Quantum Proofs of Knowledge

Dominique Unruh
University of Tartu

February 11, 2015

Abstract

We motivate, define and construct quantum proofs of knowledge, that is, proofs of knowledge secure against quantum adversaries. Our constructions are based on a new quantum rewinding technique that allows us to extract witnesses in many classical proofs of knowledge. We give criteria under which a classical proof of knowledge is a quantum proof of knowledge. Combining our results with Watrous' results on quantum zero-knowledge, we show that there are zero-knowledge quantum proofs of knowledge for all languages in NP (assuming quantum 1-1 one-way functions).

Contents

1 Introduction

Cryptographic protocols, with few exceptions, are based on the assumption that certain problems are computationally hard. Typical examples include specific number-theoretic problems such as the difficulty of finding discrete logarithms, and general problems such as inverting one-way functions. It is well-known, however, that many such problems

would become easy in the advent of quantum computers. For example, Shor's algorithm [Sho94] efficiently solves the discrete logarithm problem and allows to factor large integers. While quantum computers do not exist today, it is not unreasonable to expect quantum computers to be available in the future. To meet this threat, we need cryptographic protocols that are secure even in the presence of an adversary with a quantum computer. We stress that this does not necessarily imply that the protocol itself should make use of quantum technology; instead, it is preferable that the protocol itself can be easily implemented on today's readily-available classical computers.

Finding such quantum-secure protocols, however, is not trivial. Even when we have found suitable complexity-theoretic assumptions such as the hardness of certain lattice problems, a classical protocol based on these assumptions may fail to be secure against quantum computers. The reason for this is that many cryptographic proofs use a technique called rewinding. This technique requires that it is possible, when simulating some machine, to make snapshots of the state of that machine and then later to go back to that snapshot. As first observed by van de Graaf [vdG98], classical rewinding-based proofs do not carry over to the quantum case. Two features unique to the quantum setting prohibit (naive) rewinding: The no-cloning theorem [WZ82] states that quantum-information cannot be copied, so we cannot make snapshots. Furthermore, measurements destroy information, so interacting with a simulated machine may destroy information that would be needed later.

This leads to the following observation: Even if a classical protocol is proven secure based on the hardness of some problem, and that problem is hard even for quantum computers, we have no guarantee that the protocol is secure against quantum computers. The reduction of the protocol's security to the problem's hardness may be based on inherently classical features such as the possibility of rewinding.

An example of a protocol construction that suffers from this difficulty are zero-knowledge proofs. Zero-knowledge proofs are interactive proofs with the special property that the verifier does not learn anything except the validity of the proven statement. Zero-knowledge proofs are inherently based on rewinding (at least as long as we do not assume additional trusted setup such as so-called common-reference strings). Yet, zero-knowledge proofs are one of the most powerful tools available to the cryptographer; a multitude of protocol constructions use zero-knowledge proofs. These protocol constructions cannot be proven secure without using rewinding. To resolve this issue, Watrous [Wat09] introduced a quantum rewinding technique. This technique allows to prove the quantum security of many common zero-knowledge proofs. One should note, however, that Watrous' technique is restricted to a specific type of rewinding: If we use Watrous' technique, whenever some machine rewinds another machine to an earlier point, the rewinding machine forgets everything it learned after that point (we call this oblivious rewinding). That is, we can use Watrous' technique to backtrack when the rewinding machine made a mistake that should be corrected, but it cannot be used to collect and combine information from different branches of an execution.

Constructing quantum zero-knowledge proofs solves, however, only half of the problem. In many, if not most, applications of zero-knowledge proofs one needs zero-knowledge

proofs of knowledge. A proof of knowledge [GMR85, BG93] is a proof system which does not only show the truth of a certain statement, but also that the prover knows a witness for that statement. This is made clearer by an example: Assume that Alice wishes to convince Bob that she (the prover) is in possession of a signature issued by some certification authority. For privacy reasons, Alice does not wish to reveal the signature itself. If Alice uses a zero-knowledge *proof*, she can only show the statement "there exists a signature with respect to the CA's public key". This does not, however, achieve anything: A signature always exists in a mathematical sense, even if it has never been computed. What Alice wishes to say is: "I *know* a signature with respect to the CA's public key." To prove such a statement, Alice needs a zero-knowledge *proof of knowledge*; a proof of knowledge would convince Bob that Alice indeed knows a witness, i.e., a signature. Very roughly, the definition of a proof of knowledge is the following: Whenever the prover can convince the verifier, one can extract the witness from the prover given oracle access to the prover. Here oracle access means that one can interact with the prover and *rewind* him. Thus, we have the same problem as in the case of quantum zero-knowledge proofs: To get proofs of knowledge that are secure against quantum adversaries, we need to use quantum rewinding. Unfortunately, Watrous' *oblivious* rewinding does not work here; proofs of knowledge use rewinding to produce two (or more) different protocol traces and compute the witness by combining the information from both traces. Thus, we are back to where we started: to make classical cryptographic protocols work in a quantum setting, we need (in many cases) quantum zero-knowledge *proofs of knowledge*, but we only have constructions for quantum zero-knowledge *proofs*.

Our contribution. We define and construct quantum proofs of knowledge. Our protocols are classical (i.e., honest parties do not use quantum computation or communication) but secure against quantum adversaries. Our constructions are based on a new quantum rewinding technique (different from Watrous' technique) that allows us to extract witnesses in many classical proofs of knowledge. We give criteria under which a classical proof of knowledge is a quantum proof of knowledge ("special soundness" and "strict soundness"). Combining our results with Watrous' results on zero-knowledge, we can show that there are zero-knowledge quantum proofs of knowledge for all languages in NP (assuming quantum 1-1 one-way functions). (We leave it as an open question whether unconditionally secure protocols exist for more restricted languages related, e.g., to lattice-problems.)

We believe that the use of our rewinding technique is not limited to QPoKs. It (or a variation of it) could find application whenever we need to show that the ability to provide any of several values implies the ability to provide all of those values simultaneously.

As a side contribution, we also generalized Watrous' analysis [Wat09] of the zero-knowledge property of Σ-procols. While Watrous applied his technique to selected examples, we have spelled out the exact requirements for a Σ-protocol to be computationally/statistically quantum zero-knowledge.

Related work. Most related work has already been discussed in the introductory paragraphs. Crépeau, Salvail, Simard, and Tapp [CSST11] independently developed a rewinding technique similar to ours for analyzing a specific two-prover commitment

scheme. Their result can be used to improve our bounds for protocols where the verifier sends only one bit (i.e., Σ-protocols with challenge space of size 2), see Section 3.1.

Follow-up work. In subsequent work, Lunemann and Nielsen [LN11] and Hallgren, Smith, and Song [HSS11] developed zero-knowledge QPoKs with the additional advantage of allowing to simultaneously simulate an interaction with the malicious prover and extract the witness; this property is necessary in some multi-party computations. (In contrast, in our setting the initial state of the prover could be lost after extracting.) We stress, however, that this powerful feature comes at a cost: They need strong assumptions, namely quantum mixed commitments (while we only need quantum 1-1 one-way functions). Both their zero-knowledge property and their extractability hold only against quantum-polynomial-time adversaries. In contrast, we get unconditional extractability and computational zero-knowledge. Finally, we note that the protocols from [LN11, HSS11] are much more involved than their classical counterparts while we only slightly modify existing classical protocols. Thus, [LN11, HSS11] give valuable alternatives to our protocols but do not supersede them. A transformation from Σ-protocols (even without strict soundness) to *non-interactive* zero-knowledge arguments of knowledge was given by Unruh [Unr14]. However, their construction is shown secure only in the random oracle model. Ambainis, Rosmanis, and Unruh [ARU14] show that the condition of strict soundness introduced in this paper is probably necessary: relative to some oracle, Σ-protocols with only special soundness are not always proofs of knowledge. Unruh [Unr15] extends our techniques to construct quantum arguments of knowledge from computationally binding commitments. Unruh [Unr13] used our protocols for constructing everlastingly secure quantum UC protocols.

Organization. In Section 1.1, we give an overview over the techniques underlying our results. In Section 2 we present and discuss the definition of quantum proofs of knowledge (QPoKs). In Section 3, we give criteria under which a proof system is a QPoK. In Section 4, we review and generalize Watrous' rewinding technique for quantum zero-knowledge [Wat09]. In Section 5, we show that zero-knowledge QPoKs exist for all languages in NP.

1.1 Our techniques

Defining proofs of knowledge. In the classical setting, proofs of knowledge are defined as follows:[1] A proof system consisting of a prover P and a verifier V is a proof of knowledge (PoK) with knowledge error κ if there is a polynomial-time machine K (the extractor) such that the following holds: For any prover P^*, if P^* convinces V with probability $\Pr_{\mathsf{V}} \geq \kappa$, then $\mathsf{K}^{\mathsf{P}^*}$ (the extractor K with rewinding black-box access to P^*) outputs a witness with probability $\Pr_{\mathsf{K}} \geq \frac{1}{p}(\Pr_{\mathsf{V}} - \kappa)^d$ for some polynomial p and some constant $d > 0$. In order to transfer this definition to the quantum setting, we need to specify

[1] This is one of different possible definitions, loosely following [HM98]. It permits us to avoid the use of expected polynomial-time. We discuss alternatives in Section 2.2 "On the success probability of the extractor".

what it means that K has quantum rewinding black-box access to P^*. We choose the following definition: Let U denote the unitary transformation describing one activation of P^* (if P^* is not unitary, we use a purification of P^*). K may invoke U (this corresponds to running P^*), K may invoke the inverse $U^\dagger$ of U (this corresponds to rewinding P^* by one activation), and K may read/write a shared register N used for exchanging messages with P^*. But K cannot make snapshots of the state of P^*. Allowing K to invoke $U^\dagger$ is justified by the fact that all quantum circuits are reversible; given a circuit for U, we can efficiently apply $U^\dagger$. Note that previous black-box constructions such as Watrous' rewinding technique and Grover's algorithm [Gro96] also make use of this fact. We can now define quantum proofs of knowledge: (P, V) is a quantum proof of knowledge (QPoK) with knowledge error κ iff there is a quantum-polynomial-time quantum algorithm K such that for all malicious provers P^*, $\mathsf{K}^{\mathsf{P}^*}$ (the extractor K with quantum rewinding black-box access to P^*) outputs a witness with probability $\Pr_{\mathsf{K}} \geq \frac{1}{p}(\Pr_{\mathsf{V}} - \kappa)^d$ for some polynomial p and constant $d > 0$. Details are given in Section 2.1.

We illustrate that QPoKs according to this definition are indeed useful for analyzing cryptographic protocols. Assume the following toy protocol: In phase 1, a certification authority (CA) signs the pair $(\texttt{Alice}, a)$ where a is Alice's age. In phase 2, Alice uses a zero-knowledge QPoK with negligible knowledge error κ to prove to Bob that she possesses a signature σ on $(\texttt{Alice}, a')$ for some $a' \geq 21$. That is, a witness in this QPoK would consist of an integer $a' \geq 21$ and a signature σ on $(\texttt{Alice}, a')$ with respect to the CA's public key. We can now show that, if Alice is underage, i.e., if $a < 21$, Bob accepts the QPoK only with negligible probability: Assume that Bob accepts with non-negligible probability ν. Then, by the definition of QPoKs, $\mathsf{K}^{\texttt{Alice}}$ will, with probability $\frac{1}{p}(\nu - \kappa)^d$, output an integer $a' \geq 21$ and a signature σ on $(\texttt{Alice}, a')$ with respect to the CA's public key ($\mathsf{K}^{\texttt{Alice}}$ is given the information learned in phase 1 as auxiliary input). Notice that $\frac{1}{p}(\nu - \kappa)^d$ is non-negligible. However, the CA only signed $(\texttt{Alice}, a)$ with $a < 21$. This implies that $\mathsf{K}^{\texttt{Alice}}$ can produce with non-negligible probability a valid signature σ of a message that has never been signed by the CA. This contradicts the security of the signature scheme (assuming, e.g., existential unforgeability [GMR88]). This shows the security of our toy protocol.

This toy protocol gives a first indication that our definition is usable in practical settings. For an example of a more complex setting where our definition is used successfully, see the commitment protocol from [Unr13] which uses quantum arguments of knowledge according as per our definition.

Relation to classical proofs of knowledge. Notice that a quantum proof of knowledge according to our definition is not necessarily a classical PoK because the quantum extractor might have more computational power. (E.g., in a proof system where the witness is a factorization, a quantum extractor could just compute this witness himself.) We stress that this "paradox" is not particular to our definition, it occurs with all simulation-based definitions (e.g., zero-knowledge [Wat09], universal composability [Unr10]). If needed, one can avoid this "paradox" by requiring the extractor/simulator to be classical if the malicious prover/verifier is. (This would actually be equivalent to requiring that the scheme is both a classical ZK PoK and a quantum one.)

Amplification. Our toy example shows that QPoKs with negligible knowledge error can be used to show the security of protocols. But what about QPoKs with non-negligible knowledge error? In the classical case, we know that the knowledge error of a PoK can be made exponentially small by sequential repetition. Fortunately, this result carries over to the quantum case; its proof follows the same lines.

Elementary constructions. In order to understand our constructions of QPoKs, let us first revisit a common method for constructing classical PoKs. Assume a protocol that consists of three messages: the commitment (sent by the prover), the challenge (picked from a set C and sent by the verifier), and the response (sent by prover). Such a protocol is called a Σ-protocol. Assume that there is an efficient algorithm K_0 that computes a witness given two conversations with the same commitment but different challenges; this property is called special soundness. Then we can construct the following (classical) extractor K: $\mathsf{K}^{\mathsf{P}^*}$ runs P^* using a random challenge ch. Then $\mathsf{K}^{\mathsf{P}^*}$ rewinds P^* to the point after it produced the commitment, and then $\mathsf{K}^{\mathsf{P}^*}$ runs P^* with a random challenge ch'. If both executions lead to an accepting conversation, and $ch \neq ch'$, K_0 can compute a witness. The probability of getting two accepting conversations can be shown to be $\Pr_{\mathsf{V}}^2$, where $\Pr_{\mathsf{V}}$ is the probability of the verifier accepting P^*'s proof. From this, a simple calculation shows that the knowledge error of the protocol is $1/\#C$.

If we translate this approach to the quantum setting, we end up with the following extractor: K runs one step of P^*, measures the commitment com, provides a random challenge ch, runs the second step of P^*, measures the response $resp$, runs the inverse of the second step of P^*, provides a random challenge ch', runs the second step of P^*, and measures the response $resp'$. If $ch \neq ch'$, and both $(com, ch, resp)$ and $(com, ch', resp')$ are accepting conversations, then we get a witness using K_0. We call this extractor the canonical extractor. The problem is to bound the probability of getting two accepting conversations. In the classical setting, one uses that the two conversations are essentially independent (given a fixed commitment), and each of them is, from the point of view of P^*, the same as an interaction with the honest verifier V. In the quantum setting, this is not the case. Measuring $resp$ disturbs the state of P^*; hence we cannot make any statement about the probability that the second conversation is accepting.

How can we solve this problem? Note that we cannot use Watrous' oblivious rewinding since we need to remember both responses $resp$ and $resp'$ from two different execution paths of P^*. Instead, we observe that, the more information we measure in the first conversation (i.e., the longer $resp$ is), the more we disturb the state of P^* used in the second conversation. Conversely, if would measure only one bit, the disturbance of P^*'s state would be small enough to still get a sufficiently high success probability. But if $resp$ would contain only one bit, it would clearly be too short to be of any use for K_0. Yet, it turns out that this conflict can be resolved: In order not to disturb P^*'s state, we only need that the response $resp$ information-theoretically contains little information. For K_0, however, even an information-theoretically determined $resp$ is still useful; it might, for example, reveal some value that P^* was already committed to. To make use of this observation, we introduce an additional condition on our proof systems, strict soundness. A proof system has strict soundness if for any commitment and challenge, there is at

most one response that makes the conversation accepting. Given a proof system with special and strict soundness, we can show that measuring *resp* does not disturb P^*'s state too much; the canonical extractor is successful with probability approximately $\Pr_{\mathsf{V}}^3$. A precise calculation shows that a proof system with special and strict soundness has knowledge error $1/\sqrt{\#C}$.

QPoKs for all languages in NP. Blum [Blu86] presents a classical zero-knowledge PoK for showing the knowledge of a Hamiltonian cycle. Using a suitable commitment scheme (it should have the property that the opening information is uniquely determined by the commitment), the proof system is easily seen to have special and strict soundness, thus it is a QPoK. By sequential repetition, we get a QPoK for Hamiltonian cycles. Using Watrous' rewinding technique, we get that the QPoK is also zero-knowledge. Using the fact that the Hamiltonian cycle problem is NP-complete, we get zero-knowledge QPoKs for all languages in NP (assuming quantum 1-1 one-way functions).

1.2 Preliminaries

General. A non-negative function μ is called negligible if for all $c > 0$ and all sufficiently large k, $\mu(k) < k^{-c}$. A non-negative function μ is called non-negligible if it is not negligible. $\mathrm{E}[X]$ denotes the expected value of X. $\#C$ is the cardinality of the set C. $\eta > 0$ always refers to the security parameter, an integer that controls the level of security of our protocols. The set $\{0,1\}^*$ is the set of all bitstring, and $\{0,1\}^{\leq \ell}$ the set of bitstrings of length at most ℓ.

Quantum systems. We can only give a terse overview over the formalism used in quantum computing. For a thorough introduction, we recommend the textbook by Nielsen and Chuang [NC10, Chap. 1–2]. A (pure) state in a quantum system is described by a unit vector $|\Phi\rangle$ in some Hilbert space $\mathcal{H}$. We always assume a designated orthonormal basis for each Hilbert space, called the computational basis. The tensor product of several states (describing a joint system) is written $|\Phi\rangle \otimes |\Psi\rangle$. We write $\langle\Psi|$ for the linear transformation mapping $|\Phi\rangle$ to the scalar product $\langle\Psi|\Phi\rangle$. The norm $\||\Phi\rangle\|$ is defined as $\sqrt{\langle\Phi|\Phi\rangle}$. A unit vector is a vector with $\||\Phi\rangle\| = 1$. The Hermitean transpose of a linear operator A is written $A^\dagger$. A is called positive if $A = A^\dagger$ and $\langle\Phi|A|\Phi\rangle \geq 0$ for all $|\Phi\rangle$. The operator norm of A is $|||A||| := \sup_{|\Phi\rangle} \|A|\Phi\rangle\|$ with $|\Phi\rangle$ ranging over unit vectors; we call A bounded if $|||A|||$ exists.

2 Quantum Proofs of Knowledge

2.1 Definitions

Interactive machines. Intuitively, an *interactive quantum machine* M (machine, for short) is a machine that maintains two quantum registers: a register S for the internal state of M, and a register N for sending and receiving messages (the network register). Upon each activation, M expects some message in N, and the state of the preceding

invocation in S. After the activation, S contains the new state of M, and N contains the message that M sends. A machine M gets as input: a security parameter η, a classical input x, and a quantum input $|\Phi\rangle$. For simplicity, we assume that the number of messages a machine sends and receives is determined by the security parameter and the classical input. The quantum input is initially stored in S. More formally, a quantum machine is described by a family of quantum circuits $(\mathsf{M}_{\eta x})_{\eta\in\mathbb{N},x\in\{0,1\}^*}$ and a family of integers $(r^{\mathsf{M}}_{\eta x})_{\eta\in\mathbb{N},x\in\{0,1\}^*}$. $\mathsf{M}_{\eta x}$ determines the unitary operation that is performed on the quantum registers S and N, and $r^{\mathsf{M}}_{\eta x}$ determines the number of messages. Note that all our machines perform only unitary operations. This does not, however, constitute a restriction since a machine with measurements can be transformed into a unitary machine by a standard purification argument. We call a machine M *quantum-polynomial-time* if the circuit $\mathsf{M}_{\eta x}$ has polynomial size in $\eta+|x|$, $r^{\mathsf{M}}_{\eta,x}$ is polynomially-bounded in $\eta+|x|$, and $r^{\mathsf{M}}_{\eta,x}$ and the circuit's description can be computed in deterministic polynomial time in $\eta+|x|$ given η, x.

Execution of interactive machines. Given a pair of machines M and M', the security parameter η, a pair of quantum states $|\Phi\rangle$ and $|\Phi'\rangle$, and a pair of classical bitstrings $x, x' \in \{0,1\}^*$, we define the execution $\langle \mathsf{M}(x,|\Phi\rangle), \mathsf{M}'(x',|\Phi'\rangle)\rangle$ by the following process:[2] Initialize quantum registers S, S', N with $|\Phi\rangle, |\Phi'\rangle, |0\rangle$, respectively. Alternatingly, apply the circuit $\mathsf{M}_{\eta x}$ to S, N and the circuit $\mathsf{M}'_{\eta x'}$ to S', N. Stop applying $\mathsf{M}_{\eta x}$ after $r^{\mathsf{M}}_{\eta x}$ applications and stop applying $\mathsf{M}'_{\eta x'}$ after $r^{\mathsf{M}'}_{\eta x'}$ applications.[3] Then measure S' in the computational basis. The random variable $\langle \mathsf{M}(x,|\Phi\rangle), \mathsf{M}'(x',|\Phi'\rangle)\rangle$ denotes the result of that measurement. In other words, $\langle \mathsf{M}(x,|\Phi\rangle), \mathsf{M}'(x',|\Phi'\rangle)\rangle$ is the classical output of M' in an interaction where M is activated first. Often, we will omit the quantum input $|\Phi\rangle$ or $|\Phi'\rangle$. In this case, we assume the input $|0\rangle$.

Oracle algorithms with rewinding. A *quantum oracle algorithm* A is an algorithm that has access to a quantum interactive machine that is given as an oracle. Besides the security parameter η and its own (classical) input x, the algorithm gets access to an interactive quantum machine M running on classical input x' and quantum input $|\Phi\rangle$. We allow A to provide messages to and read messages from $\mathsf{M}_{\eta x'}$ and to execute the (unitary) quantum circuit $\mathsf{M}_{\eta x'}$ that describes M. Furthermore, A may execute the inverse of $\mathsf{M}_{\eta x'}$, this corresponds to the classical notion of rewinding the machine M. We also allow that A is in a superposition between executing $\mathsf{M}_{\eta x'}$ and not executing it.[4] We will not, however, allow A to directly access the state of M or to its quantum input. (I.e., A has no access to the internal state and the quantum input of the prover. Any access to this information is done by communicating with M.) Formally, a quantum oracle algorithm A is described by a family of circuits $(\mathsf{A}_{\eta x})_{\eta\in\mathbb{N},x\in\{0,1\}^*}$ operating on three quantum registers S_{A}, N and S_{M}. (S_{A} and S_{M} contain the states of A and M, respectively. N is used for

[2] Note that we keep the security parameter η implicit in this notation for brevity. The reader should keep in mind that the behavior of M and M' depends on η.

[3] If r^{M}_{x} and $r^{\mathsf{M}'}_{x'}$ do not match, it may happen that the circuit of one machine is executed several times in a row after the other machine already stopped.

[4] The ability of A to execute $\mathsf{M}_{\eta x}$ in superposition is not, however, necessary for the results presented in this work.

communication between A and M.) The circuit $\mathsf{A}_{\eta x}$ may contain normal gates (from some fixed universal set of gates) operating on S_{A} and N (but not S_{M}), as well as two special gates $\square$ and $\square^\dagger$. (These represent an application of the oracle given to A.) Both operate on one qubit of S_{A} (the control qubit) and on the whole of (N, S_{M}). We define an execution $\mathsf{A}^{\mathsf{M}(x',|\Phi\rangle)}(x)$ as follows:[5] Initialize $S_{\mathsf{A}}, N, S_{\mathsf{M}}$ with $|0\rangle, |0\rangle, |\Phi\rangle$. Execute the circuit $\mathsf{A}_{\eta x}$. When the gate $\square$ is to be applied on C, N, S_{M} where C is a qubit in S_{A}, apply the unitary transformation U defined by $U(|0\rangle \otimes |\psi\rangle \otimes |\varphi\rangle) := |0\rangle \otimes |\psi\rangle \otimes |\varphi\rangle$ and $U(|1\rangle \otimes |\psi\rangle \otimes |\varphi\rangle) := |1\rangle \otimes \mathsf{M}_{\eta x'}(|\psi\rangle \otimes |\varphi\rangle)$ where $\mathsf{M}_{\eta x'}$ is the unitary transformation describing one activation of M. (Intuitively, $\mathsf{M}_{\eta x'}$ is applied if C contains $|1\rangle$.) The gate $\square^\dagger$ is treated analogously, except that we use $\mathsf{M}^\dagger_{\eta x'}$ instead of $\mathsf{M}_{\eta x'}$. Finally, we measure S_{A} in the computational basis. The random variable $\mathsf{A}^{\mathsf{M}(x',|\Phi\rangle)}(x)$ describes the outcome of that measurement. We call an algorithm A *quantum-polynomial-time* if the circuit $\mathsf{A}_{\eta x}$ has polynomial-size in $|x| + \eta$ and its description can be computed in deterministic time in $|x| + \eta$ given η, x.

Proof systems. In the following, we consider relations parametrized by the security parameter η. That is, a relation in our sense R is actually a family $(R_\eta)_{\eta\in\mathbb{N}}$ of relations $R_\eta \subseteq \{0,1\}^* \times \{0,1\}^*$. A *quantum proof system* for a relation R is a pair of two machines (P, V). We call P the prover and V the verifier. The prover expects a classical input (x, w) with $(x, w) \in R$, the verifier expects only the input x. We call (P, V) *complete* iff there is a negligible function μ such that for all $\eta \in \mathbb{N}$ and all $(x, w) \in R$, we have that $\Pr[\langle \mathsf{P}(x, w), \mathsf{V}(x)\rangle = 1] \geq 1 - \mu(\eta)$. (Remember that, if we do not explicitly specify a quantum input, we assume the quantum input $|0\rangle$.) Although we allow P and V to be quantum machines, and in particular to send and receive quantum messages, we will not need this property in the following; all protocols constructed in this paper will consist of classical machines. We call a (P, V) *sound* with soundness error s iff for all malicious provers P^*, all $\eta \in \mathbb{N}$, all auxiliary inputs $|\Phi\rangle$, and all x with $\nexists w : (x, w) \in R$, we have $\Pr[\langle \mathsf{P}^*(x, |\Phi\rangle), \mathsf{V}(x)\rangle = 1] \leq s(\eta)$.

Quantum Proofs of Knowledge. We can now define quantum proofs of knowledge (QPoKs). Roughly, a quantum proof system (P, V) is a QPoK iff there is a quantum oracle algorithm K (the extractor) that achieves the following: Whenever some malicious prover P^* convinces V that a certain statement holds, the extractor $\mathsf{K}^{\mathsf{P}^*}$ with oracle access to P^* is able to return a witness for that statement. Here, we allow a certain knowledge error κ: if P^* convinces V with a probability smaller than κ, we do not require anything. Furthermore, we also do not require that the success probability of $\mathsf{K}^{\mathsf{P}^*}$ is as high as the success probability of P^*; instead, we only require that it is polynomially related. Finally, to facilitate the use of QPoKs as subprotocols, we give the malicious prover an auxiliary input $|\Phi\rangle$. We get the following definition:

Definition 1 (Quantum Proofs of Knowledge) *We call a proof system* (P, V) *for a relation* R quantum extractable with knowledge error κ *iff there exists a constant* $d > 0$*, a polynomially-bounded function* $p > 0$*, and a quantum-polynomial-time oracle*

[5] Again, the security parameter η is left implicit. The behavior of both A and M may depend on η.

machine K *such that for any interactive quantum machine* P^**, any polynomial* ℓ*, any security parameter* $\eta \in \mathbb{N}$*, any state* $|\psi\rangle$*, and any* $x \in \{0,1\}^{\leq \ell(\eta)}$*, we have that*

$$\Pr[\langle \mathsf{P}^*(x,|\psi\rangle), \mathsf{V}(x)\rangle = 1] \geq \kappa(\eta) \implies$$
$$\Pr[(x,w) \in R : w \leftarrow \mathsf{K}^{\mathsf{P}^*(x,|\psi\rangle)}(x)] \geq \tfrac{1}{p(\eta)} \Big(\Pr\big[\langle \mathsf{P}^*(x,|\psi\rangle), \mathsf{V}(x)\rangle = 1\big] - \kappa(\eta) \Big)^d.$$

A quantum proof of knowledge *for* R *with knowledge error* κ *(QPoK, for short) is a complete quantum extractable proof system for* R *with knowledge error* κ*.*

Note that by quantifying over all unitary provers P^*, we implicitly quantify over *all* purifications of *all* possible non-unitary provers. Note that extractability with knowledge error κ implies soundness with soundness error κ. We thus do not need to explicitly require soundness in Definition 1. The knowledge error κ can be made exponentially small by sequential repetition, see Section 2.3.

2.2 Discussion

In this section, we motivate various design choices made in the definition of QPoKs.

The security parameter. All our definitions include a security parameter η, and the behavior of all machines, and even the relation R can depend on it. All bounds (e.g., the runtimes, the soundness error $s(\eta)$, the knowledge error κ) also depend on η. In complexity theory, it is more common not to have an explicit security parameter, but to let all bounds depend on $|x|$. E.g., (P,V) is sound with soundness error s iff for all malicious provers P^*, all auxiliary inputs $|\Phi\rangle$, and all x with $\nexists w : (x,w) \in R$, we have $\Pr[\langle \mathsf{P}^*(x,|\Phi\rangle), \mathsf{V}(x)\rangle = 1] \leq s(|x|)$. E.g., [Wat09, Unr12] use this convention. This makes notation simpler, but may necessitate artificial padding, leading to unnatural protocols, especially if a proof is used as a subprotocol of a larger protocol. We remark that our definitions also captures the definitions without explicit security parameter if we use the relation $R_\eta := \{(x,w) \in R : |x| = \eta\}$, and let $\mathsf{P}(x,w)$ and $\mathsf{V}(x)$ abort (return 0) for $|x| \neq \bot$.

Access to the black-box prover's state and input. The extractor has no access to the prover's state nor to its quantum input. (This is modeled by the fact that an oracle algorithm may not apply any gates except for $\Box, \Box^\dagger$ to the register containing the oracle's state and quantum input.) In this, we follow [BG93] who argue in Section 4.3 that a proof of knowledge is supposed to "capture the knowledge of the prover *demonstrated by the interaction*" and that thus the extractor is not supposed to see the internal state of the prover. We stress, however, that our results are independent of this issue; they also hold if we allow the extractor direct access to the prover's state.

Unitary & invertible provers – technical view. Probably the most important design choice in our definition is to require the prover to be a unitary operation, and to allow the extractor to also execute the inverse of this operation. We begin with a discussion of this design choice from a technical point of view. First, we stress that

it seems that these assumptions are necessary: Since in a quantum world, making a snapshot/copy of a state is not possible or even well-defined, we have to allow the extractor to run the prover "backwards". But the inverse of a non-unitary quantum operation does not, in general, exist. Thus rewinding seems only possible with respect to unitary provers. Second, the probably most important question is: Does the definition, from an operational point of view, make sense? That is, does our definition behave well in cryptographic, reduction-based proofs? A final answer to this question can only be given when more protocols using QPoKs have been analyzed. The toy protocol discussed on page 5 gives a first indication that our definition can be used in a similar fashion to classical proofs of knowledge. Third, we would like to remind the reader that any non-unitary prover can be transformed into a unitary one by purification before applying the definition of QPoKs. Thus, allowing only unitary malicious provers does not seem to be a restriction in practice.

Unitary & invertible provers – "philosophical" view. Intuitively, a QPoK should guarantee that a prover that convinces the verifier "knows" the witness.[6] The basic idea is that if an extractor can extract the witness using only what is available to the prover, then the prover "knew" the witness (or could have computed it). In particular, we may allow the extractor to run a purified (unitary) version of the prover because the prover himself could have done so. Similarly for the inverse of that operation. Of course, this leaves the question why we give these two capabilities to the extractor but not others (e.g., access to the circuit of the prover)? We would like to stress that analogous questions are still open (from a philosophical point) even in the classical case: Why is it natural to allow an extractor to rewind the prover? Why is it natural to give a trapdoor for a common reference string to the extractor? We would like to point out one justification for the assumption that the prover is unitary, though: [BG93] suggests that we "capture the knowledge of the prover *demonstrated by the interaction*". A prover that performs non-unitary operations is identical in terms of its interaction to one that is purified. Thus, by restricting to unitary provers, we come closer to only capturing the interaction but not the inner workings of the prover.

On the success probability of the extractor. We require the extractor to run in quantum-polynomial-time and to succeed with probability $\frac{1}{p}(\Pr_{\mathsf{V}} - \kappa)^d$ where $\Pr_{\mathsf{V}}$ is the probability that the prover convinces the verifier. This follows [HM98]. An alternative definition would be to require the prover to run in expected time $p/(\Pr_{\mathsf{V}} - \kappa)^d$ and to extract a witness with probability 1 (or negligibly to 1). In the classical setting, the former definition is easily seen to imply the latter: to increase the success probability to 1, one repeatedly runs the extractor until it succeeds. This multiplies the expected running time with the inverse success probability, as required.

However, in the quantum setting, this does not seem possible. If the extractor fails on the first run, we do not have access to the original initial state to run the extractor again.[7] Even for the specific constructions analyzed in Section 3, we do not know how to

[6] We believe, though, that this issue is secondary to the technical suitability; it is much more important that a QPoK is useful as a cryptographic subprotocol.

[7] The oblivious rewinding technique by Watrous [Wat09] would seem to help here, but when trying to

amplify the success probability of the extractor.

There are applications where our definition is not strong enough and one needs an extractor that succeeds at least with overwhelming probability. For example, when we use the proof of knowledge property in the construction of a simulator who needs the witness to perform his simulation correctly. One such case is the graph non-isomorphism proof from [GMW91] where the zero-knowledge property of the graph non-isomorphism proof relies on the proof of knowledge property of the graph isomorphism proof. A similar case is the GMW-compiler for multi-party computation (as presented in [Gol04, Chapter 7]).

Arguments of knowledge. Definition 1 considers computationally unlimited malicious provers P. In many situations it is useful to weaken security and to consider only quantum-polynomial-time provers. This leads to the notion of quantum *arguments* of knowledge (a.k.a. quantum computationally sound proofs of knowledge). In this article, we do not investigate quantum arguments of knowledge. In fact, as shown in [ARU14], the constructions of proofs of knowledge described here (Section 3) do not directly translate to quantum arguments of knowledge. They show that relative to a suitable oracle, in the computational case, the constructions from Section 3 do not even constitute quantum arguments (and thus in particular not quantum arguments of knowledge).[8] However, for completeness, we state the definition of quantum arguments of knowledge here:

Definition 2 (Quantum Argument of Knowledge) *We call a proof system* (P,V) *for a relation* R quantum-computationally extractable with knowledge error κ *iff there exists a constant* $d>0$*, a polynomially-bounded function* $p>0$*, and a quantum-polynomial-time oracle machine* K *such that for any quantum-polynomial-time machine* P^* *and any polynomial* ℓ*, there exists a negligible function* μ*, such that any security parameter* $\eta\in\mathbb{N}$*, any state* $|\psi\rangle$*, and any* $x\in\{0,1\}^{\leq\ell(\eta)}$*, we have that*

$$\Pr[\langle\mathsf{P}^*(x,|\psi\rangle),\mathsf{V}(x)\rangle=1]\geq\kappa(\eta)\implies$$
$$\Pr[(x,w)\in R:w\leftarrow\mathsf{K}^{\mathsf{P}^*(x,|\psi\rangle)}(x)]\geq\tfrac{1}{p(\eta)}\Big(\Pr\big[\langle\mathsf{P}^*(x,|\psi\rangle),\mathsf{V}(x)\rangle=1\big]-\kappa(\eta)\Big)^d-\mu(\eta).$$

A quantum argument of knowledge *for* R *with knowledge error* κ *is a complete quantum-computationally extractable proof system for* R *with knowledge error* κ*.*

Note that in comparison to Definition 1, we added an additional negligible error μ that may depend on the malicious prover P^*. This is because κ is not allowed to depend on the choice of P^*, but there usually is a negligible attack probability μ that depends on P^* (e.g., $\mu=2^{-\eta}T$ where T is the runtime of P^*).

Unruh [Unr15] generalizes our technique using "collapse-binding" commitments instead of "strict binding" commitments to construct quantum arguments of knowledge.

apply that technique one gets the requirement that the invoked extractors' success probability must be independent of the auxiliary input. This condition is not necessarily fulfilled.

[8] More precisely (using language from Section 3): A Σ-protocol that has quantum-computational special soundness and quantum-computational strict soundness is not necessarily a quantum argument.

2.3 Amplification

In some cases, elementary constructions only yield QPoKs with constant knowledge error κ. Yet, in most cases we need QPoKs with negligible knowledge error. One possibility to construct these is to sequentially iterate a QPoK with constant knowledge error, the knowledge error of the resulting QPoK then becomes exponentially small. This result is well-known in the classical case [BG93]; the proof in the quantum case follows the same lines.

Theorem 3 *Let $n = n(\eta)$ be a polynomially bounded and efficiently computable function. Let (P, V) be extractable with knowledge error κ. Let $(\mathsf{P}', \mathsf{V}')$ be the proof system consisting of n sequential executions of (P, V). Then $(\mathsf{P}', \mathsf{V}')$ is extractable with knowledge error κ^n.*

Proof. We write short n, κ, p for $n(\eta), \kappa(\eta), p(\eta)$. We call (P, V) the atomic proof and $(\mathsf{P}', \mathsf{V}')$ the composed proof. Fix a malicious prover P^* (that is supposed to interact with V'), a security parameter η, a statement x, and an auxiliary input $|\Phi\rangle$ for P^*. In the execution of the composed proof with prover P^*, we call each execution of the atomic proof a round. Without loss of generality, we can assume that P^* consists of n sequentially executed provers P_i^* such that P_i^* executes the i-th round of the composed proof. For $i \geq 2$, P_i^* expects as quantum input the state that was output by P_{i-1}^*. Let K be the knowledge extractor for the atomic proof. We construct a knowledge extractor K' for the composed proof as follows: First, K' picks a random $i \in \{1, \ldots, n\}$. Then K' internally simulates the first $i-1$ rounds of the composed proof (with provers $\mathsf{P}_1^*, \ldots, \mathsf{P}_{i-1}^*$). Let $|\Phi'\rangle$ denote the state output by P_{i-1}^*. (And $|\Phi'\rangle := |\Phi\rangle$ if $i = 1$.) Then K' runs $w \leftarrow \mathsf{K}^{\mathsf{P}_i^*(x, |\Phi'\rangle)}(x)$ and outputs w.[9]

For the remainder of the proof, fix the security parameter η. We use the following notation: a_i is the probability that the first i rounds of the composed proof succeed (with prover P^*). We stress that a_{i-1} is also the probability that in an execution of K', the internal simulation of the first $i-1$ rounds succeeds. Let c_i denote the probability that the i-th round of the composed proof succeeds, conditioned on the event that the first $i-1$ rounds succeed. We have $a_0 = 1$ and $a_i = c_i a_{i-1}$ for $i = 1, \ldots, n$.

Let $\Pr_{\mathsf{V}'}$ denote the probability that the composed proof succeeds, and let $\Pr_{\mathsf{K}'}$ denote the probability that K' succeeds (i.e., returns a valid witness). Fix some i. Let $\Pr_{\mathsf{K}'}^{(i)}$ denote the probability that K' succeeds, conditioned on the fact that K' chooses that i. Then, by construction of K', we have that $\Pr_{\mathsf{K}'} = \sum_{i=1}^n \frac{1}{n} \Pr_{\mathsf{K}'}^{(i)} \geq \max_i \frac{1}{n} \Pr_{\mathsf{K}'}^{(i)}$. We will show that there exists an i (dependent on P^*, $|\Phi\rangle$, η, and x), as well as a polynomially-bounded $p > 0$ and an integer $d > 0$ (independent of i, P^*, $|\Phi\rangle$, η, and x) such that $\Pr_{\mathsf{K}'}^{(i)} \geq \frac{1}{p}(\Pr_{\mathsf{V}'} - \kappa^n)^d$ if $\Pr \mathsf{V}' \geq \kappa^n$. This implies that $\Pr_{\mathsf{K}'} \geq \frac{1}{pn}(\Pr_{\mathsf{V}'} - \kappa^n)^d$. Thus $(\mathsf{P}', \mathsf{V}')$ has knowledge error κ^n.

We proceed to bound $\Pr_{K'}^{(i)}$ in terms of a_{i-1} and c_i. Let $|\Phi'\rangle$ be the output state of P_{i-1}^* in the event that the first $i-1$ rounds of the composed proof succeed. Let

[9] Note that K as defined and analyzed here is not a unitary algorithm, but instead performs random choices and measurement. Since any such K can be converted into a unitary one by purification, we can use a non-unitary K without loss of generality.

$\Pr^{(i)}_{\mathsf{K}}(|\Phi'\rangle)$ denote the probability that $\mathsf{K}^{\mathsf{P}^*(x,|\Phi'\rangle)}(x)$ succeeds (outputs a witness), and $\Pr^{(i)}_{V}(|\Phi'\rangle)$ the probability that the atomic proof with prover P^* and auxiliary input $|\Phi'\rangle$ succeeds. Then the probability that K' succeeds, conditioned on the event that the first $i-1$ rounds succeed, is $\Pr^{(i)}_{\mathsf{K}}(|\Phi'\rangle)$. Hence $\Pr^{(i)}_{\mathsf{K}'} = a_{i-1}\Pr^{(i)}_{\mathsf{K}}(|\Phi'\rangle)$. Since the atomic proof has knowledge error κ, there are a polynomially-bounded $p > 0$ and an integer $d > 0$ such that $\Pr^{(i)}_{\mathsf{K}}(|\Phi'\rangle) \geq \frac{1}{p}\big(\Pr^{(i)}_{\mathsf{V}}(|\Phi'\rangle) - \kappa\big)^d$ for all $|\Phi'\rangle$. Without loss of generality, we can pick $d \geq 1$. We stress that p and d are independent of i, P^*, $|\Phi\rangle$, η, and x. It follows that

$$\Pr^{(i)}_{\mathsf{K}'} = a_{i-1}\Pr^{(i)}_{\mathsf{K}}(|\Phi'\rangle) \geq a_{i-1}\tfrac{1}{p}\big(\Pr^{(i)}_{\mathsf{V}}(|\Phi'\rangle) - \kappa\big)^d = a_{i-1}\tfrac{1}{p}\big(c_i - \kappa\big)^d.$$

Summarizing, at this point we know that $\Pr_{\mathsf{K}'} \geq \max_i \frac{1}{n}\Pr^{(i)}_{\mathsf{K}'} \geq \max_i \frac{a_{i-1}}{pn}\big(c_i-\kappa\big)^d$, that $a_i = c_i a_{i-1}$ for all i, and that $\Pr_{\mathsf{V}'} = a_n$.

Let $\delta := \Pr_{\mathsf{V}'} - \kappa^n$. Assume that $\delta > 0$ and $\kappa \leq 1$, otherwise nothing needs to be shown. Since $a_0 \leq 1 = \kappa^0 + \frac{0\delta}{n}$ and $a_n \geq \Pr_{\mathsf{V}'} = \kappa^n + \frac{n\delta}{n}$, we have that for some $i \in \{1,\ldots,n\}$, $a_{i-1} \leq \kappa^{i-1} + \frac{(i-1)\delta}{n}$ and $a_i \geq \kappa^i + \frac{i\delta}{n}$. For that i, we have

$$a_{i-1}(c_i-\kappa) = a_i - a_{i-1}\kappa \geq (\kappa^i + \tfrac{i\delta}{n}) - (\kappa^i + \tfrac{(i-1)\delta\kappa}{n}) \overset{(\kappa\leq 1)}{\geq} (\kappa^i + \tfrac{i\delta}{n}) - (\kappa^i + \tfrac{(i-1)\delta}{n}) = \tfrac{\delta}{n}$$

and hence

$$\Pr_{\mathsf{K}'} \geq \max_i \frac{a_{i-1}}{pn}\big(c_i-\kappa\big)^d \overset{(d\geq 1)}{\geq} \max_i \frac{1}{pn} a_{i-1}^d\big(c_i-\kappa\big)^d \geq \frac{1}{pn}(\tfrac{\delta}{n})^d = \frac{1}{pn^{d+1}}(\Pr_{\mathsf{V}'} - \kappa^n)^d.$$

Since pn^{d+1} is polynomially-bounded in η, it follows that the composed proof $(\mathsf{P}', \mathsf{V}')$ has knowledge error κ^n. □

Parallel amplification. We have no results whether the knowledge error of quantum proofs of knowledge decreases in general under parallel composition. (This is not even clear in the classical case.) However, the knowledge error in the constructions from Section 3 decreases exponentially under parallel composition. This is because the assumptions for the construction (special soundness and strict soundness) are easily seen to be preserved under parallel composition, and the size of the challenge space increases exponentially. Unfortunately, the zero-knowledge property is not preserved under parallel composition. Still, parallel composition can be useful if only witness indistinguishability is required.

In particular, this implies (using the constructions from Section 5 below) that quantum-computationally witness indistinguishable three-round QPoK with negligible knowledge error exist.

3 Elementary constructions

In this section, we show that under certain conditions, a classical PoK is also a QPoK (i.e., secure against malicious quantum provers). The first condition refers to the outer form of the protocol; we require that the proof systems is a protocol with three messages

(commitment, challenge, and response) with a public-coin verifier. Such protocols are called Σ-protocols. Furthermore, we require that the proof system has special soundness. This means that given two accepting conversations between prover and verifier that have the same commitment but different challenges, we can efficiently compute a witness. Σ-protocols with special soundness are well-studied in the classical case; many efficient classical protocols with these properties exist. The third condition (strict soundness) is non-standard. We require that given the commitment and the challenge of a conversation, there is at most one response that would make the verifier accept. We require strict soundness to ensure that the response given by the prover does not contain too much information; measuring it will then not disturb the state of the prover too much. Not all known protocols have strict soundness (the proof for graph isomorphism [GMW91] is an example). Fortunately, many protocols can be modified to satisfy strict soundness; a slight variation of the proof for Hamiltonian cycles [Blu86] is an example (see Section 5).

Definition 4 (Σ-protocol) *A proof system (P,V) is called a Σ-protocol iff P and V are classical, the interaction consists of three messages $com, ch, resp$ (sent by P, V, and P, respectively, and called* commitment*,* challenge*, and* response*), and ch is uniformly chosen from some set $C_{\eta x}$ (the* challenge space*) that may only depend on the statement x and the security parameter η. Furthermore, the verifier decides whether to accept or not by a deterministic polynomial-time computation on $x, com, ch, resp$. (We call $(com, ch, resp)$ an* accepting conversation for x *if the verifier would accept it.) We also require that it is possible in probabilistic polynomial time to sample uniformly from $C_{\eta x}$ up to negligible error,*[10] *and that membership in $C_{\eta x}$ should be decidable given η, x in deterministic polynomial time in $\eta + |x|$.*

Definition 5 (Special soundness) *We say a Σ-protocol (P,V) for a relation R has* special soundness *iff there is a deterministic polynomial-time algorithm K_0 (the* special extractor*) such that the following holds: For any two accepting conversations $(com, ch, resp)$ and $(com, ch', resp')$ for x such that $ch \neq ch'$ and $ch, ch' \in C_{\eta x}$, we have that $w := \mathsf{K}_0(x, com, ch, resp, ch', resp')$ satisfies $(x, w) \in R$.*

Definition 6 (Strict soundness) *We say a Σ-protocol (P,V) has* strict soundness *iff for any two accepting conversations $(com, ch, resp)$ and $(com, ch, resp')$ for x, we have that $resp = resp'$.*

Canonical extractor. Let (P,V) be a Σ-protocol with special soundness and strict soundness. Let K_0 be the special extractor for that protocol. We define the *canonical extractor* K for (P,V). K will use measurements, even though our definition of quantum

[10] That is, there should be a probabilistic Turing machine M that runs in polynomial time in $|x| + \eta$ such hat the output of $M(\eta, x)$ has negligible (in η) statistical distance from the uniform distribution on $C_{\eta x}$.

We allow negligible error since otherwise it is not even possible to sample from, say, $C_{\eta x}$. See [KSU13] for more discussion of this issue.

oracle algorithms only allows for unitary operations. This is only for the sake of presentation; by purifying K one can derive a unitary algorithm with the same properties. Given a malicious prover P^*, $\mathsf{K}^{\mathsf{P}^*(x,|\Phi\rangle)}(x)$ operates on two quantum registers N, S_{P^*}. N is used for communication with P^*, and S_{P^*} is used for the state of P^*. As described in the definition of quantum oracle machines, the registers N, S_{P^*} are initialized with $|0\rangle, |\Phi\rangle$. Let $\mathsf{P}^*_{\eta x}$ denote the unitary transformation describing a single activation of P^*. First, K applies $\mathsf{P}^*_{\eta x}$ to N, S_{P^*}. (This can be done using the special gate $\square$.) This corresponds to running the first step of P^*; in particular, N should now contain the commitment. Then K measures N in the computational basis; call the result *com*. Then K initializes N with $|0\rangle$. Then K chooses uniformly random values $ch, ch' \in C_{\eta x}$. Let U_{ch} denote the unitary transformation operating on N such that $U_{ch}|x\rangle = |x \oplus ch\rangle$. (Here $\oplus$ denotes the bitwise XOR of bitstrings.) Then K applies $\mathsf{P}^*_{\eta x} U_{ch}$. (Now N is expected to contain the response for challenge *ch*.) Then K measures N in the computational basis; call the result *resp*. Then K applies $(\mathsf{P}^*_{\eta x} U_{ch})^\dagger$ (we "rewind" the prover). Then $\mathsf{P}^*_{\eta x} U_{ch'}$ is applied.[11] (Now N is expected to contain the response for challenge ch'.) Then N is measured in the computational basis; call the result $resp'$. Then $(\mathsf{P}^*_{\eta x} U_{ch'})^\dagger$ is applied. Finally, K outputs $w := \mathsf{K}_0(x, com, ch, resp, ch', resp')$.

Analysis of the canonical extractor. In order to analyze the canonical extractor (Theorem 9 below), we first need a lemma that bounds the probability that two consecutive binary measurements P_{ch} and $P_{ch'}$ with random $ch \neq ch'$ succeed in terms of the probability that a single such measurement succeeds. In a classical setting, the answer is simple: the outcomes of the measurements are independent; thus the probability that two measurements succeed is the square of the probability that a single measurement succeeds. (And similar reasoning applies in the quantum case if the measurements commute.) In the general quantum case, however, the first measurement may disturb the state; this makes the analysis considerably more involved. We first prove some inequalities needed in the proof:

Lemma 7 *Let C be a set with $\#C = c$. Let $(P_i)_{i\in C}$ be orthogonal projectors on a Hilbert space $\mathcal{H}$. Let $|\Phi\rangle \in \mathcal{H}$ be a unit vector. Let $V := \sum_{i\in C} \frac{1}{c}\||P_i|\Phi\rangle\|^2$ and $F := \sum_{i,j\in C} \frac{1}{c^2}\|P_iP_j|\Phi\rangle\|^2$. Then $F \geq V^3$.*

Proof. To prove the lemma, we first show two simple facts:

Claim 1 *For any positive operator A on $\mathcal{H}$ and any unit vector $|\Phi\rangle \in \mathcal{H}$, we have that $(\langle\Phi|A|\Phi\rangle)^3 \leq \langle\Phi|A^3|\Phi\rangle$.*

Since A is positive, it is diagonalizable. Thus we can assume without loss of generality that A is diagonal (by applying a suitable basis transform to A and $|\Phi\rangle$). Let a_i be the i-th diagonal element of A, and let f_i be the i-th component of $|\Phi\rangle$. Then

$$(\langle\Phi|A|\Phi\rangle)^3 = \Big(\sum_i |f_i|^2 a_i\Big)^3 \overset{(*)}{\leq} \sum_i |f_i|^2 a_i^3 = \langle\Phi|A^3|\Phi\rangle.$$

[11] This step has no effect on the observable behavior of K, but it makes the analysis of K more pleasant.

Here $(*)$ uses Jensen's inequality [Jen06] and the facts that $a_i \geq 0$, and that $a_i \mapsto a_i^3$ is a convex function on nonnegative numbers, and that $\sum_i |f_i|^2 = 1$. This concludes the proof of Claim 1.

Claim 2 *For vectors $|\Psi_1\rangle, \ldots, |\Psi_c\rangle \in \mathcal{H}$, it holds that $\big\|\frac{1}{c}\sum_i |\Psi_i\rangle\big\|^2 \leq \frac{1}{c}\sum_i \big\||\Psi_i\rangle\big\|^2$.*

To show the claim, let $|\bar{\Psi}\rangle := \sum_i \frac{1}{c}|\Psi_i\rangle$. Then

$$\begin{aligned}
\sum_i \Big(\big\||\Psi_i\rangle\big\|^2 - \big\||\bar{\Psi}\rangle\big\|^2\Big) &= \sum_i \Big(\big\||\Psi_i\rangle\big\| - \big\||\bar{\Psi}\rangle\big\|\Big)\Big(\big\||\Psi_i\rangle\big\| - \big\||\bar{\Psi}\rangle\big\| + 2\big\||\bar{\Psi}\rangle\big\|\Big) \\
&= \sum_i \Big(\big\||\Psi_i\rangle\big\| - \big\||\bar{\Psi}\rangle\big\|\Big)^2 + 2\big\||\bar{\Psi}\rangle\big\| \sum_i \Big(\big\||\Psi_i\rangle\big\| - \big\||\bar{\Psi}\rangle\big\|\Big) \\
&\geq 2\big\||\bar{\Psi}\rangle\big\| \sum_i \Big(\big\||\Psi_i\rangle\big\| - \big\||\bar{\Psi}\rangle\big\|\Big) = 2\big\||\bar{\Psi}\rangle\big\| \Big(\sum_i \big\||\Psi_i\rangle\big\| - \big\|c|\bar{\Psi}\rangle\big\|\Big) \\
&= 2\big\||\bar{\Psi}\rangle\big\| \Big(\sum_i \big\||\Psi_i\rangle\big\| - \Big\|\sum_i |\Psi_i\rangle\Big\|\Big) \qquad (1)
\end{aligned}$$

From the triangle inequality, it follows that $\sum_i \big\||\Psi_i\rangle\big\| \geq \big\|\sum_i |\Psi_i\rangle\big\|$, hence with (1), we have $\sum_i \Big(\big\||\Psi_i\rangle\big\|^2 - \big\||\bar{\Psi}\rangle\big\|^2\Big) \geq 0$. Then $\frac{1}{c}\sum_i \big\||\Psi_i\rangle\big\|^2 - \big\|\frac{1}{c}\sum_i |\Psi_i\rangle\big\|^2 = \frac{1}{c}\sum_i \Big(\big\||\Psi_i\rangle\big\|^2 - \big\||\bar{\Psi}\rangle\big\|^2\Big) \geq 0$. Claim 2 follows.

We proceed to prove Lemma 7. Let $A := \sum_i \frac{1}{c} P_i$, let $|\Psi_{ij}\rangle := P_j P_i |\Phi\rangle$. Then A is positive. Furthermore,

$$\begin{aligned}
V^3 &= \Big(\sum_i \tfrac{1}{c}\langle\Phi|P_i|\Phi\rangle\Big)^3 = \big(\langle\Phi|A|\Phi\rangle\big)^3 \overset{(*)}{\leq} \langle\Phi|A^3|\Phi\rangle = \sum_{i,j,k} \tfrac{1}{c^3}\langle\Phi|P_i P_j P_k|\Phi\rangle \\
&= \sum_{i,j,k} \tfrac{1}{c^3}\langle\Psi_{ij}|\Psi_{kj}\rangle = \sum_j \tfrac{1}{c}\Big(\sum_i \tfrac{1}{c}\langle\Psi_{ij}|\Big)\Big(\sum_k \tfrac{1}{c}|\Psi_{kj}\rangle\Big) = \sum_j \tfrac{1}{c}\Big\|\sum_i \tfrac{1}{c}|\Psi_{ij}\rangle\Big\|^2 \\
&\overset{(**)}{\leq} \sum_j \tfrac{1}{c}\sum_i \tfrac{1}{c}\big\||\Psi_{ij}\rangle\big\|^2 = F.
\end{aligned}$$

Here $(*)$ uses Claim 1 and $(**)$ uses Claim 2. Thus we have $F \geq V^3$ and Lemma 7 follows.

Lemma 8 *Let C be a set with $\#C = c$. Let $(P_i)_{i\in C}$ be orthogonal projectors on a Hilbert space $\mathcal{H}$. Let $|\Phi\rangle \in \mathcal{H}$ be a unit vector. Let $V := \sum_{i\in C} \frac{1}{c}\|P_i|\Phi\rangle\|^2$ and $E := \sum_{i,j\in C, i\neq j} \frac{1}{c^2}\|P_i P_j|\Phi\rangle\|^2$. Then, if $V \geq \frac{1}{\sqrt{c}}$, we have $E \geq V(V^2 - \frac{1}{c})$.*

Proof. Let F be as in Lemma 7. Then

$$\begin{aligned}
E &= \sum_{\substack{i,j\in C \\ i\neq j}} \tfrac{1}{c^2}\big\|P_i P_j|\Phi\rangle\big\|^2 = \sum_{i,j\in C} \tfrac{1}{c^2}\big\|P_i P_j|\Phi\rangle\big\|^2 - \sum_{i\in C} \tfrac{1}{c^2}\big\|P_i P_i|\Phi\rangle\big\|^2 \\
&\overset{(*)}{=} \sum_{i,j\in C} \tfrac{1}{c^2}\big\|P_i P_j|\Phi\rangle\big\|^2 - \sum_{i\in C} \tfrac{1}{c^2}\big\|P_i|\Phi\rangle\big\|^2 = F - \tfrac{V}{c} \overset{(**)}{\geq} V^3 - \tfrac{V}{c} = V(V^2 - \tfrac{1}{c}).
\end{aligned}$$

Here $(*)$ uses that $P_i = P_i P_i$ since P_i is a projection, and $(**)$ uses Lemma 7. □

Theorem 9 *Let (P,V) be a Σ-protocol for a relation R with challenge space $C_{\eta x}$. Fix a function c such that for all $\eta \in \mathbb{N}$, $x \in \{0,1\}^*$ we have $\#C_{\eta x} \geq c(\eta)$. Assume that (P,V) has special soundness and strict soundness. Then (P,V) is quantum extractable with knowledge error $1/\sqrt{c}$.*

Proof. To show that (P,V) is extractable, we will use the canonical extractor K. Fix a malicious prover P^*, a statement x, and an auxiliary input $|\Phi\rangle$. Let $\Pr_{\mathsf{V}}$ denote the probability that the verifier accepts when interacting with P^*. Let $\Pr_{\mathsf{K}}$ denote the probability that $\mathsf{K}^{\mathsf{P}^*(x,|\Phi\rangle)}(x)$ outputs some w with $(x,w) \in R$. We will show that $\Pr_{\mathsf{K}} \geq \Pr_{\mathsf{V}} \cdot (\Pr_{\mathsf{V}}^2 - \frac{1}{\#C_{\eta x}})$ for $\Pr_V \geq \frac{1}{\sqrt{\#C_{\eta x}}}$. Hence for $\Pr_{\mathsf{V}} \geq \frac{1}{\sqrt{c(\eta)}} \geq \frac{1}{\sqrt{\#C_{\eta x}}}$, we have that

$$\Pr_{\mathsf{V}}(\Pr_{\mathsf{V}}{}^2 - \tfrac{1}{\#C_{\eta x}}) \geq \Pr_{\mathsf{V}}(\Pr_{\mathsf{V}}{}^2 - \tfrac{1}{c(\eta)}) = \Pr_{\mathsf{V}}\big(\Pr_{\mathsf{V}} + \tfrac{1}{\sqrt{c(\eta)}}\big)\big(\Pr_{\mathsf{V}} - \tfrac{1}{\sqrt{c(\eta)}}\big) \geq \big(\Pr_{\mathsf{V}} - \tfrac{1}{\sqrt{c(\eta)}}\big)^3.$$

Since furthermore K is quantum-polynomial-time, this implies that (P,V) is extractable with knowledge error $1/\sqrt{c}$.

In order to show $\Pr_{\mathsf{K}} \geq \Pr_{\mathsf{V}} \cdot (\Pr_{\mathsf{V}}^2 - \frac{1}{\#C_{\eta x}})$, we will use a short sequence of games. Each game will contain an event Succ, and in the first game, we will have $\Pr[\mathsf{Succ} : \text{Game } 1] = \Pr_{\mathsf{K}}$. For any two consecutive games, we will have $\Pr[\mathsf{Succ} : \text{Game } i] \geq \Pr[\mathsf{Succ} : \text{Game } i+1]$, and for the final game, we will have $\Pr[\mathsf{Succ} : \text{Game } 7] \geq \Pr_{\mathsf{V}} \cdot (\Pr_{\mathsf{V}}^2 - \frac{1}{\#C_{\eta x}})$. This will then conclude the proof. The description of each game will only contain the changes with respect to the preceding game.

Game 1. An execution of $\mathsf{K}^{\mathsf{P}^*(x,|\Phi\rangle)}(x)$. Succ denotes the event that K outputs a valid witness for x. By definition, $\Pr_{\mathsf{K}} = \Pr[\mathsf{Succ} : \text{Game } 1]$.

Game 2. Succ denotes the event that $(com, ch, resp)$ and $(com, ch', resp')$ are accepting conversations for x and $ch \neq ch'$. (The variables $(com, ch, resp)$ and $(com, ch', resp')$ are as in the definition of the canonical extractor.) Since (P,V) has special soundness, if Succ occurs, K outputs a valid witness. Thus $\Pr[\mathsf{Succ} : \text{Game } 1] \geq \Pr[\mathsf{Succ} : \text{Game } 2]$.

Game 3. Before K measures $resp$, it first measures whether measuring $resp$ would yield an accepting conversation. More precisely, it measures N with the orthogonal projector P_{ch} projecting onto $V_{ch} := \operatorname{span}\{|resp\rangle : (com, ch, resp) \text{ is accepting}\}$. Analogously for the measurement of $resp'$ (using the projector $P_{ch'}$.) Since a complete measurement (of $resp$ and $resp'$, respectively) is performed on N after applying the measurement P_{ch} and $P_{ch'}$, introducing the additional measurements does not change the outcomes $resp$ and $resp'$ of these complete measurements, nor their post-measurement state. Thus $\Pr[\mathsf{Succ} : \text{Game } 2] = \Pr[\mathsf{Succ} : \text{Game } 3]$.

Game 4. Succ denotes the event that $ch \neq ch'$ and both measurements P_{ch} and $P_{ch'}$ succeed. By definition of these measurements, this happens iff $(com, ch, resp)$ and $(com, ch', resp')$ are accepting conversations and $ch \neq ch'$. Thus $\Pr[\mathsf{Succ} : \text{Game } 3] = \Pr[\mathsf{Succ} : \text{Game } 4]$.

Game 5. We do not execute K_0, i.e., we stop after applying $(\mathsf{P}^*_{\eta x}U_{ch'})^\dagger$. Since at that point, Succ has already been determined, $\Pr[\mathsf{Succ} : \text{Game 4}] = \Pr[\mathsf{Succ} : \text{Game 5}]$.

Game 6. We remove the measurements of *resp* and *resp*′. Note that the outcomes of these measurements are not used any more. Since (P, V) has strict soundness, $V_{ch} = \operatorname{span}\{|resp_0\rangle\}$ for a single value $resp_0$ (depending on *com* and *ch*, of course). Thus if the measurement P_{ch} succeeds, the post-measurement state in N is $|resp_0\rangle$. That is, the state in N is classical at this point. Thus, measuring N in the computational basis does not change the state. Hence, the measurement of *resp* does not change the state. Analogously for the measurement of *resp*′. It follows that $\Pr[\mathsf{Succ} : \text{Game 5}] = \Pr[\mathsf{Succ} : \text{Game 6}]$.

Game 7. First, N and S_{P^*} are initialized with $|0\rangle$ and $|\Phi\rangle$. Then the unitary transformation $\mathsf{P}^*_{\eta x}$ is applied. Then *com* is measured (complete measurement on N), and N is initialized to $|0\rangle$. Random $ch, ch' \in C_{\eta x}$ are chosen. Then $\mathsf{P}^*_{\eta x}U_{ch}$ is applied. Then the measurement P_{ch} is performed. Then $(\mathsf{P}^*_{\eta x}U_{ch})^\dagger$ is applied. Then $\mathsf{P}^*_{\eta x}U_{ch'}$ is applied. Then the measurement $P_{ch'}$ is performed. Then $(\mathsf{P}^*_{\eta x}U_{ch'})^\dagger$ is applied. The event Succ holds if $ch \neq ch'$ and both measurements succeed. Games 6 and 7 are identical; we have just recapitulated the game for clarity. Thus, $\Pr[\mathsf{Succ} : \text{Game 6}] = \Pr[\mathsf{Succ} : \text{Game 7}]$.

In Game 7, for some value d, let p_d denote the probability that $com = d$ is measured. Let $|\Phi_d\rangle$ denote the state of N, S_{P^*} after measuring $com = d$ and initializing N with $|0\rangle$. (I.e., the state directly before applying $\mathsf{P}^*_{\eta x}U_{ch}$.) Let K_d denote the probability that starting from state $|\Phi_d\rangle$, both measurements P_{ch} and $P_{ch'}$ succeed and $ch \neq ch'$. Then we have that $\Pr[\mathsf{Succ} : \text{Game 7}] = \sum_d p_d K_d$ and

$$K_d = \sum_{\substack{ch,ch'\in C_{\eta x}\\ ch\neq ch'}} \frac{1}{\#C_{\eta x}^{\,2}} \big\|(\mathsf{P}^*_{\eta x}U_{ch'})^\dagger P_{ch'}(\mathsf{P}^*_{\eta x}U_{ch'})(\mathsf{P}^*_{\eta x}U_{ch})^\dagger P_{ch}(\mathsf{P}^*_{\eta x}U_{ch})|\Phi_d\rangle\big\|^2$$
$$= \sum_{\substack{ch,ch'\in C_{\eta x}\\ ch\neq ch'}} \frac{1}{\#C_{\eta x}^{\,2}} \big\|P^*_{ch'}P^*_{ch}|\Phi_d\rangle\big\|^2$$

where $P^*_{ch} := (\mathsf{P}^*_{\eta x}U_{ch})^\dagger P_{ch}(\mathsf{P}^*_{\eta x}U_{ch})$. Since P_{ch} is an orthogonal projector and $\mathsf{P}^*_{\eta x}U_{ch}$ is unitary, P^*_{ch} is an orthogonal projector. Let $\varphi(v) := v(v^2 - \frac{1}{\#C_{\eta x}})$ for $v \in \big[\frac{1}{\sqrt{\#C_{\eta x}}}, 1\big]$ and $\varphi(v) := 0$ for $v \in [0, \frac{1}{\sqrt{\#C_{\eta x}}}]$. Then, by Lemma 8, $K_d \geq \varphi(V_d)$ for $V_d := \sum_{ch\in C_{\eta x}} \frac{1}{\#C_{\eta x}}\|P^*_{ch}|\Phi_d\rangle\|^2$.

Furthermore, by construction of the honest verifier V, we have that

$$\Pr_{\mathsf{V}} = \sum_d p_d \sum_{ch\in C_{\eta x}} \tfrac{1}{\#C_{\eta x}} \big\|P_{ch}\mathsf{P}^*_{\eta x}U_{ch}|\Phi_d\rangle\big\|^2$$
$$\overset{(*)}{=} \sum_d p_d \sum_{ch\in C_{\eta x}} \tfrac{1}{\#C_{\eta x}} \big\|(\mathsf{P}^*_{\eta x}U_{ch})^\dagger P_{ch}(\mathsf{P}^*_{\eta x}U_{ch})|\Phi_d\rangle\big\|^2 = \sum_d p_d V_d.$$

where $(*)$ uses that $(\mathsf{P}^*_{\eta x}U_{ch})^\dagger$ is unitary. Finally, we have

$$\Pr_{\mathsf{K}} = \Pr[\mathsf{Succ} : \text{Game 1}] \geq \Pr[\mathsf{Succ} : \text{Game 7}] = \sum_d p_d K_d \geq \sum_d p_d \varphi(V_d) \overset{(*)}{\geq} \varphi(\Pr_{\mathsf{V}}).$$

Here $(*)$ uses Jensen's inequality [Jen06] and the fact that φ is convex on $[0,1]$. As discussed in the beginning of the proof, $\Pr_{\mathsf{K}} \geq \varphi(\Pr_{\mathsf{V}}) = \Pr_{\mathsf{V}} \cdot (\Pr_{\mathsf{V}}^2 - \frac{1}{\#C_{\eta x}})$ for $\Pr_{\mathsf{V}} \geq \frac{1}{\sqrt{\#C_{\eta x}}}$ implies that (P,V) is a QPoK with knowledge error $1/\sqrt{c}$. □

Arguments of knowledge. On might be tempted to think that the results from this section carry over directly to the computational case. That is, if a Σ-protocol has computational special soundness and computational strict soundness,[12] then it is an argument of knowledge. Unfortunately, this is not true. In the proof of Theorem 9, in the transition from Game 5 to Game 6, we used that a measurement of *resp* will not disturb the state. This was the case because strict soundness implied that there exists only one such *resp*. However, if we only use computational strict soundness, the register might contain several different *resp* in superposition. (Computational strict soundness merely implies that we cannot simultaneously find two of those *resp*.) Hence measuring *resp* could disturb the state.

In fact, [ARU14] shows that, relative to some oracle, this indeed happens. More precisely, a proof system with special soundness and *computational* strict soundness exists that is not a quantum argument of knowledge. And a proof system with *computational* special soundness and *computational* strict soundness exists that is not even a quantum argument. This excludes, at the very least, any relativizing proof of our results in the computational case. Most likely, the requirements will need to be strengthened.

3.1 On using existing bounds from the literature

A rewinding technique similar to ours has occurred in [CSST11] in the context of a specific two-prover commitment-scheme. Their proof is specific to the case where there are only two possible measurements (i.e., $\#C = 2$ in our language). However, in that specific case, their calculations allow us to derive a better bound also for our setting.

The following lemma is implicitly proven in [CSST11] in the proof of their Lemma 1:

Lemma 10 (Rewinding of mBQKW commitment [CSST11]) *Consider two projectors P_0 and P_1 of the form $P_i = U_i^\dagger(|\hat{w}_i\rangle\langle\hat{w}_i| \otimes I)U_i$. (Here U_0, U_1 are unitaries, and $\hat{w}_0, \hat{w}_1 \in \{0,1\}^n$ for some n.) Consider a state $|\psi\rangle$. Let $p_i := \||P_i|\psi\rangle\|^2$. (That is, p_i is the probability of measuring $\hat{w}_i$ in the first register after applying U_i to $|\psi\rangle$.) Let $p_\oplus := \|P_1P_0|\psi\rangle\|^2$. (That is, $p_\oplus$ is the probability of measuring $\hat{w}_0$ after applying U_0 to $|\psi\rangle$ and subsequently measuring $\hat{w}_1$ after applying $U_1U_0^\dagger$.)*

Assume that $p_0 + p_1 \geq 1 + \varepsilon$ for some $\varepsilon \geq 0$. Then $p_\oplus \geq \varepsilon^2/4$.

We can restate this in a form analogous to Lemma 8:

[12] Computational special soundness means that it is computationally hard to find a tuple $(x, com, ch, resp, ch', resp')$ with $ch \neq ch'$ such that K_0 does not output a valid witness. Computational strict soundness means that it is computationally hard to find accepting conversations $(com, ch, resp)$ and $(com, ch, resp')$ with $resp \neq resp'$. For precise definitions see [ARU14].

Lemma 11 *Let $C = \{0,1\}$. Let P_0, P_1 be projectors as in Lemma 10. Let $|\Phi\rangle$ be a unit vector. Let $V := \sum_{i=0,1} \frac{1}{2}\|P_i|\Phi\rangle\|^2$ and $E := \|P_1P_0|\Phi\rangle\|^2$. Then, if $V \geq \frac{1}{2}$, we have $E \geq (V - \frac{1}{2})^2$.*

Proof. Let $\varepsilon := 2V - 1$ and observe that $p_0 + p_1 = 2V = 1 + \varepsilon$ and $E = p_\oplus \geq \varepsilon^2/4$. □

We can use this lemma to show a variant of Theorem 9 for $\#C = 2$ with knowledge error $\frac{1}{2}$ (and not $\frac{1}{\sqrt{2}}$ as in Theorem 9).

Corollary 12 *Let (P, V) be a Σ-protocol for a relation R with challenge space $C_{\eta x}$. Assume $\#C_{\eta x} = 2$ for all η, x. Assume that (P, V) has special soundness and strict soundness. Then (P, V) is extractable with knowledge error $1/2$.*

Proof. Without loss of generality, we can assume that $C_{\eta x} = \{0,1\}$. Instead of using the canonical extractor, we use the extractor K that always chooses $ch = 0, ch' = 1$ (instead of randomly from $\{0,1\}$).[13] Besides that, K behaves like the canonical extractor.

Almost identically as in Theorem 9, we get $K_d = \|P_1^* P_0^* |\Phi_d\rangle\|^2$ (instead of $K_d = \sum_{\substack{ch,ch' \in C_{\eta x} \\ ch \neq ch'}} \frac{1}{\#C_{\eta x}^2} \|P_{ch'}^* P_{ch}^* |\Phi_d\rangle\|^2$). Let $\varphi(v) := (v - \frac{1}{2})^2$ for $v \geq \frac{1}{2}$ and $\varphi(v) := 0$ for $v < \frac{1}{2}$. By Lemma 11, we get $K_d \geq \varphi(V_d)$ for $V_d := \sum_{ch=0,1} \frac{1}{2}\|P_i^*|\Phi\rangle\|^2$.

Note: Lemma 11 applies only to projectors of the special form from Lemma 10. However, the projectors P_0^*, P_1^* are of that form since P_{ch} is a rank-1 projector on the register N (by strict soundness), and $P_{ch}^* = (\mathsf{P}_{\eta x}^* U_{ch})^\dagger P_{ch} (\mathsf{P}_{\eta x}^* U_{ch})$.

Finally, by convexity of φ and Jensen's inequality, we get as in Theorem 9:

$$\Pr_{\mathsf{K}} = \sum_d p_d K_d \geq \sum_d p_d \varphi(V_d) \geq \varphi\Big(\sum_d p_d V_d\Big) = \varphi(\Pr_{\mathsf{V}}).$$

Hence $\Pr_{\mathsf{K}} \geq (\Pr_{\mathsf{V}} - \frac{1}{2})^2$ if $\Pr_{\mathsf{V}} \geq \frac{1}{2}$. Thus (P, V) is extractable with knowledge error $1/2$. □

Other bounds in the literature that bound the disturbance of a state after a measurement are the "Almost As Good As New Lemma" [Aar05, Lemma 2.2] and the "Tender Measurement Lemma" [Win99, Lemma 1.5]. However, we did not get better bounds using those lemmas.

And in the case $\#C = c \gg 2$, all of [CSST11, Aar05, Win99] seem only to give bounds where the knowledge error does not decrease much with increasing c.

4 Zero knowledge

In the preceding sections, we have studied the question what quantum proofs of knowledge are, and when a Σ-protocol is a quantum proof of knowledge. However, a protocol that is just a proof of knowledge is not very useful, one usually additionally requires the protocol

[13] We could also use the canonical extractor here, but that would yield only half the success probability because we have $ch = ch'$ with probability $\frac{1}{2}$. The knowledge error would not be affected by this, however.

to be zero-knowledge. In this section, we study the zero-knowledge property. Basically, we somewhat generalize the result from [Wat09]. The rewinding technique we use in this section is the one from [Wat09]. The difference to [Wat09] is that [Wat09] applies the technique to selected examples while we make the requirements explicit that a Σ-protocol needs to have. (In addition, we present the reasoning in more detail, leaving out less of the computations. See in particular Footnote 15 below.)

The property that we will require from a Σ-protocol is honest-verifier zero-knowledge (HVZK), which is fully analogous to the classical definition:

Definition 13 (Honest-verifier zero-knowledge (HVZK)) *We call Σ-protocol (P,V)* honest-verifier zero-knowledge (HVZK) *iff there is a quantum-polynomial-time algorithm S_Σ (the simulator) such that the transcript of the interaction $\langle \mathsf{P}(x,w),\mathsf{V}(x)\rangle$ is quantum computationally indistinguishable from the output of $S_\Sigma(x)$. More precisely, we require that there exists a quantum-polynomial-time S_Σ such that for any quantum-polynomial-time algorithm D_Σ (the distinguisher) and any polynomial ℓ, there is a negligible μ such that for all $(x,w)\in R$ with $|x|,|w|\leq \ell(\eta)$, and for all states $|\Psi\rangle$:*

$$\begin{aligned}&\big|\Pr[b=1 : (com,ch,resp)\leftarrow\langle \mathsf{P}(x,w),\mathsf{V}(x)\rangle, b\leftarrow D_\Sigma(|\Psi\rangle,com,ch,resp)]\\ &-\Pr[b=1 : (com,ch,resp)\leftarrow S_\Sigma(x), \qquad b\leftarrow D_\Sigma(|\Psi\rangle,com,ch,resp)]\big|\leq\mu(\eta)\end{aligned}$$

Here $(com,ch,resp)\leftarrow\langle \mathsf{P}(x,w),\mathsf{V}(x)\rangle$ means that $com,ch,resp$ are the messages sent in an interaction between P and V.

Note that we allow S_Σ to be quantum here. The resulting notion is sufficient for our purposes because it implies quantum computational zero-knowledge (see below). Alternatively, one can strengthen the definition by requiring S_Σ to be classical. Then HVZK will also imply classical zero-knowledge. For the purposes of this work, it does not matter which definition is chosen.

We also consider a statistical variant of HVZK:

Definition 14 (Statistical honest-verifier zero-knowledge (SHVZK)) *We call Σ-protocol (P,V)* ***statistical honest-verifier zero-knowledge (SHVZK)*** *iff there is a quantum-polynomial-time algorithm S_Σ (the simulator) such that the transcript of the interaction $\langle \mathsf{P}(x,w),\mathsf{V}(x)\rangle$ is statistically indistinguishable from the output of $S_\Sigma(x)$. That is, the definition is like Definition 13, except that we quantify over all (possibly unlimited) D_Σ.*

We can now state the definition of quantum computational zero-knowledge:

Definition 15 (Quantum computational zero-knowledge) *An interactive proof system (P,V) for relation R is* quantum computational zero-knowledge *iff for every quantum-polynomial-time algorithm (verifier) V^* there is a quantum-polynomial-time algorithm (simulator) S such that for any quantum-polynomial-time D (distinguisher) and*

any polynomial ℓ there is a negligible μ such that for any $(x,w) \in R$ with $|x|, |w| \leq \ell(\eta)$, and for any quantum state $|\Psi\rangle$, we have:

$$\Big|\Pr[b = 1 : ZE \leftarrow |\Psi\rangle, \langle \mathsf{P}(x,w), \mathsf{V}^*(Z)\rangle,\ b \leftarrow \mathsf{D}(Z,E)] - \Pr[b = 1 : ZE \leftarrow |\Psi\rangle,\ \mathsf{S}(x,Z),\ b \leftarrow \mathsf{D}(Z,E)]\Big| \leq \mu(\eta).$$

Here $ZE \leftarrow |\Psi\rangle$ denotes that the quantum registers Z, E are initialized jointly with state $|\Psi\rangle$. And $\langle \mathsf{P}(x,w), \mathsf{V}^*(Z)\rangle$ denotes an interaction between prover P and verifier V^* where V^* gets access to the quantum register Z. Note that after that execution V^* may have changed the state of Z. $\mathsf{S}(x,Z)$ also gets access to and may change Z.

There is one important modifications in this definition with respect to the one from [Wat09]: We give the honest prover access to the witness P while in [Wat09] the honest prover is required to find the witness himself (i.e., in [Wat09] we cannot have efficient honest provers except for trivial languages).

We again have a statistical variant:

Definition 16 (Quantum statistical zero-knowledge) *Like quantum statistical zero-knowledge, except that we quantify over all (possibly unlimited) distinguishers D.*

The following corollary is a reformulation of Watrous' quantum rewinding technique [Wat09].

Corollary 17 (Quantum Rewinding Lemma with small perturbations) *Let C, Z, E, Y be quantum registers, where C is one qubit. Let S_1 be a unitary operation operating on C, Z, Y. Let $\mathbf{M}$ be a measurement in the computational basis on register C.*

For a quantum state $|\Psi\rangle$, let $p(|\Psi\rangle) := \Pr[succ = 1 : S_1(CZY), succ \leftarrow \mathbf{M}(C)]$ where Z, E are jointly initialized with $|\Psi\rangle$ and Y, C are initialized with $|0\rangle$. In the same situation, let the density operator ρ^1_Ψ denote the state of ZE in the case of $succ = 1$.

Let $\varepsilon \in (0, 1/2)$. Let $q \in (\varepsilon, 1/2]$. Assume that for all $|\Psi\rangle$, $|p(|\Psi\rangle) - q| \leq \varepsilon$.

Then there is a quantum circuit S operating on Z of size $O\big(\frac{\log(1/\varepsilon) size(S_1)}{(q-\varepsilon)(1-q+\varepsilon)}\big)$. ($S$ is a general quantum circuit, that it, S may create auxiliary qubits, destroy them, and perform measurements.) The description of S can be computed in time $O\big(\frac{\log(1/\varepsilon) size(S_1)}{(q-\varepsilon)(1-q+\varepsilon)}\big)$ given the description of S_1. And for any $|\Psi\rangle$,

$$\mathrm{TD}(\rho^1_\Psi, \rho^2_\Psi) \leq 4\sqrt{\varepsilon}\,\frac{\log(1/\varepsilon)}{(q-\varepsilon)(1-q+\varepsilon)}$$

where the density operator ρ^2_Ψ denotes the state of ZE after execution of S when ZE is initialized with $|\Psi\rangle$.

Proof. By [Wat09, Lemma 9], with $p_0 := q - \varepsilon$ and $Q := S_1$, we get in time $O\big(\frac{\log(1/\varepsilon) size(S_1)}{p_0(1-p_0)}\big)$ a quantum circuit R of size $O\big(\frac{\log(1/\varepsilon) size(S_1)}{p_0(1-p_0)}\big)$ operating on $CZEY$ such that:

Let the density operator $\rho(\Psi)$ denote final state of $CZEY$ after executing R. Let the pure state $|\phi(\Psi)\rangle$ denote the state of $CZEY$ after executing $S_1(ZY), succ \overset{R}{\leftarrow} \mathbf{M}(Z)$ and getting $succ = 1$. Then

$$F(\rho(\Psi), |\phi(\Psi)\rangle\langle\phi(\Psi)|)^2 \geq 1 - 16\varepsilon \frac{\log^2(1/\varepsilon)}{p_0^2(1-p_0)^2}$$

where $F(\cdot,\cdot)$ denotes the fidelity. Furthermore, by inspection of the construction from [Wat09] we see that if $Q = S_1$ applies no gates to E, neither does R. Thus R is a circuit on CZY.

By [NC10, (9.101)] this implies

$$\mathrm{TD}(\rho(\Psi), |\phi(\Psi)\rangle\langle\phi(\Psi)|) \leq 4\sqrt{\varepsilon}\, \frac{\log(1/\varepsilon)}{p_0(1-p_0)}.$$

From R we construct the circuit S operating on ZE by first initializing auxiliary register C, E with $|0\rangle$, running R, and then destroying C, E. Then $\rho_\Psi^2 = \mathrm{tr}_{CE}\, \rho(\Psi)$. And by definition, $\rho_\Psi^1 = \mathrm{tr}_{CE}|\phi(\Psi)\rangle\langle\phi(\Psi)|$. Since the trace distance cannot increase under application of partial trace, we get

$$\mathrm{TD}(\rho_\Psi^1, \rho_\Psi^2) \leq 4\sqrt{\varepsilon}\, \frac{\log(1/\varepsilon)}{p_0(1-p_0)} = 4\sqrt{\varepsilon}\, \frac{\log(1/\varepsilon)}{(q-\varepsilon)(1-q+\varepsilon)}. \qquad \square$$

We can now state the main result of this section:

Theorem 18 *Let (P, V) be a Σ-protocol. Assume that P gives a fixed response* `error` *when receiving a challenge $ch \notin C_{\eta x}$.*[14]

If $|C_{\eta x}|$ is polynomially-bounded in $\eta + |x|$ and Σ is SHVZK, then (P, V) is quantum statistical zero-knowledge.

If $|C_{\eta x}|$ is polynomially-bounded in $\eta + |x|$ and Σ is HVZK, then (P, V) is quantum computational zero-knowledge.

Proof. We do the proof of both parts of the theorem simultaneously. The text contains the wording for the statistical case, with annotations giving the changes needed for the computational case, like ⟦this⟧.

Without loss of generality, we can assume that V^* never sends $ch \notin C_{\eta x}$. Namely, if V^* does send $ch \notin C_{\eta x}$, we can transform it into a verifier $\tilde{\mathsf{V}}^*$ that runs V^*, with the following modification: When V^* sends $ch \notin C_{\eta x}$, then $\tilde{\mathsf{V}}^*$ sends some $ch_0 \in C_{\eta x}$. In this case, instead of the prover's response *resp*, $\tilde{\mathsf{V}}^*$ passes `error` to V^*. We then have that $\langle \mathsf{P}(x,w), \mathsf{V}^*(Z)\rangle$ and $\langle \mathsf{P}(x,w), \tilde{\mathsf{V}}^*(Z)\rangle$ have the same final state in Z, hence $\tilde{\mathsf{V}}^*$ is as successful as V^*, and $\tilde{\mathsf{V}}^*$ never sends $ch \notin C_{\eta x}$. Thus we can assume V^* never to send $ch \notin C_{\eta x}$.

[14] The HVZK/SHVZK property does not guarantee anything when $ch \notin C_{\eta x}$. So P could be HVZK/SHVZK but still reveal the witness when sent an invalid challenge.

By Definition 16 ⟦Definition 15⟧, to show that Σ is quantum statistical ⟦computational⟧ zero-knowledge, for any quantum-polynomial-time V^* and polynomial ℓ, we need to find a quantum-polynomial-time simulator S such that for any ⟦quantum-polynomial-time⟧ D, the following is negligible for $|x|, |w| \leq \ell(\eta)$:

$$\Big|\Pr[b = 1 : ZE \leftarrow |\Psi\rangle, \langle \mathsf{P}(x,w), \mathsf{V}^*(Z)\rangle,\ b \leftarrow \mathsf{D}(Z,E)] - \Pr[b = 1 : ZE \leftarrow |\Psi\rangle,\ \mathsf{S}(x,Z),\ b \leftarrow \mathsf{D}(Z,E)]\Big|. \tag{2}$$

Since a sigma protocol is a three round protocol, we can represent the prover P as two quantum-polynomial-time algorithms P_1, P_2 such that: The commitment *com* sent by the prover P is computed by $com \leftarrow P_1(x,w)$, and the prover's response *resp* to the challenge *ch* is computed by $resp \leftarrow P_2(x,w,ch)$. P_1 and P_2 may share state. Similarly, the malicious verifier V^* can be represented by two quantum-polynomial-time algorithms V_1^*, V_2^* such that the challenge *ch* is produced by $V_1^*(com, Z)$, and, given the response *resp*, the verifier runs $V_2^*(resp, Z)$. (Note that V_2^* does not give output because V^* does not. However, both V_1^*, V_2^* can have side effects on the quantum register Z.) V_1^*, V_2^* may share state.

With that notation,

$$\langle \mathsf{P}(x,w), \mathsf{V}^*(Z)\rangle \qquad \text{is the same as} \qquad com \leftarrow P_1(x,w), ch \leftarrow V_1^*(com, Z), resp \leftarrow P_2(x,w,ch), V_2^*(resp, Z). \tag{3}$$

Σ is SHVZK ⟦HVZK⟧. Hence there is a quantum-polynomial-time simulator S_Σ such that for any ⟦quantum-polynomial-time⟧ D_Σ:

$$\Big|\Pr[b = 1 : com \leftarrow P_1(x,w), ch \xleftarrow{R} C_{\eta x}, resp \leftarrow P_2(x,w,ch), b \leftarrow D_\Sigma(|\Psi\rangle, com, ch, resp)] - \Pr[b = 1 : (com, ch, resp) \leftarrow S_\Sigma(x), b \leftarrow D_\Sigma(|\Psi\rangle, com, ch, resp)]\Big| \leq \varepsilon_D \tag{4}$$

where $\varepsilon_D = \varepsilon_D(\eta)$ is a negligible function depending on D_Σ.

Let $[ch = ch^*] := 1$ iff $ch = ch^*$.

Then:

$$\begin{aligned}
&\Pr[succ = 1 \wedge b = 1 : ZE \leftarrow |\Psi\rangle, \langle \mathsf{P}(x,w), \mathsf{V}^*(Z)\rangle,\ b \leftarrow \mathsf{D}(Z,E),\\
&\quad ch^* \xleftarrow{R} C_{\eta x},\ succ := [ch = ch^*]]\\
&\overset{(3)}{=} \Pr[succ = 1 \wedge b = 1 : ZE \leftarrow |\Psi\rangle, com \leftarrow P_1(x,w),\ ch \leftarrow V_1^*(com, Z),\ resp \leftarrow P_2(x,w,ch),\\
&\quad V_2^*(resp, Z),\ b \leftarrow \mathsf{D}(Z,E),\ ch^* \xleftarrow{R} C_{\eta x},\ succ := [ch = ch^*]]\\
&\overset{(*)}{=} \Pr[succ = 1 \wedge b = 1 : com \leftarrow P_1(x,w),\ ch^* \xleftarrow{R} C_{\eta x},\ resp \leftarrow P_2(x,w,ch^*),\ ZE \leftarrow |\Psi\rangle,\\
&\quad ch \leftarrow V_1^*(com, Z),\ succ := [ch = ch^*],\ V_2^*(resp, Z),\ b \leftarrow \mathsf{D}(Z,E)]\\
&\overset{\varepsilon}{\approx} \Pr[succ = 1 \wedge b = 1 : (com, ch^*, resp) \leftarrow S_\Sigma(x),\ ZE \leftarrow |\Psi\rangle,\\
&\quad ch \leftarrow V_1^*(com, Z),\ succ := [ch = ch^*],\ V_2^*(resp, Z),\ b \leftarrow \mathsf{D}(Z,E)]\\
&= \Pr[succ = 1 \wedge b = 1 : ZE \leftarrow |\Psi\rangle,\ Y, C \leftarrow |0\rangle,\ S_1(x, CZY),\ succ \leftarrow \mathbf{M}(C),\ b \leftarrow \mathsf{D}(Z,E)]
\end{aligned} \tag{5}$$

Here $(*)$ uses the fact that $P_2(x,w,ch)$ and $P(x,w,ch^*)$ get the same arguments when $succ = 1$. And $\overset{\varepsilon}{\approx}$ means a difference of at most ε. The $\overset{\varepsilon}{\approx}$ follows from (4) with the quantum-polynomial-time adversary $D_\Sigma(|\Psi\rangle, com, ch^*, resp)$ that runs "$ZE \leftarrow |\Psi\rangle, ch \leftarrow V_1^*(com, Z), succ := [ch = ch^*], V_2^*(resp, Z), b \leftarrow \mathsf{D}(Z,E)$" and returns $b \wedge succ$. And $\varepsilon := \varepsilon_D$ is negligible with ε_D as in (4). And in the last line, S_1 is the unitary quantum circuit (depending on x) constructed as follow: Transform the steps "$(com, ch^*, resp) \leftarrow S_\Sigma(x), ch \leftarrow V_1^*(com, Z), succ := [ch = ch^*], V_2^*(resp, Z)]$" into a *unitary* quantum circuit S_1 operating on registers C, Z and Y (using Y for auxiliary qubits). The value of *succ* is stored in the one-qubit quantum register C. $succ \leftarrow \mathbf{M}(C)$ then retrieves *succ* by a measurement $\mathbf{M}$ on C in the computational basis.

Furthermore, for all quantum states $|\Psi\rangle$,

$$\begin{aligned}\frac{1}{|C_{\eta x}|} &= \Pr[succ = 1 : ZE \leftarrow |\Psi\rangle, \langle \mathsf{P}(x,w), \mathsf{V}^*(Z)\rangle,\ ch^* \overset{R}{\leftarrow} C_{\eta x},\ succ := [ch = ch^*]] \\ &\overset{\varepsilon'}{\approx} \Pr[succ = 1 : ZE \leftarrow |\Psi\rangle,\ Y, C \leftarrow |0\rangle,\ S_1(x, CZY),\ succ \leftarrow \mathbf{M}(C)] \qquad (6)\end{aligned}$$

Here $\overset{\varepsilon'}{\approx}$ is shown analogously to (5) for negligible $\varepsilon' = \varepsilon'(\eta)$. Without loss of generality, we can choose $\varepsilon' \geq 2^{-\eta}$.

Note that, since $|C_{\eta x}|$ is polynomially-bounded in $\eta + |x|$, and $|x| \leq \ell(\eta)$, we have that $q := 1/|C_{\eta x}|$ is noticeable in η.

By (6) and Corollary 17 (with $q := 1/|C_{\eta x}|$) we get that there is an algorithm S such that for sufficiently large η:[15]

$$\mathrm{TD}(\rho_\Psi^1, \rho_\Psi^2) \leq 4\sqrt{\varepsilon'}\frac{\eta}{(q-\varepsilon')(1-q+\varepsilon')} =: \delta \qquad (7)$$

Here ρ_Ψ^1 is the final state of ZE after executing $S_1(x, CZY), succ \leftarrow \mathbf{M}(C)$ and getting $succ = 1$. And ρ_Ψ^2 is the final state of ZE after executing $S(x, CZY)$. (Corollary 17 requires $q > \varepsilon$. Since q is noticeable and ε is negligible, this is the case for sufficiently large η.)

The algorithm has running time $O\big(\frac{\log(1/\varepsilon) size(S_1)}{(q-\varepsilon)(1-q+\varepsilon)}\big)$, which is polynomially-bounded since S_1 has polynomially-bounded size and $q - \varepsilon$ is noticeable and $q \leq \frac{1}{2}$.

From (7) and the fact that two states cannot be distinguished better than their trace distance, we get

$$\begin{aligned}&\Pr[b = 1 : ZE \leftarrow |\Psi\rangle,\ Y, C \leftarrow |0\rangle,\ S_1(x, CZY),\ succ \leftarrow \mathbf{M}(C),\ b \leftarrow \mathsf{D}(Z,E) \mid succ = 1] \\ &\overset{\delta}{\approx} \Pr[b = 1 : ZE \leftarrow |\Psi\rangle,\ Y, C \leftarrow |0\rangle,\ S(x, CZY),\ b \leftarrow \mathsf{D}(Z,E)]. \qquad (8)\end{aligned}$$

Before we continue, we prove an auxiliary claim:

[15] Note here that to apply Corollary 17 we indeed need (6) to hold for *all* quantum states $|\Psi\rangle$. In particular, that means that Corollary 17 would not be applicable in the computational case if HVZK was defined uniformly (i.e., with $|\psi\rangle$ being chosen by a quantum-polynomial-time algorithm). We leave the security guarantees obtained from uniform HVZK for future research.

Claim 3 *Consider two probability spaces* $\Pr_1, \Pr_2$ *with events* B, S *each. Assume that in* $\Pr_1$*,* B *and* S *are independent. Let* $N := 1/\Pr_1[S]$*. Assume that* $\Pr_1[B \wedge S] \stackrel{\varepsilon}{\approx} \Pr_2[B \wedge S]$ *and* $\Pr_1[S] \stackrel{\varepsilon'}{\approx} \Pr_2[S]$*. Then* $\Pr_1[B] \stackrel{N\varepsilon+N\varepsilon'}{\approx} \Pr_2[B|S]$*.*

The claim follows from the following calculation:

$$\begin{aligned}
&\big|\Pr_1[B] - \Pr_2[B|S]\big| \stackrel{(*)}{=} \Big|\frac{\Pr_1[B \wedge S]}{\Pr_1[S]} - \frac{\Pr_2[B \wedge S]}{\Pr_2[S]}\Big| \\
&= \Big|\frac{\Pr_1[B \wedge S]}{\Pr_1[S]} - \frac{\Pr_2[B \wedge S]}{\Pr_1[S]} + \frac{\Pr_2[B \wedge S]}{\Pr_1[S]} - \frac{\Pr_2[B \wedge S]}{\Pr_2[S]}\Big| \\
&\leq \Big|\frac{\Pr_1[B \wedge S]}{\Pr_1[S]} - \frac{\Pr_2[B \wedge S]}{\Pr_1[S]}\Big| + \Big|\frac{\Pr_2[B \wedge S]}{\Pr_1[S]} - \frac{\Pr_2[B \wedge S]}{\Pr_2[S]}\Big| \\
&= \underbrace{\frac{1}{\Pr_1[S]}}_{=N}\underbrace{\big|\Pr_1[B \wedge S] - \Pr_2[B \wedge S]\big|}_{\leq\varepsilon} + \underbrace{\Pr_2[B \wedge S]}_{\leq \Pr_2[S]}\Big|\frac{1}{\Pr_1[S]} - \frac{1}{\Pr_2[S]}\Big| \\
&\leq N\varepsilon + \Big|\frac{\Pr_2[S]}{\Pr_1[S]} - 1\Big| = N\varepsilon + \frac{\big|\Pr_2[S] - \Pr_1[S]\big|}{\Pr_1[S]} \leq N\varepsilon + N\varepsilon'.
\end{aligned}$$

Here $(*)$ uses that B and S are independent in $\Pr_1$. This shows the claim.

We now instantiate the claim. B is the event $b = 1$, and S the event $succ = 1$. $\Pr_1$ refers to the probability space defined by $ZE \leftarrow |\Psi\rangle, \langle \mathsf{P}(x,w), \mathsf{V}^*(Z)\rangle,\ b \leftarrow \mathsf{D}(Z,E),\ ch^* \stackrel{R}{\leftarrow} C_{\eta x},\ succ := [ch = ch^*]$, and $\Pr_2$ refers to the probability space defined by the game $ZE \leftarrow |\Psi\rangle,\ Y, C \leftarrow |0\rangle,\ S_1(x, CZY),\ succ \leftarrow \mathbf{M}(C),\ b \leftarrow \mathsf{D}(Z,E)$. Then B and S are obviously independent in $\Pr_1$. The condition $\Pr_1[B \wedge S] \stackrel{\varepsilon}{\approx} \Pr_2[B \wedge S]$ is satisfied by (5). And the condition $\Pr_1[S] \stackrel{\varepsilon'}{\approx} \Pr_2[S]$ by (6) (using that the invocation $b \leftarrow \mathsf{D}(Z,E)$ does not influence *succ*). And $N = 1/\Pr_1[S] = |C_{\eta x}|$. Then Claim 3 implies the $\stackrel{N\varepsilon+N\varepsilon'}{\approx}$ step of the following calculation:

$$\begin{aligned}
&\Pr[b = 1 : ZE \leftarrow |\Psi\rangle, \langle \mathsf{P}(x,w), \mathsf{V}^*(Z)\rangle,\ b \leftarrow \mathsf{D}(Z,E)] \\
&= \Pr[b = 1 : ZE \leftarrow |\Psi\rangle, \langle \mathsf{P}(x,w), \mathsf{V}^*(Z)\rangle,\ b \leftarrow \mathsf{D}(Z,E), \\
&\qquad ch^* \stackrel{R}{\leftarrow} C_{\eta x},\ succ := [ch = ch^*]] \\
&\stackrel{N\varepsilon+N\varepsilon'}{\approx} \Pr[b = 1 : ZE \leftarrow |\Psi\rangle,\ Y, C \leftarrow |0\rangle,\ S_1(x, CZY), \\
&\qquad succ \leftarrow \mathbf{M}(C),\ b \leftarrow \mathsf{D}(Z,E) \mid succ = 1] \\
&\stackrel{\delta}{\approx} \Pr[b = 1 : ZE \leftarrow |\Psi\rangle,\ Y, C \leftarrow |0\rangle,\ S(x, CZY),\ b \leftarrow \mathsf{D}(Z,E)] \\
&= \Pr[b = 1 : ZE \leftarrow |\Psi\rangle,\ \mathsf{S}(x, Z),\ b \leftarrow \mathsf{D}(Z,E)].
\end{aligned}$$

Here we defined the simulator $\mathsf{S}(x, Z)$ to run "$Y, C \leftarrow |0\rangle, S(x, CZY)$". And the $\stackrel{\delta}{\approx}$ step follows from (8).

Since S is quantum-polynomial-time, ε, ε', and δ are negligible, and N is polynomially-bounded, we have that (2) is bounded by the negligible $N\varepsilon + N\varepsilon' + \delta$. This shows that Σ is quantum statistical [[computational]] zero-knowledge. □

5 QPoKs for all languages in NP

In Section 3, we saw that complete proof systems with strict and special soundness are QPoKs. The question that remains to be asked is: do such proof systems, with the additional property of being zero-knowledge, exist for interesting languages? In this section, we will show that for any language in NP (more precisely, for any NP-relation), there is a zero-knowledge QPoK. (Assuming the existence of quantum 1-1 one-way functions.) Here and in the following, by zero-knowledge we mean quantum computational zero-knowledge.

The starting point for our construction will be the Blum's zero-knowledge PoK for Hamiltonian cycles [Blu86]. In this Σ-protocol, the prover commits to the vertices of a graph using a perfectly binding commitment scheme. In the prover's response, some of these commitments are opened. That is, the response contains the opening information for some of the commitments. The problem is that standard definitions of commitment schemes do not guarantee that the opening information is unique; only the actual content of the commitment has to be determined by the commitment. This means that the prover's response is not unique. Thus, with a standard commitment scheme we do not get strict soundness. Instead, we need a commitment scheme such that the sender of the commitment scheme is committed not only to the actual content of the commitment, but also to the opening information.

Definition 19 (Strict binding) *A commitment scheme* COM *is a deterministic polynomial-time function taking two arguments a, y, the opening information a (a.k.a. the randomness) and the message y. We say* COM *is* strictly binding *if for all a, y, a', y' with $(a, y) \neq (a', y')$, we have that* $\mathrm{COM}(a, y) \neq \mathrm{COM}(a', y')$.

Furthermore, in order to get the zero-knowledge property, we will need that our commitment schemes are quantum computationally concealing. We refer to [Wat09] for a precise definition of this property. In [AC02], an unconditionally binding, quantum computationally concealing commitment scheme based on quantum 1-1 one-way function is presented.[16] (We discuss the existence of quantum 1-1 one-way functions on page 30 below.) Their definitions differ somewhat from those of [Wat09], but as mentioned in [Wat09], their proof carries over to the definitions from [Wat09]. Furthermore, in the scheme from [AC02], the commitment contains the image of the opening information under a quantum 1-1 one-way function. Thus the strict binding property is trivially fulfilled. Thus strictly binding, quantum computationally concealing commitment schemes exist under the assumption that quantum 1-1 one-way functions exist.

Given such a commitment scheme COM, we construct the proof system (P, V) presented in Figure 1. Besides using a strictly binding commitment, (P, V) differs in one

[16] In [AC02], the result is stated for quantum one-way permutations $f : \{0,1\}^n \to \{0,1\}^n$. (To the best of our knowledge, no candidates for quantum one-way permutations are known.) Inspection of their proof reveals, however, that the result also holds for families of quantum 1-1 one-way functions $f_i : \{0,1\}^n \to D$ for arbitrary domain D and efficiently samplable indices i, assuming that given an index i, it can be efficiently verified that f_i is injective.

Inputs: A directed graph x (the statement) with vertices W, and a Hamiltonian cycle w in x (the witness).
Subprotocols: A strictly binding, quantum computationally concealing commitment scheme COM.
Protocol:

1. P picks a random permutation π on W. Let A be the adjacency matrix of the graph $\pi(x)$. Let $H := \{(\pi(i), \pi(j)) : (i,j) \in w\}$. Using COM, P commits to π, H, and to each entry A_{ij} of A. P sends the resulting commitments to V.
2. V picks $ch \in \{0,1\}$ and sends ch to P.
3. If $ch = 0$, P opens the commitments to π and A. If $ch = 1$, P opens the commitments to H and to all A_{ij} with $(i,j) \in H$.
4. If $\mathsf{ch} = 0$, V checks that the commitments are opened correctly, that π is a permutation, and that A is the adjacency matrix of $\pi(x)$. If $ch = 1$, V checks that the commitments are opened correctly, that H is a cycle, that exactly the A_{ij} with $(i,j) \in H$ are opened, and that $A_{ij} = 1$ for all $(i,j) \in H$. If all checks succeed, V outputs 1.

Figure 1: A QPoK (P, V) for Hamiltonian cycles.

other aspect from the proof system in [Blu86]: The prover does not only commit to the vertices in the graph $\pi(x)$, but also to the permutation π and the cycle H. Without these additional commitments, we would not get strict soundness; there might be several permutations leading to the same graph, or the graph might contain several Hamiltonian cycles.

Theorem 20 *Let $(x,w) \in R$ iff w is a Hamiltonian cycle of the graph x. Assume that* COM *is a strictly binding, quantum computationally concealing commitment scheme. Then the proof system (P, V) is a quantum computational zero-knowledge QPoK for R with knowledge error $\frac{1}{2}$.*

Proof. We need to show completeness, extractability (via special and strict soundness), and zero-knowledge (via HVZK). Completeness is straightforward by inspection of the protocol.

Special soundness. Let $(com, ch, resp)$ and $(com, ch', resp')$ be two accepting conversations for x with $ch \neq ch'$. Without loss of generality, $ch = 0$ and $ch' = 1$. Then $resp$ contains a permutation π and the adjacency matrix A of $\pi(x)$. And $resp'$ contains a cycle H such that $\tilde{A}_{ij} = 1$ for all $(i,j) \in H$ where $\tilde{A}_{ij}$ are the committed values opened in $resp'$. Since ch is strictly binding, $1 = \tilde{A}_{ij} = A_{ij}$ for all $(i,j) \in H$, thus H is a Hamiltonian cycle of $\pi(x)$. Then $w := \mathsf{K}_0(x, com, ch, resp, ch', resp') := \pi^{-1}(H)$ is a Hamiltonian cycle of x, i.e., $(x,w) \in R$.

Strict soundness. Fix an accepting conversation $(com, ch, resp)$. If $ch = 0$, $resp$ consists only of the opening of commitments. Since COM has strict binding, it follows that $resp$ is uniquely determined by com, ch. If $ch = 1$, COM consists of an opening

of the commitment to H, and of the commitments to A_{ij} with $(i,j) \in H$. Hence H and its opening information are uniquely determined since COM has strict binding, and thus it is also determined which A_{ij} are opened. Again by strict binding, the values A_{ij} and corresponding opening information are uniquely determined. Thus *resp* is uniquely determined by com, ch.

Extractability. Since (P, V) has special and strict soundness, and a challenge space of size 2, by Corollary 12, we have that (P, V) is extractable with knowledge error $\frac{1}{2}$.

Honest-verifier zero-knowledge. We describe the simulator S_Σ (cf. Definition 13). $S_\Sigma(x)$ first picks a random $ch \in \{0,1\}$. If $ch = 0$, S_Σ chooses a random permutation π, computes the adjacency matrix A of $\pi(x)$, and picks an arbitrary H. If $ch = 1$, S_Σ chooses a random permutation π, sets A to be the all-one matrix, and lets H be a random cycle. Let *com* consist of the commitments to π, A, H. Let *resp* be the openings as specified in Figure 1, Step 4. Then S_Σ outputs $(com, ch, resp)$.

Since COM is quantum computationally conceiling, we easily see that the output from S_Σ is indistinguishable from $\langle \mathsf{P}, \mathsf{V} \rangle$ in the sense of Definition 13. Thus (P, V) is HVZK.

Zero-knowledge. The challenge space $C_{\eta x}$ is $\{0,1\}$, hence $|C_{\eta x}| = 2$ is polynomially-bounded. As shown above, (P, V) is HVZK. By Theorem 18, this implies that (P, V) is quantum computational zero-knowledge.

Corollary 21 (QPoKs for all languages in NP) *Let R be an NP-relation.*[17] *Then there is a zero-knowledge QPoK for R with negligible knowledge error.*

Proof. Using the fact that the Hamiltonian cycle problem is NP-complete, from Theorem 20 it follows that there is a zero-knowledge QPoK for R with knowledge error $\frac{1}{2}$. By sequential repetition, we get a QPoK for R with negligible knowledge error (Theorem 3). Sequential repetition preserves the zero-knowledge property (see [Wat09]). □

Quantum 1-1 one-way functions. Our construction relies on quantum 1-1 one-way functions. Classical 1-1 one-way functions include the RSA function (assumping the RSA assumption) and exponentiation in a finite group (assuming the hardness of discrete logarithms in that group). Both are not secure in the quantum setting due to Shor's algorithm [Sho94]. To the best of our knowledge, these are all the candidates for 1-1 one-way functions described in the literature. However, we present the following two functions as candidates for quantum 1-1 one-way functions:

- $f_1(x) := H(1\|x)\| \dots \|H(n\|x)$ for a hash function H (say, SHA-3) and n large enough so that f_1 becomes injective.
- $f_2(k) := E(k, m_1)\| \dots \|E(k, m_n)$ for a block cipher E (say, AES). Here $m_1, \dots, m_n$ is a sufficiently long fixed list of publicly known plaintexts such that there are no two keys encrypting them to the same list of ciphertexts. (I.e., n should be the unicity distance of E.)

[17] An NP-relation is a relation R such that $(x,w) \in R$ is decidable in deterministic polynomial time, and there is a polynomial p such that for all $(x,w) \in R$, $|w| \leq p(|x|)$.

One easily verifies that f_1 and f_2 are one-way if $H(\cdot\|x)$ and $E(k,\cdot)$, respectively, are quantum pseudorandom functions. (This is a common assumption since functions like SHA-3 and AES do not seem to have algebraic structure that can be exploited by quantum algorithms.) Proving that f_1 and f_2 are actually injective seems hard (due to that same lack of structure). But for sufficiently large n, it seems a reasonable assumption that both functions are injective. We therefore propose those two constructions as candidates for quantum 1-1 one-way functions.

Note however, that f_1 and f_2 are not quantum one-way *permutations*. In fact, we do not know any candidates for those, since the only known candidates in the classical setting are RSA and exponentiation. Adcock and Cleve [AC02] and Watrous [Wat09] both assume quantum one-way permutations in their constructions. However, an inspection of their proofs shows that they actually only require quantum 1-1 one-way functions. Note however that not all schemes stay secure when replacing a one-way permutation by a 1-1 one-way function. For example, the commitment from [DMS00] loses the conceiling property if the underlying one-way function is not a permutation.

6 Open questions

We list several natural open questions related to the results of this paper:

Tight bounds on the knowledge error. In this work, we show that a Σ-protocol with strict and special soundness has knowledge error $1/\sqrt{c}$ where c is the size of the challenge space. In the classical setting, the knowledge error is $1/c$. Can we improve the quantum bound to match the classical bound?[18] Possibly by using a different construction for the extractor? Or can we show the bound to be tight? (At least relative to some oracle.) And related, we achieve an exponent $d = 3$ in Definition 1, while in the quantum case, the exponent is $d = 2$. Is $d = 3$ tight? (Note that, at least for the case $c = 2$, we have a better bound in Corollary 12.)

Arguments and arguments of knowledge. We show that Σ-protocols with special and strict soundness are quantum proof of knowledge. But if the Σ-protocol has only computational special soundness and computational strict soundness? Do we get an argument of knowledge? [ARU14] gives evidence against this; at least relative to some oracle the resulting scheme will not always be a quantum argument of knowledge (nor even a quantum argument). Does this result also hold in a non-relativised setting? Or can we somewhat modify the definition of computational strict soundness to get arguments of knowledge? In the case of a polynomially-bounded challenge space, Unruh [Unr15] gives a positive answer to the last question, using "collapse-binding" commitments.

Specific protocols without strict soundness. [ARU14] shows that strict soundness seems necessary (with respect to some oracle) to get quantum proof of knowledge. But this does not mean that specific protocols might not be quamtum proofs of knowledge.

[18] The classical bound can easily be seen to be tight: If the Σ-protocol is HVZK, then a malicious prover can succeed in the protocol with probability $\frac{1}{2} \pm$ *negligible* by executing the simulator and sending the commitment and response provided by the simulator.

For example, the protocol for graph isomorphism [GMW91]: is that one a quantum proof of knowledge?

Amplifying the success probability of the extractor. Our constructions show the existence of an extractor with success probability $\Pr_K \geq \frac{1}{p}(\Pr_V - \kappa)^d$ where $\Pr_{\mathsf{V}}$ is the success probability of the prover. If κ is negligible, this implies that a non-negligible $\Pr_{\mathsf{V}}$ implies a non-negligible extraction success probability $\Pr_{\mathsf{E}}$. However, sometimes it is needed to have a success probability close to 1. For example, when using the proof of knowledge property in the construction of a simulator who needs the witness to perform his simulation correctly. One such case is the graph non-isomorphism proof from [GMW91] where the zero-knowledge property of the graph non-isomorphism proof relies on the proof of knowledge property of the graph isomorphism proof. A similar case is the GMW-compiler for multi-party computation (as presented in [Gol04, Chapter 7]). In the classical setting, the success probability of the extractor can be increased at the expense of runtime by running the extractor many times. This does not work in the quantum setting: if the extractor fails upon the first invocation, its initial state will be destroyed. Is there a possibility to amplify the success probability of extractors in the quantum setting? Can we show the graph non-isomorphism protocol from [GMW91] to be zero-knowledge?

Quantum 1-1 one-way functions. Our construction from Section 5 assume quantum 1-1 one-way functions. (And so does [Wat09].) We gave one candidate for such functions. Are there more candidates? Can we modify our results to avoid the use of quantum 1-1 one-way functions? (Basically, this boils down to finding other constructions of quantum conceiling strict binding commitment schemes.)

Acknowledgements.

We thank the anonymous referees, Märt Põldvere, and Rainis Haller for suggestions on how to significantly simplify the proof of Lemma 8. We thank Dennis Hofheinz, Chris Peikert, and Vinod Vaikuntanathan for discussions on candidates for quantum 1-1 one-way functions. We thank Claude Crépeau and Louis Salvail for inspiring discussions on the difficulties of quantum proofs of knowledge. We thank Ehsan Ebrahimi Targhi and Mayuresh Anand for pointing out inconsistencies in the discussion of alternative definitions. This research was supported by the Cluster of Excellence "Multimodal Computing and Interaction", by the European Social Fund's Doctoral Studies and Internationalisation Programme DoRa, by the European Regional Development Fund through the Estonian Center of Excellence in Computer Science, EXCS, by European Social Fund through the Estonian Doctoral School in Information and Communication Technology, and by the Estonian ICT program 2011-2015 (3.2.1201.13-0022).

References

[Aar05] Scott Aaronson. Limitations of quantum advice and one-way communication. *Theory of Computing*, 1(1):1–28, 2005. Online available at `http:`

//www.theoryofcomputing.org/articles/v001a001.

[AC02] Mark Adcock and Richard Cleve. A quantum Goldreich-Levin theorem with cryptographic applications. In *STACS 2002*, volume 2285 of *LNCS*, pages 323–334, Berlin, Heidelberg, 2002. Springer.

[ARU14] Andris Ambainis, Ansis Rosmanis, and Dominique Unruh. Quantum attacks on classical proof systems (the hardness of quantum rewinding). In *FOCS 2014*, pages 474–483. IEEE, 2014. Preprint on IACR ePrint 2014/296.

[BG93] Mihir Bellare and Oded Goldreich. On defining proofs of knowledge. In Ernest F. Brickell, editor, *CRYPTO '92*, volume 740 of *LNCS*, pages 390–420. Springer-Verlag, 1993. Extended version online available at `http://www-cse.ucsd.edu/users/mihir/papers/pok.ps`.

[Blu86] Manuel Blum. How to prove a theorem so no one else can claim it. In *Proceedings of the International Congress of Mathematicians*, pages 1444–1451, Berkeley, 1986.

[CSST11] Claude Crépeau, Louis Salvail, Jean-Raymond Simard, and Alain Tapp. Two provers in isolation. In Dong Hong Lee and Xiaoyun Wang, editors, *Asiacrypt 2011*, volume 7072 of *LNCS*, pages 407–430. Springer, 2011.

[DMS00] Paul Dumais, Dominic Mayers, and Louis Salvail. Perfectly concealing quantum bit commitment from any quantum one-way permutation. In *Eurocrypt 2000*, volume 1807 of *LNCS*, pages 300–315. Springer, 2000.

[GMR85] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof-systems. In *Proceedings of the seventeenth annual ACM symposium on Theory of computing*, pages 291–304. ACM Press, 1985.

[GMR88] Shafi Goldwasser, Silvio Micali, and Ronald L. Rivest. A digital signature scheme secure against adaptive chosen-message attacks. *SIAM Journal on Computing*, 17(2):281–308, 1988. Online available at `http://theory.lcs.mit.edu/~rivest/GoldwasserMicaliRivest-ADigitalSignatureSchemeSecureAgainstAdaptiveChosenMessageAttacks.ps`.

[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(3):690–728, 1991. Online available at `http://www.wisdom.weizmann.ac.il/~oded/X/gmw1j.pdf`.

[Gol04] Oded Goldreich. *Foundations of Cryptography: Basic Applications*, volume 2. Cambridge University Press, Cambridge, UK, 2004.

[Gro96] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *STOC*, pages 212–219, 1996.

[HM98] Shai Halevi and Silvio Micali. More on proofs of knowledge. IACR ePrint 1998/015, 1998.

[HSS11] Sean Hallgren, Adam Smith, and Fang Song. Classical cryptographic protocols in a quantum world. In *Crypto 2011*, volume 6841 of *LNCS*, pages 411–428. Springer, 2011.

[Jen06] Johan L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1):175–193, 1906. In French.

[KSU13] Lee Klingler, Rainer Steinwandt, and Dominique Unruh. On using probabilistic turing machines to model participants in cryptographic protocols. *Theoretical Computer Science*, 501:49–51, 2013.

[LN11] Carolin Lunemann and Jesper Buus Nielsen. Fully simulatable quantum-secure coin-flipping and applications. In *Africacrypt 2011*, volume 6737 of *LNCS*, pages 21–40. Springer, 2011.

[NC10] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 10th anniversary edition, 2010.

[Sho94] Peter W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *FOCS 1994*, pages 124–134. IEEE Computer Society, 1994.

[Unr10] Dominique Unruh. Universally composable quantum multi-party computation. In *Eurocrypt 2010*, LNCS, pages 486–505. Springer, 2010. Preprint on arXiv:0910.2912 [quant-ph].

[Unr12] Dominique Unruh. Quantum proofs of knowledge. In *Eurocrypt 2012*, volume 7237 of *LNCS*, pages 135–152. Springer, April 2012.

[Unr13] Dominique Unruh. Everlasting multi-party computation. In *Crypto 2013*, volume 8043 of *LNCS*, pages 380–397. Springer, 2013. Preprint on IACR ePrint 2012/177.

[Unr14] Dominique Unruh. Non-interactive zero-knowledge proofs in the quantum random oracle model. In *Eurocrypt 2014*, LNCS. Springer, 2014. To appear.

[Unr15] Dominique Unruh. Computationally binding quantum commitments. Unpublished manuscript, 2015.

[vdG98] Jeroen van de Graaf. *Towards a formal definition of security for quantum protocols*. PhD thesis, Départment d'informatique et de r.o., Université de Montréal, 1998. Online available at `http://www.cs.mcgill.ca/~crepeau/PS/these-jeroen.ps`.

[Wat09] John Watrous. Zero-knowledge against quantum attacks. *SIAM J. Comput.*, 39(1):25–58, 2009.

[Win99] Andreas Winter. *Coding Theorems of Quantum Information Theory*. PhD thesis, Universität Bielefeld, 1999. arXiv:quant-ph/9907077v1.

[WZ82] W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.