not wish to make computational assumptions? A first motivation for this is that we might simply not wish to rely on relatively untested assumptions — after all, isn't it likely that a few decades of algorithmic research (and even less for quantum algorithms) have barely scratched the surface of the possible ways of approaching a problem such as LWE or even e.g. factoring? A second motivation is that we could aim for more: firstly, in terms of complexity — almost by definition the class $\mathrm{IP}[\mathrm{BQP}]$ lies in BQP (and giving any more power to the prover risks breaking the computational assumption); what if we are interested in languages outside BQP?[2] Secondly, in terms of structural characterizations — in particular, remember how we stopped short of showing that the prover in the Mahadev protocol "has n qubits": can we achieve such a characterization in a different model?

In the next three lectures we switch gears and replace the use of computational assumptions by an assumption of spatial isolation. In this new model the verifier has the ability to interact with two (or more) provers that are restricted to acting locally on their respective quantum systems. This is the model that we already encountered in Section **??**. As we will see this physical (and, once properly formalized, mathematical) limitation on prover strategies will allow us to go further along the two motivating directions outlined in the preceding paragraph.

In this lecture we first introduce the model from a complexity-theoretic standpoint, discuss the recent characterization $\mathrm{MIP}^* = \mathrm{RE}$, and examine some consequences. In the following two lectures we introduce techniques that build towards a proof of the equality $\mathrm{MIP}^* = \mathrm{RE}$ by developing efficient tests for increasing numbers of qubits and increasingly complex computations.

8.1 Multiprover interactive proofs with entangled provers

We start with the main complexity-theoretic definition. Recall that a *promise language*[3] $L = (L_{yes}, L_{no})$ is specified by a pair of disjoint subsets L_{yes}, L_{no} of $\{0,1\}^*$ and that a *complexity class* is a collection of

[1] To be precise we should state what variant of the LWE assumption the result relies on.

[2] The practicality of a protocol in which the honest prover lies outside of BQP is almost entirely besides the point — our goal here is to study the problem of verification *per se*, and exploring it in the "high-complexity" regime is certain to yield useful insights which, who knows, may eventually lead to practical consequences of their own.

[3] Here the *promise* refers to the fact that it is not required that $L_{yes} \cup L_{no} = \{0,1\}^*$

languages.

Definition 8.1. The class MIP* is the class of promise languages $L = (L_{yes}, L_{no})$ such that there is a classical polynomial-time Turing machine M that on input 1^n returns the description of classical circuits for the verifier V_n in an interactive protocol with *two* quantum provers A and B such that:

- (Completeness:) There is a family of quantum provers $\{A_n, B_n\}_{n\in\mathbb{N}}$ such that for all $x \in L_{yes}$ the interaction of $V_{|x|}$ and $A_{|x|}, B_{|x|}$ on common input x accepts with probability at least $\frac{2}{3}$.
- (Soundness:) For any family of quantum provers $\{A_n, B_n\}_{n\in\mathbb{N}}$, for all $x \in L_{no}$ the interaction of $V_{|x|}$ and $A_{|x|}, B_{|x|}$ on common input x accepts with probability at most $\frac{1}{3}$.

Some comments on the definition are in order. Following tradition we called the provers A and B rather than P_1 and P_2; A stands for "Alice" and B for "Bob", a personification that is inspired from cryptography.[4] In general one may allow interaction with more than two provers; however the two-prover setting is sufficiently interesting for our purposes. (Furthermore, it can be shown that in purely complexity-theoretic terms there is no gain to considering more than 2 provers.) The number of rounds of interaction is left implicit in the definition; since V_n is polynomial-size there can be at most polynomially many rounds of interaction. Soon we will restrict ourselves to single-round protocols, which consist of a message from the verifier to each prover followed by an answer from each prover; again both for our purposes and in terms of complexity-theoretic expressive power this is without loss of generality.

Note that we did not (and will not) restrict the computational power of the provers — in fact, we did not even precisely specify what collection of strategies they may employ. For the time being we stay with the informal prescription that the provers may employ any quantum strategy that can be implemented *locally*, in *finite dimension*, and *without communication* — typically, local operations augmented with measurements on a shared entangled state that may have been agreed on prior to the protocol execution. We will see later how to formalize this more precisely.

The goal in complexity theory is to relate different classes of languages . This is especially interesting when the classes are defined in very different terms, as relations between them can provide insights into different models of computation. A pertinent example is the famous equality IP $=$ PSPACE due to [LFKN92, Sha92]. Among the two classes, PSPACE is the simplest to define: this is the class of all languages that can be decided using a polynomial amount of space, and arbitrary time. A complete problem for PSPACE is the *quantified Boolean formula* (QBF) problem, which is to decide if a formula of the form $\exists x_1 \forall x_2 \exists x_3 \cdots (x_1 \wedge x_2 \wedge \neg x_3) \vee (\cdots)$ is satisfiable. Clearly this can be done in polynomial space by trying out all possibilities; it is also possible to show that any problem that is solvable in PSPACE can be reduced to this one, and so we say that QBF is *complete* for PSPACE. The class IP is defined very differently: it is the class of languages L such that membership $x \in L_{yes}$ can be decided efficiently by a randomized polynomial-time verifier interacting with a single infinitely powerful prover (so this is the single-prover analogue of MIP*). While it is not too hard to show that IP $\subseteq$ PSPACE, the converse inclusion is not easy at all — to see why, try coming up with a verification protocol for the QBF problem, and keep in mind that the prover is not to be trusted!

Our goal is to characterize the complexity of MIP* in terms of other complexity classes, with the hope of gaining insights about computation, entanglement, and verification of quantum devices. Before we do this let's first review what is known about the classical analogue of MIP*, in which the provers are restricted

[4] Indeed the model of multi-prover interactive proof systems is first introduced by a team of cryptographers [BOGKW19] motivated by the development of *zero-knowledge* proof systems.

to classical strategies. This restriction affects both the completeness and soundness requirements in Definition 8.1, and so generally any stipulation of the set of allowed strategies for the provers will lead to a different complexity class.

8.1.1 Classical multiprover interactive proof systems

The * in MIP* refers to the fact that provers are allowed to use entanglement. If we omit it we get the class MIP of languages that have classical multiprover interactive proof systems. It was shown by Babai, Fortnow and Lund in the early 1990s that MIP $=$ NEXP. This was shown shortly after the aforementioned result IP $=$ PSPACE, which characterizes the unexpectedly large verification power of single-prover interactive proof systems.

Let's recall how MIP $=$ NEXP is shown. The inclusion of MIP $\subseteq$ NEXP is not hard to obtain. To show it we give a non-deterministic exponential time algorithm that exactly computes the maximum acceptance probability of the verifier in an MIP protocol. This algorithm can therefore, given an instance x and a description of the verifier $V_{|x|}$, determine whether $x \in L_{yes}$ (the maximum success probability is $\geq \frac{2}{3}$) or $x \in L_{no}$ (the maximum success probability is $\leq \frac{1}{3}$), promised that one of them is the case, and thus decide any language $L \in$ MIP; thus MIP $\subseteq$ NEXP follows. To devise such an algorithm first observe that in order to do so it suffices to consider the maximum over deterministic strategies, as for any randomized strategy there is a deterministic one that succeeds with at least the same probability. Now note that a deterministic strategy is specified by a list of answers to each possible question for each of the provers. There are at most exponentially many questions because the bit representation of each question must have polynomial length (since the verifier runs in polynomial time) and similarly for answers. Finally, the success probability of a deterministic strategy can be computed exactly in exponential time simply by executing the verification procedure on each possible tuple of questions, weighted by the probability of the question being asked. Therefore, a non-deterministic algorithm can, in exponential time and space, guess an optimal strategy and compute its success probability.

The reverse inclusion, NEXP $\subseteq$ MIP, is harder. To get a hint of how it is shown, consider the problem of verifying that an exponential-size graph is 3-colorable. Formally, an instance x of this problem is specified by a pair $x = (1^n, C)$ where 1^n denotes an integer n written in unary, and C is the description of a classical circuit $C : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. Any x that does not take this form is neither in L_{yes} nor in L_{no}, and need not be considered any further.[5] The circuit C implicitly specifies a graph $G_x = (V_x, E_x)$ with vertex set $V_x = \{0,1\}^n$ and edge set E_x such that $(i,j) \in E_x$ if and only if $C(i,j) = 1$. Then L_{yes} (resp. L_{no}) is the set of all strings x such that G_x is well-defined and is 3-colorable (resp. not 3-colorable). It is known that the language $L = (L_{yes}, L_{no})$ is complete for NEXP; intuitively this is a "scaled-up" version of the result that 3-coloring of n-vertex graphs is NP-complete. Now consider the following description for the actions of the verifier in a candidate multiprover interactive proof system for L:

1. The verifier parses its input x as $x = (1^n, C)$.

2. The verifier selects a pair of vertices (i,j) uniformly at random in $\{0,1\}^n \times \{0,1\}^n$. She sends i to Alice and j to Bob.

3. Alice and Bob each reply with a color $a, b \in \{0,1,2\}$.

4. The verifier accepts if and only if any of the following conditions hold: $C(i,j) = 0$ (there is no edge); $i = j$ and $a = b$ (same color for identical vertices); $C(i,j) = 1$ and $a \neq b$ (different colors for

[5] As usual, we consider that circuits are represented in some given, fixed manner, e.g. as a list of gates and bits that they act on.

neighboring vertices).[6]

It is clear that this protocol has completeness 1: whenever G_x is 3-colorable there is a winning strategy for the provers. Moreover, a moment's thought will reveal that if G_x is not 3-colorable then there is no perfect winning strategy; hence the maximum probability of success in this case is at most $1 - 2^{-\Omega(n)}$ (because any strategy must fail on at least one question). While this is a separation between the two cases, it is not sufficient to establish soundness, which requires that the maximum probability of success for an $x \in L_{no}$ be at most $\frac{1}{3}$.

What the proof of the inclusion NEXP $\subseteq$ MIP shows is that there is in fact a much better verifier, somewhat more involved than the one that we described here, which is such that whenever the graph is not 3-colorable then the maximum success probability is at most $\frac{1}{3}$. Achieving such a protocol essentially entails finding an efficient method that, informally, maps any graph to another graph of polynomially related size such that graphs that are 3-colorable are mapped to graphs that remain 3-colorable, but graphs that are not 3-colorable are mapped to graphs that are *very far* from 3-colorable. Achieving this can be done using advanced tools from the theory of error-correcting codes; we will not be able to say more in this lecture and refer the interested reader to e.g. [AB09].[7]

8.1.2 Interactive proof systems with entangled provers

Our focus is the class MIP*. What does the characterization MIP $=$ NEXP say about it? Not much! The most important point to realize here is that allowing the provers to use entanglement is a double-edged sword:

First, it can affect the soundness property by allowing the provers to "cheat", meaning achieve a higher success probability. We already saw a good example of this with the Magic Square game. While this game doesn't quite look like the 3-coloring protocol we introduced in the previous section, by transforming it it is possible to come up with explicit instances of the latter that are associated with non-colorable graphs but such that there nevertheless exists a quantum strategy which succeeds with probability 1; see for example [Ji13].

As a result we are unable to transfer the lower bound NEXP $\subseteq$ MIP in a "black-box" manner, and the only trivial lower bound on MIP* is PSPACE, as clearly the verifier can ignore all but one of the provers and execute any classical IP protocol with the remaining prover. In fact it is interesting to note that such a "collapse" to IP does take place when one allows even more power to the provers, in the form of arbitrary non-signaling strategies as defined in Section **??**. Indeed it is not hard to see that the non-signaling constraints are linear, so that it is possible to write the optimal success probability of non-signaling provers in a multiprover interactive proof system as an exponential-size linear program (LP). Using that linear programs can be solved in time polynomial in their size it can be shown that the class of interactive proof systems with non-signaling provers, denoted MIP^{ns}, lies in EXP. Furthermore, if the number of provers is fixed to 2 and the number of rounds of interaction to 1 then the class "collapses" even further to PSPACE, because the associated LP can be solved more efficiently than a general LP; see [Ito10].

Second, entanglement can also affect the completeness property by increasing the power of the provers in the "honest" case. If we start with a classical protocol for a problem in NEXP this is not so interesting, because we already know that the provers have a good strategy without entanglement — we are not making

[6] One may modify this protocol by having the verifier only send pairs (i,j) such that either $i = j$ or (i,j) is an edge, since the other case is an automatic "free ride" for the provers; we gloss over this point here.

[7] Technically such a reduction is not obviously necessary, because the definition of MIP allows more complicated protocols than the 3-coloring game described here. Nevertheless, using appropriate manipulations it is possible to show that any proof of NEXP $\subseteq$ MIP does imply such a reduction.

use of the fact that they can do even better with entanglement, and indeed this fact is a new nuisance that we have to deal with in order to establish the soundness property. But what if we start from a more complex problem, that does not necessarily lie in NEXP, and attempt to design a protocol such that completeness *requires* the use of entanglement?

To see how far one might hope to go in this direction we ought to think about *upper bounds* on MIP*. Recall from the previous section that for MIP we simply enumerated over all possible strategies. In the quantum setting it is not so direct: since we do not place a priori bounds on the complexity of the provers, it is unclear what dimension one should choose in order to find an optimal strategy. If one was able to show an upper bound on the dimension that is sufficient to approach the optimal success probability (as a function of the size of the protocol) then one would automatically get a corresponding upper bound on the complexity of MIP*. However, no such bound is known! The only upper bound on MIP* is the following folklore result:

Lemma 8.2. MIP$^* \subseteq$ RE*, the set of recursively enumerable languages.*

Proof. Recall that a language $L = (L_{yes}, L_{no})$ is recursively enumerable if there exists a Turing machine such that on input x, if $x \in L_{yes}$ then the Turing machine eventually halts and accepts, whereas if $x \in L_{no}$ then the Turing machine may either halt and reject, or it may never halt.

Consider the Turing machine M that on input x specifying a verifier $V_{|x|}$ searches in increasing dimension and with increasing accuracy for a good strategy in the associated protocol. Since we have not introduced a precise formalism for strategies in MIP* protocols — we will do so for two-prover one-round protocols in Section 8.2.1 — we cannot make this too precise. At present it is sufficient to think intuitively that each prover is specified by a dimension of the Hilbert space on which they act, and for each possible question they may receive, in any round, a POVM on their space that is used to determine an answer; these POVM act on an initial quantum state that lies in the tensor product of the prover's Hilbert spaces. (Any unitary actions the provers may take can be incorporated in the POVMs.) For any given dimension d and accuracy ε the space of strategies in dimension at most d can be discretized to a finite set such that the optimum success probability over elements of that set will be within an additive ε of the optimum over all strategies in dimension at most d.

If $x \in L_{yes}$ by definition there must exist a finite dimension d and a strategy in dimension d that succeeds with probability at least (say) $\frac{2}{3} - \frac{1}{100}$; eventually, taking into account discretization errors M will identify a strategy that succeeds with probability at least $\frac{2}{3} - \frac{2}{100}$ and halt with acceptance, having successfully ruled out the case that $x \in L_{no}$. However, in case $x \in L_{no}$ the Turing machine will never find a strategy with success larger than $\frac{1}{3} + \frac{1}{100}$ (where the $\frac{1}{100}$ accounts for possible discretization errors and can be made arbitrarily small), but it will not be able to rule out the existence of such a strategy either; indeed, for all it knows such a strategy may exist in "just one more dimension". □

For a long time it was unclear where the complexity of MIP* lies, between the two "trivial" extremes of IP and RE. In 2012 Ito and the author showed that NEXP $\subseteq$ MIP* by adapting the proof of NEXP $\subseteq$ MIP by Babai et al. In the past few years better lower bounds were obtained. Quite astonishingly, in 2018 Natarajan and Wright [NW19] showed that NEEXP $\subseteq$ MIP*. One reason that this is "astonishing" is because NEEXP is a strictly (unconditionally) larger class that NEXP, and so their result established unconditionally that the presence of entanglement *increases* the verifier's ability to verify languages, even though the latter's complexity has not changed at all (it remains classical polynomial-time)! Building on this result in 2020 Ji et al. [JNV$^+$20a] obtained the following characterization.

Theorem 8.3. MIP* = RE.

A complete problem for the class RE is the *halting problem*: given the description of a Turing machine M as input, does M eventually halt? What Theorem 8.3 shows is that this problem, even though it is *not decidable*, can be efficiently *verified* by asking questions to two provers sharing entanglement. In purely complexity-theoretic terms this is an extremely surprising result in and for itself; note that RE contains *any* bounded time or space complexity class — and much more. The following two lectures will be devoted to a sketch of the main arguments that go in the proof of the theorem; these arguments involve the design of tests for multiple qubits as well as delegation protocols and so we will be on familiar terrain. Aside from the complexity theory it turns out that the characterization MIP$^* =$ RE has some interesting consequences in the foundations of quantum mechanics as well as in the theory of operator algebras which we discuss next.

8.2 Consequences

Theorem 8.3 is related to a problem in the foundations of quantum non-locality called *Tsirelson's problem*, itself connected to a problem in the theory of von Neumann's algebra usually referred to as *Connes' Embedding Problem* (CEP). Even though they have no bearing on the remainder of the course, for motivation in this section we explain those connections. We start by (re-)introducing the language of nonlocal games that we already encountered in lecture **??** and which we will generally use to talk about multiprover interactive proof systems.

8.2.1 Nonlocal games

As we will see the proof of the "hard" part of Theorem 8.3 shows that RE $\subseteq$ MIP$^*(2,1)$, where the $(2,1)$ refers to verifiers that are restricted to interacting with two provers in a single round. From now on we only only consider protocols that fall in this category. In this case once an input x has been fixed the associated verifier $V_{|x|}$ together with x itself implicitly define a two-player one-round game in the sense of lecture **??**. To be fully explicit as well as set notation, a two-player one-round game is specified by a distribution π on $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}, \mathcal{Y}$ are finite sets of *questions*, and a predicate $R : \mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B} \to \{0,1\}$, for finite sets of *answers* $\mathcal{A}$ and $\mathcal{B}$, usually written as $R(a,b|x,y)$. Using this terminology the specification of the verifiers $\{V_n\}$ in a one-round two-prover interactive proof system for a language $L = L_{yes} \cup L_{no}$ is equivalent to an implicit specification of a family of games $\{G_x\}_{x \in L_{yes} \cup L_{no}}$. Explicitly, for each x we have $G_x = (\pi_x, R_x)$ where the distribution π_x is the distribution according to which the interactive proof verifier V on input x selects questions to the players and R_x denotes the decision that V makes, accept or reject, as a function of the questions sent and the answers received.

Remark 8.4. It is known that any multiprover interactive proof system that decides a language L can be "parallelized" to a single round of interaction [KKMV09]. However, this transformation in general requires the addition of a prover. It is not known how to modify a proof system with more than two provers to one with only two provers that decides the same language; it is only known, as a consequence of the inclusions MIP$^* \subseteq$ RE (Lemma 8.2) and RE $\subseteq$ MIP$^*(2,1)$ (which follows from the proof of Theorem 8.3), that such a transformation *exists*. It is an open question whether there is a simple, or efficient, such transformation.

In general given a game game $G = (\pi, R)$, a *strategy* S for G is specified by a family of distributions $\{p(\cdot,\cdot|x,y)\}_{(x,y)\in\mathcal{X}\times\mathcal{Y}}$ on $\mathcal{A} \times \mathcal{B}$. The *success probability* of the strategy S in the game G is

$$\omega(G;S) = \sum_{x,y} \pi(x,y) \sum_{a,b} R(a,b|x,y)p(a,b|x,y) .$$

Informally this quantity is the average, over the referee's choice of questions and the player's probabilistic strategy, that the players provide valid answers to the referee. If one fixes a collection of possible strategies

$\mathcal{S}$ then one can define an associated *value* $\omega(G;\mathcal{S})$ for the game, which is the supremum success probability achievable using strategies $S \in \mathcal{S}$:

$$\omega(G;\mathcal{S}) \,=\, \sup_{S\in\mathcal{S}} \omega(G;S)\ .$$

For example, if $\mathcal{S}$ is the set of classical local strategies, i.e. all those families of distributions that take the form (3.3), then $\omega(G;\mathcal{S})$ is called the *classical value* of the game and is usually denoted $\omega(G)$. If $\mathcal{S}$ is the set of (tensor) quantum strategies, i.e. all those families of distributions that take the form (3.5), then $\omega(G;\mathcal{S})$ is called the *entangled value* of the game and is usually denoted $\omega^*(G)$. Explicitly,

$$\omega^*(G) \,=\, \sup_{|\psi\rangle,\{A_a^x\},\{B_b^y\}} \sum_{x,y} \pi(x,y) \sum_{a,b} R(a,b|x,y) \langle\psi|A_a^x \otimes B_b^y|\psi\rangle\ , \tag{8.1}$$

where the supremum is taken over all quantum states $|\psi\rangle \in \mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$ for finite-dimensional $\mathcal{H}_{\mathsf{A}}$ and $\mathcal{H}_{\mathsf{B}}$ and collections of POVMs $\{A_a^x\}$ and $\{B_b^y\}$ on $\mathcal{H}_{\mathsf{A}}$ and $\mathcal{H}_{\mathsf{B}}$ respectively, one POVM for each question x or y to Alice or Bob respectively. Using the language of games the question of characterizing the complexity of the class $\text{MIP}^*(2,1)$ boils down to determining the complexity of approximating the optimum of the optimization problem (8.1). This is because for a language $L \in \text{MIP}^*(2,1)$ the problem of determining if some input $x \in L_{yes}$ reduces to evaluating the value $\omega^*(G_x)$, where G_x is the game associated with x and the verifier in a protocol for L, with good enough approximation to differentiate between the cases where $x \in L_{yes}$ ($\omega^*(G_x) \geq \frac{2}{3}$) and $x \in L_{no}$ ($\omega^*(G_x) \leq \frac{1}{3}$).[8]

8.2.2 Computing upper bounds on $\omega^*(G)$

Let's put Theorem 8.3 aside for a moment and aim instead to contradict it by devising an algorithm that approaches the maximum success probability of quantum provers sharing entanglement in a two-prover one-round interactive proof system with verifier V. Equivalently, suppose that given an explicit game G we aim to approximate the value $\omega^*(G)$ defined in (8.1). In the proof of Lemma 8.2 we already saw an algorithm, let's call it algorithm A, that returns an increasing sequence of lower bounds

$$v_1 \leq v_2 \leq \cdots \leq v_k \leq \cdots \leq \omega^*(G)$$

by enumerating strategies in increasing dimension and with increasing level of accuracy. Using the definition (8.1) of the entangled value it is clear that $v_k \to_{k\to\infty} \omega^*(G)$. To make algorithm A into an actual approximation algorithm we need to have a sense of when to stop, e.g. when can we guarantee that $|v_k - \omega^*(G)| \leq \frac{1}{100}$?[9] A natural approach is to construct a companion algorithm B that constructs a decreasing sequence of *upper* bounds

$$w_1 \geq w_2 \geq \cdots \geq w_k \geq \cdots \geq \omega^*(G)\ .$$

Given algorithms A and B consider a third algorithm C that given a game G as input runs both algorithms in an interleaved fashion, computing v_1, w_1, v_2, w_2, etc., halts whenever $|v_k - w_k| \leq \frac{1}{100}$ and returns "YES" if and only if $\frac{1}{2}(v_k + w_k) > \frac{1}{2}$. Now suppose that both (v_k) and (w_k) converge to $\omega^*(G)$. Then C

[8] The correspondence is not entirely exact because complexity is measured as a function of the input size; for MIP^* protocols the input is directly x, whereas for games G we think of the input as an explicit description of the underlying distribution π and predicate R. In particular it is possible that the description length of G_x is exponential in the description length of $V_{|x|}$, since the latter only specifies π and V implicitly through a circuit that computes them.

[9] The bound $\frac{1}{100}$ is arbitrary; we want it to be small enough to guarantee that the algorithm can eventually distinguish $\omega^*(G) \geq \frac{2}{3}$ from $\omega^*(G) \leq \frac{1}{3}$, so any bound $< \frac{1}{6}$ would do.

always terminates. Moreover, if $\omega^*(G) \geq \frac{2}{3}$ then $w_k \geq \frac{2}{3}$ for all k and so the value returned is at least $\frac{1}{2}((\frac{2}{3}-\frac{1}{100})+\frac{2}{3}) = \frac{2}{3}-\frac{1}{50} > \frac{1}{2}$, whereas if $\omega^*(G) \leq \frac{1}{3}$ it is at most $\frac{1}{2}(\frac{1}{3}+(\frac{1}{3}+\frac{1}{100})) = \frac{1}{3}+\frac{1}{50} < \frac{1}{2}$. Thus C correctly distinguishes between the two cases.

So how do we determine such a sequence of upper bounds (w_k)? A general approach to finding an upper bound on the optimum of some optimization problem is to consider *relaxations* of the problem, i.e. optimization problems whose optimum is easier to find and is guaranteed to be at least as large as the original optimum. For example, consider the following relaxation

$$\begin{aligned} \omega^*(G) &= \sup_{|\psi\rangle,\{A_a^x\},\{B_b^y\}} \sum_{x,y} \pi(x,y) \sum_{a,b} R(a,b|x,y) \langle\psi|A_a^x \otimes B_b^y|\psi\rangle \\ &\leq \sup_{|u_a^x\rangle,|v_b^y\rangle} \sum_{x,y} \pi(x,y) \sum_{a,b} R(a,b|x,y) \langle u_a^x|v_b^y\rangle\ , \end{aligned} \tag{8.2}$$

where the supremum on the second line is over all families of vectors $|u_a^x\rangle, |v_b^y\rangle \in \mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$ such that for every x, the $\{|u_a^x\rangle\}_a$ are orthogonal and $\sum_a \||u_a^x\rangle\|^2 = 1$; similarly for the $|v_b^y\rangle$. The inequality (8.2) is verified by setting $|u_a^x\rangle = A_a^x \otimes \mathrm{Id}\,|\psi\rangle$ and $|v_b^y\rangle = \mathrm{Id} \otimes B_b^y|\psi\rangle$. So (8.2) is a relaxation of (8.1). What did we gain in the process? Crucially, since the objective function in (8.2) only depends on the inner products between the vectors, without loss of generality we can restrict the vectors to lie in a Hilbert space $\mathcal{H}$ such that $\dim(\mathcal{H}) \leq \min(|\mathcal{X}||\mathcal{A}|, |\mathcal{Y}||\mathcal{B}|)$; this is true even if the original $\mathcal{H}_{\mathsf{A}}$ and $\mathcal{H}_{\mathsf{B}}$ were much larger. This means that by exhaustive search we can find arbitrarily good approximations to the optimum (8.2), without having to go beyond a certain fixed dimension that is determined by the size of the game. In fact, (8.2) is an optimization problem that falls in the class of *semidefinite programs* (informally, linear optimization problems over affine sections of the positive semidefinite cone) and can be solved in time polynomial in its size (as opposed to exponential for exhaustive search).

So the optimum (8.2) *can* be determined efficiently. How useful is it, i.e. how good is the inequality (8.1) $\leq$ (8.2)? Unfortunately, in general there can be an arbitrarily large (multiplicative) gap between the two [JP11], and in particular it can be that $\omega^*(G) \leq \frac{1}{3}$ but (8.2)$\geq \frac{2}{3}$.[10] The relaxation we have devised is thus too coarse for us to obtain a good algorithm right away. But maybe we can do better? What we did so far consists in adding a vector variable to represent $A_a^x \otimes \mathrm{Id}\,|\psi\rangle$ and $\mathrm{Id} \otimes B_b^y|\psi\rangle$. Each of these can be thought of as a degree-1 monomial in the matrix variables $\{A_a^x, B_b^y\}$, evaluated on $|\psi\rangle$. Considering vectors obtained from higher-degree monomials would allow us to impose more constraints, as for example we could require that

$$\big(\langle\psi|A_a^x \otimes \mathrm{Id}\big) \cdot \big(A_a^x \otimes B_b^y|\psi\rangle\big) \,=\, \big(\langle\psi|\,\mathrm{Id} \otimes \mathrm{Id}\big) \cdot \big(A_a^x \otimes B_b^y|\psi\rangle\big)\ ,$$

due to $\{A_a^x\}_a$ being projective. It is not hard to think of other such constraints. For any integer $k \geq 1$ let's define

$$w_k \,=\, \sup_{\Gamma^{(k)} \geq 0} \sum_{x,y} \pi(x,y) \sum_{a,b} R(a,b|x,y) \Gamma^{(k)}_{xa,yb}\ , \tag{8.3}$$

where the supremum is taken over all positive semidefinite matrices $\Gamma^{(k)}$ of dimension $\binom{|\mathcal{X}||\mathcal{A}|+|\mathcal{Y}||\mathcal{B}|}{k} \times \binom{|\mathcal{X}||\mathcal{A}|+|\mathcal{Y}||\mathcal{B}|}{k}$. Here we think of the entries of $\Gamma^{(k)}$ as being labeled by sequences $(z_1,c_1),\ldots,(z_k,c_k)$ where $z_i \in \mathcal{X} \cup \mathcal{Y}$ and $c_i \in \mathcal{A} \cup \mathcal{B}$, and $\Gamma^{(k)}$ is the Gram matrix of the associated vectors

$$|u_{(z_i,c_i)_{1\leq i\leq k}}\rangle \,=\, C_{c_k}^{z_k} \cdots C_{c_1}^{z_1}|\psi\rangle\ ,$$

[10] This fact is not obvious, and constructing such "bad examples" is quite difficult. For some restricted types of games, such as XOR games or unique games, the inequality can be shown to be exact or close to exact respectively.

where $C_c^z = A_c^z \otimes \mathrm{Id}$ if $z \in \mathcal{X}$ and $c \in \mathcal{A}$, $C_c^z = \mathrm{Id} \otimes B_c^z$ if $z \in \mathcal{Y}$ and $c \in \mathcal{B}$, and $C_c^z = 0$ otherwise. In addition, we add any linear constraint on the entries of Γ that follows from the facts that $\{A_a^x\}$ and $\{B_b^y\}$ are projective measurements for all x, y, and that they act on different tensor factors and hence commute.

With this definition we can verify that $w_1 = (8.2)$; this follows since any positive semidefinite matrix Γ has a factorization as a matrix of inner produces. Moreover, $w_1 \geq w_2 \geq \cdots \geq w_k \geq \omega^*(G)$ since each successive level in the "hierarchy" consists in adding additional variables and constraints. Finally, using standard algorithms for semidefinite programs the optimization problem at the k-th level can be solved in time polynomial in its size, i.e. time $(|\mathcal{X}||\mathcal{A}| + |\mathcal{Y}||\mathcal{B}|)^{O(k)}$. Let's call Algorithm B the algorithm that on input k returns w_k.[11]

8.2.3 The commuting value and Tsirelson's problem

Unfortunately it is not the case that $w_k \to_{k\to\infty} \omega^*(G)$. Indeed, if this were the case algorithm C would return arbitrarily good approximations to $\omega^*(G)$ and thus contradict Theorem 8.3. Nevertheless, since (w_k) is non-increasing and larger than $\omega^*(G)$ the sequence must converge to some value. Interestingly, this value is a natural quantity that is referred to as the *commuting value* of the game and defined as

$$\omega^{com}(G) = \sup_{|\psi\rangle, \{A_a^x\}, \{B_b^y\}} \sum_{x,y} \pi(x,y) \sum_{a,b} R(a,b|x,y) \langle \psi | A_a^x B_b^y | \psi \rangle \,, \tag{8.4}$$

where the supremum is taken over all states $|\psi\rangle \in \mathcal{H}$ where $\mathcal{H}$ is a (possibly infinite-dimensional) separable Hilbert space and families of projective measurements $\{A_a^x\}$ and $\{B_b^y\}$ on $\mathcal{H}$ such that for all x, y, a, b, A_a^x and B_b^y commute. Since $A \otimes \mathrm{Id}$ and $\mathrm{Id} \otimes B$ always commute it always holds that $\omega^*(G) \leq \omega^{com}(G)$. The hierarchy of values (w_k) is introduced in [NPA08], where they show the following convergence result.

Lemma 8.5. *For any game G it holds that* $\lim_{k\to\infty} w_k = \omega^{com}(G)$.

Proof. First note that by definition $\omega^{com}(G) \leq \lim_{k\to\infty} w_k$, since none of the constraints imposed on the definition (8.3) of w_k makes use of the tensor product structure other than to say that $A_a^x \otimes \mathrm{Id}$ and $\mathrm{Id} \otimes B_b^y$ commute.

The remainder of the proof shows the reverse inequality. For any $k \geq 1$ fix a feasible solution $\Gamma^{(k)}$ to the optimization problem (8.3). The entries of $\Gamma^{(k)}$ are indexed by pairs of monomials m in non-commutative variables $\{A_a^x, B_b^y\}$ of degree at most k. Crucially, the constraints on the optimization problem require that (i) $\Gamma^{(k)} \geq 0$, and (ii) this matrix satisfies $\Gamma^{(k)}_{m_1,m_2} = \Gamma^{(k)}_{n_1,n_2}$ whenever both entries are well-defined and $m_1 m_2^* = n_1 n_2^*$ as monomials in $\{A_a^x, B_b^y\}$, because by definition any such constraint is imposed on the optimization problem.

For any monomial m and integer k at least as large as the degree of m let $\tau_k(m) = \Gamma^{(k)}_{m,1}$. Extend τ_k to a linear form on all non-commutative polynomials by setting $\tau_k(m) = 0$ if m has degree larger than k and extending by linearity. Since $|\tau_k| \leq 1$ for each k (this can be verified because the diagonal entries of $\Gamma^{(k)}$ are all constrained to equal 1, so using (i) all entries of $\Gamma^{(k)}$ must have modulus at most 1) by the Banach-Aleoglu theorem the sequence $(\tau_k)_{k\geq 1}$ admits a pointwise convergent subsequence $(\tau_{k_i})_{k_1 \leq k_2 \leq \cdots}$; let τ be the pointwise limit. Now crucially we observe that τ is a positive linear form. Indeed, for any polynomial

[11] Technically we need to allow B to return an approximation to w_k. Since well-behaved semidefinite programs such as (8.3) can be solved in time polynomial in their size and in the logarithm of the desired accuracy we could e.g. require that B returns an additive approximation of w_k that is within error at most 2^{-k}; this will suffice for our purposes.

$p = \sum_m \alpha_m m$ where m ranges over monomials we have

$$\begin{aligned}\tau(p^*p) &= \lim_i \; \tau_{k_i}(p^*p)\\ &= \lim_i \sum_{m,m'} \alpha_m^* \alpha_{m'} \, \tau_{k_i}(m^*m')\\ &= \lim_i \; \alpha^\dagger \Gamma^{(k_i)} \alpha\\ &\geq 0 \,,\end{aligned}$$

where for the first line we used linearity of τ_{k_i}, for the second line we used the definition of τ_{k_i} (the equality holds for all i such that $k_i \geq \deg(p)$), for the third line we let $\alpha = (\alpha_m)$ and used property (ii), and for the last we used property (i).

At this point we may conclude in a single abstract step by invoking the GNS construction from C^*-algebra theory: for any positive linear functional τ on a C^*-algebra $\mathcal{A}$ there is a $*$-representation π of $\mathcal{A}$ on a Hilbert space $\mathcal{H}$ and a unit vector $|\xi\rangle \in \mathcal{H}$ such that

$$\forall a \in \mathcal{A} \,, \qquad \tau(a) \;=\; \langle \xi | \pi(a) | \xi \rangle \;. \tag{8.5}$$

For us $\mathcal{A}$ is the algebra of non-commutative polynomials in $\{A_a^x, B_b^y\}$ with complex coefficients satisfying the POVM and commutation conditions, and so the image $\tilde{A}_a^x = \pi(A_a^x)$, $\tilde{B}_b^y = \pi(B_b^y)$, together with the state $|\xi\rangle$, immediately gives us a commuting strategy for G with value $\lim_k w_k$:

$$\begin{aligned}\lim_{k\to\infty} w_k = \lim_{i\to\infty} \; w_{k_i} &= \lim_{i\to\infty} \sum_{x,y} \pi(x,y) \sum_{a,b} R(a,b|x,y) \Gamma^{(k_i)}_{xa,yb}\\ &= \lim_{i\to\infty} \sum_{x,y} \pi(x,y) \sum_{a,b} R(a,b|x,y) \Gamma^{(k_i)}_{(xa,yb}\\ &= \lim_{i\to\infty} \sum_{x,y} \pi(x,y) \sum_{a,b} R(a,b|x,y) \tau_{k_i}((xa,yb))\\ &= \sum_{x,y} \pi(x,y) \sum_{a,b} R(a,b|x,y) \tau((xa,yb))\\ &= \sum_{x,y} \pi(x,y) \sum_{a,b} R(a,b|x,y) \langle \xi | \pi(xa,yb) | \xi \rangle\\ &= \sum_{x,y} \pi(x,y) \sum_{a,b} R(a,b|x,y) \langle \xi | \pi(xa)\pi(yb) | \xi \rangle\\ &= \sum_{x,y} \pi(x,y) \sum_{a,b} R(a,b|x,y) \langle \xi | \tilde{A}_a^x \tilde{B}_b^y | \xi \rangle \,,\end{aligned}$$

where the first line is by definition of w_{k_i}, the second line by the linear constraints (ii), the third by definition of τ_{k_i}, the fourth by definition of τ, the fitfth by (8.5), the sixth because π is a representation and the last by definition of $\tilde{A}_a^x$ and $\tilde{B}_b^y$.

It is also possible to finish the construction more concretely by defining an infinite-dimensional matrix $\Gamma = \lim_i \Gamma^{(k_i)}$, where for the limit to make sense we embed each $\Gamma^{(k_i)}$ as the top left corner of an infinite-dimensional matrix by padding with zeroes. Since all finite minors of Γ are positive semidefinite, it is positive semidefinite and therefore admits a factorization $\Gamma_{m,m'} = \langle m | m' \rangle$ for some $\{|m\rangle\}$ in a Hilbert space $\mathcal{H}$. We can then define $\tilde{A}_a^x$ as the projection on the span of all $|m\rangle$ such that $m = A_a^x m'$ for some m', i.e. the first variable of monomial m is A_a^x. Using the relations satisfied by the inner products between the vectors $|m\rangle$ (i.e. condition (ii) above) it is possible to verify that the $\tilde{A}_a^x$ together with analogously defined

$\tilde{B}_b^y$ and $|\psi\rangle = |1\rangle$ satisfy the required conditions for a commuting strategy, and that the associated value is once again $\lim_k w_k$. □

The two values $\omega^*(G)$ and $\omega^{com}(G)$ were introduced by Tsirelson in a series of papers laying the foundations for the mathematical study of non-locality [Tsi93]. Rather than using the language of games (which at the time was not much in use yet), Tsirelson directly studied the underlying *correlation sets* defined as

$$\begin{aligned} C^*(n,k) = \Big\{ \big(\langle \psi, A_a^x \otimes B_b^y \psi\rangle\big)_{a,b,x,y} \,:\, &\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B} \text{ Hilbert spaces}, \psi \in \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}, \|\psi\| = 1, \\ &\forall (x,y) \in \{1,\dots,n\}^2, \{A_a^x\}_{a\in\{1,\dots,k\}}, \{B_b^y\}_{b\in\{1,\dots,k\}} \text{ POVM on } \mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B} \text{ resp.}\Big\}\,, \end{aligned} \tag{8.6}$$

$$\begin{aligned} C^{com}(n,k) = \Big\{ \big(\langle \psi, A_a^x B_b^y \psi\rangle\big)_{a,b,x,y} \,:\, &\mathcal{H} \text{ Hilbert space}, \psi \in \mathcal{H}, \|\psi\| = 1, \\ &\forall (x,y) \in \{1,\dots,n\}^2, \{A_a^x\}_{a\in\{1,\dots,k\}}, \{B_b^y\}_{b\in\{1,\dots,k\}} \text{ PVOM on } \mathcal{H} \\ &\text{s.t. } [A_a^x, B_b^y] = 0 \; \forall (a,b) \in \{1,\dots,k\}^2 \Big\}\,.^{12} \end{aligned} \tag{8.7}$$

By taking direct sums of POVMs and scaled vectors it is not hard to see that both sets are convex subsets of $[0,1]^{n^2k^2}$. Note that in the definition of $C^*(n,k)$ we did not restrict the dimension of $\mathcal{H}_\mathsf{A}$ and $\mathcal{H}_\mathsf{B}$ to be finite. This is to match Tsirelson's presentation; for our purposes the distinction is not important as it is not hard to see that allowing infinite-dimensional strategies in the definition of the entangled value $\omega^*(G)$ does not change the supremum.[13][14] However, in case the Hilbert spaces in *both* definitions are taken to be finite-dimensional then the two sets can be shown to coincide. (This fact essentially follows from von Neumann's Double Commutant Theorem, though it can also be shown directly; we skip the proof.) In his paper Tsirelson states as "fact" the claim that $C^*(n,k) = C^{com}(n,k)$ for arbitrary separable Hilbert spaces and all $n,k \geq 1$. Having realized that a proof of the claim seemed elusive (with the inclusion $C^*(n,k) \subseteq C^{com}(n,k)$ that we already observed being the only obvious one), in a subsequent note[15] Tsirelson reformulates the "fact" as an open problem and, realizing that the answer may be negative, formulates as an "even more important" problem the question of whether the closure $\overline{C^*(n,k)} = C^{com}(n,k)$. (Here the overline designates closure in the usual topology for $\mathbb{R}^{n^2k^2}$. It is not hard to verify that C^{com} is closed.) Two and a half decades after its introduction Tsirelson's first problem was solved by Slofstra [Slo19], who used techniques from the theory of nonlocal games to show the existence of finite n,k such that $C^*(n,k) \neq C^{com}(n,k)$. Until the proof of Theorem 8.3, an apparently purely complexity-theoretic result, Tsirelson's "even more important problem" remained open. However, we can now observe the following corollary to Theorem 8.3.

Corollary 8.6. *There exists finite $n,k \geq 1$ such that $\overline{C^*(n,k)} \subsetneq C^{com}(n,k)$.*

Proof. Suppose for contradiction that $\overline{C^*(n,k)} = C^{com}(n,k)$ for all $n,k \geq 1$. As an immediate consequence, for any game G it holds that $\omega^*(G) = \omega^{com}(G)$. Therefore, algorithm C described in Section 8.2.2 always converges in finite time to a correct answer. This contradicts Theorem 8.3, which implies that the problem "Given a game G, is $\omega^*(G) \geq \frac{2}{3}$ or $\omega^*(G) \leq \frac{1}{3}$?" is undecidable. □

[13] To show this, observe that any state $|\psi\rangle \in \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$, even in infinite dimensions, always has a Schmidt decomposition $|\psi\rangle = \sum_i \lambda_i |u_i\rangle|u_i\rangle$ such that $\sum_i \lambda_i^2 = 1$. $|\psi\rangle$ can be arbitrarily well approximated in finite dimension by truncating the coefficients; using that the restriction of a POVM to a subspace is a POVM we find arbitrarily good approximations to the game value in finite dimension.

[14] It does change the definition of the set however: as shown in [CS18] some elements of $C^*(n,k)$ cannot be represented in finite dimensions.

[15] "Bell inequalities and operator algebras", available at `https://www.tau.ac.il/~tsirel/download/bellopalg.pdf`.

Note how indirect the proof of Lemma 8.6 is! In particular, while it asserts the existence of n, k there is no obvious way to determine what these integers are, or even upper bounds on them, from the proof. In fact it is possible to tweak the argument to get an explicit construction; we refer to [JNV$^+$20a] for more.

8.2.4 Connes Embedding Problem

Quantum mechanics and the theory of operator algebras have been intertwined since their origin. In the 1930s [VN32] von Neumann laid the foundations for the theory of (what are now known as) von Neumann algebras with the explicit goal of establishing Heisenberg's matrix mechanics on a rigorous footing (quoting from the preface, in the translation by Beyer: "The object of this book is to present the new quantum mechanics in a unified representation which, so far as it is possible and useful, is mathematically rigorous"). Following the initial explorations of Murray and von Neumann the new theory took on a life of its own, eventually leading to multiple applications unrelated to quantum mechanics, such as to free probability or noncommutative geometry.

In his 1976 paper completing the classification of injective von Neumann algebras [Con76] Connes made a casual remark that has become a central conjecture in the theory of operator algebras. Since we do not have the mathematical language to express it precisely, I will paraphrase Connes' remark as the comment that "any finite von Neumann algebra *ought to* be well-approximated by finite-dimensional matrix algebras." (In more formal terms, CEP states that every von Neumann algebra type II_1 factor embeds into an ultrapower of the hyperfinite II_1 factor.) Although this conjecture may at first seem rather specific (and in fact as far as I know Connes himself did not pursue the question any further than the remark made in his paper), in the two decades that followed the problem rose to prominence thanks to the work of other mathematicians, such as Kirchberg and Voiculescu, who gave equivalent reformulations of the conjecture in operator algebras and free probability. (See e.g. [Cap15] for more reformulations.) Kirchberg's formulation is closest to us: Kirchberg showed that CEP is equivalent to the *QWEP conjecture* about the equivalence of the minimal and maximal tensor products on the full group C^* algebra of a nonabelian free group [Kir93].[16] Informally, the minimal and maximal tensor products of two $C*$ algebras provide two ways to define the closure of the algebraic tensor product, with respect to two different norms, the minimal norm

$$\|x\|_{min} = \sup_{\pi_A, \pi_B} \left\|\pi_A \otimes \pi_B(x)\right\|$$

where the supremum ranges over all pairs of representations $\pi_A : C^*(\mathbb{F}_2) \to \mathcal{B}(\mathcal{H}_\mathsf{A})$ and $\pi_B : C^*(\mathbb{F}_2) \to \mathcal{B}(\mathcal{H}_\mathsf{B})$, whereas

$$\|x\|_{max} = \sup_{\pi} \|\pi(x)\|$$

where here $\pi : C^*(\mathbb{F}_2) \otimes C^*(\mathbb{F}_2) \to \mathcal{B}(\mathcal{H})$ is any representation that is such that $\pi(a \otimes b) = \pi_A(a)\pi_B(b)$ where $\pi_A, \pi_B : C^*(\mathbb{F}_2) \to \mathcal{B}(\mathcal{H})$ are representations with commuting range. Clearly, $\|x\|_{min} \leq \|x\|_{max}$ always, and these two norms can be seen to be the smallest and largest "reasonable" norms that one may put on the tensor product of two C^*-algebras.

With this reformulation it may not be surprising that Kirchberg's QWEP is directly related to Tsirelson's problem, and indeed building on work of Fritz [Fri12] and Junge et al. [JNP$^+$11] Ozawa [Oza13a] showed that Tsirelson's "even more important" problem is equivalent to CEP. This brings us to a second corollary of Theorem 8.3.

[15] The brief discussion in this section is adapted from [Vid19].

[16] Concretely, a C^* algebra can always be represented as a sub-algebra of the algebra of bounded linear operators on a Hilbert space that is closed under taking adjoints, and closed under the norm topology. A von Neumann algebra is further restricted to be closed under the weak operator topology.

Corollary 8.7. *CEP has a negative answer, i.e. there exists a von Neumann algebra that is not hyperfinite.*

For more background on the relation between Tsirelson's problem and Kirchberg's conjecture, presented in an accessible way, I recommend [Fri12]. For additional results and the connection to CEP, presented in a less accessible way, I recommend [Oza13b].

Lecture 9

Compression of nonlocal games

9.1 An overview of the proof of RE $\subseteq$ MIP*

9.1.1 A cartoon version

At the highest level our proof strategy is as follows. Recall from the previous lecture that for the case of classical protocols one can show the inclusion NEXP $\subseteq$ MIP. While by itself this is already non-trivial, let's take as our starting point the assumption that we are able to show an analogous inclusion for quantum interactive proofs, i.e. NEXP $\subseteq$ MIP$^*(2,1)$.[1] Observe that this inclusion can be recast as a form of delegation. Paraphrasing, the inclusion states that any language $L \in$ NEXP has a multiprover interactive proof systems with quantum provers. Now for any Turing Machine M the language L_M that consists of all n such that there is a string $y \in \{0,1\}^*$ such that M accepts (n,a) in time at most $\exp(n)$ lies in NEXP;[2] moreover for some choices of M it is NEXP-complete.[3] Thus there is an efficient reduction from Turing machines M to verifiers V such that for all integer n, on input z such that $|z| = n$,

$$\exists a \;:\quad M \text{ accepts } (z,a) \text{ in time} \leq \exp(n) \tag{9.1}$$

then $\omega^*(V_n(z)) \geq \frac{2}{3}$, and if no such a exists then $\omega^*(V_n(z)) \leq \frac{1}{3}$. Now suppose that we're able to achieve a somewhat stronger reduction, where for the starting point we replace the condition (9.1) by

$$\textit{On average over } x \sim \mathcal{U}_N\,, \quad \Pr_x\Big(\exists a \;:\; M \text{ accepts } (z,x,a) \text{ in time} \leq \exp(n)\Big) \geq \frac{2}{3}\,, \tag{9.2}$$

where $N = 2^n$ and for every n, $\mathcal{U}_N$ is the uniform distribution on $\{0,1\}^N$. (Suppose also that a symmetric condition holds for soundness.) This would be a form of delegation for (exponential-time) AM ("Arthur-Merlin") protocols, where an AM protocol is one in which the verifier can send a uniformly random string as question to the prover before receiving the proof. Note that the step we just made is highly non-trivial because of the introduction of a distribution on x; delegating randomized computations like this is hard because there is no easy means to verify that the computation is being performed with the "right choice" of the random string x—indeed, we need to make sure to detect cases where it might be that there exists (x,a)

[1] This inclusion is shown in [IV12] for 5 provers. The 2-prover version follows from the work in [JNV$^+$20b].

[2] This formulation is a bit unusual due to the use of the letter n to represent the input, which is usually called x; this is for later convenience. Here n is written in binary. Note that the time bound implies that without loss of generality $|y| \leq \exp(n)$.

[3] An example would be to take M the Turing machine that parses n as an implicitly represented graph $n = (1^{n'}, C)$ and expects y to be an explicit coloring for the $2^{n'}$ vertices of the graph; see Section 8.1.1.

such that M accepts (n,x,a), but it is still very unlikely to be the case when x is chosen at random. As we will see later the use of quantum provers and entanglement will be useful to achieve this.

Let's do one last leap of faith and suppose that we have an even stronger reduction, that applies directly to exponential-size multiprover interactive proofs. Precisely, we'd replace the condition (9.2) by

$$\textit{On avg over } (x,y) \sim \mathcal{U}_N \times \mathcal{U}_N\,,\ \Pr_{(x,y)}\Big(\exists (a,b)\ :\ M \text{ accepts } (z,x,y,a,b) \text{ in time} \leq \exp(n)\Big) \geq \frac{2}{3}\,, \tag{9.3}$$

where in addition we'd require that (a,b) are generated locally by quantum provers sharing entanglement, such that the provers are given x and y respectively; formally, given (x,y) the pair (a,b) should be distributed as $\langle\psi|A_a^x \otimes B_b^y|\psi\rangle$ for some state $|\psi\rangle$ (independent of (x,y)) and POVM $\{A_a^x\}$ and $\{B_b^y\}$. Once again we'd also require a symmetric condition with probabilities $\leq \frac{1}{3}$ for soundness.

Let's call the resulting reduction a "compression" procedure: it takes as input an exponential-time verifier V and returns a polynomial-time verifier V^{COMPR} that has the same completeness and soundness properties: if there is a good strategy for V_n there is also one for $V\textsc{Compr}_n$ and vice-versa. Then I claim that by iterating this compression procedure we could obtain progressively stronger inclusions, from EXP $\subseteq$ MIP* to EEXP $\subseteq$ MIP* to .... any time complexity that is a finite tower of exponentials.[4] Recall that for well-chosen M, the problem of given an integer n, does M halt in at most 2^n steps is EXP-complete. Now suppose that e.g. we have a family of verifiers $\{V_n\}$, implicitly depending on M, such that $\omega^*(V_n) \geq \frac{2}{3}$ if M halts in $\leq 2^n$ steps, and $\omega^*(V_n) \leq \frac{1}{3}$ otherwise; such a family follows from EXP $\subseteq$ MIP$^*(2,1)$. Now define $\{V_n^{\text{COMPR}}\} = \textsc{Compr}(\{V_{2^n}\})$. Then by definition $\omega^*(V_n^{\text{COMPR}}) \geq \frac{2}{3}$ if $\omega^*(V_{2^n}) \geq \frac{2}{3}$ if M halts in $\leq 2^{2^n}$ steps, and similarly $\omega^*(V_n^{\text{COMPR}}) \leq \frac{1}{3}$ otherwise. Thus EEXP $\subseteq$ MIP*. Iterating this procedure and stretching things a little bit, this would give us the inclusion $\mathsf{TIME}(T(n)) \subseteq$ MIP* for any computable function T. And then taking the "limit", we'd get RE $\subseteq$ MIP*...?

Obviously there's a lot of moving pieces in this description. The goal in this lecture is to make them sufficiently precise as to be believable, and eventually arrive at a core "nugget" that encapsulates the key step that needs to be proven—which we'll do in the next lecture. For now we focus on, first, setting things up so that the above sketch can be made more precise, and second, discussing in more detail the "compression" procedure, which is the key part where the use of quantum provers is essential.

9.1.2 The Halting problem

Recall from the last lecture that the two main consequences of RE $\subseteq$ MIP* that we discussed, negative answers to Tsirelson's problem and to Connes' Embedding Conjecture, both follow from the fact that the problem "Given a two-player one-round game G such that $\omega^*(G) \geq \frac{2}{3}$ or $\omega^*(G) \leq \frac{1}{3}$, which is the case?" is undecidable. In this lecture we will show that there is a computable map $\mathcal{F}$ from Turing Machines M to games $G = G_M$ such that if M halts then $\omega^*(G) \geq \frac{2}{3}$, whereas if M does not halt then $\omega^*(G) \leq \frac{1}{3}$. To argue that this indeed shows that the aforementioned problem is undecidable, we recall the proof that the Halting problem is undecidable.

Definition 9.1. The language L_{HALT} is the set of all $x \in \{0,1\}^*$ such that x is the description of a Turing Machine M such that M, when it is executed on an empty tape, eventually halts.

For convenience we use the notation $\overline{M} \in \{0,1\}^*$ to denote the description of a Turing Machine M, using some canonical representation. Recall that there exists a "universal" Turing machine $\mathcal{U}$ that on input

[4] We could do the same argument for non-deterministic time complexities, but it is easier to present in the deterministic case. The place where we do need non-determinism is for the compression procedure.

$\overline{M}$ and x simulates the execution of M on input x.

Lemma 9.2. *The language L_{HALT} is undecidable.*

Proof. Suppose for contradiction that there exists a Turing Machine A such that given as input $\overline{M}$, A halts with "YES" in case M halts on the empty tape, and A halts with "NO" otherwise. Now consider the following Turing Machine B. When run on an empty tape, B first executes A on $\overline{B}$. If A halts with "YES" then B enters an infinite loop. If A halts with "NO" then B halts with "YES". Does B halt? We have reached a contradiction, therefore A does not exist. □

Note that in the proof of Lemma 9.2 we designed a Turing Machine B that at some point performs an instruction that depends on its own "source code" $\overline{B}$. That this is allowed is a consequence of Kleene's recursion theorem, which is basically a generalization of the standard diagonalization argument. We will use this possibility again later.

9.1.3 Compression

We make more precise what we need of the magical "compression procedure" discussed in Section 9.1.1. First we introduce a restricted class of verifiers.

Definition 9.3. A *normal form verifier* is a Turing Machine V that on input n returns the description of a Turing Machine R_n (the "referee," or "decision procedure") that on input $(x,y,a,b) \in \{0,1\}^{4n}$ returns a value $d \in \{0,1\}$. To R_n we associate a two-player one-round game G_n whose question and answer sets are $X = Y = A = B = \{0,1\}^n$ and such that the question distribution π is uniform on $\{0,1\}^n \times \{0,1\}^n$ and the referee predicate is given by R_n.

We let $\mathsf{TIME}_V(n)$ be the worst-case running time of R_n over all inputs (x,y,a,b). If $\mathsf{TIME}_V(n) \leq (\lambda n)^\lambda$ for some integer $\lambda \geq 1$ and all $n \geq 1$ then we say that V is *λ-bounded*.

Note that in the definition we fixed the question distribution used for the game G_n to the uniform distribution. At this stage this is mostly for convenience. Later we will realize that this is too restrictive, and so one should bear in mind that the definition can be generalized to allow various classes of distributions, where the key point is that the distribution should be fixed and independent of V.

We need one last definition.

Definition 9.4. For a two-player one-round game G and a probability $p \in [0,1]$ let $\mathcal{E}(G,p)$ denote the smallest integer $d \geq 1$ such that there exists a strategy $(|\psi\rangle, \{A_a^x\}, \{B_b^y\})$ for the players in G that has success probability at least p and such that $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$. If no such strategy exists, then $\mathcal{E}(G,p) = \infty$.

For example, for the Magic Square game (Section 3.3.1) it is possible to show that $\mathcal{E}(G,1) = 4$ (two qubits per player), and in fact there is a $c < 1$ such that $\mathcal{E}(G,p) = 4$ for all $p \in [c,1]$. For a game that has a perfect classical strategy we have $\mathcal{E}(G,p) = 1$ for all $p \in [0,1]$.

Let's make the following specification for a compression procedure.

Claim 9.5. *There is a polynomial-time computable mapping* COMPR *that takes as input a Turing machine description $\overline{V}$ and an integer λ written in unary and returns a Turing machine description $\overline{V^{\mathrm{COMPR}}} = \mathrm{COMPR}(\overline{V},\lambda)$ such that the following conditions hold:*

(a) V^{COMPR} is always a normal form verifier such that $\mathsf{TIME}_{V^{\mathrm{COMPR}}}(n) \leq p_{\mathrm{COMPR}}(\lambda + n)$, for some universal polynomial p_{COMPR} independent of V.

(b) If V is a normal form λ-bounded verifier then for every $n \geq 1$ letting $N = 2^n$ the following hold:

(b.i) If $\omega^*(R_N) = 1$ *then* $\omega^*(R_n^{\textsc{Compr}}) = 1$.

(b.ii) $\mathcal{E}(G_n^{\textsc{Compr}}, \frac{1}{2}) \geq \max\{\mathcal{E}(G_N, \frac{1}{2}), N\}$.

The key point about Claim 9.5 is that the running time of $R_n^{\textsc{Compr}}$ can be much smaller than that of R_N, yet it preserves essential properties of it, stated in (b.i) and (b.ii).

We make a few comments on the requirements stated in the claim. First of all, even though eventually we only need to create a *computable* mapping $\mathcal{F}$ from Turing Machines to games, it will be important that here COMPR is required to run in polynomial time. Second, it will also be essential that for any input $(\overline{V}, \lambda)$ to COMPR the output $\overline{V^{\textsc{Compr}}}$ is the description of a time-bounded verifier. Note that this is not hard to enforce in practice by hard-coding some kind of time-out mechanism in the definition of $V^{\textsc{Compr}}$. Finally, observe that condition (b.ii) states something a little stronger (strictly speaking, incomparable) than the "soundness preservation" condition we considered in Section 9.1.1. Indeed the fact that we are able to make a statement about entanglement will play an important role in the final argument. (On the other hand, that the conditions apply to $N = 2^n$ as opposed to e.g. $N = n + 1$ is not important; since it is what comes out of the proof we keep it here—what matters is that $V_n^{\textsc{Compr}}$ reproduces properties of V_N for some $N > n$ while having complexity comparable to V_n, not V_N.) Finally, note that due to condition (b) running COMPR on a trivial input that always accepts already yields an interesting family of games: due to (b.i) we will have $\omega^*(V_n^{\textsc{Compr}}) = 1$ for all n, and due to (b.ii) achieving any value larger than $\frac{1}{2}$ will necessarily require a quantum state of local dimension at least $N = 2^n$.

These observations show that designing a procedure COMPR that fulfills all conditions will likely not be an easy task. Nevertheless, let's put that task aside for the time being and see how the desired reduction can be completed assuming the validity of Claim 9.5.

9.1.4 A self-referential verifier

Fix a Turing Machine M and an integer $\lambda \geq 1$ and consider the following normal form verifier $V = V_{M,\lambda}$, that implicitly depends on M and λ (and in fact is efficiently computable from the pair $(\overline{M}, \lambda)$). For any $n \geq 1$, we describe the decision procedure R_n using high-level "pseudocode". Given $(x, y, a, b) \in \{0,1\}^{4n}$, R_n does the following:

1. R_n simulates M on the empty tape for n steps. If M halts then R_n accepts (i.e. it returns the value '1', irrespective of its inputs (x, y, a, b). Otherwise, if M has not halted in n steps then R_n proceeds to the next item.

2. R_n computes $\overline{V^{\textsc{Compr}}} = \textsc{Compr}(\overline{V}, \lambda)$.

3. R_n returns the decision $(R^{\textsc{Compr}})_n(x, y, a, b) \in \{0,1\}$.

Note that in giving this high-level description of a Turing Machine V, that on input 1^n returns the description $\overline{R_n}$, we have referred to the description $\overline{V}$ itself. That this is possible, i.e. V is a well-defined Turing Machine, is a consequence of Kleene's recursion theorem—this is similar to the self-referential call we made for the definition of algorithm B in the proof of Lemma 9.2.

The following three claims establish the key properties of this construction.

Claim 9.6. *For any Turing Machine M there is an integer $\lambda \geq 1$ which is computable from $|M|$ and such that V is λ-bounded.*

Proof. By definition V on input 1^n returns a decision procedure R_n that takes four inputs of length n each, so it is a normal form verifier. It remains to estimate its running time. First we estimate $|\overline{V}|$. Clearly, the actions to be performed in each of the three steps can be described using $\mathrm{poly}(|\overline{M}|, \lambda)$ bits. Note that the description $\overline{\text{COMPR}}$ does not depend on anything, so its size is a constant.

Next we estimate the running time of R_n. The first step, the simulation of M for n steps, takes time $p_1(n, \overline{M})$ for some universal polynomial p_1. The second step, the computation of $\overline{V^{\text{COMPR}}}$, takes time $p_2(\overline{V}, \lambda)$, for some universal polynomial p_2 that bounds the running time of COMPR. The last step, the evaluation of $R_n^{\text{COMPR}}(x, y, a, b)$, takes time $p_{\text{COMPR}}(\lambda + n)$ by property (a) in Claim 9.5.

Overall the running time is $\mathrm{poly}(n, \overline{M}, \lambda)$ for some universal polynomial. This can be bounded above by the expression $(\lambda n)^\lambda$ for all $n \geq 1$ provided λ is large enough compared to $\overline{M}$. □

For the remaining two claims we fix λ to the value promised in Claim 9.6 and let $\{R_n\}$ and $\{G_n\}$ be the family of decision procedures and games respectively implied by the verifier V specified from M and λ.

Claim 9.7. *Suppose that M halts on an empty input tape. Then $\omega^*(R_n) = 1$ for all n.*

Proof. Let T be the number of steps taken by M to halt. Then for all $n \geq T$ the decision procedure R_n always accepts its inputs at step 1. Therefore $\omega^*(R_n) = 1$ for all $n \geq T$. Now we show by (strong) downwards induction from $m = T$ to 1 that $\omega^*(R_m) = 1$. We showed the induction hypothesis for $m = T$ already. Suppose it true up to some value $m > 1$. Then since M does not halt in $(m-1)$ steps, the decision procedure R_{m-1} proceeds to step 2. and executes $(R^{\text{COMPR}})_{m-1}$. Since $2^{m-1} > m - 1$, it follows from the induction hypothesis that $\omega^*(R_{2^{m-1}}) = 1$. Using property (b.i) in Claim 9.5 we have that $\omega^*(R_{m-1}) = 1$, as desired. □

Claim 9.8. *Suppose that M does not halt. Then $\omega^*(R_n) \leq \frac{1}{2}$ for all $n \geq 1$.*

Proof. We show that $\mathcal{E}(G_n, \frac{1}{2}) = \infty$ for all $n \geq 1$. This shows that no finite strategy can achieve a success probability larger than $\frac{1}{2}$, and taking the limit that $\omega^*(R_n) \leq \frac{1}{2}$, as desired. Since M does not halt, for any n, R_n proceeds to step 2. and returns the desision of R_N^{COMPR} where $N = 2^n$. By property (b.ii) in Claim 9.5 it follows that for all $n \geq 1$,

$$\mathcal{E}(G_n, \frac{1}{2}) \geq \max\left\{\mathcal{E}(G_N, \ \frac{1}{2}), N\right\} .$$

By straightforward induction, $\mathcal{E}(G_n, \frac{1}{2}) \geq T$ for any integer T, so it must be ∞. □

9.1.5 A game for the halting problem

We now describe the reduction $\mathcal{F}$. On input $\overline{M}$, $\mathcal{F}$ first computes the integer λ whose existence is promised in Claim 9.6. Then $\mathcal{F}$ computes a description of the decision procedure R_1 from the start of Section 9.1.4, based on $\overline{M}$ and λ. Finally, $\mathcal{F}$ returns a description of the associated game $G = G_1$.[5]

Suppose first that M halts. Then by Claim 9.7 it holds that $\omega^*(G) = 1$. Suppose now that M does not halt. It follows from Claim 9.8 that $\omega^*(G) \leq \frac{1}{2}$. This completes the reduction.[6]

[5] It is interesting that ultimately we only need G_1, but to arrive at constructing it we had to consider infinite families of games.

[6] While we promised to obtain a separation between values $\frac{2}{3}$ and $\frac{1}{3}$, we only obtained one between 1 and $\frac{1}{2}$. There is nothing in the line of argument that is special about $\frac{1}{2}$ and we could have done the same replacing it by $\frac{1}{3}$, giving us an even stronger reduction than desired. In general, the values $\frac{2}{3}$ and $\frac{1}{3}$ can be amplified towards 1 and 0 respectively by applying techniques from parallel repetition; see e.g. [Yue16].

Remark 9.9. It is worth pausing to appreciate the significance of this reduction. Beyond the stated inclusion of complexity classes, it makes quite a striking statement about the complexity that may lurk behind simple, finite, observable phenomena in quantum mechanics. What the existence of $\mathcal{F}$ states is that for *any* problem that can be encoded in the halting of a Turing machine there is a game, that moreover is easily computable from the Turing machine, that "witnesses" this fact. Consider for example the Riemann Hypothesis (RH). There is a simple Turing Machine M that halts if and only if RH is provable in ZFC (Zermelo-Fraenkel set theory with the axiom of choice included). Indeed M simply enumerates over all possible proofs in ZFC and checks if they are (i) valid proofs and (ii) prove RH. Moreover, the Turing machine M is large, but not absurdly so; probably a few millions of characters are more than enough. This means that we can, in principle but also in practice, write a simple computer program that will return the rules for a moderately-sized nonlocal game $G = G_{RH}$ such that $\omega^*(G) = 1$ if and only if RH is provable in ZFC. Isn't this amazing?

Remark 9.10. A natural question is what is the *commuting value* ω_{com} of the games $G = G_{M,\lambda}$. Naturally this value is always at least $\omega^*(M)$, and moreover for *some* infinite family of M it must be the case that $\omega_{com}(G) = 1$ even if M does not halt (and hence $\omega^*(G) \leq \frac{1}{2}$), as otherwise using algorithm C from the previous lecture we would be able to solve the Halting problem. We do not know if $\omega_{com}(G) = 1$ for all games G in the range of the reduction $\mathcal{F}$.

9.2 The compression procedure

Based on the work completed in the previous section it "only" remains to show the existence of an appropriate compression procedure, that satisfies the requirements of Claim 9.5. As it turns out this is quite a tall order, and we won't be able to give full details here. Instead we very roughly describe how the design of the compression procedure eventually boils down to a much more manageable "nugget" that takes the form of a nonlocal game with good "self-testing" properties.

9.2.1 A test for n qubits

In the next lecture we will give a proof of the following.

Claim 9.11. *There is a family of games $\{G_n\}_{n\geq 1}$ such that for each $n \geq 1$, the game G_n tests n qubits.*

We gave a quite succinct statement for the claim; let's unpack it a little bit. First, by "the game G_n tests n qubits" we mean the following: there is a constant $c < 1$, independent of n, such that for any n and any two players that succeed in the game G_n with probability at least c, each player must "have n qubits". Recalling our definition of n qubits what we mean by this is that to each strategy we should be able to associate $2n$ observables $(X_i, Z_i)_{i=1,\ldots,n}$ such that whenever the strategy succeeds with high enough probability the (X_i, Z_i) are "close" to satisfying the appropriate commutation and anti-commutation relations, where closeness is measured in the right state-dependent norm. As it turns out, for the construction that we give next time a stronger condition will hold, giving us the following somewhat more formal variant of Claim 9.11.

Claim 9.12. *There is a family of games $\{G_n\}_{n\geq 1}$ with the following properties. For each $n \geq 1$, in the game G_n there are questions of the form (X, a) and (Z, b) where X, Z are labels and $a, b \in \{0,1\}^n$ such that the answer expected from a player upon such a question is a single bit. In addition there are two questions X and Z such that the expected answer on such a question is n bits long. Furthermore, for each $n \geq 1$*

and $\varepsilon > 0$, for any strategy $(|\psi\rangle, \{A_a^x\}, \{B_b^y\})$ that succeeds with probability tat least $1-\varepsilon$ in the game there is an isometry $V_A : \mathcal{H}_{\mathsf{A}} \to (\mathbb{C}^2)^{\otimes n} \otimes \mathcal{H}'_{\mathsf{A}}$ such that, if $X(a)$, $Z(b)$ and $\{A_a^X\}_a, \{A_b^Z\}_b$ are Alice's observables and POVM on the aforementioned questions then

$$\max\Big\{ \mathrm{E}_a \big\| (V_A X(a) - (\sigma_X(a) \otimes \mathrm{Id}_{\mathsf{A}'}) V_A) \otimes \mathrm{Id}_{\mathsf{B}} |\psi\rangle \big\|^2 , \\ \mathrm{E}_b \big\| (V_A Z(b) - (\sigma_Z(b) \otimes \mathrm{Id}_{\mathsf{A}'}) V_A) \otimes \mathrm{Id}_{\mathsf{B}} |\psi\rangle \big\|^2 \Big\} \le O(\varepsilon^d) ,$$

where the expectation is taken over uniformly random $a, b \in \{0,1\}^n$, and letting $\{\sigma_a^X\}_a$ and $\{\sigma_b^Z\}$ denote POVMs representing an n-qubit measurement in the Hadamard and computational basis respectively,

$$\max\Big\{ \sum_a \big\| (V_A A_a^X - (\sigma_a^X \otimes \mathrm{Id}_{\mathsf{A}'}) V_A) \otimes \mathrm{Id}_{\mathsf{B}} |\psi\rangle \big\|^2 , \\ \sum_b \big\| (V_A A_b^Z - (\sigma_b^Z \otimes \mathrm{Id}_{\mathsf{A}'}) V_A) \otimes \mathrm{Id}_{\mathsf{B}} |\psi\rangle \big\|^2 \Big\} \le O(\varepsilon^d) ,$$

for some universal constant $d > 0$. A similar statement holds for Bob's observables. Finally, there is a state $|aux\rangle \in \mathcal{H}'_A \otimes \mathcal{H}'_B$ such that

$$\big\| V_A \otimes V_B |\psi\rangle - |\phi^+\rangle^{\otimes n} |aux\rangle \big\|^2 \le O(\varepsilon^d) ,$$

where recall that $|\phi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ denotes the state of an EPR pair.

The important part of the claim is that the game G_n allows us to "command" either player to measure n halves of shared EPR pairs in the computational or Hadamard basis, and that whenever asked to do so — they faithfully do it, up to a local isometry. What is particularly non-trivial about the claim as we stated it is the fact that the stated error bounds do not depend on n, and this crucial for its usefulness. We will give the proof of Claim 9.12 in the next lecture; let's now verify that it encapsulates (almost...) all that we need to design the compression procedure.

9.2.2 How to delegate a nonlocal game

Recall that the goal of the compression procedure is to turn a verifier Turing Machine V given as input into a verifier Turing Machine V^{COMPR} such that V_n^{COMPR} "simulates" V_N^{COMPR} for $N = 2^n$, in the sense described in Claim 9.5. Let's start by reviewing how the game specified by V_N proceeds, assuming V_N is normal form:

1. Generate a pair of questions $(X, Y) \in \{0,1\}^N \times \{0,1\}^N$ uniformly at random;
2. Execute each prover's strategy to obtain answers $(A, B) \in \{0,1\}^N \times \{0,1\}^N$;
3. Return $V_N(X, Y, A, B)$.

Here, the first and last steps are executed by V_N and take time polynomial in N, hence exponential in n. The second step is executed by the provers and can take an arbitrary amount of time. The main idea for designing V_n^{COMPR} is to "delegate" these steps to the provers and limit the verifier's computation to some form of "verification" that the provers are performing the right computation:

1. Generate a pair of questions $(x, y) \in \{0,1\}^n \times \{0,1\}^n$ uniformly at random.
2. Provers do the following:

(a) Locally generate $X \in \{0,1\}^N$ and $Y \in \{0,1\}^N$ respectively;

(b) Execute original prover's strategy to obtain answers (A, B);

(c) Return "proofs" a and b respectively that (i) X and Y have been selected uniformly at random; (ii) A and B can be obtained locally; (iii) $V_N(X, Y, A, B) = 1$.

3. Given (x, y, a, b), check that (a, b) are a valid proof of (i), (ii) and (iii).

As it turns out the most delicate step in implementing this scheme is the verification of (i). Indeed this is where we will use the fact that the provers can be forced to share entanglement in an essential way. The verification of (ii) would be automatic if we asked each prover to report A and B respectively; because we only ask them for shorter "proofs" a and b this requires more care, but it is not a real problem. Finally, while the verification of (iii) requires quite some work at its heart it is not substantially different from the efficient verification of an exponential-time computation; here again the part that is a bit delicate is that the inputs X, Y, A, B to the computation are not known explicitly to the verifier but only through the "proofs" of items (i) and (ii).

Focusing on (i) now, observe that the game G_N from Claim 9.12 provides us with the means to "force" the provers to generate uniformly random X and Y: we simply send them, say, question Z in the game, and request that they use the outcome obtained as their X and Y respectively.

There is one good thing about this, and many bad things. The good thing is that based on Claim 9.12 we know that if the provers succeed in the game G_N, a condition which we can imagine checking separately (meaning that with some probability the verifier V_n^{COMPR} executes G_N and accepts if and only if accepts), then when one prover is sent the label Z and the other the label X they obtain outcomes a, b which are statistically close to uniformly distributed and independent from each other; both facts are a consequence of the fact that computational and Hadamard basis measurements on two halves of the same EPR pair yield independent uniformly distributed outcomes.

As for the bad things, we point out a few of them:

- While we can request the provers to perform computational basis measurements and thereby obtain uniformly random X, Y, it is unclear how we can require that these are precisely the inputs that they use for the evaluation of V_N, unless they report X, Y to us—which is not possible since these strings are too long;
- Furthermore, X, Y are not uniformly random but only close to being so—we need to make sure that this possible error cannot be maliciously used by the provers to too large an extent;
- The game G_N as we described it involves N-bit long questions; however, the verifier V_n^{COMPR} is only allowed n-bit long questions;
- We have not examined the running time of V_n^{COMPR}: this should be bounded by some polynomial in n and λ, and independent of the running time of V_N.

These difficulties are substantial difficulties. They can all be overcome by introducing a more involved variant of Claim 9.12 that is based on the use of what are called "low-degree tests' in complexity theory. Informally, this consists in combining the presentation of V^{COMPR} that we gave with ideas from the theory of error-correcting codes so as to (a) reduce the complexity of the referee in the game, and (b) ensure that the little information that the players provide about X, Y is sufficient to "lock" them into using X, Y, or to the least strings that are sufficiently close to them, in the other steps (i.e. the proofs of items (ii) and (iii)). We will not be able to dive into these difficulties here, and instead allocate the last lecture to a proof of Claim 9.12.

Remark 9.13. For the entire lecture we assumed that we could work with verifiers that use a uniform distribution on their questions. In fact, this is provably not sufficient, and once must necessarily use more complex question distributions. For this reason the fact that Claim 9.12 allows us to certify measurements in either computational or Hadamard basis for the provers is crucial: in this way we are able not only to request that they generate uniform independent (X, Y), by sending them different bases, but also uniform *equal* (X, Y), by sending them the same basis. By introducing further variations on this theme it is possible to "command" the provers to generate (X, Y) according to quite a broad class of distributions.

Lecture 10

A test for n qubits

Recall from the last lecture that we reduced the proof of $\mathrm{MIP}^* = \mathrm{RE}$ to showing a form of "n-qubit test", stated as Claim 9.12 and reformulated as a theorem here:

Theorem 10.1. *There is a family of games $\{G_n\}_{n\geq 1}$ with the following properties. For each $n \geq 1$, in the game G_n there are questions of the form (X,a) and (Z,b) where X, Z are labels and $a,b \in \{0,1\}^n$ such that the answer expected from a player upon such a question is a single bit. In addition there are two questions X and Z such that the expected answer on such a question is n bits long. Furthermore, for each $n \geq 1$ and $\varepsilon > 0$, for any strategy $(|\psi\rangle, \{A_a^x\}, \{B_b^y\})$ that succeeds with probability tat least $1-\varepsilon$ in the game G_n there is an isometry $V_A : \mathcal{H}_{\mathsf{A}} \to (\mathbb{C}^2)^{\otimes n} \otimes \mathcal{H}'_{\mathsf{A}}$ such that, if $X(a)$, $Z(b)$ and $\{A_a^X\}_a, \{A_b^Z\}_b$ are Alice's observables and POVM on the aforementioned questions then*

$$\max\Big\{\mathrm{E}_a \big\|(V_A X(a) - (\sigma_X(a)\otimes \mathrm{Id}_{\mathsf{A}'})V_A)\otimes \mathrm{Id}_{\mathsf{B}}|\psi\rangle\big\|^2 , \\ \mathrm{E}_b \big\|(V_A Z(b) - (\sigma_Z(b)\otimes \mathrm{Id}_{\mathsf{A}'})V_A)\otimes \mathrm{Id}_{\mathsf{B}}|\psi\rangle\big\|^2\Big\} \leq O(\varepsilon^d) , \tag{10.1}$$

where the expectation is taken over uniformly random $a,b \in \{0,1\}^n$, and letting $\{\sigma_a^X\}_a$ and $\{\sigma_b^Z\}$ denote POVMs representing an n-qubit measurement in the Hadamard and computational basis respectively,

$$\max\Big\{\sum_a \big\|(V_A A_a^X - (\sigma_a^X\otimes \mathrm{Id}_{\mathsf{A}'})V_A)\otimes \mathrm{Id}_{\mathsf{B}}|\psi\rangle\big\|^2 , \\ \sum_b \big\|(V_A A_b^Z - (\sigma_b^Z\otimes \mathrm{Id}_{\mathsf{A}'})V_A)\otimes \mathrm{Id}_{\mathsf{B}}|\psi\rangle\big\|^2\Big\} \leq O(\varepsilon^d) , \tag{10.2}$$

for some universal constant $d > 0$. A similar statement holds for Bob's observables. Finally, there is a state $|aux\rangle \in \mathcal{H}'_A \otimes \mathcal{H}'_B$ such that

$$\big\|V_A \otimes V_B|\psi\rangle - |\phi^+\rangle^{\otimes n}|aux\rangle\big\|^2 \leq O(\varepsilon^d) , \tag{10.3}$$

where recall that $|\phi^+\rangle \in \mathbb{C}^2\otimes\mathbb{C}^2$ denotes the state of an EPR pair.

Remark 10.2. The games G_n in Theorem 10.1, as well as all games considered in this lecture, are *symmetric*: $\mathcal{X} = \mathcal{Y}$ and $\mathcal{A} = \mathcal{B}$, the distribution on questions π is invariant under permutation of the two questions, $\pi(x,y) = \pi(y,x)$ for all (x,y), and the verification predicate is symmetric as well, i.e. $R_n(a,b,x,y) = R_n(b,a,y,x)$ for all (x,y,a,b). Define a strategy $(|\psi\rangle, \{A_a^x\}, \{B_b^y\})$ to be *symmetric* if

$\mathcal{H}_\mathsf{A} = \mathcal{H}_\mathsf{B}$, $|\psi\rangle$ is invariant under exchange of the two subsystems, and $A_a^x = B_a^x$ for all x, a. It is not hard to verify that whenever a game G is symmetric then for any strategy $(|\psi\rangle, \{A_a^x\}, \{B_b^y\})$ that succeeds with some probability $1 - \varepsilon$ in the game there is a symmetric strategy $(|\tilde{\psi}\rangle, \{\tilde{A}_a^x\})$ that succeeds with the same probability, such that moreover "rigidity" statements such as the conclusion of Theorem 10.1 can be "lifted" from the symmetric strategy back to the original strategy. (The symmetrized strategy uses a state

$$|\tilde{\psi}\rangle = |\psi\rangle_{\mathsf{AB}} \otimes \frac{1}{\sqrt{2}}\big(|0\rangle_{\mathsf{A}'}|1\rangle_{\mathsf{B}'} + |1\rangle_{\mathsf{A}'}|0\rangle_{\mathsf{B}'}\big) \in \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B} \otimes \mathbb{C}^2_{\mathsf{A}'} \otimes \mathbb{C}^2_{\mathsf{B}'} ,$$

and each player uses its additional qubit to decide whether to compute its answer a to question x using the "Alice" POVM $\{A_a^x\}$ or the "Bob" POVM $\{B_a^x\}$; formally, for all x, a, $\tilde{A}_a^x = A_a^x \otimes |0\rangle\langle 0| + B_a^x \otimes |1\rangle\langle 1|$.) Due to this observation in the lecture we limit ourselves to the analysis of symmetric strategies.

There is one important aspect in which the theorem falls short of what we needed in the previous lecture: its verification procedure is too complex. In the theorem, n qubits are being tested using —as we will see— questions of length $O(n)$. In order for the test to be of use in the design of the compression procedure (Claim 9.5) we need the test verifier to run in time $\ll n$, so in particular questions cannot be of length $O(n)$. Achieving such a test is a challenge beyond what we can expose in a single lecture, and so we will focus on the less efficient version provided by Theorem 10.1. The more efficient version combines the kind of techniques introduced in this lecture with techniques similar to the use of polynomial error-correcting codes in the proof of the PCP theorem. For details we refer to [JNV$^+$20b] and [JNV$^+$20a, Appendix A].

Our strategy for the proof of Theorem 10.1 is to derive it from two ingredients. Firstly, in Section 10.1 we introduce a general theory of approximate group representations and note that the consequences of Theorem 10.1 follow provided that the players' strategy is, in some sense, an approximate representation of the n-qubit Pauli group (a finite group which we will define). Secondly, in Section 10.2 we design a game, or "test", that enforces that any successful strategy for the players in the game specifies an approximate representation of the n-qubit Pauli group, so that the results from the first part can be applied to it to conclude the theorem.

10.1 Approximate group representations

10.1.1 Definitions

We first make a small detour through the theory of group representations. For d-dimensional matrices A, B and σ such that σ is positive semidefinite, write

$$\langle A, B\rangle_\sigma = \mathrm{Tr}(AB^*\sigma) ,$$

where we use B^* to denote the conjugate-transpose. Note that the matrix trace inner product is recovered for $\sigma = \mathrm{Id}$. If σ is the totally mixed state, then we obtain a dimension-normalized variant of the trace inner product. We will also write $\|A\|_\sigma = \langle A, A\rangle_\sigma^{1/2}$.

Given an arbitrary finite group G (not necessarily abelian), a group representation of G is a map $f : G \to U_d(\mathbb{C})$, the group of $d \times d$ unitary matrices, such that f is a homomorphism: for any $x, y \in G$, $f(x^{-1}y) = f(x)^* f(y)$, where we used * to denote the conjugate transpose (which, for unitary matrices, corresponds to taking the inverse). The following definition introduces a notion of *approximate* group representation.

Definition 10.3. Given a finite group G, an integer $d \geq 1$, $\varepsilon \geq 0$, and a d-dimensional positive semidefinite matrix σ with trace 1, an (ε, σ)-representation of G is a function $f : G \to U_d(\mathbb{C})$, the unitary group of $d \times d$ matrices, such that

$$\mathrm{E}_{x,y \in G} \Re\big(\langle f(x)^* f(y), f(x^{-1}y)\rangle_\sigma\big) \geq 1 - \varepsilon , \tag{10.4}$$

where the expectation is taken under the uniform distribution over G.

Note that the condition (10.4) is equivalent to

$$\mathrm{E}_{x,y \in G} \big\| f(x^{-1}y) - f(x)^* f(y) \big\|_\sigma^2 \leq 2\varepsilon . \tag{10.5}$$

Taking $\varepsilon = 0$ and σ any invertible positive definite matrix, we see that the case $\varepsilon = 0$ corresponds to an *exact* representation of G.

Example 10.4. Consider the Weyl-Heisenberg group $\mathcal{P}$, which is the group generated by the Pauli σ_X and σ_Z matrices. It is not hard to verify that this group has 8 elements, which can be decomposed as $(-1)^c \sigma_X^a \sigma_Z^b$ for $a, b, c \in \{0, 1\}$. A qubit (X, Z) according to our first definition (Definition 1.1) can be used to specify a $(0, \sigma)$ representation of $\mathcal{P}$ for *any* σ as follows:

$$f\big((-1)^c \sigma_X^a \sigma_Z^b\big) = (-1)^c X^a Z^b , \tag{10.6}$$

for all $a, b, c \in \{0, 1\}$. It is immediate to verify that for all $(x, y) \in \mathcal{P}$ we have $f(x)^* f(y) = f(x^{-1}y)$ and so (10.4) holds with $\varepsilon = 0$ for any σ.

The example makes explicit the connection between our notion of qubit with group representation theory. In this lecture we will leverage the connection to make use of powerful theorems from representation theory towards the analysis of an n-qubit test. As a warm-up, the following simple and recommended exercise asks you to generalize the example to the case of a single ε-approximate qubit.

Exercise 10.1. Let $(|\psi\rangle, X, Z)$ be a qubit such that $|\psi\rangle \in \mathcal{H}$. Define $f : \mathcal{P} \mapsto U(\mathcal{H})$ as in (10.6). Show that f is an $(0, \sigma)$-representation of $\mathcal{P}$ for $\sigma = |\psi\rangle\langle\psi|$. Now suppose that $(|\psi\rangle, X, Z)$ is only an ε-approximate qubit in the sense of Exercise 2.1. Show that f is an $(O(\sqrt{\varepsilon}), \sigma)$-representation of $\mathcal{P}$.

Remark 10.5. The condition (10.4) in Definition 10.3 is very closely related to Gowers' U^2 norm

$$\|f\|_{U^2}^4 = \mathrm{E}_{xy^{-1} = zw^{-1}} \langle f(x) f(y)^*, f(z) f(w)^* \rangle_\sigma .$$

While a large Gowers norm implies closeness to an affine function, we are interested in testing homomorphisms, and the condition (10.4) will arise naturally from our calculations in the next section.

10.1.2 The Gowers-Hatami theorem

There are many possible notions of approximate group representation. Traditionally the most frequently considered one replaces the norm in Definition 10.3 by the operator norm. An inconvenience of that variant is that in general approximate representations are not always close to exact representations (see, for example, the famous problem on "approximately commuting" versus "nearly commuting" operators). In contrast Gowers and Hatami [GH17] showed that in the case of Definition 10.3, approximate group representations can always be "rounded" to a nearby exact representation.[1] We state and prove a slightly more general, but quantitatively weaker, variant of the Gowers-Hatami result.

[1] We are barely scratching the surface of a growing theory of "stability" for group homomorphisms; see e.g. [BC20] for an introduction and discussion of related notions.

Theorem 10.6 (Gowers-Hatami)**.** *Let G be a finite group, $\varepsilon \geq 0$, and $f : G \to U_d(\mathbb{C})$ an (ε,σ)-representation of G. Then there exists a $d' \geq d$, an isometry $V : \mathbb{C}^d \to \mathbb{C}^{d'}$, and a representation $g : G \to U_{d'}(\mathbb{C})$ such that*

$$\mathrm{E}_{x\in G} \left\| f(x) - V^* g(x) V \right\|_\sigma^2 \leq 2\,\varepsilon.$$

Gowers and Hatami limit themselves to the case of $\sigma = d^{-1} I_d$, which corresponds to the dimension-normalized Frobenius norm. In this scenario they in addition obtain a tight control of the dimension d', and show that one can always take $d' = (1 + O(\varepsilon))d$ in the theorem. We will see a much shorter proof than theirs (the proof is implicit in their argument) that does not seem to allow to recover this estimate. The extension to general σ, however, will be necessary for our purposes.

Note that Theorem 10.6 does not in general hold with $d' = d$. The reason is that it is possible for G to have an approximate representation in some dimension d, but no exact representation of the same dimension: to obtain an example of this, take any group G that has all non-trivial irreducible representations of large enough dimension, and create an approximate representation in e.g. dimension one less by "cutting off" one row and column from an exact representation. For sufficiently "smooth" σ (no disproportionately large singular values) the dimension normalization induced by the norm $\|\cdot\|_\sigma$ will make this barely noticeable, but it will be impossible to "round" the approximate representation obtained to an exact one without modifying the dimension.

Example 10.7. Continuing with Example 10.4 we consider the example of $G = \mathcal{P}$. In Example 10.4 we observed that a qubit $(|\psi\rangle, X, Z)$ can be used to specify a $(0,\sigma)$ representation f of $\mathcal{P}$ such that moreover $f(-1) = -\,\mathrm{Id}$. We now check that the converse holds: for any $(0,\sigma)$-representation of $\mathcal{P}$ for invertible σ, if $X = f(\sigma_X)$ and $Z = f(\sigma_Z)$ then using (10.5), taking $x = y = \sigma_X$ and $x = y = \sigma_Z$ it follows that $X^2 = Z^2 = f(1) = \mathrm{Id}$, and taking $x = \sigma_X$ and $y = \sigma_Z$ we get that $XZ = f(\sigma_X\sigma_Z)$ while

$$ZX = f(\sigma_Z\sigma_X) = f(-\sigma_X\sigma_Z) = f(-1)f(\sigma_X\sigma_Z)\,,$$

where the second equality uses that the Pauli anti-commute and the last equality again uses (10.5). Thus if $f(-1) = -\,\mathrm{Id}$ then $\{X, Z\} = 0$, so the $(0,\sigma)$-representation f specifies a qubit and in particular Lemma 2.3 on the structure of a qubit applies. As a result, we have shown that there exists a single representation of $\mathcal{P}$ such that $f(-1) = -\,\mathrm{Id}$, and that it is given by the Pauli matrices in dimension 2. [2]

Exercise 10.2. Using Exercise 10.1 and Example 10.7, show that Theorem 10.6 for the case where $G = \mathcal{P}$ implies Exercise 2.1.

Exercise 10.3. Show Theorem 10.6 for the case where $G = \mathcal{P}$. *[Hint: Adapt the proof "by calculation" of Proposition 6.2 to take into account ε-approximations.]*

The main ingredient for the proof of Theorem 10.6 is an appropriate notion of Fourier transform over non-abelian groups. Given an irreducible representation $\rho : G \to U_{d_\rho}(\mathbb{C})$, define

$$\hat{f}(\rho) = \mathrm{E}_{x\in G}\, f(x) \otimes \overline{\rho(x)}. \tag{10.7}$$

In case G is abelian, we always have $d_\rho = 1$, the tensor product is a product, and (10.7) reduces to the usual definition of Fourier coefficient. The only properties we will need of irreducible representations is that they satisfy the relation

$$\sum_\rho d_\rho \,\mathrm{Tr}(\rho(x)) = |G|\delta_{xe}\,, \tag{10.8}$$

[2] The condition $f(-1) = -\,\mathrm{Id}$ is necessary, as there are four 1-dimensional representations of $\mathcal{P}$: all combinations $f(\sigma_X) = \pm 1$ and $f(\sigma_Z) = \pm 1$. We have found the right number of irreps: $1 \cdot 2^2 + 4 \cdot 1 = 8 = |\mathcal{P}|$.

for any $x \in G$. Note that plugging in $x = e$ (the identity element in G) yields $\sum_\rho d_\rho^2 = |G|$.

Proof of Theorem 10.6. Our first step is to define an isometry $V : \mathbb{C}^d \to \mathbb{C}^d \otimes (\oplus_\rho \mathbb{C}^{d_\rho} \otimes \mathbb{C}^{d_\rho})$ by

$$V:\; u \in \mathbb{C}^d \;\mapsto\; \bigoplus_\rho d_\rho^{1/2} \sum_{i=1}^{d_\rho} \big(\hat{f}(\rho)(u \otimes e_i)\big) \otimes e_i,$$

where the direct sum ranges over all irreducible representations ρ of G and $\{e_i\}$ is the canonical basis.[3] Note what V does: it "embeds" any vector $u \in \mathbb{C}^d$ into a direct sum, over irreducible representations ρ, of a d-dimensional vector of $d_\rho \times d_\rho$ matrices. Each (matrix) entry of this vector can be thought of as the Fourier coefficient of the corresponding entry of the vector $f(x)u$ associated with ρ. The fact that V is an isometry follows from the appropriate extension of Parseval's formula:

$$\begin{aligned} V^*V &= \sum_\rho d_\rho \sum_i (I \otimes e_i^*) \hat{f}(\rho)^* \hat{f}(\rho)(I \otimes e_i) \\ &= \mathrm{E}_{x,y}\, f(x)^* f(y) \sum_\rho d_\rho \sum_i (e_i^* \rho(x)^T \overline{\rho(y)} e_i) \\ &= \sum_\rho \frac{d_\rho^2}{|G|} I = I, \end{aligned}$$

where for the second line we used the definition (10.7) of $\hat{f}(\rho)$ and for the third we used (10.8) and the fact that f takes values in the unitary group.

Next define

$$g(x) = \bigoplus_\rho \big(I_d \otimes I_{d_\rho} \otimes \rho(x)\big),$$

a direct sum over all irreducible representations of G (hence itself a representation). Lets' first compute the "pull-back" of g by V: following a similar calculation as above, for any $x \in G$,

$$\begin{aligned} V^* g(x) V &= \sum_\rho d_\rho \sum_{i,j} (I \otimes e_i^*) \hat{f}(\rho)^* \hat{f}(\rho)(I \otimes e_j) \otimes e_i^* \rho(x) e_j) \\ &= \mathrm{E}_{z,y}\, f(z)^* f(y) \sum_\rho d_\rho \sum_{i,j} (e_i^* \rho(z)^T \overline{\rho(y)} e_j)\big(e_i^* \rho(x) e_j\big) \\ &= \mathrm{E}_{z,y}\, f(z)^* f(y) \sum_\rho d_\rho \mathrm{Tr}\big(\rho(z)^T \overline{\rho(y)} \rho(x)^T\big) \\ &= \mathrm{E}_{z,y}\, f(z)^* f(y) \sum_\rho d_\rho \mathrm{Tr}\big(\rho(z^{-1} y x^{-1})\big) \\ &= \mathrm{E}_z\, f(z)^* f(zx), \end{aligned}$$

where the last equality uses (10.8). It then follows that

$$\mathrm{E}_x \big\langle f(x), V^* g(x) V \big\rangle_\sigma \;\; = \mathrm{E}_{x,z} \mathrm{Tr}\big(f(x) f(zx)^* f(z) \sigma\big).$$

This relates correlation of f with $V^* g V$ to the quality of f as an approximate representation and proves the theorem. □

[3] Observe that this expression directly generalizes (6.1).

10.1.3 Application: rigidity for the Magic Square game

Recall the Magic Square game from Section 3.3.1. In Lemma 3.12 we analyzed perfect strategies in this game and showed that any perfect strategy must "have a qubit". Remembering the proof, we had seen that Bob's observables in a perfect strategy must form an operator solution to the underlying system of equations, and that any such operator solution must contain two anti-commuting observables. Using Theorem 10.6 we can extend our earlier result to the case of approximate strategies with very little extra work. (This was first shown in [WBMS16] using a more direct proof.)

Theorem 10.8. *Let $\varepsilon > 0$, and suppose that a strategy for the players in the Magic Square game, using a bipartite state $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ and observables $B_1, \ldots, B_9$ for Bob succeeds with probability $1 - \varepsilon$, for some $\varepsilon \geq 0$. Then $(|\psi\rangle, B_2, B_4)$ is an $O(\sqrt{\varepsilon})$-approximate qubit. Moreover, there are local isometries $V_A, V_B : \mathbb{C}^d \to \mathbb{C}^2 \otimes \mathbb{C}^{d'}$ such that*

$$\big\| V_A \otimes V_B |\psi\rangle - |\phi^+\rangle \otimes |\psi'\rangle \big\| = O(\sqrt{\varepsilon}) \,, \tag{10.9}$$

and

$$\big\| (V_A \otimes V_B)(\mathrm{Id} \otimes B_2)|\psi\rangle - (\sigma_Z \otimes \mathrm{Id})|\phi^+\rangle \otimes |\psi'\rangle \big\| \;=\; O\big(\sqrt{\varepsilon}\big) \,, \tag{10.10}$$

and a similar relation holds with B_2 replaced by B_4 and σ_Z replaced by σ_X.

Note that this is a slightly weaker variant of Theorem 3.16 that we stated without proof, because it only characterizes a single qubit of the strategy, instead of two for the case of Theorem 3.16. This will suffice for our purposes.

Proof. For the first step of the proof we follow the proof of Claim 3.11, except that all equalities must be made approximate equalities. Assume without loss of generality that Alice's strategy is specified by six 4-outcomes projective measurements $\{A^i_{a_1,a_2,a_3}\}$, where $a_1, a_2, a_3 \in \{\pm 1\}$ range over the four possible assignments that satisfy the constraint associated with the i-th row $(i \in \{1,2,3\})$ or $(i-3)$-th column $(i \in \{4,5,6\})$. (We can assume that Alice always returns a valid assignment because she knows that she will lose if not, so we do not even need to include a symbol for such evidently wrong answers in the game.)

For any $i \in \{1, \ldots, 9\}$, in addition to Bob's observable B_i associated with question i we define two observables from Alice's strategy, obtained by recording her answer associated to entry i when she is asked the unique row containing i — call this observable C_i — or the unique column containing i — call this observable C'_i. Formally, $C_1 = \sum_{a_1,a_2,a_3 \in \{\pm 1\}} a_1 A^1_{a_1,a_2,a_3}$, and similar relations can be used to define each C_i and C'_i. Due to the parity constraints enforced on Alice's answers it is always the case that $C_1 C_2 C_3 = +\mathrm{Id}, \ldots, C_3 C_6 C_9 = -\mathrm{Id}$.

By definition, success of the strategy in the game implies 18 equations of the form

$$\langle \psi | C_i \otimes B_i |\psi\rangle \;\geq\; 1 - 18\varepsilon \qquad \text{and} \qquad \langle \psi | C'_i \otimes B_i |\psi\rangle \;\geq\; 1 - 18\varepsilon \,, \tag{10.11}$$

for all $i \in \{1, \ldots, 9\}$. Indeed, each such equation represents the probability that the players return valid answers, conditioned on each of the 18 possible pairs of questions in the game. These relations allow us to mimic the proof of Claim 3.11, as follows:

$$\begin{aligned}
B_2B_4|\psi\rangle &\approx C_4 \otimes B_2|\psi\rangle\\
&= C_5C_6 \otimes B_2|\psi\rangle\\
&\approx C_5C_6C_2 \otimes \mathrm{Id}\,|\psi\rangle\\
&= C_5C_6C_3C_1 \otimes \mathrm{Id}\,|\psi\rangle\\
&\approx C_5C_6C_3 \otimes B_1|\psi\rangle\\
&\approx C_5C_6 \otimes B_1B_3|\psi\rangle\\
&\approx C_5C_6' \otimes B_1B_3|\psi\rangle\\
&\approx C_5C_6'C_3' \otimes B_1|\psi\rangle\\
&= C_5C_9' \otimes B_1|\psi\rangle\ ,
\end{aligned}$$

where here we use the notation $|u\rangle \approx |v\rangle$ to mean $\||u\rangle - |v\rangle\|^2 = O(\varepsilon)$. Here, each of the approximations is obtained by bounding the squared norm of the difference using the Cauchy-Schwarz inequality and the required relation; for example, for the first approximation we write

$$\begin{aligned}
\big\|(\mathrm{Id} \otimes B_2B_4 - C_4 \otimes B_2)|\psi\rangle\big\|^2 &= \big\|(\mathrm{Id} \otimes B_2)(\mathrm{Id} \otimes B_4 - C_4 \otimes \mathrm{Id})|\psi\rangle\big\|^2\\
&= \big\|(\mathrm{Id} \otimes B_4 - C_4 \otimes \mathrm{Id})|\psi\rangle\big\|^2\\
&= 2 - 2\langle\psi|B_4 \otimes C_4|\psi\rangle\\
&\leq 36\varepsilon\ ,
\end{aligned}$$

where the derivation uses that B_2, B_4 and C_4 are Hermitian and square to identity. Using a similar chain of approximations starting from $B_4B_2|\psi\rangle$, it follows that $(|\psi\rangle, B_2, B_4)$ is an $O(\sqrt{\varepsilon})$-approximate qubit.

As already observed in Exercise 10.1, it follows that B_2 and B_4 induce an approximate representation of $\mathcal{P}$ by setting

$$f(\pm\,\mathrm{Id}) = \pm\,\mathrm{Id}, \quad f(\pm\sigma_Z) = \pm B_2, \qquad f(\pm\sigma_X) = \pm B_4, \quad f(\pm\sigma_X\sigma_Z) = \pm B_4B_2\ .$$

Note that this is a legal definition, since B_2, B_4, and B_2B_4 are all unitary. Moreover, using only the approximate anti-commutation and the fact that B_2 and B_4 are observables it is immediate to verify that the conditions of Theorem 10.6 are satisfied, i.e. f is an $(O(\varepsilon), \rho_\mathsf{B})$-representation of $\mathcal{P}$, where ρ_B denotes the reduced density of $|\psi\rangle$ on $\mathcal{H}_\mathsf{B}$.

Applying the theorem, there must exist an exact representation g of $\mathcal{P}$ to which f is close. However, as we saw in Example 10.7 the representation theory of $\mathcal{P}$ is not complicated. It has four 1-dimensional irreducible representations, but all of them map $-\,\mathrm{Id}$ to 1, so they cannot be close to f. Since any representation is a direct sum of irreducible representations, and all irreducible representations of $\mathcal{P}$ are far from f, g must be a direct sum of multiple copies of the unique irreducible 2-dimensional representation of $\mathcal{P}$, which is precisely given by the Pauli matrices, together with possibly a small (relative to the total dimension) number of 1-dimensional relations. Ignoring the presence of such representations for simplicity,[4] $g(\sigma_Z) = \sigma_Z \otimes \mathrm{Id}$ and $g(\sigma_X) = \sigma_X \otimes \mathrm{Id}$, which gives (10.10).

To conclude (10.9) we observe that the relations (10.11) imply that also $(\rho_\mathsf{A}, C_2, C_4)$ is an $O(\sqrt{\varepsilon})$-approximate qubit. This allows us to define an approximate representation of $\mathcal{P}$ on $\mathcal{H}_\mathsf{A}$, and apply Theorem 10.6 again to obtain on isometry V_A which maps $(\rho_\mathsf{A}, C_2, C_4)$ close to an exact qubit. Letting

$$|\psi'\rangle = V_\mathsf{A} \otimes V_\mathsf{B}|\psi\rangle \in (\mathbb{C}^2 \otimes \mathcal{H}_\mathsf{A}') \otimes (\mathbb{C}^2 \otimes \mathcal{H}_\mathsf{B}')$$

[4] To account for them we would select the “corner” of the range of the isometry V that includes only copies of the 2-dimensional irrep; this is a simple technicality.

we see from (10.11) and (10.10) that

$$\big|\langle\psi|\big(\sigma_X\otimes\sigma_X\otimes \mathrm{Id}_{\mathsf{A'B'}} + \sigma_Z\otimes\sigma_Z\otimes \mathrm{Id}_{\mathsf{A'B'}}\big)|\psi\rangle\big| \geq 1-O(\varepsilon)\;.$$

To conclude note that $\frac{1}{2}(\sigma_X\otimes\sigma_X+\sigma_Z\otimes\sigma_Z)$ is an observable with a single eigenvalue 1, with associated eigenvector $|\phi^+\rangle$, and all other eigenvalues equal to 0 or -1. □

10.2 Testing n qubits

The intuition we gained from analyzing the Magic Square game provides us with a clear roadmap for the design of a test for n qubits: (i) design a game such that success in the game requires the players to share observables that satisfy all relations that we expect from elements of the group generated by n quits $(X_1,Z_1),\dots,(X_n,Z_n)$, (ii) apply Theorem 10.6 to obtain closeness of the strategy to an exact representation of this group, and (iii) use that, hopefully, all non-trivial representations gives us what we want, i.e. n exact qubits.

We start in Section 10.2.1 by studying the underlying group of n qubits. Then in Section 10.2.2 we design "tests" for the group product relation.

10.2.1 The Weyl-Heisenberg group

Denote by $\mathcal{P}_n$ the "n-qubit Weyl-Heisenberg group," i.e. the matrix group generated by n-fold tensor products of single-qubit σ_X and σ_Z matrices. The group $\mathcal{P}_n$ has cardinality $2\cdot 4^n$, and elements of $\mathcal{P}_n$ have a unique representative of the form $\pm\sigma_X(a)\sigma_Z(b)$ for $a,b\in\{0,1\}^n$.

The irreducible representations of $\mathcal{P}_n$ are easily computed from those of $\mathcal{P}$; for us the only thing that matters is that the only irreducible representation which satisfies $g(-\mathrm{Id})=-g(\mathrm{Id})$ has dimension 2^n and is given by the defining matrix representation (in fact, it is the only irreducible representation in dimension larger than 1).

With the upcoming application to a qubit test in mind, we state a version of the Gowers-Hatami theorem tailored to the group $\mathcal{P}_n$ and a specific choice of presentation for the group relations.

Corollary 10.9. *Let n,d be integer, $\varepsilon\geq 0$, $|\psi\rangle\in\mathbb{C}^d\otimes\mathbb{C}^d$ a permutation-invariant state, σ the reduced density of $|\psi\rangle$ on either system, and $f:\{X,Z\}\times\{0,1\}^n\to U(\mathbb{C}^d)$. For $a,b\in\{0,1\}^n$ let $X(a)=f(X,a)$, $Z(b)=f(Z,b)$, and assume $X(a)^2=Z(b)^2=I_d$ for all a,b. Suppose that the following inequalities hold: consistency*

$$\mathrm{E}_a\,\langle\psi|\big(X(a)\otimes X(a)\big)|\psi\rangle\;\geq\;1-\varepsilon\;,\qquad \mathrm{E}_b\,\langle\psi|\big(Z(b)\otimes Z(b)\big)|\psi\rangle\;\geq\;1-\varepsilon\;,\tag{10.12}$$

linearity

$$\mathrm{E}_{a,a'}\,\big\|X(a)X(a')-X(a+a')\big\|_\sigma^2\leq\varepsilon\;,\qquad \mathrm{E}_{b,b'}\,\big\|Z(b)Z(b')-Z(b+b')\big\|_\sigma^2\leq\varepsilon\;,\tag{10.13}$$

and anti-commutation

$$\mathrm{E}_{a,b}\,\big\|X(a)Z(b)-(-1)^{a\cdot b}X(a)Z(b)\big\|_\sigma^2\;\leq\;\varepsilon\;.\tag{10.14}$$

Then there exists a $d'\geq d$, an isometry $V:\mathbb{C}^d\to\mathbb{C}^{d'}$, and a representation $g:\mathcal{P}_n\to U_{d'}(\mathbb{C})$ such that $g(-I)=-I_{d'}$ and

$$\mathrm{E}_{a,b}\,\big\|X(a)Z(b)-V^*g(\sigma_X(a)\sigma_Z(b))V\big\|_\sigma^2\;=\;O(\varepsilon)\;.$$

Note that the conditions (10.13) and (10.14) in the corollary are very similar to the conditions required of an approximate representation of the group $\mathcal{P}_n$; in fact it is easy to convince oneself that their exact analogue suffices to imply all the group relations. The reason for choosing those specific relations is that they can be checked using games; see the next subsection for this. Condition (10.12) is necessary to derive the conditions for the application of the Gowers-Hatami theorem from (10.13) and (10.14), and is also testable; see the proof. The conclusion of the corollary implies the conclusion of Theorem 10.1 simply by using the aforementioned fact on non-trivial representations of $\mathcal{P}_n$ (see details at the end of Section 10.2.3).

Remark 10.10. Corollary 10.9 can be seen as an extension of the Blum-Luby-Rubinfeld linearity test [BLR93]. The latter makes a similar statement, but for the commutative group $\{\pm\sigma_X(a) | \, a \in \{0,1\}^n\}$.

Proof. To apply the Gowers-Hatami theorem we need to construct an (ε, σ)-representation f of the group $\mathcal{P}_n$. Using that any element of $\mathcal{P}_n$ has a unique representative of the form $\pm\sigma_X(a)\sigma_Z(b)$ for $a, b \in \{0,1\}^n$, we define $f(\pm\sigma_X(a)\sigma_Z(b)) = \pm X(a)Z(b)$. Next we need to verify that f is an approximate representation. Let $x, y \in \mathcal{P}_n$ be such that $x = \sigma_X(a_x)\sigma_Z(b_x)$ and $y = \sigma_X(a_y)\sigma_Z(b_y)$ for n-bit strings (a_x, b_x) and (a_y, b_y) respectively. Up to phase, we can exploit successive cancellations to decompose $(f(x)f(y)^* - f(xy^{-1})) \otimes I$ as

$$\begin{aligned}
&\big(X(a_x)Z(b_x)X(a_y)Z(b_y) - (-1)^{a_y \cdot b_x} X(a_x + a_y)Z(b_x + b_y)\big) \otimes I \\
&\quad = X(a_x)Z(b_x)X(a_y)\big(Z(b_y) \otimes I - I \otimes Z(b_y)\big) \\
&\qquad + X(a_x)\big(Z(b_x)X(a_y) - (-1)^{a_y \cdot b_x} X(a_y)Z(b_x)\big) \otimes Z(b_y) \\
&\qquad + (-1)^{a_y \cdot b_x}\big(X(a_x)X(a_y) \otimes Z(b_y)\big)\big(Z(b_x) \otimes I - I \otimes Z(b_x)\big) \\
&\qquad + (-1)^{a_y \cdot b_x}\big(X(a_x)X(a_y) \otimes Z(b_y)Z(b_x) - X(a_x + a_y) \otimes Z(b_x + b_y)\big) \\
&\qquad + (-1)^{a_y \cdot b_x}\big(X(a_x + a_y) \otimes I\big)\big(I \otimes Z(b_x + b_y) - Z(b_x + b_y) \otimes I\big).
\end{aligned}$$

(It is worth staring at this sequence of equations for a little bit. In particular, note the "player-switching" that takes place in the 2nd, 4th and 6th lines; this is used as a means to "commute" the appropriate unitaries, and is the reason for including (10.12) among the assumptions of the corollary.) Evaluating each term on the state $|\psi\rangle$, taking the squared Euclidean norm, and then the expectation over uniformly random a_x, a_y, b_x, b_y, the inequality $\|AB|\psi\rangle\| \leq \|A\|\|B|\psi\rangle\|$ and the assumptions of the theorem let us bound the overlap of each term in the resulting summation by $O(\varepsilon)$. Using $\|(A \otimes I)|\psi\rangle\| = \|A\|_\sigma$ by definition and the triangle inequality we have obtained the bound

$$\mathrm{E}_{x,y} \left\| f(x)f(y)^* - f(xy^{-1}) \right\|_\sigma^2 = O(\varepsilon).$$

We are now in a position to apply the Gowers-Hatami theorem, which gives an isometry V and exact representation g such that

$$\mathrm{E}_{a,b} \left\| X(a)Z(b) - \frac{1}{2}V^*\big(g(\sigma_X(a)\sigma_Z(b)) - g(-\sigma_X(a)\sigma_Z(b))\big)V \right\|_\sigma^2 = O(\varepsilon). \tag{10.15}$$

Using that g is a representation, $g(-\sigma_X(a)\sigma_Z(b)) = g(-I)g(\sigma_X(a)\sigma_Z(b))$. It follows from (10.15) that $\|g(-I) + I\|_\sigma^2 = O(\varepsilon)$, so we may restrict the range of V to the subspace where $g(-I) = -I$ without introducing much additional error. □

10.2.2 Testing the Weyl-Heisenberg group relations

Corollary 10.9 makes three assumptions about the observables $X(a)$ and $Z(b)$: that they satisfy approximate consistency (10.12), linearity (10.13), and anti-commutation (10.14). To complete our test, we need to show how these relations can be "certified" in a two-player game. There are multiple ways this can be done; we give one. We start by introducing two stand-alone "tests" that we later combine to define the game G_n.

Linearity test:

(a) The referee selects $W \in \{X, Z\}$ and $a, a' \in \{0,1\}^n$ uniformly at random. She sends (W, a, a') to one player and (W, a), (W, a'), or $(W, a + a')$ to the other.[5]

(b) The first player replies with two bits, and the second with a single bit. The referee accepts if and only if the player's answers are consistent.

As always in this section, the test treats both players symmetrically. As a result (see Remark 10.2 we can assume that the players' strategy is symmetric, and is specified by a permutation-invariant state $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ and a measurement for each question: an observable $W(a)$ associated to questions of the form (W, a), and a four-outcome measurement $\{W_{a,a'}\}$ associated with questions of the form (W, a, a').

The linearity test described above is almost identical to the BLR linearity test, except for the use of the basis label $W \in \{X, Z\}$. The following lemma states conditions that a strategy must satisfy in order to succeed with high probability in the test.

Lemma 10.11. *Suppose that a family of observables $\{W(a)\}$ for $W \in \{X, Z\}$ and $a \in \{0,1\}^n$, generates outcomes that succeed in the linearity test with probability $1 - \varepsilon$, when applied on a symmetric bipartite state $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ with reduced density matrix σ. Then the following hold: approximate consistency*

$$\mathrm{E}_a \langle\psi|\big(X(a) \otimes X(a)\big)|\psi\rangle \;=\; 1 - O(\varepsilon), \qquad \mathrm{E}_b \langle\psi|\big(Z(b) \otimes Z(b)\big)|\psi\rangle \;\geq\; 1 - O(\varepsilon),$$

and linearity

$$\mathrm{E}_{a,a'} \big\|X(a)X(a') - X(a + a')\big\|_\sigma^2 = O(\varepsilon), \qquad \mathrm{E}_{b,b'} \big\|Z(b)Z(b') - Z(b + b')\big\|_\sigma^2 \;=\; O(\varepsilon).$$

Exercise 10.4. Prove the lemma. (In the case of classical strategies, the conditions are an immediate reformulation of the test. The proof for quantum strategies is not much harder.)

Testing anti-commutation can be done using the Magic Square game.

Anti-commutation test:

(a) The referee selects $a, b \in \{0,1\}^n$ uniformly at random under the condition that $a \cdot b = 1$. She plays the Magic Square game with both players, with the following modifications: if the question to the second player is 2 or 4 she sends (X, a) or (Z, b) instead; in all other cases he sends the original label of the question in the Magic Square game together with both strings a and b.

(b) Each player provides answers as in the Magic Square game. The referee accepts if and only if the player's answers would have been accepted in the game.

Using Theorem 10.8 it is straightforward to show the following.

[5] Elements such as (W, a) are labels sent to the players as their question, and carry no other intrinsic meaning.

Lemma 10.12. *Suppose a strategy for the players succeeds in the anti-commutation test with probability at least $1-\varepsilon$, when performed on a symmetric bipartite state $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ with reduced density matrix σ. Then the observables $X(a)$ and $Z(b)$ applied by the player upon receipt of questions (X,a) and (Z,b) respectively satisfy*

$$\mathrm{E}_{a,b:\,a\cdot b=1}\,\big\|X(a)Z(b)-(-1)^{a\cdot b}Z(b)X(a)\big\|_\sigma^2 \,=\, O\big(\sqrt{\varepsilon}\big). \tag{10.16}$$

10.2.3 Application: an n-qubit test

We are ready to put all the pieces together and make explicit the game G_n that underlies Theorem 10.1. For historical reasons we call this game the "n-qubit Pauli braiding test".

> **n-qubit Pauli braiding test:** With probability $1/3$ each,
>
> (a) Execute the linearity test;
>
> (b) Execute the anti-commutation test;
>
> (c) Execute the following consistency test: Send one player a label $W\in\{X,Z\}$ uniformly at random, and to the other (W,a) for $a\in\{0,1\}^n$ chosen uniformly at random. Expect answers $c\in\{0,1\}^n$ and $c'\in\{0,1\}$ respectively. Accept if and only if $a\cdot c=c'$.

Proof sketch of Theorem 10.1. For any integer $n\geq 1$ let G_n be the n-qubit Pauli braiding test. Suppose that a family of observables $W(a)$, for $W\in\{X,Z\}$ and $a\in\{0,1\}^n$, together with projective measurements $\{A_a^X\}_{a\in\{0,1\}^n}$ and $\{A_b^Z\}_{b\in\{0,1\}^n}$ and a state $|\psi\rangle\in\mathbb{C}^d\otimes\mathbb{C}^d$ specify a symmetric strategy that succeeds in with probability at least $1-\varepsilon$ in the game.

Using Lemma 10.11 and Lemma 10.12, success with probability $1-\varepsilon$ implies that conditions (10.12), (10.13) and (10.14) in Corollary 10.9 are all satisfied, up to error $O(\sqrt{\varepsilon})$. (In fact, Lemma 10.12 only implies (10.14) for strings a,b such that $a\cdot b=1$. The condition for string such that $a\cdot b=0$ follows from the other conditions.) The conclusion of the corollary is that there exists an isometry V such that the observables $X(a)$ and $Z(b)$ satisfy

$$\mathrm{E}_{a,b}\,\big\|X(a)Z(b)-V^*g(\sigma_X(a)\sigma_Z(b))V\big\|_\sigma^2 \,=\, O(\sqrt{\varepsilon})\;,$$

for some non-trivial representation g of $\mathcal{P}_n$. Using what we know of non-trivial representations, (10.1) follows. To obtain (10.2) we use success in part (c) of the test, which using $\sigma_a^X=\mathrm{E}_b(-1)^{a\cdot b}\sigma_X(b)$ and similarly for σ_b^Z immediately translates into the desired relation.

Finally, using the consistency relations (10.12) that follow from part (a) of the test together with the above we get

$$\mathrm{E}_{a,b}\,\langle\psi|(V\otimes V)^*\big(\sigma_X(a)\sigma_Z(b)\otimes\sigma_X(a)\sigma_Z(b)\big)(V\otimes V)|\psi\rangle \,=\, 1-O(\sqrt{\varepsilon}).$$

Using similar reasoning as the end of the proof of Theorem 10.8, (10.3) follows. □

Bibliography

[Aar10] Scott Aaronson. Bqp and the polynomial hierarchy. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 141–150. ACM, 2010.

[AB09] Sanjeev Arora and Boaz Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[ACGK17] Scott Aaronson, Alexandru Cojocaru, Alexandru Gheorghiu, and Elham Kashefi. On the implausibility of classical client blind quantum computing. *arXiv preprint arXiv:1704.08482*, 2017.

[AG17] Dorit Aharonov and Ayal Green. A quantum inspired proof of $\mathrm{P}^{\backslash\#\mathrm{p}} \subseteq \mathrm{IP}$. *arXiv preprint arXiv:1710.09078*, 2017.

[AGV09] Adi Akavia, Shafi Goldwasser, and Vinod Vaikuntanathan. Simultaneous hardcore bits and cryptography against memory attacks. In *Theory of Cryptography Conference*, pages 474–495. Springer, 2009.

[AV12] Dorit Aharonov and Umesh Vazirani. Is quantum mechanics falsifiable? A computational perspective on the foundations of quantum mechanics. *arXiv preprint arXiv:1206.3686*, 2012.

[AV13] Dorit Aharonov and Umesh Vazirani. *Is quantum mechanics falsifiable? A computational perspective on the foundations of quantum mechanics*. Computability: Turing, Gödel, Church, and Beyond. MIT Press, 2013.

[BC20] Oren Becker and Michael Chapman. Stability of approximate group actions: uniform and probabilistic. *arXiv preprint arXiv:2005.06652*, 2020.

[BCM^{+}18] Zvika Brakerski, Paul Christiano, Urmila Mahadev, Umesh Vazirani, and Thomas Vidick. A cryptographic test of quantumness and certifiable randomness from a single quantum device. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 320–331. IEEE, 2018.

[Bel64] John S Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of computer and system sciences*, 47(3):549–595, 1993.

[BOGKW19] Michael Ben-Or, Shafi Goldwasser, Joe Kilian, and Avi Wigderson. Multi-prover interactive proofs: How to remove intractability assumptions. In *Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali*, pages 373–410. 2019.

[Bre06] Frédéric Brechenmacher. *Histoire du théorème de Jordan de la décomposition matricielle (1870-1930). Formes de représentation et méthodes de décomposition.* PhD thesis, 2006.

[BV14] Zvika Brakerski and Vinod Vaikuntanathan. Efficient fully homomorphic encryption from (standard) lwe. *SIAM Journal on Computing*, 43(2):831–871, 2014.

[Cap15] Valerio Capraro. *Connes' Embedding Conjecture*, pages 73–107. Springer International Publishing, Cham, 2015.

[CCKW19] Alexandru Cojocaru, Léo Colisson, Elham Kashefi, and Petros Wallden. Qfactory: classically-instructed remote secret qubits preparation. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 615–645. Springer, 2019.

[CLS17] Richard Cleve, Li Liu, and William Slofstra. Perfect commuting-operator strategies for linear system games. *Journal of Mathematical Physics*, 58(1):012202, 2017.

[CM14] Richard Cleve and Rajat Mittal. Characterization of binary constraint system games. In *International Colloquium on Automata, Languages, and Programming*, pages 320–331. Springer, 2014.

[CM16] Toby Cubitt and Ashley Montanaro. Complexity classification of local hamiltonian problems. *SIAM Journal on Computing*, 45(2):268–316, 2016.

[Con76] Alain Connes. Classification of injective factors cases II_1, II_∞, III_λ, $\lambda \neq 1$. *Annals of Mathematics*, pages 73–115, 1976.

[CR20] Rui Chao and Ben W Reichardt. Quantum dimension test using the uncertainty principle. *arXiv preprint arXiv:2002.12432*, 2020.

[CRSV17] Rui Chao, Ben W Reichardt, Chris Sutherland, and Thomas Vidick. Overlapping qubits. *arXiv preprint arXiv:1701.01062*, 2017.

[CRSV18] Rui Chao, Ben W Reichardt, Chris Sutherland, and Thomas Vidick. Test for a large amount of entanglement, using few measurements. *Quantum*, 2:92, 2018.

[CS17] Andrea Coladangelo and Jalex Stark. Robust self-testing for linear constraint system games. *arXiv preprint arXiv:1709.09267*, 2017.

[CS18] Andrea Coladangelo and Jalex Stark. Unconditional separation of finite and infinite-dimensional quantum correlations. *arXiv preprint arXiv:1804.05116*, 2018.

[DFPR14] Vedran Dunjko, Joseph F Fitzsimons, Christopher Portmann, and Renato Renner. Composable security of delegated quantum computation. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 406–425. Springer, 2014.

[EPR35] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical review*, 47(10):777, 1935.

[Fri12] Tobias Fritz. Tsirelson's problem and Kirchberg's conjecture. *Reviews in Mathematical Physics*, 24(05):1250012, 2012.

[Gen09] Craig Gentry. Fully homomorphic encryption using ideal lattices. In *STOC*, volume 9, pages 169–178, 2009.

[GH17] William Timothy Gowers and Omid Hatami. Inverse and stability theorems for approximate representations of finite groups. *Sbornik: Mathematics*, 208(12):1784, 2017.

[GKK19] Alexandru Gheorghiu, Theodoros Kapourniotis, and Elham Kashefi. Verification of quantum computation: An overview of existing approaches. *Theory of computing systems*, 63(4):715–808, 2019.

[GKR08] Shafi Goldwasser, Yael Tauman Kalai, and Guy N Rothblum. Delegating computation: interactive proofs for muggles. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 113–122. ACM, 2008.

[GKW17] Rishab Goyal, Venkata Koppula, and Brent Waters. Lockable obfuscation. In *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on*, pages 612–621. IEEE, 2017.

[GKW18] Rishab Goyal, Venkata Koppula, and Brent Waters. Collusion resistant traitor tracing from learning with errors. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 660–670. ACM, 2018.

[GMR85] Shafi Goldwasser, Silvio Micali, and Ronald Rivest. A "paradoxical" solution to the signature problem. In *Advances in Cryptology*, pages 467–467. Springer, 1985.

[GV19] Alexandru Gheorghiu and Thomas Vidick. Computationally-secure and composable remote state preparation. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1024–1033. IEEE, 2019.

[GVW01] Oded Goldreich, Salil Vadhan, and Avi Wigderson. On interactive proofs with a laconic prover. In *International Colloquium on Automata, Languages, and Programming*, pages 334–345. Springer, 2001.

[GVW15] Sergey Gorbunov, Vinod Vaikuntanathan, and Hoeteck Wee. Attribute-based encryption for circuits. *Journal of the ACM (JACM)*, 62(6):45, 2015.

[Ito10] Tsuyoshi Ito. Polynomial-space approximation of no-signaling provers. In *International Colloquium on Automata, Languages, and Programming*, pages 140–151. Springer, 2010.

[IV12] Tsuyoshi Ito and Thomas Vidick. A multi-prover interactive proof for nexp sound against entangled provers. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 243–252. IEEE, 2012.

[Ji13] Zhengfeng Ji. Binary constraint system games and locally commutative reductions. *arXiv preprint arXiv:1310.3794*, 2013.

[JNP$^+$11] Marius Junge, Miguel Navascues, Carlos Palazuelos, David Perez-Garcia, Volkher B Scholz, and Reinhard F Werner. Connes' embedding problem and Tsirelson's problem. *Journal of Mathematical Physics*, 52(1):012102, 2011.

[JNV$^+$20a] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. MIP*= RE. *arXiv preprint arXiv:2001.04383*, 2020.

[JNV$^+$20b] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. Quantum soundness of the classical low individual degree test. *arXiv preprint arXiv:2009.12982*, 2020.

[JP11] Marius Junge and Carlos Palazuelos. Large violation of bell inequalities with low entanglement. *Communications in Mathematical Physics*, 306(3):695, 2011.

[Kir93] Eberhard Kirchberg. On non-semisplit extensions, tensor products and exactness of group C^*-algebras. *Inventiones mathematicae*, 112(1):449–489, 1993.

[KKMV09] Julia Kempe, Hirotada Kobayashi, Keiji Matsumoto, and Thomas Vidick. Using entanglement in quantum multi-prover interactive proofs. *Computational Complexity*, 18(2):273–307, 2009.

[KMW17] Elham Kashefi, Luka Music, and Petros Wallden. The quantum cut-and-choose technique and quantum two-party computation. *arXiv preprint arXiv:1703.03754*, 2017.

[KRR14] Yael Tauman Kalai, Ran Raz, and Ron D Rothblum. How to delegate computations: the power of no-signaling proofs. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 485–494. ACM, 2014.

[LFKN92] Carsten Lund, Lance Fortnow, Howard Karloff, and Noam Nisan. Algebraic methods for interactive proof systems. *Journal of the ACM (JACM)*, 39(4):859–868, 1992.

[Mah18] Urmila Mahadev. Classical verification of quantum computations. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 259–267. IEEE, 2018.

[Mer90] N David Mermin. Simple unified form for the major no-hidden-variables theorems. *Physical review letters*, 65(27):3373, 1990.

[Mer93] N David Mermin. Hidden variables and the two theorems of john bell. *Reviews of Modern Physics*, 65(3):803, 1993.

[MF16] Tomoyuki Morimae and Joseph F Fitzsimons. Post hoc verification with a single prover. *arXiv preprint arXiv:1603.06046*, 2016.

[Mor18] Tomoyuki Morimae. Blind quantum computing can always be made verifiable. *arXiv preprint arXiv:1803.06624*, 2018.

[MP12] Daniele Micciancio and Chris Peikert. Trapdoors for lattices: Simpler, tighter, faster, smaller. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 700–718. Springer, 2012.

[MYS12] Matthew McKague, Tzyh Haur Yang, and Valerio Scarani. Robust self-testing of the singlet. *Journal of Physics A: Mathematical and Theoretical*, 45(45):455304, 2012.

[NC02] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.

[NPA08] Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008.

[NW19] Anand Natarajan and John Wright. Neexp is contained in mip. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 510–518. IEEE, 2019.

[Oza13a] Narutaka Ozawa. About the Connes embedding conjecture. *Japanese Journal of Mathematics*, 8(1):147–183, 2013.

[Oza13b] Narutaka Ozawa. About the connes embedding conjecture. *Japanese Journal of Mathematics*, 8(1):147–183, 2013.

[P$^+$16] Chris Peikert et al. A decade of lattice cryptography. *Foundations and Trends® in Theoretical Computer Science*, 10(4):283–424, 2016.

[Reg09] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *Journal of the ACM (JACM)*, 56(6):34, 2009.

[RRR16] Omer Reingold, Guy N Rothblum, and Ron D Rothblum. Constant-round interactive proofs for delegating computation. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 49–62. ACM, 2016.

[RT19] Ran Raz and Avishay Tal. Oracle separation of BQP and PH. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 13–23, 2019.

[RUV13] Ben W Reichardt, Falk Unger, and Umesh Vazirani. Classical command of quantum systems. *Nature*, 496(7446):456–460, 2013.

[Sha92] Adi Shamir. IP= PSPACE. *Journal of the ACM (JACM)*, 39(4):869–877, 1992.

[Slo19] William Slofstra. The set of quantum correlations is not closed. In *Forum of Mathematics, Pi*, volume 7. Cambridge University Press, 2019.

[SW87] 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, 1987.

[Tsi93] Boris S Tsirelson. Some results and problems on quantum Bell-type inequalities. *Hadronic Journal Supplement*, 8(4):329–345, 1993.

[Unr16] Dominique Unruh. Computationally binding quantum commitments. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 497–527. Springer, 2016.

[Vid19] Thomas Vidick. From operator algebras to complexity theory and back. *Notices of the American Mathematical Society*, 66(10), 2019.

[Vid20] Thomas Vidick. Verifying quantum computations at scale: A cryptographic leash on quantum devices. *Bulletin of the American Mathematical Society*, 57(1):39–76, 2020.

[VN32] J Von Neumann. Mathematische grundlagen der quantenmechanik. 1932.

[VW16] Thomas Vidick and John Watrous. Quantum proofs. *Foundations and Trends® in Theoretical Computer Science*, 11(1-2):1–215, 2016.

[VZ20] Thomas Vidick and Tina Zhang. Classical proofs of quantum knowledge. *arXiv preprint arXiv:2005.01691*, 2020.

[WBMS16] Xingyao Wu, Jean-Daniel Bancal, Matthew McKague, and Valerio Scarani. Device-independent parallel self-testing of two singlets. *Physical Review A*, 93(6):062121, 2016.

[WZ17] Daniel Wichs and Giorgos Zirdelis. Obfuscating compute-and-compare programs under LWE. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 600–611. IEEE, 2017.

[Yue16] Henry Yuen. A parallel repetition theorem for all entangled games. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.