Witness-Indistinguishability Against Quantum Adversaries

6.845 Quantum Complexity Theory – Project Report

Raluca Ada Popa

1 Introduction

Proof systems are a central concept in complexity theory and cryptography. Zero-knowledge and witness-indistinguishability are useful security properties of proof systems. Considering the increased power of quantum computation, it comes as a natural question to understand what happens to these security properties when quantum computation becomes feasible.

Zero-knowledge [GMR89] is the security property of proof systems that has received the most attention. Intuitively, a proof system is zero-knowledge if the prover does not leak any information to the verifier other than the veracity of the statement to be proved. There has been a significant amount of research aimed at characterizing what happens to zero-knowledge when quantum adversaries are possible (e.g., [Wat09, Wat02, CK08, HKSZ08, Kob08]) and we will survey the main such results. A weaker security property of proof systems is witness-indistinguishability. Introduced by Feige and Shamir [FS90] in 1990, witness-indistinguishability roughly means that the verifier cannot distinguish which witness the prover used among the possible witnesses.

In this report, we characterize witness-indistinguishability against quantum adversaries. To the best of our knowledge, quantum witness-indistinguishability has not been studied so far; moreover, witness-indistinguishability is worthwhile to study because:

- Witness-indistinguishability has the beneficial property of being preserved under arbitrary composition of protocols, unlike zero-knowledge.
- It has been a useful tool in building various kinds of zero-knowledge protocols in classical cryptography (e.g., concurrent zero-knowledge [RK99], non black-box simulation techniques [Bar01]).
- It is a weaker notion of zero-knowledge that may sometimes suffice in practice.

We give definitions of witness-indistinguishability in the quantum setting and prove some properties:

- Under some reasonable assumptions, all of IP (and thus NP) has a computational witness-indistinguishable proof against quantum adversaries.
- Any language with a statistical public-coin witness-indistinguishable proof against classical adversaries will have such a proof against quantum adversaries.

2 Preliminaries

We assume the reader is familiar with the notion of interactive proofs and with basic quantum computation notions.

2.1 Notation and Terminology

If (P, V) is a proof system in which P is the prover and V is the verifier, $(P, V(w)))(x)$ indicates that x is an input to both P and V and V takes as auxiliary input w with $|w|$ polynomially-bounded in $|x|$.

By $a \leftarrow A$, we mean that a is a random sample from the distribution A.

Some **abbreviations** will serve brevity and simplicity. We will use $a.Q.$ to mean that a classical protocol is secure against quantum attacks. PPT will denote a probabilistic polynomial-time algorithm.

We will use the same name for a property of a proof-system and for the class of languages having such a property, e.g., a proof system can be ZK and a language can be in ZK.

A *public-coin proof system* is an interactive proof system where the verifier acts as follows: in each round, it tosses a prescribed number of coins and sends the output to the prover; at the end, it decides whether to accept.

A proof system satisfies a property for the *honest verifier* if the property holds with respect to one prescribed verifier (which also satisfies completeness and soundness).

2.2 Classical Interactive Zero-Knowledge

Definition 1 (Computational indistinguishability)**.** *Two probability ensembles indexed by a set of strings S, $A \overset{def}{=} \{A_x\}_{x\in S}$ and $B \overset{def}{=} \{B_x\}_{x\in S}$, are computationally indistinguishable if, for every PPT D, every positive polynomial $p(\cdot)$ and all sufficiently long $x \in S$,*

$$|\Pr[D(A_x, x) = 1] - \Pr[D(B_x, x) = 1]| < 1/p(|x|).$$

Definition 2 (Verifier view)**.** *The view of V in an interactive protocol (P, V) with input x and auxiliary input w, denoted $\mathsf{VIEW}_{(P,V(w)))}(x)$, is a distribution over strings each consisting of all the information available at V during the interaction with P, over the random coins of P and V.*

Definition 3 (Statistical distance)**.** *For D and S two distributions on the same alphabet Σ, the statistical distance $\Delta(D, S) = 1/2 \sum_{x\in\Sigma} |Pr[x \leftarrow D] - \Pr[x \leftarrow S]|$.*

Definition 4 (Zero-knowledge)**.** *Interactive proof system (P, V) is zero-knowledge if, for a polynomial q,*

$$\forall\ \mathsf{PPT}\ V',\ \exists\ \mathsf{PPT}\ S,\ \forall\ x \in L, w \in \{0, 1\}^{q(x)} : \mathsf{VIEW}_{(P,V'(w))}(x) \approx S(w, x).$$

For computational zero-knowledge (denoted CZK), "$\approx$" denotes computational indistinguishability (Def. 1); for statistical zero-knowledge (denoted SZK), "$\approx$" indicates that the statistical distance is negligible: $\Delta(\mathsf{VIEW}_{(P,V'(w))}(x), S(w, x)) < \delta(|x|)$, for some negligible function δ.

Intuitively, the verifier can simulate in polynomial time without the help of the prover the same information it got while interacting with the prover, hence, not learning any information it did not know.

2.3 Classical witness-indistinguishability

Witness-indistinguishability (WI) is a notion introduced by Feige and Shamir [FS90]. Intuitively, WI prevents the verifier from learning which witness the prover used out of the possible witnesses for the common input. It is a weaker notion than zero-knowledge because it may leak some information about the witnesses of x; nevertheless, such information will not enable the verifier to distinguish which witness the prover used. Another useful property of WI is that, if a language has at least two witnesses for every value x, the verifier cannot learn any witness of x [FS90] (the witness is hidden).

Let L be a relation $\{(x, w)\}$ testable in polynomial time, where $|x| = |w|$. For any x, its witness set $w(x)$ is the set w such that $(x, w) \in L$.

Definition 5 (Witness-indistinguishability)**.** *A proof system (P, V) is witness-indistinguishable (WI) over some relation L, if for all PPT V', for all $x \in L$ sufficiently large in size, for all w_1, $w_2 \in w(x)$, for all y auxiliary to V',*

$$\mathsf{VIEW}_{(P(w_1),V'(y))(x)} \approx \mathsf{VIEW}_{(P(w_2),V'(y))(x)}.$$

As in the case of zero-knowledge, for computational witness indistinguishability (denoted CWI), "$\approx$" denotes computational indistinguishability (Def. 1); for statistical witness-indistinguishability (denoted SWI), "$\approx$" indicates that statistical distance (Def. 3) is negligible.

3 Interactive Zero-Knowledge against Quantum Adversaries

In this section, we present existing work in zero-knowledge against quantum adversaries, which will provide background for our witness-indistinguishability study. We focus on classical zero-knowledge protocols against quantum adversaries rather than quantum zero-knowledge protocols. The reason is that it is important to understand whether the protocols we have today will remain secure when quantum computation becomes feasible.

3.1 Definitions

A linear super-operator $\mathbf{\Phi} : L(\mathcal{X}) \to L(\mathcal{Y})$ is *admissible* if it is completely positive and preserves trace. An admissible superoperator maps density matrices to density matrices in a way that is physically-realizable. The diamond norm [Kit97] is used as a distance measure between two admissible maps.

Definition 6 (Diamond norm)**.** *Let $\mathbf{\Phi}_0$, $\mathbf{\Phi}_1 : L(\mathcal{X}) \to L(\mathcal{Y})$ be two admissible superoperators. The diamond norm of the difference between $\mathbf{\Phi}_0$ and $\mathbf{\Phi}_1$ is*

$$||\mathbf{\Phi}_0 - \mathbf{\Phi}_1||_\diamond = \max\{||(\mathbf{\Phi}_0 \otimes \mathbb{I}_{L(\mathcal{W})})(\rho) - (\mathbf{\Phi}_1 \otimes \mathbb{I}_{L(\mathcal{W})})(\rho)||_1 : \rho \in D(\mathcal{X} \otimes \mathcal{W})\}.$$

where $\mathcal{W}$ has the same dimension as $\mathcal{X}$.

The following definitions are due to Watrous [Wat09]. We present the main insights differentiating from classical definitions.

Definition 7 (Polynomial-time generated –**p.g.**)**.** *A collection $\{Q_x : x \in \Sigma^*\}$ is polynomial-time generated if there exists a deterministic polynomial-time Turing machine that, on input $x \in \Sigma^*$, outputs a description of Q_x.*

Let (P, V) be a proof system, such that, for each input x, V takes $q(x)$ input qubits and the total output qubits are $r(x)$, for q and r bounded by some polynomial. The interaction between P and V on input x is an admissible superoperator denoted $\mathbf{\Phi}_x^{(P,V)}$ from density matrices of $q(x)$ qubits to density matrices of $r(x)$ qubits. Similarly, for a simulator S simulating a verifier's view with the same q and r input lengths, for each x, S induces an admissible superoperator $\mathbf{\Psi}_x^S$.

Definition 8 (Statistical zero-knowledge a.Q. – SZKQ)**.** *An interactive proof system (P, V) for a promise problem $A = (A_{\mathsf{yes}}, A_{\mathsf{no}})$ is statistical zero-knowledge a.Q. if, for all p.g. quantum V', there exists a p.g. quantum S such that:*

1. *V' and S have the same polynomially-bounded number of input and output qubits for the same input.*
2. *There exists a negligible function δ such that $||\mathbf{\Phi}_x^{(P,V')} - \mathbf{\Psi}_x^S||_\diamond < \delta(|x|)$, for all $x \in A_{\mathsf{yes}}$.*

This definition seems analogous to the classical one, except for one key difference. By using the diamond norm, the definition includes the case when the auxiliary inputs to the verifier are entangled with some qubits of an external party. In fact, it is possible to define $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ such that $||\mathbf{\Phi}_1 - \mathbf{\Phi}_2||_\diamond = 2$, while $||\mathbf{\Phi}_1(\rho) - \mathbf{\Phi}_2(\rho)||_1$ is exponentially small (in the number of input and output qubits) for all $\rho \in D(\mathcal{W})$. This means that an external physical process could distinguish between a simulator output and the actual view, so the definition of zero-knowledge must prevent against this case.

To define computational zero-knowledge a.Q., we first need to define quantum computational indistinguishability a.Q..

Definition 9 (Polynomially quantum indistinguishable admissible maps)**.** *Assume $S \subseteq \Sigma^*$ is an infinite set of strings, q and r are polynomially bounded functions, and $\mathbf{\Phi}_x$ and $\mathbf{\Psi}_x$ are admissible superoperators from $q(|x|)$ to $r(|x|)$ qubits for each $x \in S$. Then, the ensembles $\{\mathbf{\Phi}_x : x \in S\}$ and $\{\mathbf{\Psi}_x : x \in S\}$ are polynomially quantum indistinguishable, if, for all polynomials p, constants s and k, for all x of sufficiently large size, for all mixed states σ on $q(|x|) + k$ qubits and all $(r(|x| + k)$-input-qubit circuit Q of size s,*

$$|\Pr[Q((\mathbf{\Phi}_x \otimes I_k)(\sigma))] - \Pr[Q((\mathbf{\Psi} \otimes I_k)(\sigma))]| < 1/p(|x|).$$

Definition 10 (Computational zero-knowledge a.Q.– CZKQ)**.** *An interactive proof system (P, V) for a promise problem $A = (A_{\mathsf{yes}}, A_{\mathsf{no}})$ is computational zero-knowledge a.Q. if, for all p.g. quantum V', there exists a p.g. quantum S such that:*

1. *V' and S have the same polynomially-bounded number of input and output qubits for the same input.*
2. *The ensembles $\{\mathbf{\Phi}_x^{(P,V')} : x \in A_{\mathsf{yes}}\}$ and $\{\mathbf{\Psi}_x^S : x \in A_{\mathsf{yes}}\}$ are polynomially quantum indistinguishable.*

This definition is similar to the classical definition (Def. 1) with the difference that the verifier and simulator are quantum and the circuit Q is allowed to give entangled inputs to the verifier.

3.2 Previous Results

The main difficulty with providing zero-knowledge proofs with quantum verifiers and simulators is that we cannot rewind the state of the verifier. Rewinding is an important technique for constructing a simulation because it allows the simulator to produce an indistinguishable transcript of the verifier's view "off-line", restarting from the beginning ("rewinding") whenever the simulation does not have a correct outcome. If the simulator had to be "online" (with no chance of rewinding), it would need to be roughly as powerful or knowledgeable as the prover. Rewinding is a problem with quantum verifiers as follows.

- No cloning theorem: every time the simulator rewinds the verifier, it needs to provide a fresh auxiliary state to the verifier, so it needs to be able to copy the state.
- Irreversibility of measurements: the verifier may perform some irreversible measurements on the state; in fact, even the simulator may perform measurements on the state when trying to decide if a simulation was successful.

Watrous [Wat09] was one of the first to provide zero-knowledge protocols against quantum adversaries. He overcomes the difficulty of rewinding the verifier using the *Quantum Rewinding Lemma*; this construction rewinds quantum states rather than the whole verifier in a black-box fashion.

Lemma 1 (Quantum Rewinding Lemma)**.** *Let Q be a quantum circuit (consisting of only unitary operations) on $n + k$ input qubits. For an arbitrary state $|\psi\rangle$ on n qubits, consider Q applied to $|\psi\rangle |0^k\rangle$ and then measuring the first qubit. Let $p(\psi)$ be the probability that the outcome of the experiment is 0. If $p(\psi) = p \in (0, 1)$ is a constant, for all $\epsilon > 0$, there is a quantum circuit R (performing only admissible operations) with*

$$\mathsf{size}(R) = O\left(\frac{\log(1/\epsilon)\mathsf{size}(Q)}{p(1-p)}\right),$$

and for all inputs $|\psi\rangle$, the output of R satisfies

$$\langle\phi_0(\psi)| R(\psi) |\phi_0(\psi)\rangle \geq 1 - \epsilon.$$

Proof. A sketch of the main proof ideas is in the appendix. □

In the case of a zero-knowledge proof, Q will play the role of the simulator, $p(\psi)$ will be the probability of success of the simulator when the verifier has auxiliary state ψ. For the rewinding lemma to apply, this probability has to be independent of the state ψ, and thus it will not apply to all proof systems.

Using the Quantum Rewinding Lemma, Watrous proves that the Graph Isomorphism protocol is zero-knowledge with respect to polynomial-time quantum verifiers [Wat09]. Recall that, in this problem, the prover wants to prove that two graphs G_0 and G_1 are isomorphic (that is, $G_0 \approx G_1$).

Theorem 2. *Graph Isomorphism (GI) has a zero-knowledge proof against quantum attacks.*

Proof. The main proof steps are described in Appendix A. This proof is insightful because it gives an example of how to overcome the problem of verifier rewinding by using the Quantum Rewinding Lemma. □

Watrous also proves the following result, thus establishing that the famous result of Goldreich et al. [GMW91] for NP still holds in a quantum world.

Theorem 3. *Assuming the existence of quantum computationally concealing schemes, every language in NP has a classical computational zero-knowledge proof against quantum adversaries.*

Proof. A proof sketch is in Appendix A; it mainly consists of applying a weaker version of the Quantum Rewinding Lemma that allows a negligible correlation between the verifier's auxiliary input and the success probability of the simulator (such negligible correlation is introduced by the intractability assumption). □

For ease of exposition, let us define the following property used by Hallgren et al. [HKSZ08].

Definition 11 (Transcript indistinguishability)**.** *A zero-knowledge proof system (P, V) with both P and V classical algorithms, and an associated classical zero-knowledge simulator S, have transcript indistinguishability, if, for positive instances of the problem, the message transcript of (P, V) and the simulated message transcript of S are quantum computationally indistinguishable.*

Another important result in zero-knowledge a.Q. is the following and is due to Hallgren et al. [HKSZ08].

Theorem 4. *Any classical zero-knowledge protocol against the honest verifier also has a classical zero-knowledge protocol against cheating and quantum verifiers, assuming transcript indistinguishability.*

Proof. A proof sketch is in Appendix A. □

For computational zero-knowledge, transcript indistinguishability need only hold for some classical honest verifier. Then, even for cheating and quantum verifiers, this theorem proves that the language will remain in zero-knowledge.

For statistical zero-knowledge, transcript indistinguishability automatically holds and the theorem is unconditional. Intuitively, this is because statistical closeness is preserved under any function (and a physical process is also a function or a superposition of functions).

4 Witness-Indistinguishability Against Quantum Attacks

In this section, we provide an initial study of witness indistinguishability against quantum attacks. We propose definitions for the quantum setting and ask similar questions about the resulting classes of languages as have been asked for zero-knowledge.

4.1 Definitions

Consider a proof system (P, V) for promise problem $A = (A_{\mathsf{yes}}, A_{\mathsf{no}})$, where the yes and no instances are of the form (x, w), and $(x, w) \in A_{\mathsf{yes}}$ is testable in polynomial time. Let $\mathbf{\Phi}_x^w$ be the admissible map induced by the interaction between P and V on some polynomially-bounded number of input and output qubits when P uses witness w to prove $x \in A_{\mathsf{yes}}$.

Definition 12 (Statistical witness indistinguishable a.Q. – SWIQ)**.** *An interactive proof system (P, V) for a promise problem $A = (A_{\mathsf{yes}}, A_{\mathsf{no}})$ is statistical witness-indistinguishable a.Q. if for all p.g. quantum V', for all $x \in A_{\mathsf{yes}}$, $\forall\ w_1,\ w_2 \in w(x)$, there exists a negligible function δ such that $||\mathbf{\Phi}_x^{w_1} - \mathbf{\Phi}_x^{w_2}||_\diamond < \delta(|x|)$.*

Define a *witness ensemble* for A, E^A, to be a function mapping each $x \in A_{\mathsf{yes}}$ to any one witness of x.

Definition 13 (Computational witness indistinguishable a.Q. – CWIQ)**.** *An interactive proof system (P, V) for a promise problem $A = (A_{\mathsf{yes}}, A_{\mathsf{no}})$ is computational witness-indistinguishable a.Q. if, for all p.g. quantum V', for all witness ensembles E_1, E_2, the ensembles $\{\mathbf{\Phi}_x^w : x \in A_{\mathsf{yes}}, w = E_1(x)\}$ and $\{\mathbf{\Phi}_x^w : x \in A_{\mathsf{yes}}, w = E_2(x)\}$ are polynomially quantum indistinguishable.*

These definitions attempt to emulate the classical definitions in a quantum setting. Moreover, they use the notions of statistical closeness and computational indistinguishability introduced by Watrous. Therefore, these also allow the verifier's auxiliary to be entangled with the qubits of an external party running the verifier.

5 Properties

In this section, we prove some properties of witness-indistinguishability a.Q..

Claim 5. *Any language in* $\mathsf{CZKQ}/\mathsf{SZKQ}$ *is in* $\mathsf{CWIQ}/\mathsf{SWIQ}$*.*

Proof. Intuitively, this property should be clear: zero-knowledge does not leak any information, and in particular, it does not leak the witness used.

Let's prove the statement for SZKQ only because the case for CZKQ is analogous. Let $A = (A_{\mathsf{yes}}, A_{\mathsf{no}}) \in \mathsf{SZKQ}$. Therefore, by the zero-knowledge property, for all p.g. V', there exists a p.g. S such that, for all $x \in A_{\mathsf{yes}}$, for all $w \in w(x)$,

$$||\mathbf{\Phi}_x^{(P(w),V')} - \mathbf{\Psi}_x^S||_\diamond < \delta(|x|), \tag{1}$$

for some negligible function δ.

Therefore, for all $x \in A_{\mathsf{yes}}$, for all $w_1, w_2 \in w(x)$,

$$\begin{aligned} ||\mathbf{\Phi}_x^{(P(w_1),V')} - \mathbf{\Phi}_x^{(P(w_2),V')}||_\diamond &= ||\mathbf{\Phi}_x^{(P(w_1),V')} - \mathbf{\Psi}_x^S + \mathbf{\Psi}_x^S - \mathbf{\Phi}_x^{(P(w_2),V')}||_\diamond \\ &\leq ||\mathbf{\Phi}_x^{(P(w_1),V')} - \mathbf{\Psi}_x^S||_\diamond + ||\mathbf{\Psi}_x^S - \mathbf{\Phi}_x^{(P(w_2),V')}||_\diamond \text{ (by triangle ineq.)} \\ &\leq 2\delta(|x|) \text{ (by Eq. (1))}, \end{aligned}$$

thus proving the required property. □

Claim 6. *Any language in* NP *is in* CWIQ*, assuming quantum computationally concealing commitments exist.*

Proof. Theorem 3 states that every language in NP is in CZKQ, assuming quantum computationally concealing commitments. Therefore, using Claim 5, under the same assumption, every language in NP is also in CWIQ. □

Adcock and Cleve [AC02] show how to build a quantum computationally concealing commitment scheme from a quantum one-way permutation. To apply their result to our case, we need the stronger assumption that a quantum one-way permutation that cannot be inverted even with auxiliary state exists.

Claim 7. *Any language in* IP *is in* CZKQ*, and therefore also in* CWIQ*, assuming (1) one-way permutations exist and (2) transcript indistinguishability of some zero-knowledge proof system for the language holds.*

Proof. Assuming the existence of a one-way permutation, a secure probabilistic encryption scheme (as defined in [BGG+90]) can be constructed [Gol88], [BGG+90]. Assuming the existence of such a secure probabilistic encryption scheme, Ben-or et al. [BGG+90] show that every language in IP has a computational zero-knowledge proof. Now, employing Theorem 4 using the second assumption, it follows that every language in IP is in fact in CZKQ. Using Claim 5, the same holds for CWIQ. □

Regarding statistical witness-indistinguishability, we likely cannot prove similar claims to Claims 6 and 7; the reason is that it is not believed for there to exist a statistical zero-knowledge proof (against classical adversaries) for every language in NP [Gol02]. Instead, we can prove the following unconditional result that shows that the existence of quantum adversaries does not decrease the set of languages having statistical (public-coin) zero-knowledge proofs.

Theorem 8. *Public-coin statistical witness-indistinguishability equals public-coin statistical witness-indistinguishability against quantum attacks.*

Unlike the previous claims, this statement no longer follows from the properties of zero-knowledge because statistical witness-indistinguishability is a weaker property than statistical zero-knowledge; even though this is also true for the computational case, we had complete results for zero knowledge in that case (namely, all NP and IP languages are in CZK), which is not the case for statistical zero-knowledge.

To prove this theorem, we first prove two lemmas.

Lemma 9. *Any language in honest verifier public-coin* SWI *is also in* SWI *against any classical verifier strategy (including unbounded and cheating ones).*

Proof. The proof proceeds by modifying the method of Goldreich, Sahai, and Vadhan [GSV98], which provides a (classical) method from transforming any honest verifier statistical zero-knowledge protocol into a cheating verifier statistical zero-knowledge protocol. We use the same transformation they use, but we have to prove a different set of properties about it: if the original construction was honest verifier SWI, the transformation is SWI against any classical strategy.

Completeness of the proof system is unaffected. Goldreich et al. [GSV98] show that soundness is also preserved for the classical verifier, which implies that it will be preserved for the quantum verifier as well; the reason is that the quantum verifier is at least as powerful as the classical verifier. This statement is true even if the prover is quantum because soundness must hold for any prover strategy (even an unbounded strategy).

By the lemma's hypothesis, there is an interactive proof system (P^o, V^o) ("o" stands for "original") that is public-coin HVSWI. Let r be the number of rounds between P^o and V^o, where r is a polynomial in the size of the input. Since (P^o, V^o) are public-coin, each round consists of V^o sending a random number to P^o and P^o sending a message based on this value. Therefore, the transcript of communication between (P^o, V^o) has the form $(\alpha_1, \beta_1, \ldots, \alpha_{r(|x|)}, \beta_{r(|x|)})$, where α_i are the coins sent by the verifier and β_i is the response of the prover in round i. Each β_i only depends on α_j for $j \leq i$, β_j, for $j < i$, any witness input at the prover, and the prover's randomness.

The proof system is SWI meaning that, for all x, for all $w_1, w_2 \in w(x)$,

$$\Delta((\alpha_1, \beta_1, \ldots, \alpha_{r(|x|)}, \beta_{r(|x|)}) \leftarrow (P^o(w_1), V^o), (\alpha_1, \beta_1, \ldots, \alpha_{r(|x|)}, \beta_{r(|x|)}) \leftarrow (P^o(w_2), V^o)) < \delta(|x|), \quad (2)$$

for some negligible function δ.

Goldreich et al. [GSV98] provide a protocol, called *Random Selection*, using which the prover and verifier can choose randomness together to replace the step in which the honest verifier would choose randomness by itself; we do not present the Random Selection protocol here and just use its proved properties in a black-box way. Using the Random Selection protocol, the authors construct a new prover P^t and verifier V^t ("t" stands for "transformed") whose interaction is of the form

$$T = (t_1, a_1, \beta_1, \ldots, t_{r(|x|)}, a_{r(|x|)}, \beta_{r(|x|)}),$$

where t_i is a three-message transcript obtained after running the Random Selection protocol. V_t does not have to behave in any specific way; if V_t does not provide a value inside a desired set, the value will be mapped in that set by the prover. Each a_i and β_i are generated by the prover; a_i only depends on t_i and prover randomness.

In Claim 6.1 of their paper [GSV98], the authors show that there exists a polynomial time algorithm S such that, for any verifier strategy V^* (cheating and potentially unbounded[1]),

$$\Delta\left(T \leftarrow (P^t, V^*), (S(\alpha_i), \alpha_i, \beta_i)_i \text{ for } (\alpha_i, \beta_i)_i \leftarrow (P^o, V^o)\right) < \delta^*(|x|), \quad (3)$$

for some negligible function δ^*. The proof of claim 6.1 relies exclusively on the properties of the Random Selection protocol so it holds in our setting as well.

Eq. (3) holds for any witness given to P^o and P^t. Consider any positive instance x and any two witnesses w_1 and w_2. Let δ_1^* and δ_2^* be the corresponding negligible functions.

Since S is a function of α_i and some randomness only, from Eq. (2), it follows that for all x, for all $w_1, w_2 \in w(x)$,

$$\Delta((S(\alpha_i), \alpha_i, \beta_i)_i \text{ for } (\alpha_i, \beta_i)_i \leftarrow (P^o(w_1), V^o), (S(\alpha_i), \alpha_i, \beta_i)_i \text{ for } (\alpha_i, \beta_i)_i \leftarrow (P^o(w_2), V^o)) < \delta(|x|), \quad (4)$$

for some negligible function δ.

Combining Eq. (3) and Eq. (4) and using the triangle inequality in the definition of statistical distance, Def. 3, we obtain that $\Delta(T \leftarrow (P^{t,w_1}, V^*)), T \leftarrow (P^{t,w_2}, V^*) < \delta_1^*(|x|) + \delta_2^*(|x|) + \delta(|x|)$, which completes our proof. □

[1] In this case, any computation S performs between black-box calls to V^* is polynomial.

Lemma 10. *Any public-coin proof system that is statistical witness-indistinguishable against any classical verifier strategy is statistical witness-indistinguishable against any quantum verifier.*

Proof. Let (P^o, V^o) be a public-coin proof system that is SWI against any classical verifier strategy.

This means that for all verifiers, for any auxiliary state they receive, the transcripts between the prover and the verifier are statistically close for any two witnesses used. Intuitively, since any quantum adversary attempting to distinguish between the two witnesses is a function (or a superposition of functions), the lemma follows because statistical closeness is preserved by any function [Gol01].

Consider any polynomial-time generated quantum verifier V. Let $\mathbf{\Phi}_x^w$ be the admissible map induced by the interaction of V with P^o when V is given w. Consider the following description of a collection of quantum circuits $\{V_x^* : x \in \Sigma^*\}$. V_x^* has hard-coded a state ρ_x, such that the maximum diamond norm of the difference of $\mathbf{\Phi}_x^{w_1}$ and $\mathbf{\Phi}_x^{w_2}$ (denoted MaxD) over all pairs w_1 and w_2 of witnesses for x is achieved when V is given ρ_x. This is well defined because the trace norm is convex and there is always such a ρ_x. V_x^* is also a physically realizable quantum circuit (though there may be no polynomial time machine that can describe it, which does not matter for our purposes). (Note that, for the purposes of the following proof, V_x^* does not need to have exactly such maximum ρ_x hard-coded; it suffices to have any other state ρ'_x whose corresponding trace norm difference in the definition of the diamond norm is within a negligible function of the maximum diamond norm, MaxD. For simplicity of the argument, consider that V_x^* has ρ_x hard-coded.)

On input x and ρ_y auxiliary state, V_x^* ignores ρ_y and instead just runs V on ρ_x. Since V_x^* is a physical process, it also generates an admissible map from the input auxiliary state to the output qubits. Let $\mathbf{\Psi}_x^{w_1}$ be the admissible map generated when the prover uses w_1 on input x. This admissible superoperator maps any input density matrix to the same density matrix, because it ignores the input auxiliary state.

Consider all x, w_1, $w_2 \in w(x)$. For simplicity, in the notation $|0\rangle\langle 0|$ we ignore the number of qubits. We have

$$\begin{array}{rcl}
||\mathbf{\Phi}_x^{w_1} - \mathbf{\Phi}_x^{w_2}||_\diamond & \leq & ||\mathbf{\Psi}_x^{w_1} - \mathbf{\Psi}_x^{w_2}||_\diamond \\
& = & ||\mathbf{\Psi}_x^{w_1} \otimes \mathbb{I}_{L(\mathcal{W})})(|0\rangle\langle 0|) - \mathbf{\Psi}_x^{w_2} \otimes \mathbb{I}_{L(\mathcal{W})})(|0\rangle\langle 0|)||_1 \text{ (by def. of } V_x^*) \\
& \leq & ||\mathbf{\Psi}_x^{w_1}(|0\rangle\langle 0|) - \mathbf{\Psi}_x^{w_2}(|0\rangle\langle 0|)||_1 + 0 \text{ (subadditivity w.r.t. trace norm),} \\
& = & ||\mathbf{\Phi}_x^{w_1}(\rho_x) - \mathbf{\Phi}_x^{w_2}(\rho_x)||_1 \\
& = & ||\sum_t \Pr[t \leftarrow \mathbf{\Phi}_x^{w_1}(\rho_x)]\, |t\rangle\langle t| - \sum_t \Pr[t \leftarrow \mathbf{\Phi}_x^{w_2}(\rho_x)]\, |t\rangle\langle t|\, ||_1 \\
& \leq & \sum_t |\Pr[t \leftarrow \mathbf{\Phi}_x^{w_1}(\rho_x)] - \Pr[t \leftarrow \mathbf{\Phi}_x^{w_2}]|||\, |t\rangle\langle t|\, ||_1 \text{ (triangle ineq.)} \\
& < & 2\delta(|x|),
\end{array}$$

where the last property follows from the witness-indistinguishability property of P^o.

□

Proof of Theorem 8. If a language is in public-coin SWIQ, it must be in public-coin SWI because a quantum verifier can simulate a classical verifier. Therefore, we only need to prove that every language in public-coin SWI is in public-coin SWIQ. A protocol that is public-coin SWI is also public-coin honest-verifier SWI because we can denote one particular cheating verifier strategy as the desired "honest" strategy. Having a public-coin honest-verifier WI protocol, we can now apply Lemma 9 and then Lemma 10 to achieve the desired proof. □

6 Conclusions

We can conclude that, under stronger but reasonable assumptions, the classes of languages having zero-knowledge or witness-indistinguishable interactive proofs against classical adversaries are not decreased when quantum adversaries become practical.

There are at least two interesting open problems:

- What is a good candidate for a one-way permutation resilient to quantum attacks? This question was already posed by [Wat09], but we reiterate it here because our results also rely on the existence of such a function.
- Assuming one-way permutations against quantum adversaries, can we remove the transcript indistinguishability assumption from Theorem 4 (and thus strengthen Claim 7)?

Acknowledgments

The author would like to thank Scott Aaronson for his great lectures and guidance, and to John Watrous and Iordanis Kerenidis for kindly answering her emails and providing useful advice.

References

[AC02] Mark Adcock and Richard Cleve, *A quantum goldreich-levin theorem with cryptographic applications*, STACS 2002, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2002.

[Bar01] Boaz Barak, *How to go beyond the black-box simulation barrier*, In 42nd FOCS, 2001, pp. 106–115.

[BGG+90] Michael Ben-Or, Oded Goldreich, Shafi Goldwasser, Johan Håstad, Joe Kilian, Silvio Micali, and Phillip Rogaway, *Everything provable is provable in zero-knowledge*, Proceedings of the 8th Annual International Cryptology Conference on Advances in Cryptology, Springer-Verlag, 1990, pp. 37–56.

[BOGG+90] Michael Ben-Or, Oded Goldreich, Shafi Goldwasser, Johan Håstad, Joe Kilian, Silvio Micali, and Phillip Rogaway, *Everything provable is provable in zero-knowledge*, Proceedings of the 8th Annual International Cryptology Conference on Advances in Cryptology, Springer-Verlag, 1990, pp. 37–56.

[CK08] André Chailloux and Iordanis Kerenidis, *Increasing the power of the verifier in quantum zero knowledge*, FSTTCS, 2008, pp. 95–106.

[DGOW95] Ivan Damgård, Oded Goldreich, Tatsuaki Okamoto, and Avi Wigderson, *Honest verifier vs dishonest verifier in public coin zero-knowledge proofs*, Proceedings of the 15th Annual International Cryptology Conference on Advances in Cryptology, CRYPTO '95, Springer-Verlag, 1995.

[FS90] Uriel Feige and Adi Shamir, *Witness indistinguishable and witness hiding protocols*, in 22nd STOC, ACM Press, 1990, pp. 416–426.

[GMR89] Shafi Goldwasser, Silvio Micali, and Charles Rackoff, *The knowledge complexity of interactive proof systems*, SIAM J. of Computing, vol. 18, 1989, pp. 186–208.

[GMW91] Oded Goldreich, Silvio Micali, and Avi Wigderson, *Proofs that yield nothing but their validity or all languages in np have zero-knowledge proof systems*, Journal of the ACM **38** (1991), 691–729.

[Gol88] O. Goldreich, *Randomness, interactive proofs, and zero-knowledge-a survey*, A half-century survey on The Universal Turing Machine, Oxford University Press, Inc., 1988, pp. 377–405.

[Gol01] Oded Goldreich, *Foundations of Cryptography*, vol. Basic Tools, Cambridge University Press, 2001.

[Gol02] Oded Goldreich, *Zero-knowledge twenty years after its invention*, Tech. report, Electronic Colloquium on Computational Complexity (http://www.eccc.uni-trier.de/eccc/), Report No, 2002.

[GSV98] Oded Goldreich, Amit Sahai, and Salil Vadhan, *Honest-verifier statistical zero-knowledge equals general statistical zero-knowledge*, In Proceedings of the 30th Annual ACM Symposium on Theory of Computing, 1998, pp. 399–408.

[HKSZ08] Sean Hallgren, Alexandra Kolla, Pranab Sen, and Shengyu Zhang, *Making classical honest verifier zero knowledge protocols secure against quantum attacks*, Proceedings of the 35th international colloquium on Automata, Languages and Programming, Part II, ICALP, Springer-Verlag, 2008, pp. 592–603.

[Kit97] A Yu Kitaev, *Quantum computations: algorithms and error correction*, Russian Mathematical Surveys **52** (1997).

[Kob08] Hirotada Kobayashi, *General properties of quantum zero-knowledge proofs*, Proceedings of the 5th conference on Theory of cryptography, TCC'08, Springer-Verlag, 2008, pp. 107–124.

[RK99] Ransom Richardson and Joe Kilian, *On the concurrent composition of zero-knowledge proofs*, Eurocrypt, Springer LNCS 1592, Springer-Verlag, 1999, pp. 415–431.

[Wat02] John Watrous, *Limits on the power of quantum statistical zero-knowledge*, Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002.

[Wat09] John Watrous, *Zero-knowledge against quantum attacks*, SIAM Journal on Computing, 2009, Earlier version appeared in STOC'06.

A Proofs for ZKQ

Proof sketch of Theorem 1. Let W represent the registers corresponding to $|\psi\rangle$ and X to the ancilla k qubits. Let B be the register corresponding to the first output qubit of Q and Y representing the rest $n_k - 1$ qubits.

The algorithm for R is:

Algorithm 1 (R)**.**

```
1. Apply $Q$ to $(W, X)$ obtaining $(B, Y)$
2. Repeat
   (a) Measure $B$ in the standard basis.
   (b) If outcome of measurement is 1
       i. Apply $Q^*$ to $(B, Y)$, obtaining $(W, X)$.
       ii. Perform a phase flip if any qubits of $X$ are 1.
       iii. Apply $Q$ to $(W, X)$ obtaining $(B, Y)$.
3. until the measurement outcome is 0 or we iterated $\lceil \log(1/\epsilon)/(4p(1-p)) \rceil$ steps.
```

Watrous shows, using some simple linear algebra, that the measurement of the outcome being 0 in the first t measurements is $1 - (1-p)(1-2p)^{2t}$, which suffices for our proof. This result relies heavily on the fact that $p(\psi)$ is a constant function. □

Proof of Theorem 2. Watrous proves that the classical GI protocol has a quantum simulator.

Let $\mathbb{P}_n$ be the space of permutations on $1, \ldots, n$. Recall the classical protocol:

Algorithm 2 (*GI*)**.** The input is a pair G_0, G_1 of simple, undirected graphs each having vertex set $\{1, 2, \ldots, n\}$.

```
$P$ step 1: Let $\sigma$ be a permutation satisfying $\sigma(G_0) = G_1$ if $G_0$ and $G_1$ are isomorphic, else the identity permutation. Choose $\pi \in \mathbb{P}_n$ uniformly at random and send $H = \pi(G_0)$ to $V$.
$V$ step 1: Choose a bit $a$ at random and send it to $P$.
$P$ step 2: Send $\tau := \pi\sigma^a$ to $V$.
$V$ step 2: Accept if $\tau(G_a) = H$, else reject.
```

This protocol has perfect completeness so it will also have perfect completeness when the verifier is quantum. Moreover, soundness is a property of the verifier, so a stronger quantum verifier can achieve the same error or less than a classical verifier because a quantum algorithm can simulate a classical algorithm in polynomial time.

The simulation procedure takes in two kinds of registers: registers corresponding to the view of the verifier (including register A indicating the verifier's random choice) and registers corresponding to the workspace of the simulator (including register B indicating the simulator's guess for the verifier's bit a).

Watrous describes a circuit Q that brings into superposition all possible transcripts of an interaction between the simulator and verifier. Then, a measurement is performed whose outcome is 0 if $A = B$ and 1 otherwise. If register B is measured, an outcome of zero corresponds to the remaining state collapsing to a superposition of all transcripts of a correct simulation with chance of a half after tracing out the simulator's qubits. This is the desired resulting state. The chance of this state occuring is $1/2$, independent of the auxiliary state at the verifier or of the verifier's actions, because (as in the classical case), there can be no correlations between the guess b of the simulator and the bit a of the verifier.

Therefore, we can apply the Quantum Rewinding Lemma (Lemma 1) and obtain, after exactly two iterations (since $p = 1/2$, the probability of success becomes one after two iterations) the same admissible map as the one of the interaction. □

Proof sketch of Theorem 3. Watrous proves this statement by proving a computational zero-knowledge proof against quantum adversaries for Graph 3-Coloring.

He also defines quantum computationally concealing schemes to signify, intuitively, commitment schemes whose committed value is hidden even to a quantum adversary.

The proof idea is similar with some technical issues to address. The simulator similarly brings into superposition all the possible transcripts between the prover and the verifier; some of these transcripts represent correct simulations and others represent unsuccessful simulations. At each iteration, a measurement is performed, whose outcome of zero collapses the state to the correct simulations, and outcome of one continue the simulation.

In order to apply the rewinding lemma, it must be the case that $p(\psi)$ is a constant independent of ψ and the verifier's strategy. This happens if the simulator's guess for the random choice of the verifier is uncorrelated with the choice of the verifier. This could be guaranteed if the commitments were unconditionally hiding which is not the case. Since they are quantum computationally hiding, there can be some negligible correlation. Therefore, Watrous presents a weaker version of the Quantum Rewinding Lemma allowing for such a small correlation and gives a proof for it.

□

Proof Sketch of Theorem 4. This proofs attempts to put together two sets of results:

1. Watrous' quantum rewinding procedure: to obtain zero-knowledge against quantum attacks.

2. A classical cryptography result: [DGOW95] to make classical honest verifier protocols resilient against classical cheating verifiers.

The technical difficulty is that the second protocol only applies to constant-round honest verifier protocols and the first protocol requires independence of simulation success from auxiliary state of the verifier which does not hold in cheating-verifier non-constant round constructions in the literature.

The main idea of the authors is to transform the initial classical zero-knowledge protocol into a *stage-by-stage* such protocol. A stage-by-stage protocol is a protocol in which the simulator's transcript consists of a sequence of stages, each with a constant number of interactions, and during which the simulator uses independent randomness. This enables the authors to apply both results mentioned above within each stage and prove correctness overall.

To obtain a stage-by-stage simulator, the authors adapt the work of Ben-Or et al. [BOGG$^+$90], which shows that every public-coin interactive protocol can be transformed in zero-knowledge and in fact give a stage-by-stage simulator construction. The construction is based on bit commitments which introduce computational assumption to be avoided for statistical zero-knowledge, but the authors show how to construct this using instance-dependent bit commitments (these commitments are hiding on the bit to be committed for positive instances of the problem and binding on the bit for negative instances of the problem.)

□