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 X 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, $\{R, G, B\}^{1000\times 1000}$, 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 $i = 1, \ldots,$ 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 $res^{(1)}, res^{(2)}, \ldots$, 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 $O_x = \sum_a \lambda_a \Pi_a^x$, where λ_a are arbitrary and Π_a^x projections that sum to identity, i.e. Π_a^x is a POVM. According to the Born rule, it obtains an answer distributed as $\Pr(a|x) = \mathrm{Tr}(\Pi_a^x \rho)$.[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 $\Pi_a^x = 1_{f(x)=a}$ 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, $\mathrm{Tr}(\Pi_a^x \rho) = \mathrm{Tr}(\Pi_a^x |\psi\rangle\langle\psi|) = \langle\psi|\Pi_a^x|\psi\rangle$.

[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 meanginful? 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*

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

Note that the lemma no longer says that X is *equal* to $\sigma_X \otimes \mathrm{Id}$ (under the isomprhism 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 for $W \in \{X,Z\}$,

$$\left\|\left(W - \pi(\sigma_W \otimes \mathrm{Id})\pi^{-1}\right)|\psi\rangle\right\|^2 \leq O(\varepsilon) .$$

[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 $\{P_a^x\}_{a\in\mathcal{A}}$ for $x \in \mathcal{X}$ such that $p(a|x) = \langle\psi|P_a^x|\psi\rangle$ 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

$$V(P_0^{x_0} - P_1^{x_0})|\psi\rangle = (\sigma_X \otimes \mathrm{Id})V|\psi\rangle \quad \text{and} \quad V(P_0^{z_0} - P_1^{z_0})|\psi\rangle = (\sigma_Z \otimes \mathrm{Id})V|\psi\rangle . \tag{2.3}$$

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 artefact 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 $P_a^x = p(a|x)$ 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.

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.

[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.

2.4.1 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*

$$|\psi\rangle_{\mathsf{AB}} = \sum_i \sqrt{\lambda_i} |u_i\rangle_\mathsf{A} |v_i\rangle_\mathsf{B} \tag{2.4}$$

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 (2.4) 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 $\mathbb{C}^2$,

$$|\phi^+\rangle = \frac{1}{\sqrt{2}}(|u_0\rangle\overline{|u_0\rangle} + |u_1\rangle\overline{|u_1\rangle}) \; . \tag{2.5}$$

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 $|u_b\rangle\langle u_b|$.

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.

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 $v \in \{0,1\}$.

[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.

[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.

4.-5. Same.

Using the observation (2.5) (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 probability in the test. By definition it is

$$\begin{aligned}\Pr(v = v') = \frac{1}{2}\big(\langle\psi|(|0\rangle\langle 0| \otimes P^0_0)|\psi\rangle + \langle\psi|(|1\rangle\langle 1| \otimes P^0_1)|\psi\rangle\big) \\ + \frac{1}{2}\big(\langle\psi|(|+\rangle\langle +| \otimes P^1_0)|\psi\rangle + \langle\psi|(|-\rangle\langle -| \otimes P^1_1)|\psi\rangle\big)\,. \qquad (2.6)\end{aligned}$$

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

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

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 (2.6) as

$$\Pr(v = v') = \frac{1}{2} + \frac{1}{4}\big(\langle\psi|\sigma_Z \otimes P^0|\psi\rangle + \langle\psi|\sigma_X \otimes P^1|\psi\rangle\big)\ . \tag{2.7}$$

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*

$$\langle\psi|\sigma_X \otimes X|\psi\rangle = \langle\psi|\sigma_Z \otimes Z|\psi\rangle = 1\ . \tag{2.8}$$

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 (2.8) 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 (2.7) to verify that (2.8) 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 $i = 1, \ldots, n$ and sends them to the prover. Once all qubits have been sent, the verifier announces $\theta_1, \ldots, \theta_n$ and expects $v'_1, \ldots, v'_n \in \{0,1\}$ 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-qbut 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 analyzis 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

$$H(R) + H(S) \geq \log_2 \frac{1}{c} . \tag{2.9}$$

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 (2.9) 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 (2.9), 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},

$$H(A|B)_\rho = H(\rho_{\mathsf{AB}}) - H(\rho_{\mathsf{B}}) , \tag{2.10}$$

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:

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

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,

$$\begin{aligned} H(A|B)_\rho &= H(\rho_{\mathsf{AB}}) - H(\rho_{\mathsf{B}}) \\ &= \Big(H(\{p_c\}) + \sum_c p_c H(\rho'_c)\Big) - \Big(H(\{p_c\} + \sum_c p_c H(\rho_c)\Big) \\ &\geq \min_c \big(H(\rho'_c) - H(\rho_c)\big) \\ &= \min_c H(A|Q)_{\rho'_c} \\ &\geq -\log \dim \mathcal{H}_{\mathsf{Q}} . \end{aligned} \tag{2.11}$$

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 $\mathcal{H}_A$ 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 $\mathcal{H}_A \otimes \mathcal{H}_B$. 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*

$$H(A|B)_{R(\rho)} + H(A|B)_{S(\rho)} \geq \log_2 \frac{1}{c} + H(A|B) . \tag{2.12}$$

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.

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.

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

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 in probability 1 in the protocol implies that necessarily the left-hand side in (2.12) 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 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 (2.12) is non-positive (and in fact, equal to 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 (2.11): the dimension of Bob's quantum memory must be at least 2^n. □

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

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 certies "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.